Proceedings of the 7th International Conference English Pronunciation: Issues and Practices (EPIP 7)

Alice Henderson, Anastazija Kirkova-Naskova, Alex Baratta, Sylvain Coulange, Kristýna Červinková Poesová, Lenka Čtvrtečková, Ondřej Fischer, Radek Skarnitzl, Ivana Duckinoska-Mihajlovska, Kizzi Edensor Costille, et al.

Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices (EPIP 7)

edited by
Alice Henderson
Anastazija Kirkova-Naskova

Université Grenoble-Alpes,
Grenoble, France
2023

Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices (EPIP 7)

edited by
Alice Henderson
Anastazija Kirkova-Naskova

Université Grenoble-Alpes,
Grenoble, France
2023

DOI 10.5281/zenodo.10205580

2023

Table of Contents

Acknowledgements

Preparing a text for publication requires the attentive, eagle-eyed efforts of numerous people, starting with the authors. We thank them all profusely for trusting us with their work. It was a true leap of faith to send us a text, with the vague promise of an online publication that would have a DOI. We hope the final result meets your expectations.

Reviewers have also played a critical role in bringing these Proceedings to life, ensuring double-blind peer-reviews. They brought much more than just another pair of eyes and their intellectual openness; they also freely contributed their specific expertise – sometimes dropping other work to engage in back-and-forth discussions about possible solutions, potential exceptions. The following colleagues worked extremely hard to bring out the best in each text:

Gemma Archer	University of Strathclyde, Glasgow, Scotland
Małgorzata Baran-Łucarz	University of Wrocław, Poland
Walcir Cardoso	Concordia University, Montréal, Canada
Vincent Chanethom	Princeton University, USA
Graeme Couper	Auckland University of Technology, New Zealand
Tracey Derwing	University of Alberta, Canada
Rias van den Doel	University of Utrecht, Netherlands
Jonás Fouz González	University of Murcia, Spain
Helen Fraser	University of Melbourne, Australia
Dan Frost	University Grenoble-Alpes, France
Francisco Gallardo Del Puerto	University of Cantabria, Spain
Esther Gomez Lacabex	University of the Basque Country, Spain
Nadine Herry-Bénit	University Paris Nanterre, France
Amanda Huensch	University of Pittsburgh, USA
Anna Jarosz	University of Łódź, Poland
Okim Kang	North Arizona University, USA
David Levey	University of Cadiz, Spain
Pekka Lintunen	University of Turku, Finland
José Mompéan-Gonzalez	University of Murcia, Spain
Charlie Nagle	University of Austin, Texas, USA
Mireia Ortega	University of Barcelona, Spain
Mirosław Pawlak	Adam Mickiewicz University, Kalisz, Poland
Arkadiusz Rojczyk	University of Silesia in Katowice, Poland
Leticia Quesada	Nebrija University, Madrid, Spain
Veronica Sardegna	Duquesne University, Pittsburgh, USA
Linda Shockey	University of Reading, England
Radek Skarnitzl	Charles University, Prague, Czech Republic
Sinem Sonsaat-Hegelheimer	Iowa State University, USA
Arizio Sweeting	University of the Sunshine Coast, Australia

Elina Tergujeff	University of Jyväskylä, Finland
Linda Terrier	University Toulouse – Jean Jaurès, France
Jan Volín	Charles University, Prague, Czech Republic
Shona Whyte	University Côte d'Azur, France
Adam Wilson	University of Lorraine – Metz, France
Magdalena Wrembel	Adam Mickiewicz University, Poznań, Poland
Elisabeth Zetterholm	Linköping University, Sweden
Beth Zielinski	Macquarie University, Australia

Online access to the Proceedings only became a reality thanks to the enthusiastic efforts of three colleagues who operate very much behind the scenes. Isabelle Rousset, the research engineer in the LIDLEM research group at University Grenoble-Alpes, lobbied hard to make sure all group members had regular, individual access to reference librarians from our university libraries. Thanks to Isabelle's foresight and tenacity, we came into contact with two formidably competent and patient colleagues: Sandrine Corvey-Biron and Anne-Christine Jacob. They repeatedly came to our rescue, explaining Creative Commons Licenses and the value of DOIs to us, as well as walking us through all the steps to create a collection of texts on HAL. We cannot thank them enough for their friendly smiles and creative calm in finding solutions to every problem we brought to them. We would also like to thank Aynur Kaso for helping us with the front cover design for the Proceedings.

Finally, we both feel a deep debt of gratitude towards our families and friends who accepted our (temporarily) hermit-like existence and inability to go out and play as much as we would have liked, especially in the final months. Without their support, it would not have been possible for us to bring these Proceedings to you.

Alice Henderson and Anastazija Kirkova-Naskova
France and North Macedonia
August 2023

Henderson, A. & Kirkova-Naskova, A. (2023). Introduction: Extending the reach of English pronunciation issues and practices. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. iii–vii). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225005

Introduction: Extending the reach of English pronunciation issues and practices

Alice Henderson
Université Grenoble-Alpes

Anastazija Kirkova-Naskova
Ss. Cyril and Methodius University, Skopje

English Pronunciation: Issues & Practices is an international bi-annual conference devoted to how English pronunciation is taught and learnt, together with the associated scientific, social, and pedagogical issues. The central goal of EPIP conferences is to bring together teachers and researchers at all stages of their careers, from undergraduate students to well-established and internationally renowned scholars.

These proceedings provide extended accounts of selected oral presentations, posters, and workshops from EPIP7,[1] held in 2022 at the Université Grenoble-Alpes in France. As our aim was to bring out people's best work and to help authors at the beginning of their career paths, we asked reviewers to provide constructive criticism, as well as analyses of each paper's strengths and weaknesses. Although the double-blind peer-review process resulted in the rejection of some contributions, the ones included here cover a range of current research themes, as shown by the words with at least 41 occurrences in the proceedings (Table 1):[2]

Table 1

Number of Occurrences (#) of Keywords in Compiled Proceedings (82,756 tokens; 6513 types)

# – Keyword	# – Keyword	# – Keyword	# – Keyword
554 pronunciation	136 vowels (pl.)	79 instruction	55 sentences
552 English	127 target	74 perception	54 errors
432 learners (pl.)	125 accent	71 orthographic	54 listeners
288 participants	102 learner	69 British	51 transcription
243 speech	99 French	66 intonation	49 syllables (pl.)

[1] Earlier EPIP conferences were held in Skopje, North Macedonia (2019), Caen, France (2017), Prague, Czech Republic (2015), Murcia, Spain (2013), Grahamstown, South Africa (2011), after being launched in Chambéry, France (2009).

[2] These results were extracted from the compiled Proceedings using the software AntConc (Anthony, 2022).

# – Keyword	# – Keyword	# – Keyword	# – Keyword
229 teaching	94 tasks	66 syllable	48 comprehensibility
214 vowel	91 intelligibility	64 Czech	45 prosody
194 teachers (pl.)	90 accents (pl.)	64 pointing	45 rising
183 learning	90 teacher	62 awareness	43 contrasts (pl.)
179 native	89 training	61 accented	42 gestures
176 language	88 sound	60 feedback	41 Spanish
150 stress	85 input	60 proficiency	
141 production	85 phonological	59 MOOC	

Note. The words followed by (pl.) also occur in the singular form.

The search terms are merely the keywords suggested by authors or terms which seemed important to the editors, so the results do not provide a representative picture of the field of L2 pronunciation research. Nevertheless, it will be interesting to see how such results change over time, and having online proceedings for EPIP makes this sort of text query much more accessible. For example, will MOOCs and gestures appear more frequently as objects of study? Will the interest in researching orthography, strategies, intelligibility or comprehensibility endure? And which populations or aspects of spoken English will dominate future studies?

As editors, we wanted to make sure that EPIP proceedings are accessible to as wide and varied a readership as possible. We also felt that it was an ideal time for the EPIP community to benefit from a new format, rather than the printed books published following previous editions of EPIP (Henderson, 2010; Kirkova-Naskova et al., 2021; Mompeán & Fouz-González, 2015; Volin & Skarnitzl, 2018). Although these books were fine accomplishments and remain tremendously useful, each one took up to 3 years to produce and was often relegated to dusty shelves. Our challenge was, therefore, to find a way of sharing research more quickly and more openly. A major source of inspiration in this quest was the online PSLLT archives[3] at Iowa State University. These archives are a recognised source of stimulating, often-cutting edge work in our field and many of us frequently browse them to update our knowledge, even if we were able to attend the actual conference in North America. It therefore seemed logical to publish proceedings to EPIP conferences online and thereby create another reliable location for publishing and sharing research freely.

One means of presenting research online is to create a website, another is to extend a conference website and host the contributions within it. We discussed the available options with the reference librarians at Université Grenoble-Alpes, and they argued that the best way to promote our field in general, as well as the individual authors and the LIDILEM[4] research group, was to opt for Open Access and to procure Digital Object Identifiers (DOI) for each text. DOIs make it easier to retrieve documents and these remain accessible over the long term, because the identifier is not linked to a conference or research group website. Moreover, some countries give more value to documents with a DOI when assessing public research institutions, so DOIs are also strategically important for many researchers' careers. Working in Europe, and given the urgent need to promote the values of freedom, peace, social justice, scientific progress, and cultural and linguistic diversity, in line with European Union aims,[5] we felt it was important to obtain the DOIs from a European institution, rather than from a commercial

[3] https://apling.engl.iastate.edu/conferences/pronunciation-in-second-language-learning-and-teaching-conference/psllt-archive/
[4] https://lidilem.univ-grenoble-alpes.fr/
[5] https://european-union.europa.eu/principles-countries-history/principles-and-values/aims-and-values_en

service. To our surprise, we learned that we could obtain DOIs from a repository called Zenodo, "a memory institution for particle physics"[6] founded in 2013 at CERN[7] and originally intended for European Community funded research. Zenodo's mindset is perfectly in sync with ours at EPIP:

> To fully understand and reproduce research performed by others, it is necessary to have all the details. In the digital age, that means all the digital artefacts, which are all welcomed in Zenodo. […] Quite literally we wish there to be no reason for researchers not to share! (Zenodo, About section, para. 7)

Zenodo is one manifestation of the international open access and open data movement, which we fully support. However, we also wanted to have an online space where we could assemble the contributions as a recognisable object. We thus chose the open science archive created by France's ministry for research and higher education. Known as HAL, the archive ensures that all uploaded documents are "well referenced by search engines and interconnected with other services (ORCID, preprint servers)."[8] Each time a reader opens or downloads a text via HAL, their actions are cross-referenced. This raises the online profiles of researchers and, more broadly, bolsters the field's representation in citation and abstract databases.

Investigating the technicalities of Open Science, DOIs, and Creative Commons licenses is one way we have tried to show respect for our contributors. Authors' trust and reviewers' generosity must be encouraged and facilitated, as they are crucial to all of us in the field and essential in preventing the commodification of academia. Indeed, many of us have been contacted by journals or publishers who will publish research work, but only for a fee. By supporting Open Science, we can foster the ideal of sharing our work widely without such fees, but it is up to us to remain our own gatekeepers while ensuring (and raising) the quality of our work. The entire field can benefit from this, but it requires us to engage in rigorous, double-blind, peer review.

The rest of this introduction briefly summarises the 23 contributions included in these proceedings and reflects the diversity of issues, approaches, and contexts. The Table of Contents lists contributions in author-alphabetical order, as we chose not to group them into sections (thematically or based on the format of the conference contribution). Because all the contributions start with a detailed abstract, the following summary presents them in relation to three aspects which often guide how we read proceedings. These aspects are not always self-evident from titles or abstracts, yet may inspire readers to seek deeper contact with each contribution:

- *contextual*: Where are the researchers working? What type of participants are involved?
- *methodological*: Is the work quantitative, qualitative, or mixed-methods – and if the latter, which methods and tools have been combined?
- *pedagogical*: Is a transfer to teaching made explicit or implied?

First, the research contexts are predominantly European, reflecting the fact that all EPIPs but one have been hosted in a European country. One contribution is about Ecuadorian students and another involves listeners from 81 countries, but otherwise the participants are all European – ranging from English teachers in Britain, to users of Catalan and Spanish, Czech, Finnish, French, Italian, Macedonian, or Polish, in their respective countries. Participants are primarily

[6] https://about.zenodo.org/
[7] https://www.home.cern/
[8] https://hal.science/

in the 19 to 25-year-old bracket and majoring in English at university, although two studies focus on teenagers in high school (Mora-Plaza; Galimberti et al.). Proficiency levels vary from intermediate to advanced (B1–C2). In future we hope to see work from other contexts and countries, and with learners who are younger or older than the average here, especially as research into ageing and language learning gains momentum. Five contributions focus on future or in-service teachers, addressing teacher cognition (Čtvrtečková et al.), degree of pedagogical content knowledge (Červinková Poesová), issues related to accent and models (Baratta), choice of teaching/learning paradigm (Messum & Young), and the value of pointing on charts (Young & Messum). Four contributions explore technology – phone apps (Coulange), a website with a real-time 3D spectrogram (Edensor-Costille), a plugin for visualising prosody (Herment), and a MOOC (Rupp et al.) – and are primarily aimed at adults, even though these tools can be accessed by people of any age and from anywhere on Earth with a good Internet connection.

The widest diversity is found in relation to methodology and tools, which is a sign of a dynamic field that brings together researchers from many academic and professional backgrounds. Mixed methods seem to be well-established, with more than half of the contributions combining qualitative and quantitative approaches. For example, numerous contributions include acoustic analyses of learner productions and combine them with questionnaire data, and/or interviews, and/or Likert-scale assessments. One exploits learner diaries of strategy use (Duckinoska-Mihajlovska & Kirkova-Naskova), and another taps into peer teaching observations and reflections, lesson plans, and feedback provided by the teacher trainer (Červinková Poesová). Such combinations enrich the explanatory power of discussion sections and provide new insights. Two contributions draw on spoken corpus data, one from tandem conversations (Horgues & Scheuer), and one from both read-aloud and spontaneous speech (Herment). One contribution tackles the methodological issues of measuring intelligibility with extremely diverse listener populations (Thir), while another provides an honest, semi-narrative account of how difficult it can be for a multinational team to agree on measures (Martin-Rubió). We are delighted to be able to include a longitudinal study (Pesantez & Dellwo), as we need to know more about what happens over the long term – what is retained and what fades, at what rate, and by whom.

Finally, it is encouraging that so many contributors address pedagogical implications or applications, as we did not require contributors to include a detailed section on this.[9] Eight intervention studies explore issues such as word stress and strategy instruction (Duckinoska-Mihajlovska & Kirkova-Naskova); input types (hand gestures, Xi et al.; audio-synchronised textual enhancement, Galimberti et al.; written forms, Mouquet & Mairano); task-based pronunciation teaching (Mora-Plaza); using lyrics to promote phonetic awareness (Nowacka); using a web-based tool (Edensor-Costille); and promoting phonological self-awareness (Kivistö-De Souza & Lintunen). These constitute solid responses to the need for more research on teaching interventions, to determine what works. Twelve contributions either imply or make clear recommendations on the following: how to analyse apps and the need for more collaboration between teachers, researchers, and engineers (Coulange); how technology can promote social, cultural, and linguistic inclusion (Rupp et al.); the need for teachers to be aware of the paradigms they adopt or set aside (Messum & Young); how reflecting on teaching practices and contexts can be empowering (Baratta; Kirkova-Naskova); the importance of individual differences in learning trajectories (Pesantez & Dellwo) and of multimodal pronunciation teaching strategies (Xi et al.); the importance of not trying to cover too much (Nowacka); the need to take into consideration the orthographic systems available to learners

[9] This was obligatory for the book which followed EPIP 6 (Kirkova-Naskova et al., 2021). Similarly, the title of a recent book from a related conference – *Accents*, founded by Prof. Ewa Waniek-Klimczak at the University of Łódź, Poland – showcases pedagogical implications (Sardegna & Jarosz, 2023).

(Mouquet & Mairano); and, finally, the influence of learners' proficiency levels when choosing how to broach intonation (Herment), vowels in nasalising contexts (Sanvicente et al.), or connected speech phenomena (Kalvodová & Skarnitzl).

To conclude, we believe that readers from many different horizons will find something of interest in these proceedings. They may discover new authors to follow in the future, or even discover work by established researchers of which they were unaware. Hopefully delving into the proceedings will be as enjoyable as the experience of editing them.

References

Anthony, L. (2022). AntConc (Version 4.2.0) [Computer Software].

Henderson, A. (Ed.). (2010). *English Pronunciation: Issues and Practices (EPIP): Proceedings of the first international conference*. Université de Savoie.

Kirkova-Naskova, A., Henderson, A., & Fouz-González, J. (Eds.). (2021). *English pronunciation instruction: Research-based insights*. John Benjamins.

Mompeán, J. A., & Fouz-González, J. (Eds.). (2015). *Investigating English pronunciation: Trends and directions*. Palgrave Macmillan.

Sardegna, V., & Jarosz, A. (Eds.). (2023). *English pronunciation teaching: Theory, practice and research findings*. Multilingual Matters.

Volin, J., & Skarnitzl, R. (Eds.). (2018). *The pronunciation of English by speakers of other languages*. Cambridge Scholars Publishing.

About the authors

Alice Henderson is an Associate Professor at Université Grenoble - Alpes, France where she teaches English for Specific Purposes to Science & Technology students. She taught English phonetics and phonology to English majors for 24 years and has been involved in training teachers in France, Norway, Poland, and Spain. In 2009 she initiated the international bi-annual conference *English Pronunciation: Issues & Practices*. Her research interests include English pronunciation teaching and learning, the perception of foreign-accented speech, and English Medium Instruction (EMI). Much of her research has focused on speakers, but she is now intrigued by listeners' roles, from an intercultural and sociolinguistic perspective.

Email: alice.henderson@univ-grenoble-alpes.fr

Anastazija Kirkova-Naskova is an Associate Professor in English phonetics and morphology in the Department of English Language and Literature, Ss. Cyril and Methodius University in Skopje, North Macedonia. She also teaches research methodology (MA level). Her research interests include English pronunciation teaching and learning, foreign-accented speech, speech perception, pronunciation learning strategies, and teacher cognition. Together with Alice Henderson and Jonás Fouz-González, she co-edited *English pronunciation instruction: Research-based insights* (John Benjamins, 2021). She was an assistant editor-in-chief for *The Journal of Contemporary Philology*.

Email: akirkova@flf.ukim.edu.mk

Baratta, A. (2023). Regional accents are OK for teaching but not *too* regional: A discussion of accent preference in British teaching. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 1–10). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8129241

Regional accents are OK for teaching but not *too* regional: A discussion of accent preference in British teaching

Alex Baratta
University of Manchester

Accent can still be a contentious topic in Britain, with accent-based preference, if not prejudice, in turn reflecting class-based prejudice. Thus, accent in Britain often functions as a linguistic proxy for class and as such, negativity ascribed to accents judged to be working class still exists. This paper reflects the views of two British teachers from a larger sample within the primary and secondary sectors of teaching, seeking to better understand the role that accent plays in the British teaching profession. In doing so, it addresses the central question: what are the implications for teachers with broad accents in a profession that champions equality, yet set against the societal reality of accent-based prejudice? From these two teachers' accounts, the results show that accents perceived as broad and/or too 'regional' are those which are targeted for modification by senior staff. However, the two teachers have different attitudes, with one agreeing to the modification, and the other in complete disagreement. From these two accounts, we can glean an insight, however small, into the role that accent plays in one's professional identity as a teacher, in a country where accent continues to be of particular relevance.

Keywords: accent, teaching, accentism, class, professionalism

1 Introduction

Accent-based prejudice is hardly a unique phenomenon. In some cases, it can be reflected in banter, perhaps not intended to offend. However, it can lead to negative judgements made of individuals, with accompanying stereotypes linked to their supposed intelligence, trustworthiness and friendliness (Baratta, 2021; Hughes et al., 2012; Mugglestone, 2003).

This has relevance for the teaching profession, in which a teacher's voice, and by implication accent, is of importance. If we consider the British context of teaching in particular, at both the primary and secondary levels, then an issue is the class-consciousness which is still present in British society (Rickett, 2023). Putting the two points together, what are the implications for British teachers who have accents judged to be working-class, accents often referred to as broad varieties (Baratta, 2021)? What might the implications be for these teachers, whose accents are stigmatised in society? That is, if these negative judgements are just as relevant in the classroom as they are on the city streets, then are teachers expected to modify their accents in some way? If so, the modification might be self-selected by the teachers, and responded to with an overall objective attitude – in principle, no different than choosing formal attire for teaching as opposed to wearing, for example, pyjamas. However, we need to consider the implications for teachers who do not agree with modifying their 'home accents', as it were, despite directives from senior staff and/or mentors.[1] When we consider the need for 'professionalism' in teaching, what might this mean on a linguistic level? Such a question is largely unaddressed in the context of British teaching, with the focus within this paper addressing this gap, and for the following reasons.

First, given the sheer diversity of accents in Britain, a study within British teaching would be useful. This is because such linguistic diversity might be represented within a profession in which one's accent – notably for phonics teaching – is relevant. Granted, while there are only two teachers whose accounts are presented within this paper, their feedback derives from larger studies which involved, or focused on, teachers' accents in British teaching (Baratta, 2016, 2018). These are, to my best knowledge, the only studies which have focused solely and exclusively on the British teaching profession, though Sharma et al. (2020) have further discussed the role of accent in professional contexts overall in Britain, discovering that accent-based prejudice is a live issue. Second, as mentioned, what might the implications be for stigmatised accents within the teaching profession? Do negative societal connotations apply to teachers' accents? And yet should this be the case, for a profession in which equality is surely encouraged and for teachers who are otherwise fully qualified to teach? Finally, in instances of accent modification, what are the specific phonological targets teachers are expected to aspire to, as part of what might be considered a professional – if not standard – accent?

2 Previous research on accentism

A dislike of certain accents can lead to negative judgements made of the speakers, based on accents which are perceived to lack societal prestige and/or attractiveness (see Coupland & Bishop, 2007). Accentism can lead to action, such as people having their job threatened, or not being hired, based on their accent (Lippi-Green, 1997; Udavant, 2020), and there are implications for credibility within the courtroom based on the accents of witnesses (Cantone et al., 2019).

In the British context, accentism has an arguably long history, based in large part on class-based assumptions. The accent referred to as Received Pronunciation (RP) is one that has

[1] In British teacher training, a mentor is responsible for guiding the trainee teacher throughout their course, offering feedback and advice, for example.

existed as a prestige accent (Snell & Andrews, 2016; Trudgill, 2002) and exists in Britain, and certainly England, as a class-based accent. Its associations with the upper-middle, and upper-class, reflects its prestige, given positive associations and connotations connected with the upper-class. RP is not a linguistic monolith, however, and comes in several varieties, with Wells (1982, p. 280) discussing U-RP, adoptive RP and mainstream RP, with U-RP having connotations of 'an elderly Oxbridge don' or 'a jolly-hockey-sticks schoolmistress at an expensive private girls' school'. More recently, Cruttenden (2014) references RP as having been influenced by more regional accents such as Cockney, which, to some extent, might reflect Estuary English (Przedlacka, 2001). While current associations of RP, certainly the more conservative varieties (e.g., U-RP), might reflect snobbery (Hughes et al., 2012), and even be regarded as outdated (Lindsey, 2019), RP nonetheless maintains, in all its varieties, connotations of upper-class status and with this, specific images that can reflect education and sophistication (Addison & Mountford, 2015), for example. Given that RP exists as a class-based accent, this in turn implies a lack of *regional*-based phonology. Herein is arguably a key to the prestige of RP – it can be spoken from Newcastle to London, and everywhere in between. By removing regional features, it frees speakers from being tied to a given city region, which subsequently means that the stereotypes associated with inner cities in Britain are also removed (Baratta, 2021).

This leads to the need to clarify what is meant by terms such as broad in relation to accents. Such a term reflects societal labels in Britain, often used to describe people in terms of indeed having broad/strong accents (as well as mild accents) (Baratta, 2018). The opposite end of the accent spectrum involves individuals being viewed as *not* having an accent, which simply means that regional sounds – those that could immediately tie the speaker to a given region, whether on a city- or local-level – are absent (Baranowski & Turton, 2015; Cardoso et al., 2019; Ramsaran, 2015; Strycharczuk et al., 2020). The implications for such neutral accents in Britain are that they – and the speakers – are perceived as middle-class (Cardoso et al., 2019; Strycharczuk et al., 2020). From here, associations connected with positive connotations, such as social mobility (Donnelly et al., 2019; Friedman, 2016), can be derived. In this manner, we are essentially dealing with a linguistic means to signal social mobility in the workplace by means of avoiding accents which might, in society's view, suggest the opposite. There is indeed evidence that noticeably regional accents can have negative implications for the workplace (see Sharma et al, 2020), including contexts such as finance (Moore et al., 2016), library work (Lippi-Green, 1997), and banking in Britain (The Sutton Trust, 2017).

In terms of the implications for accent within the British teaching profession, information is limited. Garner (2013) reports on an OFSTED[2] inspection in which a teacher from Cumbria (in Northern England), but teaching in the South of England, was told at an inspection to sound 'less Northern'. Addison and Mountford (2015) investigate the speech of both teachers and students in higher education settings in England, notably in terms of class perceptions and the ways in which broad regional accents are perceived to be disfavoured in higher education, lacking social capital, and the overall perception of a need to 'fit in', with regional accents sometimes perceived as not being an appropriate fit. In terms of the particular phonemes which serve as linguistic examples of broad accents, examples are now provided. To illustrate, words such as *back, face*, and *baby*, if respectively realised with broad Liverpool, Yorkshire and Newcastle accents, would sound thus [bax], [fe:s], and [bɪəbi]. Such identifiable features are referred to as phonological giveaways (Baratta, 2021), or as salient features (Labov, 1972). But regardless of who is doing the listening, if individuals seek to present an image of being upwardly mobile, such broad accents might be perceived incompatible for certain professions.

[2] OFSTED stands for Office for Standards in Education, Children's Services and Skills, an organisation in Britain dedicated to maintaining (and improving where necessary) standards for schools and teaching.

Likewise, it is accents that are perceived, to some extent, as being regionless (and simultaneously not reflecting a working-class background), that are arguably perceived as standard (Bibby et al., 2017).

3 Research question

This research is guided by the question: what are the implications for two British trainee teachers of having a broad regional accent?

4 Research methodology

I conducted three studies from 2014 – 2015. My first study (Baratta, 2016) obtained the views of 92 British participants from a range of locations and professions, using a questionnaire to explore their accents in terms of: a) how accent defined them and contributed to their personal identity; b) instances in which they had chosen to modify their accent, or been told to, and the rationale for such modification; and c) how accent modification impacted on their personal identity, if at all. Six of the responses came from teachers and the nature of their responses led to the following two studies, which focused solely on teachers in Britain. Study two interviewed 11 trainee teachers from two Northern English universities, with study three again using a questionnaire to obtain the views of 15 trainee teachers from two Southern English universities. For both studies two and three (Baratta, 2018), the questions remained the same, but with a questionnaire being used for study three. This was a necessary compromise for data collection, which otherwise would have involved multiple instances of travel and expense.

A notable departure from study one in terms of focus is that studies two and three did not focus on identity but instead focused solely on instances in which teachers had been told to modify their accents as part of their teaching role. From here, rather than ask participants to explain how modification impacted on personal identity, I instead focused on something more concrete – the teachers' immediate feelings on the matter. These three studies provided me with a great number of accents, including from the Republic of Ireland, Northern Ireland and Scotland, though the majority of accents were English.

Out of a total of 32 teachers, only two teachers' accounts have been provided in this paper; the first teacher is from study two, self-identified as having a Huddersfield accent; the second teacher is from study three, self-identified as having a South London accent (see Table 1 below, which provides an overview of each study, in terms of identifying key aspects such as the teachers' accents, and whether they were told to modify or not). The two teachers will be identified in the paper as T1 and T2, respectively. This decision to focus on just two teachers is first and foremost based on purely practical concerns, as the kind of qualitative depth this paper seeks to provide within the word limits cannot be accomplished for larger numbers of teachers' accounts. Second, the two accounts have been selected as they illustrate the mentors' rationale for accent modification for all the teachers in studies two and three (Baratta, 2018) – the need to be understood. Again, this does not suggest, nor can it imply, that this is a nationwide issue. Nor can the accounts of just two teachers hope to speak for other teachers who have been provided with accent-based guidance or directives. But what can hopefully be gleaned from these two accounts are the ways in which accent is relevant in the context of teaching, more so when teachers are in training to become fully-fledged teachers and enter the profession; as such, what is deemed to be a 'professional' accent?

Finally, it is acknowledged that the presentation of these two accounts might seem one-sided. That is, we have two teachers who were told by mentors to modify their accents during training, yet there are thirteen teachers from the three studies (Baratta, 2016, 2018) who were *not* given accent-based guidance or directives. However, there are two points to mention in this

Table 1

Presentation of Teachers' Backgrounds

Study #	Self-identified accent	Teaching level	Asked to modify language use?
Study 1	Rosendale	Primary	Yes
	Manchester	Primary	No
	Glaswegian	University	No
	Stockport	University	No
	Barnsley	University	No
	Rochdale	EFL, secondary	Yes
Study 2	Huddersfield	Primary	Yes
	Stoke-onTrent	Primary	Yes
	Manchester	Primary	Yes
	Broad Manchester	Primary	Yes
	Derbyshire - Yorkshire mix	Primary	Yes
	Warrington, so a Manchester-Liverpool mix	Primary	No
	Nottingham	Primary	Yes
	Rochdale	Primary	Yes
	'Not quite posh enough for RP' (the teacher is from Portsmouth)	Secondary	No
	Liverpool	Secondary	Yes
	Derry, Northern Ireland	Secondary	Yes
Study 3	Midlands	Primary	Yes
	Medway	Primary	Yes
	South London, quite strong	Secondary	Yes
	Irish, Dublin	Secondary	No
	Estuary English	Secondary	No
	RP/Lancastrian	Primary	No
	'Standard English'	Secondary	No
	A mixture of Cockney and Irish (Republic of Ireland)	Secondary	Yes
	Mostly RP with some Estuary English	Primary	Yes
	A happy medium between Estuary English and RP	Primary	No
	Mild Belfast	Secondary	No
	Southern English	Primary	No
	Well-spoken Northern Irish	Primary	Yes
	Irish, but not strong	Primary	Yes
	Newcastle	Secondary	Yes

regard. First, it is arguably the case that accent-based guidance provides us with a clearer picture in terms of what the particular issues might be in the minds of mentors – and the teachers themselves. The act of providing guidelines for teachers' accents suggests that the accents are, to some extent, judged to be problematic. Once this is explained in phonological terms, then we have a more precise discussion as to what the salient features in question might be for

certain accents and more broadly, for teaching; this is something currently lacking in the specific context of the British teaching profession. Second, the two accounts provided here nonetheless offer variety in that we have two teachers provided with accent directives, who reacted to the guidance in very different ways, based in large part on how these two individuals understood what is meant by the development and projection of a 'professional' teacher identity.

5 Data analysis and results

The first account is from T1, an individual from Huddersfield in the Northern English county of Yorkshire, who teaches at the primary level. The second account comes from T2, a secondary Art teacher from Croydon, South London, who described her accent as strong South London. Both teachers also identified as being working class, with their accents arguably a reflection of such.

T1 was told to modify his accent by his mentor in specific phonological terms that indeed involved removal of the more salient features – the phonological giveaways tied to his region of origin. T1 illustrated this with the expression *go home*, in which, as he put it, the 'teacher me' would realise this expression as [goʊ hoʊm]. This still allows for an overall Yorkshire accent to be used for teaching, but a variety that is "shined up a bit" (Baratta, 2018, p. 129), as T1 explained. T1 then referenced his self-described 'home me', which would realise the expression thus [gə hə:m]; the use of the schwa vowel in place of the long [o] sound is clearly identifiable to Yorkshire (Finnegan, 2015). T1 also referenced removing glottal stops from his speech, a feature often thought of in connection with Cockney speech, but now found throughout England (Baranowski & Turton, 2015). However, T1 was in complete agreement with the need to modify his accent and regarded the directive in largely neutral terms. As he explained, there is a constant need to make students "enunciate and speak properly" (Baratta, 2018, p. 127). While T1 also explained that there is a "richness to be had" (p. 128) regarding the variety of accents in Britain, his view regarding accent within teaching is that teachers need to be understood by their students and in turn, students need to be understood in society. Thus, T1 explained the need to speak with a less regional accent as a means to be better understood by his students but also, to act as a linguistic role model of sorts, so that the students would in turn emulate his speech. T1 even felt that his students respected him based on his teacher accent but also based on his own instruction to students to adjust their accents, such as avoiding glottal stops. Given that T1 was teaching in his home region during his training placement, it seems unlikely that a broad accent would be difficult to understand. But if we consider the wider societal considerations, then a broad Yorkshire accent might, outside this region, be less understood. Though speculative, it may well be that T1 is exhorting his students – by self-demonstration – to avoid broad accents so as to avoid negative stereotyping for their future. This reflects the views of a primary school principal in England as reported in several British news outlets, who instructed his students to avoid 10 expressions which were effectively banned from the classroom, some of which were accent-based (e.g., *somefink*), as a means to create a more professional persona for their future careers (Fricker, 2013). That T1 used the word 'proper' in relation to his, and students' speech, is telling. Notions of 'proper' speech can be a controversial issue, but for T1, and the school principal referenced, this idea is tied to avoidance of specific regional sounds to create more positive perceptions in the future workplace.

T2 explained that she was "very proud" (Baratta, 2018, p. 156) to be from Croydon, and as such, she was proud of her accent as a phonological symbol of her regional – and class-based – origins, acknowledging that Croydon is referred to as "the ghetto" (p. 156). As with T1, she took pride in her origins associated with a specific part of a city; thus, we can see the

intersection of class and region (Donnelly et al., 2022). For T2, her accent served to benefit her teaching, as she explained that she can "easily relate to students…because I sound more like them" (p. 156). Thus, an accent shared by teachers and students, especially if it is a stigmatised accent, can serve to build a strong rapport, as T2 believes. T2's mentor was somewhat of an outlier, in that in addition to referencing the need to be understood, he had also told T2 that she needed to use an accent deemed 'professional' for teaching. As a phonological example of an 'unprofessional' accent by implication, T2's mentor instructed her to write the word *water* with a capital < t >, as in *waTer*, in order to remind her to avoid the glottal stop. Here, then, is a clear phonological giveaway, one which serves to signal pride in T2's regional and class origins, but signals unprofessionalism, at least in speech, according to her mentor. T2 further explained that her mentor "was very patronizing and tried everything to change my accent" (p. 155), causing T2 to feel that she "was never good enough" (p. 155). For T2, however, her accent *is* professional, because she believed it helped her students to approach her and perceive her more positively.

6 Discussion of results and key findings

Findings based on the discussion of just two individuals do not constitute a firm base on which to generalise to the wider population of British teachers, nor do the full results from 32 British teachers (Baratta, 2016, 2018). However, the purpose of these studies was never to make generalisations in the first instance but rather, to simply investigate the role that accent plays in British teaching, whether positive, negative or neutral, and from here, determine what insights, if any, can be gleaned from the sample. On that basis, the results suggest that for these two teachers, their regional accents per se are not necessarily proscribed for teaching; it is the broad nature of their accents that appears to be the issue. This reference to broad accents (and by implication, neutral or mild accents) is arguably reflective of terms used by people within society, as well as being referenced in studies focused on accent (Baratta, 2018; Strycharczuk et al., 2020). The phonological implications are a removal of the more readily distinctive sounds, those that identify an individual immediately to a region, be it a city region or indeed a locality within a city, and with this, suggested working-class origins. The *societal* implications of such modification are that individuals are arguably perceived better, precisely because their modified accents give less clues to regional – or class-based – origin.

Though speculative, it might be the case that within the teaching profession, and other professions connected to social mobility, individuals might choose to lessen their broad accents (Baratta, 2018; Donnelly et al., 2019, 2022; Sharma et al., 2020). This in turn might signal to interlocutors – in this case mentors – that such individuals seek to present a more 'fitting' image for the context. Phonologically-speaking, this means modifying accents which suggest working-class status, precisely by removing the more salient features, or, phonological giveaways. The need to 'fit in' and develop linguistic capital as a means to do so was clearly illustrated in the study of Addison and Mountford (2015), which was focused on higher education. Within my paper, while there was no specific discussion from mentors or teachers regarding 'fitting in' or developing more linguistic capital, it is strongly suggested that these notions are implied, given the directives for 'professional' language, a point over which T1 and T2 disagree.

Overall, however, there is little linguistic clarity for teachers in Britain on any official level, beyond the need to use standard English (which can be spoken using any accent). The linguistic guidance that was provided to trainee teachers in particular within the entirety of my studies appears to be based, officially, on mentors' desire to ensure the teachers they train are understood by their students (Baratta, 2016, 2018). Again, however, by implication students are being prepared for the world beyond school, and thus accent intelligibility, as demonstrated

by their teachers, might be regarded as code for speaking with a less stigmatised accent. As Addison and Mountford (2015) make clear, "being able to 'talk the talk' can have a significant return on investment in this game of worth because of how talking and accent is coded with symbolic value" (p. 10). While their study provided more depth and specificity in terms of class perceptions within higher education, my study has sought, in part, to provide phonological clarity regarding accents perceived as working-class, those perceived to be notably regional.

7 Conclusion

Based on the results from a sample of merely two teachers, there is no room to generalise. What can be stated, however, is that we have an insight into the experiences of these two teachers in particular, even though they are teaching different subjects, at different levels and using different accents. The commonality is that they both have accents self-described as 'strong' and 'working-class'. Thus, rather than generalise, I instead seek to merely ask whether or not the experiences of these two teachers might be similar for others with broad accents? While this appears to be the case when taking into consideration the results of my overall studies on this matter (Baratta, 2016, 2018), only a large-scale study across the UK and involving a variety of taught subjects could give us a more comprehensive picture as to the reality when broadly-accented teachers enter the teaching profession. For the one mentor who explained to T2 the need for 'professional' speech, perhaps it could be argued that while this is somewhat of a blunt directive, it could be seen as nonetheless more honest – a reflection of what society, and even students' parents, might indeed believe. If so, this could mean that accents deemed compatible for British teachers are those which avoid the broader varieties and instead seek to sound more neutral, as has been described. The result would be that the negative connotations associated with such accents, often based on class assumptions are diminished, if not removed, and teachers are perceived to be more professional as a result.

However, in the absence of any specific or official accent-based guidance in Britain, teachers' accents may continue to be discussed between themselves and mentors during training. A joint discussion between teachers and mentors could be a useful starting point to initiate a conversation on this topic, leading to updated standards, to include a full description of what exactly the expected accent standards are. On the other hand, this might also appear prescriptive, which is arguably what some teachers wish to avoid regarding their accents. In either case, whether official or unofficial, this small-scale study at least implies that there are expectations for accent within British teaching beyond these two individuals. For some teachers, being told to modify their accent is regarded as an inconvenience at least, prejudicial at worst (Baratta, 2018). A vast array of accents is heard outside the school gates, so to include those accents within the classroom – including broad varieties – could help students with similar accents feel empowered, because 'sir/miss sounds like *me*'.

References

Addison, M., & Mountford, V. (2015). Talking the talk and fitting in: Troubling the practices of speaking and 'what you are worth' in higher education in the UK. *Sociological Research Online*, *20*(2), 1–13. https://doi.org/10.5153/sro.3575

Baranowski, M., & Turton, D. (2015). Manchester English. In R. Hickey (Ed.), *Researching Northern Englishes* (pp. 293–316). John Benjamins.

Baratta, A. (2016). Keeping it real or selling out: The effects of accent modification on personal identity. *Pragmatics and Society*, *7*(2), 291–319. https://doi.org/10.1075/ps.7.2.06bar

Baratta, A. (2018). *Accent and teacher identity in Britain: Linguistic favouritism and imposed identities*. Bloomsbury.

Baratta, A. (2021). Varieties of 'standard accents' among teachers in contemporary Britain. *World Englishes,* 1–16. Advance online publication. https://doi.org/10.111/weng.12561

Bibby, T., Lupton, R., & Raffo, C. (2017). *Responding to poverty and disadvantage in schools*. Palgrave Macmillan.

Cantone, J., Martinez, L., Esqueda, C., & Miller, T. (2019). Sounding guilty: How accent bias affects juror judgments of culpability. *Journal of Ethnicity in Criminal Justice*, *17*(3), 1–26. https://doi.org/10.1080/15377938.2019.1623963

Cardoso, A., Levon, E., Sharma, D., Watt, D., & Ye, Y. (2019). *Inter-speaker variation and the evaluation of British English accents in employment contexts* [Conference presentation]. *19th International Congress of Phonetic Sciences, Melbourne*, Australia 2019. Canberra: Australasian Speech Science and Technology Association Inc.

Coupland, N., & Bishop, H. (2007). Ideologised values for British accents. *Journal of Sociolinguistics*, *11*(1), 74–93. https://doi.org/10.1111/j.1467-9841.2007.00311.x

Cruttenden, A. (2014). *Gimson's pronunciation of English*. Penguin.

Donnelly, M., Baratta, A., & Gamsu, S. (2019). A sociolinguistic perspective on accent and social mobility in the UK teaching profession. *Sociological Research Online*, *24*(4), 496–513. https://doi.org/10.1177/1360780418816335

Donnelly, M., Gamsu, S., & Baratta, A. (2022). Accent and the manifestation of spatialised class structure. *The Sociological Review, 70*(6), 1100–1118. https://doi.org/10.1177/00380261221076188

Finnegan, K. (2015). Sheffield. In R. Hickey (Ed.), *Researching Northern English* (pp. 227–250). John Benjamins.

Friedman, S. (2016). Habitus clivé and the emotional imprint of social mobility. *The Sociological Review*, *64*(1), 129–147. https://doi.org/10.1111/1467-954X.12280

Fricker, M. (2013, November 14). Yow cor spaek lyuke that! West Midlands school bans pupils from using 'damaging' regional slang. *The Mirror.* http://mirror.co.uk/news/uk-news/black-country-slang-banned-halesowen-2791066.

Garner, R. (2013, November 18). Teacher 'told to sound less Northern' after Southern Ofsted inspection. *The Independent*. https://www.independent.co.uk/news/uk/home-news/teacher-told-sound-less-northern-after-southern-ofsted-inspection-8947332.html

Hughes, A., Trudgill, P., & Watt, D. (2012). *English accents and dialects*. Hodder.

Labov, W. (1972). *Sociolinguistic patterns*. Blackwell.

Lindsey, G. (2019). *English after RP: Standard British pronunciation today*. Palgrave Macmillan.

Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. Routledge.

Moore, J., Higham, L., Mountford-Zimdars, A., Ashley, L., Birkett, H., Duberley, J., & Kenny, E. (2016). *Socio-economic diversity in life sciences and investment banking*. Social Mobility Commission.

Mugglestone, L. (2003). *Talking proper: The rise of accent as social symbol*. Clarendon Press.

Przedlacka, J. (2001). Estuary English and RP: Some recent findings. *Studia Anglica Posnaniensia*, *36*, 35–50.

Ramsaran, S. (2015). RP: Fact and fiction. In S. Ramsaran (Ed.), *Studies in the pronunciation of English: A commemorative volume in honour of A. C. Gimson* (pp. 178–190). Routledge.

Rickett, B. (2023). "It's not what you say, it's how you say it." The discursive (un)stitching of classed (un)belonging in 'elite' ideals. [Presentation]. *Psychology of classed (un)belonging in 'elite' work Symposium.* University of Leeds, January 27th, 2023.

Sharma, D., Erez, E. Watt, D., & Perry, C. (2020). Accent bias in Britain. Project report, Queen Mary University of London & the University of York.

Snell, J., & Andrews, R. (2016). To what extent does a regional dialect and accent impact on the development of reading and writing skills? *Cambridge Journal of Education*, *47*(3), 1–17. https://doi.org/10.1080/0305764X.2016.1159660

Strycharczuk, P., López-Ibáñez, M., Brown, G., & Leemann, A. (2020). General Northern English: Exploring regional variation in the North of England with machine learning. *Frontiers in Artificial Intelligence*, *48*, 1–18. https://doi.org/10.3389/frai.2020.00048

The Sutton Trust (2017). Four-fifths of senior figures in finance think presentation at interview affects disadvantaged people's chances of getting a job. Sutton Trust press release.
Trudgill, P. (2002). *Sociolinguistic variation and change*. Edinburgh University Press.
Udavant, S. (2020, November 1). Stop discriminating against me because of my accent – how I talk doesn't affect my ability to do the job. *The Independent*. https://www.independent.co.uk/voices/accent-discrimination-work-job-prejudice-b1399561.html
Wells, J. (1982). *Accents of English: The British Isles*, Vol. 2. Cambridge University Press.

About the author

Alex Baratta is a sociolinguist, researching language and identity; linguistic prejudice; and linguistic rights within educational contexts. His research has included topics such as the ways in which certain accents are proscribed in British teacher training; the societal means of codifying expanding circle varieties of English, such as Korean English; and even the role that accent plays in the cinematic character of James Bond. Alex has presented his work globally at prestigious conferences in locations such as South Africa, Singapore and Australia, published in high-ranking journals including *World Englishes* and *Linguistics and Education*; and his upcoming book with Routledge (2024) is focused on the use of China English in Chinese students' academic writing.

Email: Alex.Baratta@manchester.ac.uk

Coulange, S. (2023). Computer-aided pronunciation training in 2022: When pedagogy struggles to catch up. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 11–22). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8137754

Computer-aided pronunciation training in 2022: When pedagogy struggles to catch up

Sylvain Coulange
Université Grenoble-Alpes

Numerous language learning applications and websites have recently introduced automatic pronunciation training features which are popular, particularly among young learners. However, language teachers may struggle to understand the capabilities, efficacy, and reliability of these programs. This chapter provides a critical overview of English pronunciation training features in mainstream applications such as DuoLingo, Memrise, Babbel, Busuu, Rosetta Stone, ELSA Speak, and IELTS Speaking Practice. It also surveys the rare studies investigating the diagnostic accuracy of these tools and whether learners make meaningful improvements by using them. We conclude that communication and collaboration need to develop among engineers, language teachers, linguists, pronunciation experts, and entrepreneurs to promote useful pedagogical outcomes for learners and to align with contemporary objectives in the field of English as a foreign language.

Keywords: CAPT, automatic pronunciation assessment, pronunciation training, language learning apps

1 Introduction

In today's globalized world, more than three quarters of English speakers are non-native (Walker et al., 2021). It is likely that a given learner of English will have more opportunities to speak with fellow non-native English speakers than native ones. Learning pronunciation by focusing on intelligibility and comprehensibility has become key for successful communication; what the listener understands and how much effort they have to make is seen as more important than how native the speaker sounds.

Although the concept of English as a Lingua Franca (ELF) became much more widespread after Jenkins' book was published (2000), only recently has there been a paradigm shift in terms of realistic goals for pronunciation training, from native-likeness to comprehensibility. For instance, Walker et al. (2021) in their Oxford University Press position paper insist on the necessity to be easily understood rather than to speak with native-like pronunciation. They suggest concentrating on very specific features to effectively improve pronunciation, such as vowel length and word stress, along with most consonant sounds. On the other hand, aspects such as vowel quality or pitch movement are not considered a priority for good comprehensibility, but rather as a means to get closer to a given native accent if that is the learner's goal (Walker et al., 2021). The 2020 update of the Common European Framework of Reference for Languages acknowledges that the former focus on native-speaker accent was "detrimental to the development of the teaching of pronunciation" (Council of Europe, 2020, p. 134). Influence from the L1 is now accepted at C2 level as long as it does not hinder comprehensibility. Furthermore, half of the updated CEFR phonological control scale is now devoted to prosodic features.

Assessment scales such as that of the CLES[1] Certification emphasize that at B2 level pronunciation and intonation should be "clear enough to be easily understood, even if an accent subsists", and that at B1 level pronunciation should be "globally understandable despite a foreign accent and/or pronunciation errors"[2]. Isaacs et al. (2018) designed a comprehensibility scale for English formative assessment, helping teachers identify the aspects of L2 that they should prioritize with their learners in order to promote the production of comprehensible English and also to help L2 learners develop awareness of their strengths and weaknesses. Their scale puts the emphasis on word stress and hesitation markers' position more than on phoneme quality, and much more than on vocabulary and grammar accuracy. Word stress is also considered a priority in Frost and O'Donnell's (2018) prosody-based descriptors for assessing oral production in English, along with intonation, unreduced vowel quality and connected speech.

With ever-advancing AI technology, it seems potentially achievable to integrate immediate feedback focusing on intelligibility features into pronunciation practice. This chapter focuses on technology made for pronunciation training in language applications (apps hereafter) and websites. Research on automated pronunciation error detection started in the early 1970s and gained in popularity in the late 1990s, but results were rather limited, and few programs made it to the commercialization stage (Witt, 2012). In the last decade however, with the ever-growing degree of computation power, the democratisation of smartphones, and enhanced speech recognition technologies, the most popular language learning apps and websites now integrate automated feedback for pronunciation training. These systems are used by a large number of people and are becoming more popular than ever, especially among young learners. However, how they function remains little known to students and teachers.

1 The CLES (*Certificat de Compétences en Langues de l'Enseignement Supérieur*) is a national, government certified test of language proficiency in France (see https://www.certification-cles.fr/).

2 The CLES Oral Interaction assessment grids are available at: https://www.certification-cles.fr/se-preparer/grilles-d-evaluation/grilles-d-evaluation-1196363.kjsp.

The following section of this chapter will give an overview of how recent mainstream CAPT tools tackle pronunciation practice. Then we will review some studies which have investigated the pedagogical efficacy of these tools, before discussing current systems limitations and suggesting potential solutions.

2 Pronunciation assessment in today's mainstream language apps

According to the website Top10.com[3], the five most widely used apps for learning languages in 2022 are DuoLingo, Memrise, Babbel, Busuu, and Rosetta Stone.[4] All of them provide automatic speaking evaluation tools as part of their contents. Apps and websites dedicated to English speaking practice, such as ELSA Speak or IELTS Speaking Practice, can also be found online. This section will describe how these apps tackle speaking production, going through their types of activities and how speech is elicited, as well as what kind of feedback is given to the learner. It is important to mention that apps constantly update, so this description covers the features available in spring 2022.

2.1 Types of activities, stimuli and focus

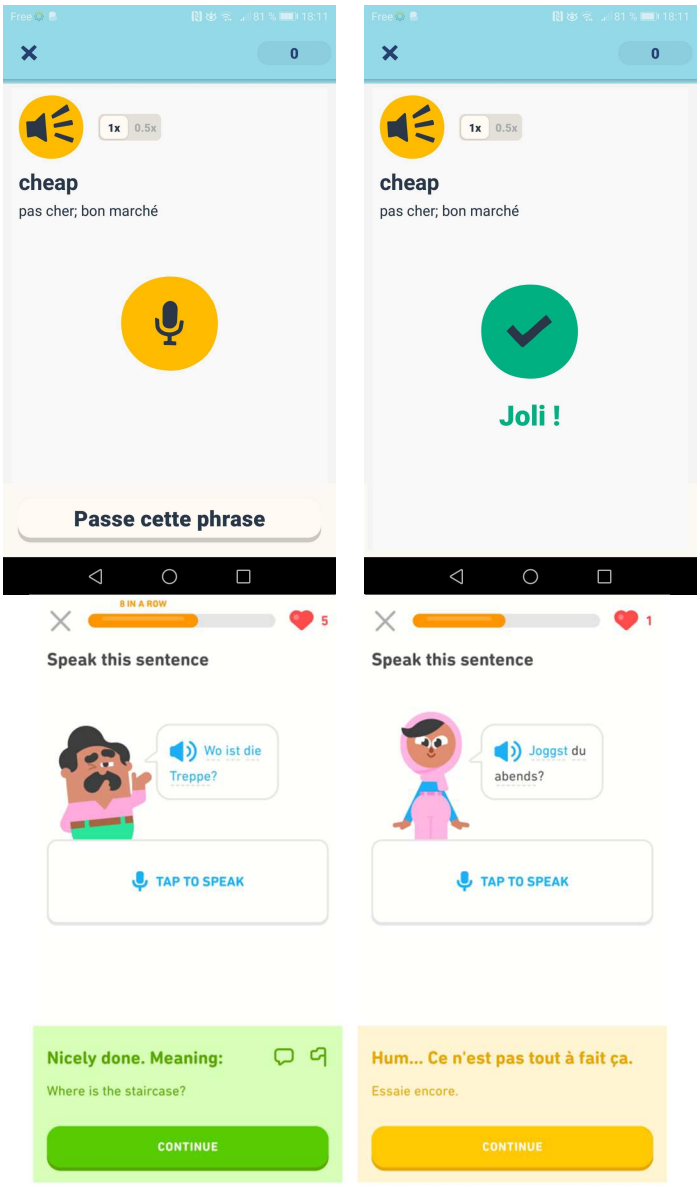
In 2022, the most frequent English-speaking practice activity is an interface with a series of words or short phrases on the screen to be read aloud while pressing a record button. Most of them also offer immediate feedback, as shown in Figure 1. In most cases, audio is also played over the written form when it appears. All apps cited above provide an audio model of what is to be said as well as the written form, except for Rosetta Stone, which sometimes does not display text until after the recording. Words and sentences are mostly practiced out of context, generated from a vocabulary bank or sentences from the current lesson, with increasing difficulty. Student productions judged to be incorrect may reappear later – and repeatedly –, until the program considers the pronunciation to be correct.

All apps presented here are limited to this same basic read-and-repeat type of exercise, except for Rosetta Stone which provides more contextualised activities; each unit ends with a complex story activity presenting a succession of pictures, telling a story which involves the learners in the first person. The text for each character appears on the screen with a picture and is read aloud. When their turn comes, the students must read aloud the text appearing on the screen or speak spontaneously according to the illustration displayed. Rosetta Stone also provides another type of exercise in which the learners see two successive questions and answers, with the second answer being hidden until they say the target words. In both cases, there is only one target sentence, and the app recognises only words that match with it, even if the learners say something different that is also correct.

3 Top10.com https://www.top10.com/language-learning-apps (November, 2022)
4 Language learning apps (presented in alphabetical order):
Babbel (2022). https://www.babbel.com/
Busuu Online S.L. (2022). https://www.busuu.com/
DuoLingo Inc. (2022). https://www.duolingo.com/
Memrise (2022). https://www.memrise.com/
Rosetta Stone Inc. (2022, v5.0.37). https://www.rosettastone.com/

Figure 1

Examples of Speaking Practice Activity on Memrise (top) and Duolingo (bottom)



Along with the written transcription and the audio, some apps provide a translation in the learner's L1, by default such as Memrise or Babbel, or on demand word by word on DuoLingo. The audio may come with a video or a picture giving contextual hints. ELSA also provides IPA transcription, but this was found in no other app. ELSA and Memrise also provide a feature to play the audio more slowly.

2.2 Types of feedback

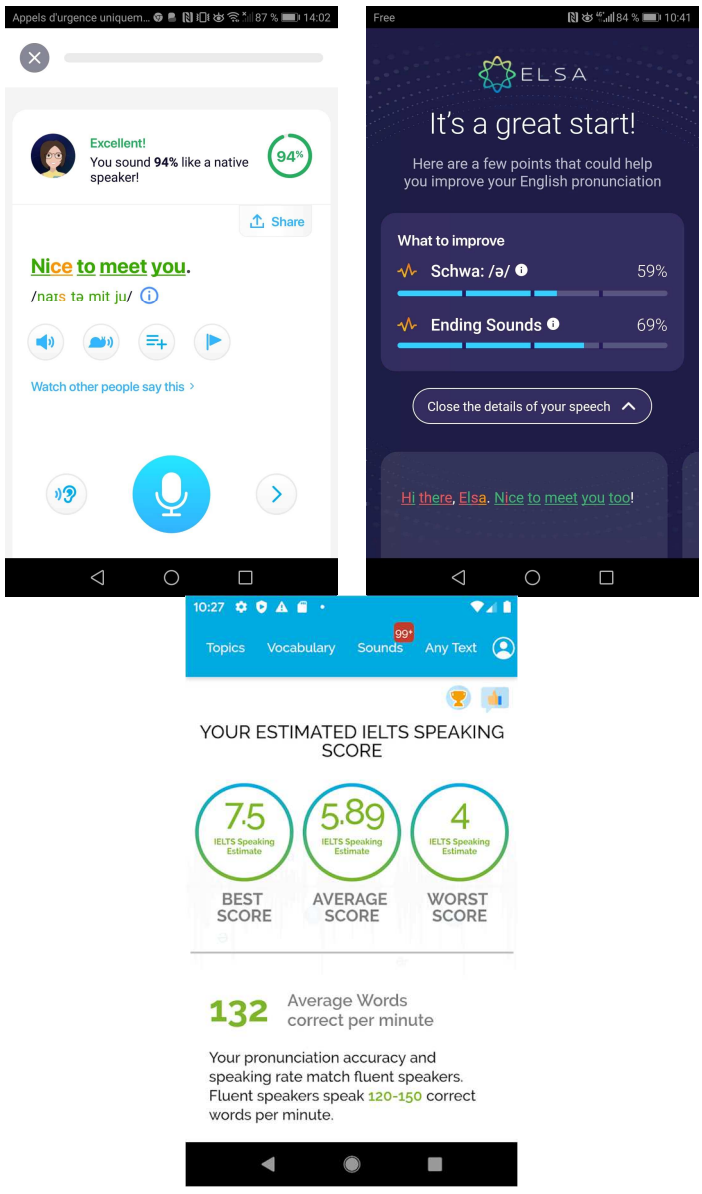
Adequate feedback provision is crucial in CAPT, with binary feedback (correct/incorrect) being the most common. In ELSA, for example, after recording, a green screen is displayed with congratulations if correct, or an orange screen with encouragement but no particular advice on how to pronounce it better (see Figure 1). Busuu and Duolingo also provide a

translation when the production is correct. Other apps show a success percentage indicating how well the words were recognised by the system.

ELSA and IELTS Speaking Practice go much further, displaying words or letters in colours: green (correct), orange (almost), red (wrong or missing). By clicking on a word, the learner can see what phonemes were expected, and which were recognized by the system. ELSA also gives explicit tips about how to pronounce incorrect segments and computes an overall English proficiency score (in percent), as well as scores for listening performance, fluency, word stress, intonation, and different categories of phonemes. ELSA and IELTS Speaking Practice calculate an estimated IELTS score, as well as the average number of correct words per minute. ELSA and IELTS Speaking Practice calculate an estimated IELTS score as well as the average correct words per minute (see Figure 2).

Figure 2

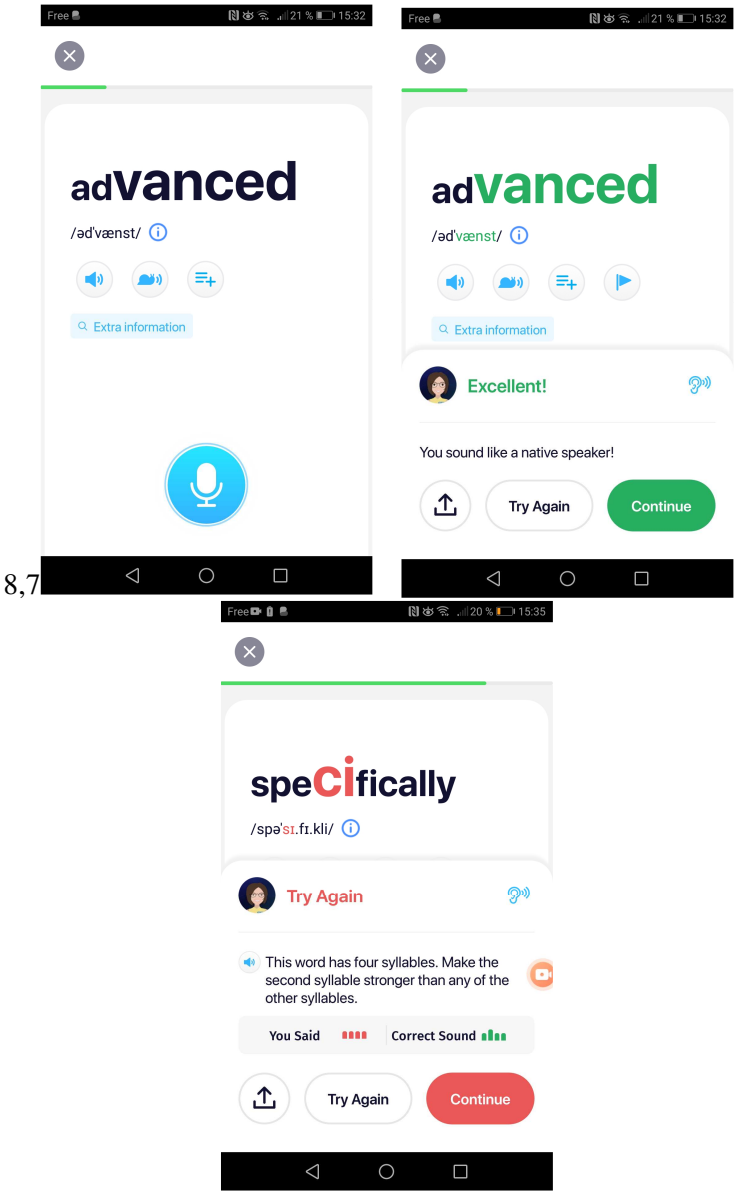
Examples of Feedback on ELSA (left), and Overall Scores on ELSA (middle) and IELTS Speaking Practice (bottom)



Feedback on lexical stress is also provided by ELSA in some specific exercises. In these activities, a single word appears on the screen with the expected stressed syllable written in larger characters. After recording, this syllable is coloured in green if the system detected the stress on that syllable in the learner's speech, or red otherwise. There is also a visual representation of the word's prosodic shape as a succession of bars, showing the syllable to stress and the syllable stressed by the student (see Figure 3).

Figure 3

Examples of a Word Stress Detection Activity on ELSA



8,7

In September 2022, ELSA launched a new premium feature on their website called Speech Analyzer[5], allowing students to record free speech and get an overall speaking score, as well as detailed scores for pronunciation, intonation, fluency, grammar and vocabulary. ELSA also provides global score predictions for IELTS, TOEFL, CEFR, and Pearson. In this feature, pronunciation feedback detects segmental errors and provides tips to pronounce them better. It also estimates intonation and gives average pitch variation, suggesting a range between 50 and 150Hz (typical of adult males), regardless of the learner's gender or age. Fluency estimations give a pace score (number of words per minute) and a pausing score as a so-called nativeness percentage.

2.3 How it works

Except for the word stress detection exercise of ELSA, all systems mentioned above work in a similar way: a speech recognition engine tries to match the target word or sentence, and compute a global, word or phoneme-level percentage of recognition confidence. The answer is considered correct when the confidence score is above a given threshold value. This value indicates how similar the user's speech is to the native model. In IELTS Speaking Practice or ELSA, non-constrained phoneme recognition allows the student to see which phonemes they actually pronounced, though this is limited to recognition of English phonemes only, and to the target number of phonemes.

In the case of ELSA's word stress detection exercises, an automatic speech recognition system (ASR) is probably used[6] to segment the word in syllables, and a stress classifier used to identify the stress pattern from acoustic measures such as duration, intensity, or pitch.

3 Do the students actually improve their pronunciation?

There are very few independent papers reviewing the effectiveness of CAPT tools for pronunciation improvement. Most studies either were funded by the company that created the app, or have one or more authors working for it; both situations present a conflict of interest. For example, the final report of Duolingo Effectiveness Study by Vesselinov and Grego (2012) claims that the vast majority of participants in the study succeeded in improving their knowledge, with a statistically significant improvement of mean average points (M = 8.1) on the WebCAPE placement test, per hour of use of the app. It would then take about 34 hours to reach 275 points on the test, 270 being the minimum to access the second semester of a university course. Based on this, Duolingo's home page claims that their app was scientifically proven to be more efficient than university courses. Krashen (2014) deplored this and took a closer look at the data. He found that there was great variability between participants (SD = 12.1, median = 3.9 points per hour), with participants studying Spanish for personal interest or school getting worse results (M = 5.7) than those studying for business (M = 11.4) or travel (M = 17.6). As for the participants, of the 156 who volunteered (having responded to an online advertisement), only 58% of them continued until the end. Learning time ranged from 2–133 hours depending on the person. This great variability across users makes it hard to generalise any results to university students, who may be obliged to take a course and often are less motivated. Anguera and Van (2016), the two creators of ELSA, did a study on 50 regular users to see how many repetitions they needed to pronounce a word without any error. They then computed their relative nativeness improvement over time; their graphs show decreasing and increasing curves, but no axis label is provided, nor is there any further textual information

5 ELSA's Speech Analyzer https://speechanalyzer.elsaspeak.com/
6 Details on ELSA's stress analysis are not publicly available.

about interpretation of the results. Doan et al. (2021), all employees of ELSA Corporation, carried out a study using ELSA at the private Chamakura Malla Reddy University, Bangalore, India. They had 206 students work with ELSA for six weeks. Improvement evaluation was done by the app itself, through the English proficiency score percentage. The improvement found in this study was 10.44%, but the paper does not specify how this number was calculated. This also means that improvement is only measured for contents studied in the app. The authors insist on the fact that every student improved their pronunciation. Other papers about ELSA's effectiveness are limited to questionnaires given to students, showing that a majority of them subjectively find the app useful and motivating: 18 university students (Kholis, 2021), five junior high school students (Pangastuti, 2021), 12 college students (Samad & Aminullah, 2019), and 25 college students (Silaen & Rangkuti, 2022).

Loewen et al.'s (2020) study on Babbel in Spanish was funded by Lesson Nine GmbH, the third author being employed by Babbel. Eighty-three university volunteers used the app 10 minutes a day for three months. The attrition rate was 32%. The evaluation was done on the 58 students who had spent at least three hours on Babbel, using an ACTFL Oral Proficiency Interview. This involves assessment by two human raters of 15 responses recorded by each student during the test. The study showed that the more time spent learning with the app, the better the results, and that most improvements were observed in grammar and vocabulary knowledge, rather than speaking ability.

Jiang et al. (2021) carried out a study specifically on Duolingo's efficacy for teaching speaking skills. It was funded by Duolingo and involved six authors from the company and one associate professor in applied linguistics at Northern Arizona University. They contacted random users of Duolingo who had completed the beginning-level course in French and Spanish. They asked them to take the Pearson Versant Test (an automatic speaking test using ASR on read and listen-and-repeat tasks, as well as say-the-opposite or reorder-the-words tasks). One hundred seventy-five learners of Spanish and 155 of French did the test. Sixty-six percent of Spanish learners and 53% of French learners achieved speaking A2 level or above, according to the Versant test, which corresponds to Duolingo's expected proficiency objectives for speaking skills. Nonetheless, it is worth noting that 28% of learners of French did not even receive a score with the Versant test, probably because of too poor pronunciation, according to the authors.

Becker and Edalatishams (2019), two independent researchers, also carried out a study on ELSA. According to them, one major shortcoming is that the app focused only on segmental aspects of pronunciation. They do not mention the stress detection activities, which might have been launched later than their study. Even if global scores of fluency and intonation are provided in today's version, no exercise allows the user to practice these aspects specifically. Becker and Edalatishams (2019) regret the fact that quantity overrides quality; hundreds of phrases are available, but all the exercises are very similar, and the feedback is the same. The paid version only gives access to more exercises about the same skills and topics. In February 2022, ELSA was claiming to offer over 3000 lessons. The two authors also note that ELSA often mistakenly identifies incorrect sounds as correct, leading to wrong and confusing feedback. They also deplore the absence of any applied linguists or English language learning experts in the ELSA development team. In February 2022, the ELSA web site[7] indicated that one speech advisor had been added to the team and he was described as "a world-class accent reduction coach with more than 35 years of training executives and Hollywood stars". In November 2022, out of 40 open positions listed on their website, 21 concerned business development and growth marketing, eight product design, seven management and marketing,

[7] https://elsaspeak.com/en/

and four speech science – not a single one was listed for pedagogical engineering, linguistics, or pedagogy.

4 Discussion

The CAPT tools reviewed here are easy to use, with a simple and intuitive design, and a great deal of graphic design rich in colours. Incorporated into gamified learning scenarios with various modalities of contents, everything is done to maintain learner motivation and to enhance the user experience. Even though most exercises are very similar, the contents and topics approached are varied. Providing this kind of automatic tool to learners allows them to get immediate feedback on their production, practice whenever and as much as they want without fear of being judged by others, and, above all, it gets them to practice speaking.

Notwithstanding, several limitations can be seen in the way these tools approach pronunciation training. These limitations can be divided into four categories: the stimulus, the focus, the model, and the feedback.

First, regarding the stimulus: with the exception of ELSA's brand new, (initially) free Speech Analyzer module, the speech stimulus of every type of exercise is a predetermined word or phrase the student has to read, repeat, or in rare cases to guess. Reading aloud may however lead to disfluencies or phonological errors, due to either pronunciation difficulties or to problems with decoding the written text before the reader's eyes. Repetition often leads to phonological imitation of the audio model, which makes it hard to diagnose student's spontaneous oral production proficiency. Furthermore, interactions and real communication situations are essential for practising speaking. Rosetta Stone was the only tool covered here to integrate student's production in a communicative situation, with much more context than the other apps (which tend to stick to micro-exercises out of context). This raises the question of the long-term effects of this predominance of micro-learning, which is observable to a larger extent in most language learning apps. Assessing spontaneous speech is a great challenge, but ELSA is proving that it is possible. Making pairs of learners speak to each other about a specific subject and providing individual feedback would be a very interesting type of activity, as well as integrating assessment features into an audio chatbot.

In terms of focus, assessment tools that go beyond mere word recognition mostly focus on phoneme quality and do not consider prosodic phenomena. ELSA Speak is the only app that seriously considers prosody in its scores and feedback. This, however, is only given as overall measures. In order to better suit students' needs, it seems important to target primarily those phenomena that hinder communication (such as speech rhythm and stress), and to give more weight to high functional load phonemes (Derwing & Munro, 2015), since their impact on speaker comprehensibility is widely accepted (Council of Europe, 2020).

All apps presented here compare students' speech production to a native model, using ASR or stress classification models. However, this tendency to base scores on the distance between one's production and a model only reflects how similar the users' production is compared to the model but not how acceptable their production is. Moreover, one might wonder how representative that model is of the diversity of voices, ages, accents, or types of discourse. This might also give more prominence to some hegemonic accents and ignore other less represented ones.

Concerning feedback, most systems still stick to binary response (good/bad) at phoneme, word or utterance level. Some apps also display a success percentage, often based on the ASR confidence score, which, in fact, can be hard to interpret. In the best cases, incorrect words or phonemes are displayed in colours along with what has been pronounced by the app user, and sometimes tips are provided on how to improve the pronunciation of incorrect phonemes. Yet, there is no error prioritization nor filtering depending on the learner's level or mother tongue,

for example. Above all, the main issue is that too many mistakes are considered correct by the apps. Memrise especially, tends to validate every attempt by the student, even when the pronunciation is extremely poor. Encouraging any answer might be seen as a positive reinforcement strategy to maintain learners' motivation, but it could become detrimental to the learning process. This issue of faulty feedback in CAPT systems was already considered a major issue by Levis (2007) and is still considered a real limitation of recent CAPT tools (O'Brien et al. 2018; Rogerson-Revell 2021). Basically, there is a need not only for more feedback, but also for more pedagogical feedback – and it needs to be prioritized instead of being displayed all at once with no hierarchy. Detey et al. (2016) call for a better consideration of diversity in learner profiles, providing adaptive feedback (as well as contents) according to intra- and inter-individual variables, such as observed difficulties, progress, and needs.

5 Conclusion

Research in CAPT seems to be focused on the technical challenges rather than the pedagogical ones. Very few studies analyse the actual impact of these technologies on language learning or how well these tools evaluate pronunciation. As for most of the commercial CAPT tools presented in this paper, it is likely that these are more often appealing technological creations intended to impress potential clients, rather than truly pedagogical tools that add value. This follows Rogerson-Revell's (2021) analysis that technological novelty "may temporarily disguise lack of pedagogic rigour" (p. 191) or even that "as technology progresses, pedagogy appears to regress".

Contemporary language learning apps are based on artificial intelligence and there is a worrying tendency to uncritically praise AI and to use it for marketing reasons rather for true pedagogical plus-value. For the learner's sake, it is important to clearly indicate how accurate the automatic evaluation is and inform them that it might not always work properly. The lack of knowledge about how the automated scoring system works makes it hard for students and teachers to interpret the scores. This is probably one reason why many teachers doubt the effectiveness of these tools (Agarwal & Chakraborty, 2019). According to Evanini and Zechner (2019), engineers who make CAPT tools are not aware of the issues in the field, such as the need for reliability, validity and fairness of evaluation, as well as for a transparent description of how scores are computed, in order to encourage a positive wash-back effect. They join Agarwal and Chakraborty (2019), O'Brien et al. (2018), Detey et al. (2016), and many others in calling for more collaboration between teachers and engineers.

Another important issue to bear in mind is data privacy. None of the apps examined here indicate what data is used, which data is stored, where it is kept, for how long, and for what purpose. It does not seem to be a real concern for companies or users.

To conclude, CAPT tools are an opportunity to complement classwork, giving learners the chance to practice and get feedback as much as they like on aspects that may not be studied in class because of a lack of time or teacher training. As Levis (2007) points out, CAPT is "tireless", "consistent", and it "can meet varied individual needs", and "promotes learner autonomy" (p. 197). The technology is here, we need to shape it according to our needs, rather than adapting our way of learning to available technology. In order to make pedagogically effective CAPT tools, engineers need to better understand the foundations of pronunciation acquisition and teaching.

References

Agarwal, C., & Chakraborty, P. (2019). A review of tools and techniques for computer aided pronunciation training (CAPT) in English. *Education and Information Technologies*, *24*(6), 3731–3743. https://doi.org/10.1007/s10639-019-09955-7

Anguera, X., & Van, V. (2016). English Language Speech Assistant. *Proceedings of InterSpeech*, 1962–1963.

Becker, K., & Edalatishams, I. (2019). ELSA Speak – Accent Reduction [Review]. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 434-438). Ames, IA: Iowa State University.

Council of Europe. (2020). *Common European framework of reference for languages*. Council of Europe.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.

Detey, S., Fontan, L., & Pellegrini, T. (2016). Traitement de la prononciation en langue étrangère : approches didactiques, méthodes automatiques et enjeux pour l'apprentissage. *TAL Traitement Automatique des Langues*, *57*, 15–39. ISSN 1248-9433 https://hal.science/hal-01919021

Doan, T., Nguyen, Q., Nguyen, B., Anguera, X. (2021). How ELSA speaking app improves learners' speaking skills - A use case from the CMR University in India. [White paper]. ELSA Corp.

English Language Speech Assistant (ELSA, 2022). https://elsaspeak.com/

Evanini, K., & Zechner, K. (2019). Overview of Automated Speech Scoring. In Zechner, K., & Evanini, K. (ed.) *Automated speaking assessment* (pp. 3–20). Routledge.

Frost, D., & O'Donnell, J. (2018). Evaluating the essentials: the place of prosody in oral production. In J. Volín & R. Skarnitzl (Eds.), *The Pronunciation of English by Speakers of Other Languages* (pp. 228–259). Cambridge Scholars Publishing.

Isaacs, T., Trofimovich, P., & Foote, J. A. (2018). Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing*, *35*(2), 193–216. https://doi.org/10.1177/0265532217703433

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.

Jiang, X., Rollinson, J., Chen, H., Reuveni, B., Gustafson, E., Plonsky, L. & Pajak, B. (2021). How well does Duolingo teach speaking skills? Duolingo Research Report DRR-21-02 June 1, 2021. https://blog.duolingo.com/how-well-does-duolingo-teach-speaking-skills/

Kholis, A. (2021). Elsa Speak app: Automatic Speech Recognition (ASR) for supplementing English pronunciation skills. *Journal of English Language Teaching*, *9*(1), 1–14. https://doi.org/10.32332/joelt.v9i1.2723

Krashen, S. (2014). Does Duolingo 'trump' university-level language learning? *The International Journal of Foreign Language Teaching*, *9*(1), 13–15. ISSN 2165-767X.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, *27*, 184–202. https://doi.org/10.1017/S0267190508070098

Loewen, S., Isbell, D. R., & Sporn, Z. (2020). The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability. *Foreign Language Annals*, *53*(2), 209–233.

O'Brien, M. G., Derwing, T. M., Cucchiarini, C., Hardison, D. M., Mixdorff, H., Thomson, R. I., Strik, H., Levis, J., Munro, M., Foote, J., Levis, G. M. (2022). Directions for the future of technology in pronunciation research and teaching. *Journal of Second Language Pronunciation*, *4*(2), 182–207. https://doi.org/10.1075/jslp.17001.obr

Pangastuti, D. (2021). The effect of ELSA Speak application on students' pronunciation in English. *Prosiding Pekan Ilmiah Mahasiswa Unis*, *1*(1), 127–133.

Rogerson-Revell, P. M. (2021). Computer-assisted pronunciation training (CAPT): Current issues and future directions. *RELC Journal*, *52*(1), 189–205. https://doi.org/10.1177/0033688220977406

Samad, I. S., & Aminullah, A. (2019). Applying ELSA Speak software in the pronunciation class: Students' perception. *Edumaspul: Jurnal Pendidikan*, *3*(1), 56–63. https://doi.org/10.33487/edumaspul.v3i1.85

Silaen, T. N., & Rangkuti, R. (2022). Elsa Speak app usage in blended learning during the Covid-19 pandemic: Students' perspectives. *Journal of Basic Education Studies*, *5*(1), 26–34. e-ISSN 2656-6702

SpeechAce (2022). IELTS speaking practice https://play.google.com/store/apps/details?id=com.ielts.speechace.ieltsace&gl=US

Vesselinov, R., & Grego, J. (2012). Duolingo effectiveness study. Final Report. https://static.duolingo.com/s3/DuolingoReport_Final.pdf

Walker, R., Low, E., & Setter, J. (2021). English pronunciation for a global world. Position paper. Oxford University Press. https://elt.oup.com/feature/global/expert/pronunciation?cc=fr&selLanguage=en

Witt, S. (2012). Automatic error detection in pronunciation training: Where we are and where we need to go. *Proceedings of the International Symposium on Audiovisual Detection of Errors in Pronunciation Training (ISADEPT)*, Stockholm, Sweden, June 2012 (pp. 1-8). KTH, Computer Science and Communication. http://www.m-a-arabia.com/wp-content/uploads/2014/11/Published-paper_isadept_ISADEPT-proceedings.pdf

About the author

Sylvain Coulange is in the second year of his interdisciplinary PhD at the Université Grenoble-Alpes, affiliated with the LIDILEM research group (Linguistics & First and Second Language Education) and the LIG (Grenoble Computer Science Laboratory), as well as the Speech and Language Processing Laboratory (SLPL) at Doshisha University in Japan. His research focuses on the automatic diagnosis of L2 pronunciation in spontaneous speech. At LIDILEM he co-directs the working group in Didactics of Languages and Multimodality and is part of the Language Acquisition Evaluation working group.

Email: sylvain.coulange@univ-grenoble-alpes.fr

Červinková Poesová, K. (2023). Peer pronunciation teaching: Initial training of Czech pre-service primary teachers. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 23–30). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8137816

Peer pronunciation teaching: Initial training of Czech pre-service primary teachers

Kristýna Červinková Poesová
Charles University, Prague

The current paper attempts to shed light on student teachers' ability to transform phonological knowledge into a pronunciation activity for primary pupils. The ad hoc study reported herein was inspired by the different extent of pre-service teachers' knowledge in the initial phase of teacher training – on the one hand, detailed knowledge of the subject matter, i.e., English phonetics and phonology, and on the other, very basic knowledge of pronunciation instruction. The study draws on data derived from multiple sources over a period of six years: peer teaching observations and reflections, lesson plans, and feedback provided by the teacher trainer. Despite the limited scope of pedagogical knowledge, the micro-teachings demonstrated elements of pedagogical content knowledge, especially in the lead-ins of pronunciation activities.

Keywords: teacher training, peer teaching, pedagogical content knowledge, English pronunciation teaching, primary education

1 Introduction

Teaching pronunciation effectively might be a dream for many foreign language teachers. One way of making this dream come true could involve the development of pedagogical content knowledge (PCK). This synthesis of content and pedagogical knowledge referred to as the ability to "do things effectively in language classrooms" (Murphy, 2017, p. 23) is believed to create the uniqueness of the teaching profession, clearly distinguishing a teacher from a content specialist, for instance a pronunciation teacher from a phonetician (Baker & Murphy, 2011; Shulman, 1987).

Although introduced by Shulman in the late 1980s, the nature of PCK has been widely debated by educational experts to date (Neumann et al., 2019). For example, Van Dijk (2009) conceptualized it as follows: "PCK is understood as topic-specific teacher knowledge that involves the transformation of content and pedagogical knowledge into instruction" (p. 19). Similarly, Bromme (1997 quoted in Neumann et al., 2019) states that "content-specific pedagogical knowledge is a necessary pre-requisite to finding adequate representations of subject matter content, and deciding about the selection and sequencing of ideas, that is, to enable transforming subject-matter structure into an instructional structure" (p. 850). Finally, Nilsson (2008, p. 1284) views PCK as dynamic knowledge generated in practice through the capability of the teacher to combine their knowledge of pedagogy, the subject matter, and contextual knowledge. The present research focuses on the interplay of these three types of knowledge, specifically in the process of transforming pre-service teachers' knowledge of the English sound system into meaningful and comprehensible forms within the context of primary pronunciation teaching.

2 The study: Research design

2.1 Motivation for the research

The first impulse to carry out the research presented in this paper arose when a five-year-long study programme for primary teachers with English specialisation[1] was cancelled permanently in the academic year 2019/2020, after approximately twenty-five years of existence. It was replaced by a newly accredited programme in which the time devoted to individual specialised fields was dramatically reduced – from nine to five semesters. In the original study programme, in the first two years the focus was put on language, linguistics, and cultural studies. In the following three years, the students were trained mainly in ELT methodology. Within the linguistic strand the study programme offered the luxury of four semesters of English phonetics and phonology spread equally over the first two years.

When the course was taken over by the author of this paper, it had been an unwritten rule to do some practically oriented activities in the last semester of English phonetics and phonology. In the first couple of years, this involved reading aloud or dramatising selected children's books, and this gradually changed to pronunciation peer teaching. Within the microcosm of the phonetics courses it seemed logical that the knowledge and skills acquired in the first three semesters[2] would be looked at from the pronunciation instruction perspective in the last

[1] The whole teacher training programme was provided by the Department of Pre-Primary and Primary Education at the Faculty of Education, Charles University in Prague. Only the specialisation (e.g., English, German, Music, Drama) chosen by the students in the first year was the responsibility of the corresponding departments e.g., the courses within the English specialisation were provided by the Department of the English Language and Literature at the same faculty.

[2] The content of the courses English phonetics and phonology I–III focused on the description of the English sound system including segmentals (vowels, consonants), processes of connected speech (assimilation, elision,

semester. This would be in line with Murphy's (2017) assumption that "teachers need at least some knowledge of phonology before the development of PCK becomes possible" (p. 23). The main task for the student teachers was to prepare a pronunciation activity according to certain instructions, peer teach it, and then reflect on that experience. Bearing in mind that it was the first peer teaching experience for the vast majority of students, the feedback towards them was both encouraging and constructive.

Taking into consideration the broader perspective of the whole study programme, it is important to state that when the students did the peer teaching, they were almost 'blank slates' as they had not attended any general pedagogy or ELT methodology courses yet; they had only spent a week observing classes at state schools, and these sessions did not necessarily include English. Therefore, the factors that probably influenced their performance were their beliefs and prior experience with English (pronunciation) learning. To summarize, on the micro level it made sense to introduce pronunciation teaching as the culmination of a three-semester-long phonetics course, even though the students' didactic skills were rather intuitive or non-existent. On the macro level, the pronunciation teaching activity was done too early, as the students were not educated or trained sufficiently and systematically in this respect. This particular situation in the English specialisation provided a rare window of opportunity for ad hoc research, looking more closely at how the students coped with the challenging task.

2.2 Treatment: Preparing teacher trainees to teach pronunciation

The student teachers were given clear guidelines on how to prepare and conduct a pronunciation activity in front of their peers. They were clearly instructed to devise a 15-minute-long activity on a pronunciation feature from one of the four areas: segmentals, connected speech, suprasegmentals, or miscellaneous (e.g., silent letters, homophones, accent differences). The obligatory parts were a lead-in (during which the topic had to be introduced interactively with pupils' active involvement) and pronunciation practice (oriented either perceptually and/or productively). The submission of a detailed activity plan was required. To provide more support, the lecturer taught one activity herself and the students were guided through a model activity plan with all its compulsory parts. Furthermore, they were given a number of practical tips, for instance to rehearse the instructions and explanations at home, avoid complicated terminology and follow the principle that "less can sometimes be more". The teacher trainees were also encouraged to adapt or create their own handouts or materials, relying on a wide range of pronunciation books, resource packs and/or recommended online sources. Thus, one of the subsidiary aims was met; the students became acquainted with some of the wealth of pronunciation teaching/learning literature.

While planning their pronunciation activity the student teachers could draw on three types of knowledge bases (Nilsson, 2008). Firstly, *pedagogical knowledge* was presented with the help of Hancock's (2014) pronunciation map. It introduces the landscape of pronunciation teaching based on three essentials: why, how and what to teach. In addition, short-term goal setting was demonstrated and a list of possible techniques for pronunciation teaching suitable for young learners was provided. For the lead-in the teacher illustrated how pupils' involvement can be secured, for instance by eliciting what they already know or having them notice certain things instead of telling them everything. Secondly, the *subject matter knowledge* was revised and briefly discussed as the students had acquired the knowledge of the English sound system in three courses of English phonetics and phonology. Lastly, the *contextual knowledge* was delimited by the age and language level of the imagined pupils: fourth or fifth

linking), syllable and suprasegmental phenomena (stress, rhythm, intonation) and simple comparison with L1. The reference variety was General British and transcription skills were developed throughout the courses.

graders (nine- or ten-year-old children). Regarding the language level it was stated what grammar, lexical fields and language functions the pupils should be able to use (e.g., present and past simple, animals, family members, hobbies, giving directions). The extent of the subject matter knowledge was greater than the other two types of knowledge, which were presented in a condensed form during one or two, 90-minute classes right before the micro-teaching began.

2.3 Participants

The total number of respondents was 73 (F = 70; M = 3), ranging in age from 19–21. They were all undergraduate students at the Faculty of Education, Charles University in Prague, and of Czech nationality. On the basis of multiple observations of primary school teacher trainees, they could be characterised as creative and playful, with a lower level of English but strong determination to stay in the teaching profession. They were didactically inexperienced and were pronunciation peer teaching for the first time. Their dominant knowledge base was the subject matter knowledge gained during three semesters of English phonetics and phonology. They all received identical pronunciation teaching basics in the fourth semester.

2.4 Research questions

The aim of the current study was to document the first encounter of pre-service teachers with the concept of pedagogical content knowledge, specifically to find out whether they were able to present the target content in a way that facilitates pupils' understanding of selected pronunciation aspects. The study addressed the following research questions:

RQ1: How do teacher trainees evaluate the pronunciation peer teaching experience?
RQ2: Does the combination of extensive, yet relatively new subject matter knowledge and rather limited pedagogical knowledge lead to the development of teacher trainees' pedagogical content knowledge (PCK)?

2.5 Research methodology, instruments, and data collection procedures

In this study a qualitative research design was adopted, namely a case study, which enables comprehensive understanding of specific situations (Schoch, 2020). Peer teaching was the major source of data collection, with on average 3-4 presentations per lesson, each followed by a 5- to 10-minute feedback session. The immediate oral feedback was given first by the participants and second by the lecturer. As the student teachers' experience giving feedback to their peers was little or non-existent, they were encouraged to say whether and why they liked the activity after each micro-teaching. Furthermore, they were asked to note anything confusing and to share their uncertainties. Building the skill of providing such critical feedback was the second subsidiary aim of the whole peer teaching experience. Delayed and detailed written feedback of the submitted activity plan was given by the lecturer and included critical remarks about the structure and content of the activity, mistake correction, and suggestions for alternative actions. After approximately one month, the peer teaching was completed, and the students were invited to take part in its written evaluation.

Data were collected by means of various research instruments: activity plans, two reflective posters, a survey, and the lecturer's written feedback. All participants had to submit an activity plan before the peer teaching took place. The teacher trainees wrote their reflections about the overall peer teaching experience on two reflective posters, one titled *What did I learn from my colleagues?* and the other with *What did I learn about myself?*. The respondents typically wrote

words, phrases, feelings, or drew happy/sad faces. A quick survey was carried out after all the peer teaching had been done, in which the respondents were invited to write individually about which three pronunciation activities they enjoyed most and to state the reasons behind their choices. While the activity plans were named, the reflective posters and surveys were anonymous. The lecturer's written feedback included comments on the lesson plans and observation notes from all micro-teachings. The four years of data were amassed over six years (between 2013–2020), with one year excluded due to the teacher's parental leave and one due to the COVID-19 pandemic. The reflective posters provided answers to the first research question, and the remaining instruments were used to answer the second research question.

2.6 Data analysis procedures

The written responses, comments, and notes were carefully read and categorized depending on the nature of the data. For data from the reflective posters, a thematic analysis was carried out, whereas data gathered via the survey were ranked from the most to the least frequently appreciated pronunciation activity. Moreover, the reasons behind the most positively evaluated activities were thematically grouped together. The last analysis involved the lecturer's observation notes and the feedback she gave students on their lesson plans. These data were analysed for a common theme, in this case for critical remarks and limitations.

3 Results of the study

3.1 Teacher trainees' reflections and evaluation of peer teaching experience

Data from the reflective posters[3] were gathered, with the first poster garnering comments from 71 student teachers about the things they had learnt from their colleagues. The responses were grouped into 23 categories, most of which were related to: general didactic principles: easy explanation; staying calm; using pictures; new activities; use of physical movements; smile is nice; engaging pupils in the lead-in; and saying all instructions. Only two categories can be considered pronunciation-specific: a) positive comments about teaching pronunciation aids (n = 7), for instance, "rubber bands are cool"; and b) a neutral comment stating "practising perception and production more than writing" (n = 1).

The second poster focused on participants' views of their personal gains from the course, in particular, what they learnt about themselves. It yielded 73 comments. The number of identified categories decreased to six, with two categories containing almost half of all comments, including comments capturing students' affective states: a) positive comments (n = 25), for example, "I enjoy teaching", "nervous but I loved it", "I want to be a teacher"; and b) negative comments (n = 16), for example, "When I'm nervous I hesitate", "feeling unprepared". The other categories included: a) acknowledgement of making mistakes (n = 7), for example, "my level of English diminishes when I teach"; b) need to be prepared (n = 7), for example, "you should always have a plan B"; c) clear and simple instructions (n = 5); and d) miscellaneous (n = 13). The only two reactions reflecting pronunciation expressed the need to improve their pronunciation.

[3] Due to the anonymous nature of the data, the author is unable to refer to individual participant codes where their responses are quoted in the text.

3.2 Teacher trainees' reasons for the most appreciated pronunciation activities

In-depth analysis of the most appreciated micro-teachings of every year is beyond the scope of this paper. However, the following list presents the rationale behind the most popular pronunciation activities, as expressed by the teacher trainees:

- using characters, invented or real, because it is easy to remember – e.g., "Lucy's Toby and Eva's bunny with hearing problems are unforgettable", "Tom and Jerry for teaching weak and strong forms", "Wanda and Victor for teaching /w/ and /v/ contrast";
- using teaching aids because they are playful and entertaining – e.g., "feather or a piece of paper for aspiration";
- providing clear illustrations and explanations of how the target sounds are created because it is easy to understand – e.g., "beaver's teeth and sending a kiss";
- being immersed in the pronunciation activity – e.g., "I did not realize I was learning; I enjoyed the activity and I didn't even notice it was focused on dental fricatives";
- using facilitating tools/helpers – e.g., pictures, photos, flags, gestures, whispering, flash cards, rubber bands, and movement;
- well-prepared presentations;
- giving clear instructions;
- continuous involvement of pupils and interactive classes.

3.3 Lecturer's evaluation of participants' lesson plans and presentation notes

Before the results in this section are summarised, it is important to realize that all teacher trainees successfully completed the micro-teaching, except for one whose pronunciation activity turned out to be extremely chaotic. Overall, a few teacher trainees did excellent or weak micro-teachings, but most were average in their performance. As opposed to the data gathered with the two reflective posters, in which positive features related to pronunciation teaching prevailed, the analysis of the submitted lesson plans and notes from the presentations revealed that there was a lack of knowledge in the participants' pronunciation teaching. The following list includes the weaknesses of the pronunciation tasks that occurred more than five times, either within one year or over several. In the lead-ins two problems reoccurred; either the tasks were rushed through (probably because of nerves) or the explanations of the rules were too lengthy and/or did not encourage pupils' engagement. Another shortcoming identified mostly in the lead-ins was the use of complicated terminology, for instance, phoneme, grapheme, and/or fricatives. Furthermore, the respondents quite often confused sounds and letters, e.g., /b/, /p/, and /h/ for <b> /biː/, <p> /piː/, and <h> /eɪtʃ/, formulated unclear aims, gave confusing instructions, and made grammatical, pronunciation, and content-related mistakes.

4 Discussion of results and key findings

The analysis of the peer and respondents' own feedback revealed two interesting findings. Firstly, the vast majority of comments were related to general principles of teaching foreign languages rather than specifically to English pronunciation, which may confirm the absence of ELT methodology in the training at that stage. Secondly, a greater variety of comments was identified when the students assessed their classmates than when they were rating themselves. When the attention was turned to their own performances, less variety occurred and the feedback was dominated by similar themes, in particular the emotions, both positive and negative, that the teaching experience provoked. Comments about being nervous and its effect

on students' skills resonated largely through all the investigated years. The other frequently mentioned themes – making mistakes and the level of preparedness – seem to be connected to student teachers' initial lack of confidence. Gaining confidence (e.g., by rehearsing at home) may lead to fewer mistakes. The pronunciation peer teaching experience can therefore be evaluated as enriching and overall positive.

Despite the great variety of micro-teachings, the reasons for choosing the best ones were narrowed down to a few items each year. The most appreciated moments involved clear understanding of a pronunciation feature, which was typically achieved with the help of an effective teaching aid, often a real or invented character (see §3.2). Subsequently, these topics seemed to be easily retained in the respondents' memory. Such reflections arguably constitute evidence of PCK development. Moreover, the notions to be taught were selected with respect to the age and level of young learners, possible complexities were reduced, for instance, the difference between American and British varieties was illustrated using only one feature, abstract concepts were visualised (e.g., smaller and bigger dots for stress patterns), and teaching materials were employed (Janík, 2008).

On the one hand we have pre-service teachers' subjective idea of what contributed to the efficacy of the presented pronunciation topics, and this is counter-balanced by the lecturer's observation, feedback, and long experience. It is also worth investigating further the phonological content-related errors identified in approximately half of the presentations each year (e.g., confusion of voiced vs. voiceless consonants). This finding suggests that the subject matter knowledge acquired in the phonetics courses might not have been completely consolidated at the time of peer teaching. In fact, the degree to which the phonological knowledge is applied – and applied correctly – in actual pronunciation teaching seems to be a window through which we can see how well the teacher trainees have actually understood the subject matter (Nilsson, 2008).

5 Conclusions and implications

Despite the participants' merely intuitive knowledge of pedagogy and ELT methodology, the findings provide examples of successful PCK for pronunciation teaching at a primary school level. Drawing on the peer evaluation, the most appreciated activities contained a playful and/or creative element, were well-prepared, and easy to understand/remember. The lecturer's feedback revealed limitations in the investigated domains of knowledge, for instance, using the wrong phonemic symbols, employing difficult terminology, poorly prepared lead-ins or long explanations without pupils' engagement. Naturally, the ability to select and structure the pronunciation activity so that it resulted in pupils' understanding varied among the teacher trainees.

On a more general note, peer teaching played a crucial role in primary teacher pronunciation training. In the assigned activity, the participants experienced first-hand the necessity to reorganize the subject matter, to transform it into meaningful forms. All four parts of the peer teaching experience – preparing for teaching and teaching itself, stepping into the pupils' shoes and reflecting – contributed equally to the "moments of revelation" which are likely to have long lasting effects, helping to form future teachers' identities and cognitions. No matter how time-consuming the inclusion of micro-teaching might be, expert literature or lecturers' well-meant recommendations can never fully replace direct experience (Murphy, 2017).

The outcomes of this research are currently being implemented into the newly accredited study programme which contains only two semesters of English phonetics and phonology. Contrary to the original intention not to include any pronunciation teaching aspects due to time constraints, short micro-teachings on selected phenomena have been included (e.g., teaching word stress) for two main reasons. First, that the ability to transform theoretical knowledge into

actual teaching can be trained to some extent even before it is thoroughly covered within ELT courses, and secondly, little time is dedicated to pronunciation teaching in the Czech educational context (Červinková Poesová & Uličná, 2016). Therefore, for some respondents the pronunciation peer teaching squeezed within the phonetic course might be, sadly, the only one they experience. In conclusion, specialized training in pronunciation pedagogy plays an important role in building the confidence of future pronunciation teachers (Burri et al., 2017; Buss, 2017) and should always have a firm position in the teacher education programmes.

References

Baker, A., & Murphy, M. (2011). Knowledge base of pronunciation teaching: Staking out the territory. *TESL Canada Journal*, *28*(2), 29–50. https://doi.org/10.18806/tesl.v28i2.1071

Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln der Lehrer. In F. E. Weinert (Ed.), *Psychologie des Unterrichts und der Schule* (pp. 177–212). Hogrefe.

Burri, M., Baker, A. & Chen, H. (2017). "I feel like having a nervous breakdown" Pre-service and in-service teachers' developing beliefs and knowledge about pronunciation instruction. *Journal of Second Language Pronunciation*, *3*(1), 109–135. https://doi.org/10.1075/jslp.3.1.05bur

Buss, L. (2017). The role of training in shaping pre-service teacher cognition related to L2 pronunciation. *Ilha Do Desterro*, *70*(3), 201–226. https://doi.org/10.5007/2175-8026.2017v70n3p201

Červinková Poesová, K., & Uličná, K. (2016). Fonetika a fonologie anglického jazyka. *Filologické studie 2016*, 68–86.

Hancock, M. (2014). A map of pronunciation teaching. *Speak Out!*, *50*, 8–11.

Janík, T. (2008). *Metodologické problémy výzkumu didaktických znalostí obsahu*. Paido.

Murphy, J. (2017). *Teaching pronunciation of English: Focus on whole courses*. The University of Michigan Press.

Neumann K., Kind, V., & Harms, U. (2019). Probing the amalgam: the relationship between science teachers' content, pedagogical and pedagogical content knowledge, *International Journal of Science Education*, *41*(7), 847–861. https://doi:org/10.1080/09500693.2018.1497217

Nilsson, P. (2008). Teaching for understanding: The complex nature of pedagogical content knowledge in pre-service education. *International Journal of Science Education*, *30*(10), 1281–1299. https://doi:org/10.1080/09500690802186993

Schoch, K. (2020). Case study research. In Burkholder, G. J., Cox, K. A., Crawford, L. M. & Hitchcock, J. H. (Eds.) *Research design and methods: An applied guide for the scholar-practitioner* (pp. 245–258). Sage Publications.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–22.

Van Dijk, E. M. (2009). Pedagogical content knowledge in sight? A comment on Kansanen. *Orbis Scholae*, *3*(2), 19–26. https://doi.org/10.14712/23363177.2018.209

About the author

Kristýna Červinková Poesová is a Lecturer at the Department of English Language and Literature, Faculty of Education, Charles University in Prague, where she teaches English phonetics and phonology and general English courses in various teacher training programmes. Her research interests include effectiveness of pronunciation teaching and learning, pronunciation instruction at a primary level and accent attitudes. She co-edits IATEFL's PronSIG bi-annual journal *Speak Out!*.

E-mail: kristyna.poesova@pedf.cuni.cz

Čtvrtečková, L., Fischer, O., & Skarnitzl R. (2023). Pronunciation of Czech secondary school English teachers and their cognition about pronunciation. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 31–45). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8174033

Pronunciation of Czech secondary school English teachers and their cognition about pronunciation

Lenka Čtvrtečková
Charles University, Prague

Ondřej Fischer
Charles University, Prague

Radek Skarnitzl
Charles University, Prague

Language teachers undoubtedly serve as important models of their learners' pronunciation. The recent growth of interest in teacher cognition also raises the question to what extent non-native English teachers' cognition about pronunciation and pronunciation teaching are aligned with their own production (from a contrastive perspective) and accent preference. In the present study, we first analysed recordings of 12 Czech secondary school teachers of English obtained directly in class. Auditory analyses revealed characteristic pronunciation features of Czech-accented English at the segmental level (e.g., vowel quality, dental fricatives, the presence of a plosive following the velar nasal, aspiration), while vowel reduction and linking were realised in a more native-like manner than by strongly accented speakers examined in previous studies. Results obtained from questionnaires investigating teachers' cognition indicated their positive attitude towards authentic input, hierarchisation of pronunciation features, and prioritisation of intelligibility in pronunciation teaching. A comparison with the teachers' overall accentedness suggested that more strongly accented participants were more convinced about the importance of native-like attainment and pronunciation teaching itself than their colleagues with a more mild-accented English. Furthermore, the results indicate only a weak relationship between teachers' accent preference and rhoticity in their own speech production.

Keywords: pronunciation teaching, teacher cognition, foreign accent, Czech accent in English, language transfer

1 Introduction

This study deals with non-native teacher cognition in teaching English as a foreign language. More specifically, it focuses on questions related to pronunciation teaching practices, aims in pronunciation instruction, the assigned value of the instruction itself and preferences in model accent. Furthermore, teachers' own pronunciation is analysed with reference to the typical features of Czech-accented English. Since teachers serve as models to their learners, the aim was to see whether their attitudes towards pronunciation are reflected in their speech production during teaching.

2 Previous research

2.1 Teaching pronunciation and teacher cognition

The current study on teacher cognition in pronunciation teaching (i.e., on their beliefs and knowledge about pronunciation teaching) explores how Czech English teachers' pronunciation and beliefs develop on an individual basis, reflecting how they are strongly tied to context and environment, and whether changes in cognition are linear, contrary to Burri and Baker's findings (2021).

Research has shown that teachers who receive training on how to teach pronunciation planned their teaching in a more systematic way (Nagle et al., 2018), and reported greater confidence when teaching it (Kochem & Levis, 2022). In Couper (2016), teachers reported that despite their knowledge of phonetics and phonology, they did not feel they had been trained enough in *how* to teach pronunciation and thus, tended to ignore the areas that were difficult for them to describe, e.g. stress and intonation. Couper also points out that teachers regarded intelligibility and accurate pronunciation on the segmental level as their objectives.

The question of how to train future teachers specifically in relation to pronunciation teaching has been gaining significant attention by researchers (Levis & Sonsaat, 2019), with studies examining, e.g., a) teachers' confidence and self-consciousness, b) constraints related to time and/or teaching materials, c) types of instruction, and d) appropriate models depending on whether the overall goal is pronunciation that is either intelligible or native-like.

Studies repeatedly report that second language teachers often lack confidence in pronunciation teaching despite their desire to address it in class (Baker, 2013; Guerra, 2017; Levis & Sonsaat, 2019). As a result, pronunciation might receive less attention in class than grammar or vocabulary skills (Burri & Baker, 2020), with feedback relegated to the level of reactive ad hoc corrections (Burri & Baker, 2020, 2021) that lack systematic approach. Tight schedules and the obligation to cover a particular syllabus are limiting factors which many teachers mention (e.g., Sicola & Darcy, 2015), but – as Darcy et al. (2020) have shown – it is possible to incorporate pronunciation activities into existing syllabi without compromising the overall course goals.

A number of varieties of English have been proposed as models for pronunciation teaching, as no single model is unequivocally suitable for all learners and purposes (Szpyra-Kozłowska, 2018). In ELT textbooks, Received Pronunciation (RP; today also referred to as Standard British English, SBE) is often adopted as a point of reference, yet generally unaccompanied by a clear reasoning (Upton, 2015). In the European context, Ivanová (2011) sees SBE as a beneficial pronunciation model for several reasons: its prestigious status; its dominance of learning and teaching materials, including dictionaries; and the notion that it is fairly easily comprehensible compared to other regional varieties of the British Isles. However, the comprehensibility of a particular variety will, to some extent, depend on the sound patterns of the learner's L1, as well as the listener's. For instance, in the Czech context, Kobák (2017)

found somewhat lower intelligibility of standard British presumably due to the absence of rhoticity. Another variety that English learners are frequently exposed to through the Internet, film and music is General American (GA). Czech learners predominantly rated GA favourably in terms of pleasantness, comprehensibility, and status (Jakšič, 2018). This begs the question of what model Czech learners should aim for, and thus, what pronunciation features teachers should devote time to, e.g., is non-rhoticity a 'necessary' feature if one wants to be intelligible and favourably perceived?

In general, addressing pronunciation in class may be even more challenging for non-native teachers due to their lack of confidence. However, as shown by Guerra's (2017) study of Portuguese university students, learners might hold both native and non-native teachers in equally high esteem, evaluating them in relation to other traits – e.g., pedagogical competence or interactional skills – instead of merely their degree of accentedness. Therefore, a useful first step is to compare teachers' cognition with their actual spoken English, which is the aim of the present study.

2.2 Czech-accented English

The nature of this study necessitates a contrastive approach, and individual pronunciation features will be treated in this paper without considering their impact on comprehensibility. Czech speakers may struggle with a number of segmental or suprasegmental aspects of English, and these were described in detail by Skarnitzl and Rumlová (2019, see also references therein); only a brief summary will be presented here.

The English vowel system is considerably more complex, especially in the open region with front /æ/ and back /ɒ/ likely to be challenging for Czech speakers of English. They are typically replaced by mid vowels [ɛ] and [o] respectively. While overall consonantal complexity seems to be greater in Czech, the English dental fricatives are typically realised incorrectly, especially the voiced /ð/ as [d]. In addition, the velar /ŋ/ is frequently followed by a velar stop /k/ in Czech-accented English.

With word stress in Czech being fixed on the first syllable of the prosodic word, Czech speakers of English may also place stress word-initially in their L2 English even in words stressed on another syllable. In addition, Czech has no systematic vowel reduction, so vowels tend to be realised as full in their quality, even in unstressed syllables in Czech-accented English. In Czech, words beginning with a vowel tend to be accompanied by a glottal stop and linking of two successive words is therefore not too common in Czech-accented English. In general, prosodic signalling is rather weak in Czech, with the stressed syllable bearing no typical signs of acoustic prominence, prosodic phrases being much longer than in English and pitch range being considerably narrower (Skarnitzl & Hledíková, 2022; Volín et al., 2015).

3 Research methodology

3.1 Research questions

While research on teacher cognition in English pronunciation teaching mostly addresses native speakers of English (e.g., Baker, 2013; Baker & Burri, 2016; Couper, 2016), more needs to be known about the thoughts, beliefs and knowledge of non-native speakers in countries where English is not the predominant L1. The present study therefore focuses on Czech L2 teachers' cognition concerning certain aspects of pronunciation teaching. In addition, we analyse teachers' own in-class pronunciation and compare relevant pronunciation features with some aspects of their cognition to see whether there is a relationship between the two. Therefore, the study aims to answer the following research questions:

RQ1: What are the cognitions of Czech teachers of English towards pronunciation teaching?

RQ2: Does their pronunciation in lessons align with the previous research of Czech-accented English? Are teachers' cognitions and their in-class production somehow interconnected?

3.2 Participants and recordings

Our sample of teachers consisted of 12 participants (F = 9; M = 3), professional teachers of English and native speakers of Czech. Half of them worked and resided in the capital city, Prague, the other half in the smaller Moravian city of Opava. Eight of these teachers taught at grammar schools, the remaining four at vocational secondary schools. Their average age was 41 years (ranging between 27 and 53), with the average length of teaching experience 15.5 years (ranging from 5–27). All but one possessed an MA degree in English.

The teachers were recorded during their regular English lessons using a lavalier microphone. This facilitated high-quality recordings, while the visual unobtrusiveness of the device also helped to mitigate both teachers' and students' potential anxiety about the recording situation. All of our participants signed an informed consent and volunteered for the research, even though they were not informed about its specific focus (i.e., pronunciation) before the recording itself, given that such information could have impaired the authenticity of their production.

The recordings were later edited to teacher talking time only, excluding especially longer pauses and any audible speech by students. We selected three approximately ninety-second excerpts from each of the recordings – one from the beginning, one from the middle and one from the final part of every lesson – to obtain a balanced sample of every teacher's pronunciation. The resulting material, consisting of about 290 seconds of speech per participant, was used in the pronunciation analysis.

3.3 Pronunciation analysis and questionnaires

Praat (Boersma & Weenink, 2021) was used to analyse pronunciation features which are typical of the Czech accent in English (see §2.2). The features were evaluated using careful auditory analysis by the first two authors of this study after a joint calibration session; for potential controversies the last author was consulted. Some of the features were analysed in a binary way (e.g., vowel reduction, linking), some in a ternary manner, with an intermediate realisation possible (e.g., /ð/ realised as dental stop [d̪]). The assessment criteria are described in Table 1. A standardised success score in the range from 0 (strongly accented) to 1 (native-like) was calculated for each of the pronunciation features analysed, with the intermediate realisation of a ternary distinction corresponding to 0.5. Each participant's comprehensive score (0 to 1) was then calculated as the average of the scores obtained for the nine individual features.

Additionally, a questionnaire comprising 14 questions (in Czech; see Appendix for the English translation) regarding cognitions about English pronunciation teaching was distributed to our participants after the in-class recording. Answers were marked on a five-point scale, with items 1–6 evaluated in terms of importance on a scale from 1 (lowest) to 5 (highest), and items 7–14 asking for a degree of agreement with a statement from "definitely not" (value 1) to "definitely yes" (value 5). Pearson correlations were used to check relationships between the teachers' pronunciation and aspects of their cognition.

Table 1

Assessment of Analysed Pronunciation Features

Feature	Native-like	Intermediate	Czech-accented
/ð/	[ð]	[d̪]	[d]
/θ/	[θ]	[t̪]	[t], [f], [s] ...
/r/	[ɹ̠], [ɹ], [ɚ]	[ɹ˜ɾ]	[ɾ], [r]
/w/	[w], in *wh-* also [ʍ]	[ʊ]	[v]
/ŋ/	[ŋ], in *-ing* also [n]	-	[ŋk], [ŋg]
/æ/	[æ]	[æ˜ɛ]	[ɛ], [e]
/ə/	[ə]	-	full vowel
aspiration	present	-	absent
linking	linked	-	glottalised

4 Data analysis and results

4.1 Teachers' pronunciation

The auditory analysis revealed that the pronunciation of our sample of secondary school teachers in the Czech Republic does not differ considerably from the strongly accented speakers examined by Skarnitzl and Rumlová (2019). As shown in Table 2, which lists the mean scores of individual pronunciation features analysed in this study in a descending order from most accurate to least accurate, the approximants /w/ and /r/ were produced most accurately by our teachers. A stark contrast appears within the dental fricatives, where the voiceless dental fricative /θ/ was accurately pronounced in nearly 90% of items, whereas its voiced counterpart /ð/ turned out to be one of the most problematic sounds for our speakers, with a comprehensive score of 0.38 accurately pronounced items. The voiced /ð/ was typically realised as an alveolar stop, [d], less frequently also as a dental stop, [d̪]; a more detailed analysis reveals that the pronunciation tended to be slightly more native-like in lexical words (e.g., *Southern*, *weather*) than grammatical words (e.g., *the*, *that*, *they*). Given previous studies which addressed the open front vowel /æ/ in Czech speakers of English, it is not surprising that this sound was mostly pronounced as a considerably higher vowel [ɛ] and scored lowest in our sample. The score of the remaining segmental features – /ŋ/ and aspiration of voiceless stops – was slightly above 0.5. As for the suprasegmental features, results show that linking did not occur frequently in the speech of the teachers, with glottalisation occurring in over one half of the possible contexts. Finally, vowel reduction occurred in about 60% of the examined items, the only value which is considerably higher than in the study reported by Skarnitzl and Rumlová (2019).

Table 2

Mean Scores of Analysed Pronunciation Features

Feature	Mean score
w	0.94
r	0.91
θ	0.89
ŋ	0.62
vowel reduction	0.61
aspiration	0.54
linking	0.46
ð	0.38
æ	0.19

To assess the pronunciation of individual teachers (T), a comprehensive accentedness score was calculated in the form of a weighted average of the analysed pronunciation features. The score reveals a range between 0.42 for T07 and 0.77 for T08 (see Table 3). Additional analysis was carried out to test whether the comprehensive accentedness scores would correlate with teachers' age, sex, length of teaching experience, type of school (grammar school vs. vocational school), or residence (the capital Prague vs. smaller Opava). However, no noteworthy relationships were revealed between the score and any of these variables.

Table 3

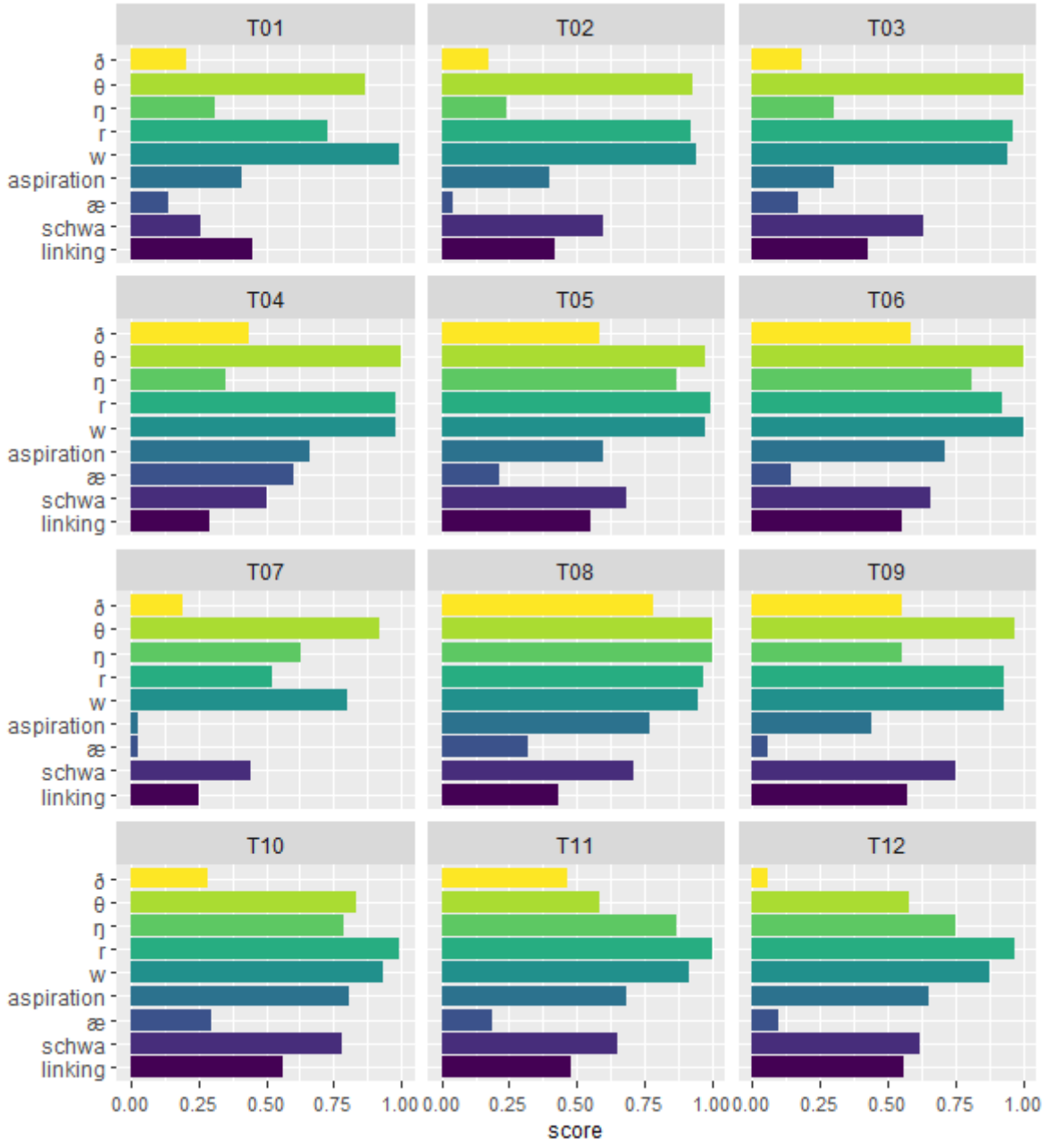
Accentedness Scores of Individual Teachers

Teacher	Comprehensive accentedness score
T01	0.48
T02	0.52
T03	0.55
T04	0.64
T05	0.72
T06	0.71
T07	0.42
T08	0.77
T09	0.64
T10	0.70
T11	0.65
T12	0.57

It is to be expected that the comprehensive accentedness scores conceal considerable individual variability, and that two similar scores are likely to be a product of different partial scores (see Skarnitzl & Rumlová, 2019). Figure 1 confirms such expectations. Speakers T04 and T09, with an identical score of 0.64, may serve as examples, with differences in the suprasegmental features but also the open front vowel. The realisation of the voiced /ð/ appears to be the feature which shows particularly high between-speaker variability, as does the pronunciation of the velar nasal /ŋ/ with or without a following plosive sound.

Figure 1

Scores of Nine Pronunciation Features in Individual Teachers

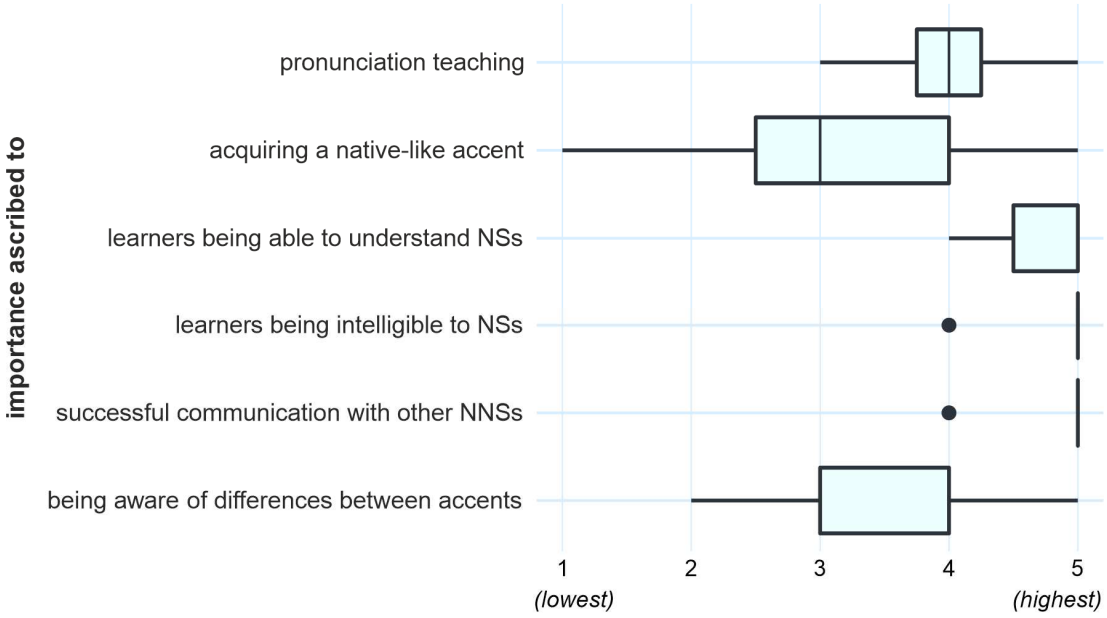


4.2 Teachers' cognition

The first set of questions addressed the importance which the teachers ascribed to various objectives in pronunciation teaching. As can be seen in the topmost boxplot of Figure 2, most of our teachers expressed belief in the importance of pronunciation teaching in general (M = 4.0): six teachers assigned pronunciation score 4, three teachers an average level of importance (M = 3) and the other three attributed the highest level to it. When we compare intelligibility vs. native-like production as the goal of pronunciation teaching, the teachers' evaluation of objectives 2–5 suggests that intelligibility is regarded as more important. Learners' acceptable intelligibility in speaking with and being understood by native speakers were both evaluated as important (M = 4.73 and M = 4.91, respectively), as was successful communication with other non-native speakers (M = 4.91). However, the importance of learners acquiring an almost native-like accent was seen as much less important (M = 3.18, with scores covering the entire evaluation range). An awareness of differences between various accents of English was also ascribed average importance (M = 3.55).

Figure 2

Scores of Importance Ascribed by the Teachers to Pronunciation Teaching Objectives



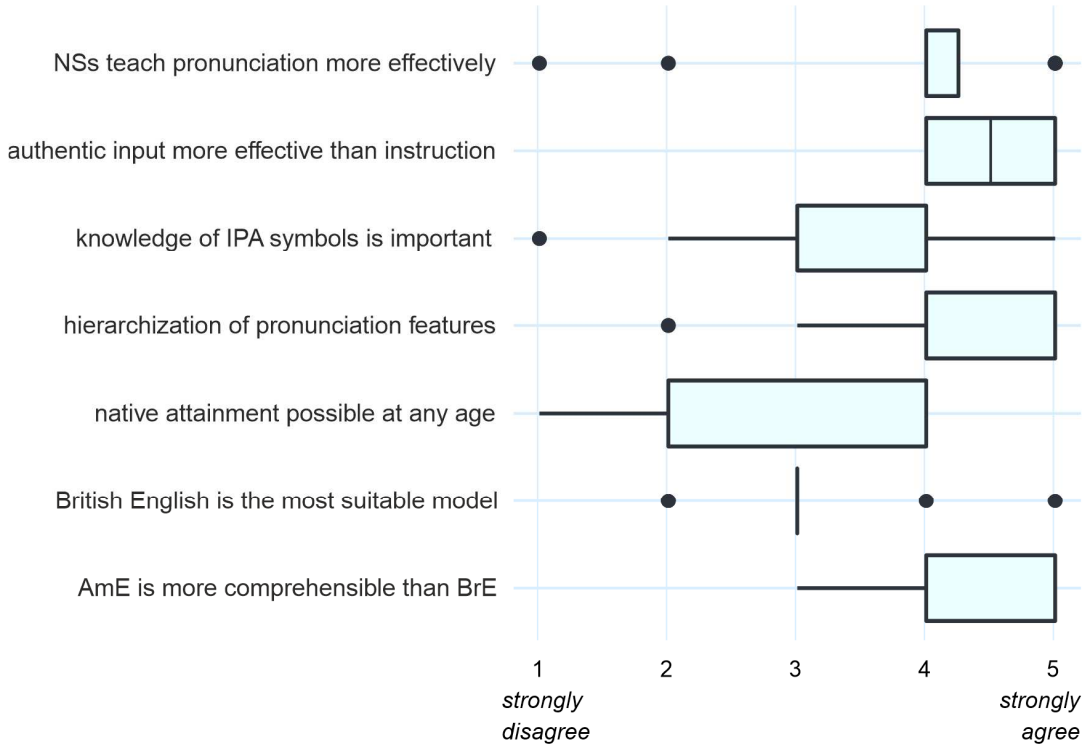
Note. NSs = native speakers, NNSs = non-native speakers.

In the second set of questions, the respondents were asked to express the degree of their agreement with various statements related to pronunciation teaching. The results are presented in Figure 3. These responses seem to be more varied. The most unequivocal question concerned input: teachers predominantly consider authentic input as a more effective means of acquiring English pronunciation than in-class instruction (M = 4.5). The opinion that some aspects of pronunciation are more important to effective communication than others, and that it is therefore important to teach pronunciation features in a hierarchical order, was also predominantly assumed (M = 4.17). Most teachers agreed that native speakers' ability to teach

pronunciation is better than that of non-native speakers (M = 3.83); in contrast, two teachers expressed disagreement with this statement. The belief that learners should be familiar with transcription symbols was acknowledged by the majority of teachers (M = 3.58), with two teachers disagreeing with the importance of transcription. The teachers were most sceptical about the possibility of acquiring a native-like accent at any age (M = 3.17). The responses related to English variety preferences showed that teachers on average do not regard British English as the most suitable pronunciation model (M = 3.08) and mostly find American English more comprehensible, i.e. easier to understand (M = 4.17).

Figure 3

Agreement with Statements Concerning Pronunciation Teaching



Note. NSs = native speakers, BrE = British English, AmE = American English.

4.3 Teachers' cognition vis-à-vis pronunciation

Additional analysis was carried out to compare teachers' pronunciation with some of the items from the cognition questionnaire and their demographics.

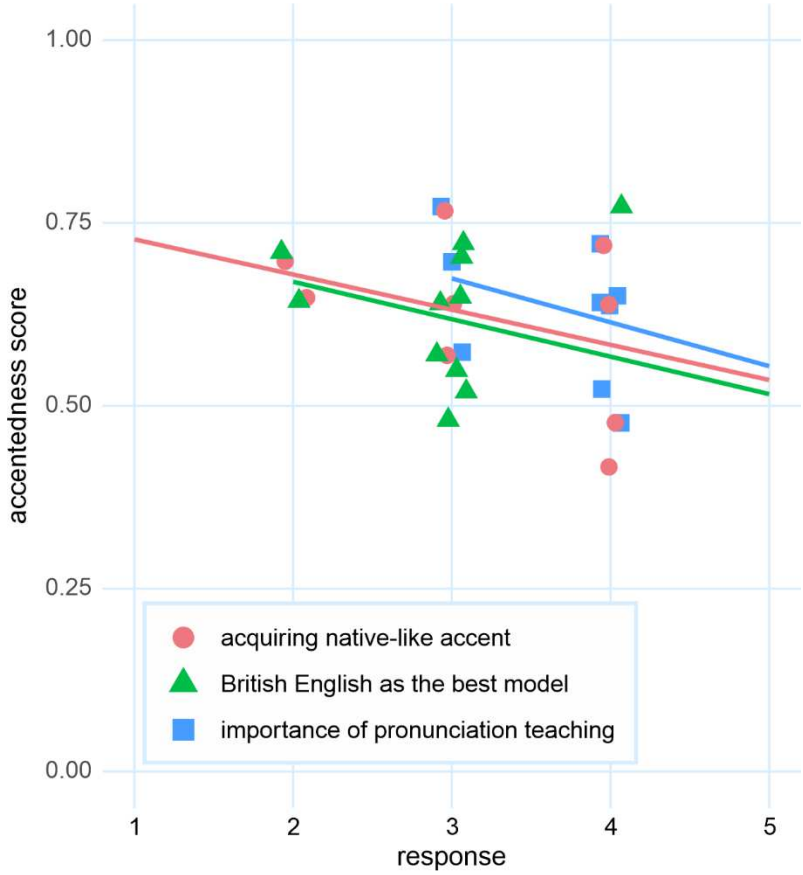
Regarding the comprehensive accentedness score, our data reveal no significant correlations of the accentedness score (see Table 2 above) with the teachers' age, experience with teaching English, the type of school, or the region where they teach. In addition, the accentedness score does not correlate with the importance ascribed to pronunciation teaching.

The results of the relationship between the accentedness score with the teachers' cognition (see Figure 4 indicate weak negative correlations ($-0.35 < r < -0.5$) between the accentedness score on the one hand and, on the other hand, the importance ascribed to the importance of pronunciation teaching, native-like attainment, and British English being the most suitable

model. In other words, more teachers with a strong Czech-accented English believed that pronunciation teaching is important and that learners should attain native-like accent, compared to those with a lower accentedness score.

Figure 4

Relationship between Accentedness Score and the Importance Ascribed to Three Phenomena



As for the relationship between the preferred accent and the teachers' rhoticity in their own pronunciation, one half of the teachers expressed preference for a general British accent. Three of these had predominantly non-rhotic pronunciation (with between 25 and 40% of post-vocalic [r] pronounced), the remaining three were rhotic (with over 90% of possible contexts pronounced with [r]). The four teachers who preferred General American were mostly rhotic, as were the two who expressed their preference for Canadian English.

5 Discussion and conclusion

The analysis of pronunciation features of Czech secondary school English teachers is in line with the results by Skarnitzl and Rumlová (2019). Concerning segmental features, it is clear that even for secondary school teachers, who are highly proficient in English, the native-like pronunciation of [æ] and [ð] remains a considerable challenge. While pronouncing [ŋ] without a subsequent plosive seems less demanding, there was considerable individual variance. As for connected speech phenomena and the prosodic domain, Czech teachers did not employ linking

very often, which resulted in choppy speech. On the other hand, their use of vowel reduction was slightly better compared to the results obtained in previous studies on strongly accented Czech speakers of English (Skarnitzl & Rumlová, 2019; Volín et al., 2013).

Data collected from questionnaires showed that, in general, Czech teachers in our sample valued the communicative dimensions of pronunciation (intelligibility and comprehensibility) rather than native-like attainment. While most of the teachers stated that they use British English as a model in class ("because of the textbook", as some noted), they do not perceive American English as inferior to the former – the majority of the participants even reported American English to be more comprehensible to them. Interestingly, strongly Czech-accented teachers seemed to ascribe higher value to British English, native-like attainment and pronunciation teaching in comparison with their less accented peers, for whom pronunciation teaching appeared to be of a lower priority, and British English along with native-like attainment were not highly valued.

The participants were also rather hesitant about the notion of the teacher serving as a model for pronunciation, about the benefit of in-class pronunciation instruction as compared to authentic input, as well as about the importance of using phonemic transcription in lessons; such views could be related to the low self-confidence of teachers mentioned earlier in the article.

There are certain limitations of the study which should be pointed out. The sample was rather small, consisting of twelve individuals only, which was caused by the ongoing COVID-19 pandemic and teachers' willingness to participate. It was also not balanced in terms of gender, geographical distribution or types of schools. The limited sample may also be the reason why we found no correlations of the teachers' age, region, the amount of teaching experience, type of secondary school or accent preferences with cognition and pronunciation in class. Alternatively, however, it is conceivable that these variables truly do have little effect on teachers' pronunciation and cognition. Finally, our results may also portray the situation in a slightly more favourable light: since only teachers who were willing to participate in the study and volunteered for it were recorded, we might expect an average secondary school English teacher in the Czech Republic to have a stronger accent. In future research, it would be interesting to relate teachers' cognition about pronunciation with their actual practices concerning pronunciation teaching.

The pronunciation of Czech teachers has confirmed the presence of many typical features of Czech-accented English. Importantly, teachers themselves proved they were aware of the importance of intelligibility in pronunciation teaching and noted that not all pronunciation features are equally significant. However, we are convinced that to serve as an effective model, teachers should aim for pronunciation which is comprehensible and does not result in negative social evaluations.

References

Baker, A. (2013). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly*, *48*(1), 136–163. https://doi.org/10.1002/tesq.99

Baker, A., & Burri, M. (2016). Feedback on second language pronunciation: A case study of EAP teachers' beliefs and practices. *Australian Journal of Teacher Education, 41*(6), 1–19. https://doi.org/10.14221/ajte.2016v41n6.1

Boersma, P., & Weenink, D. (2021). *Praat: doing phonetics by computer* (Version 6.1.42). http://www.praat.org

Burri, M., & Baker, A. (2020). "A big influence on my teaching career and my life": A longitudinal study of learning to teach English pronunciation. *TESL-EJ: The Electronic Journal for English as a Second Language, 23*(4), 1–24.

Burri, M., & Baker, A. (2021). 'I feel … slightly out of touch': A longitudinal study of teachers learning to teach English pronunciation over a six-year period. *Applied Linguistics, 42*(4), 791–809. https://doi.org/10.1093/applin/amab009

Couper, G. (2016). Teacher cognition of pronunciation teaching: Teachers' concerns and issues. *TESOL Quarterly, 51*(4), 820–843. https://doi.org/10.1002/tesq.354

Darcy, I., Rocca, B., & Hancock, Z. (2020). A window into the classroom: How teachers integrate pronunciation instruction. *RELC Journal*, *52*(1), 110–127. https://doi.org/10.1177/0033688220964269

Guerra, L. (2017). Students' perceptions and expectations of native and non-native speaking teachers. In J. D. Martínez Agudo (Ed.), Native and non-native teachers in English language classrooms: Professional challenges and teacher education (pp. 183–204), De Gruyter Mouton. https://doi.org/10.1515/9781501504143-010

Ivanová, J. (2011). *Hodnocení orálních dovedností budoucích učitelů anglického jazyka se zaměřením na výslovnost.* PhD Dissertation. Palacký University Olomouc.

Jakšič, J. (2018). *Comprehensibility and pleasantness of different varieties of English as judged by Czech listeners*. [Unpublished master's thesis. Charles University.

Kobák, A. (2017). *(Non)rhoticity in English pronunciation teaching*. [Unpublished master's thesis. Charles University in Prague.

Kochem, T., & Levis, J. M. (2022). Evolving teaching practices of pre-service language teachers in L2 pronunciation: A case study. *Ranam, 55*, 11–30.

Levis, J. M., & Sonsaat, S. (2019). Quality teacher education for pronunciation teaching in L2 classrooms. In J. D. Martínez Agudo (Ed.), *Quality in TESOL and teacher education: From a results culture towards a quality culture* (pp. 213–222). Routledge.

Nagle, C., Sachs, R., & Zárate-Sández, G. (2018). Exploring the intersection between teachers' beliefs and research findings in pronunciation instruction. *Modern Language Journal, 102*(3), 512–532. https://doi.org/10.1111/modl.12493

Sicola, L., & Darcy, I. (2015). Integrating pronunciation into the language classroom. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 471–487). Wiley Blackwell. https://doi.org/10.1002/9781118346952.ch26

Skarnitzl, R., & Hledíková, H. (2022). Prosodic phrasing of good speakers in English and Czech. *Frontiers in Psychology, 13*, 857647. https://doi.org/10.3389/fpsyg.2022.857647

Skarnitzl, R., & Rumlová, J. (2019). Phonetic aspects of strongly-accented Czech speakers of English. *AUC Philologica, 2019*(2), 109–128. https://doi.org/10.14712/24646830.2019.21

Szpyra-Kozłowska, J. (2018). Instructional models in the global context. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 232–246). Routledge.

Upton, C. (2015). British English. In M. Reed; J. M. Levis (Eds.), *The Handbook of English Pronunciation* (pp. 251–268). Wiley Blackwell. https://doi.org/10.1002/9781118346952.ch14

Volín, J., Poesová, K., & Weingartová, L. (2015). Speech melody properties in English, Czech and Czech English: Reference and interference. *Research in Language, 13*, 107–123. http://dx.doi.org/10.1515/rela-2015-0018

Volín, J., Weingartová, L., & Skarnitzl, R. (2013). Spectral characteristics of schwa in Czech accented English. *Research in Language, 11*, 31–39. https://doi.org/10.2478/v10015-012-0008-6

Appendix

Questionnaire

Gender: **Age:** **Attained level of education:**
City of pedagogical practice:
Type of school: _____

How long have you been teaching?_____

1) If you studied English at university, where was it? (multiple answers are possible)

a faculty of education a faculty of arts Other: _____

2) Have you stayed, studied or worked abroad (English-speaking or otherwise) for an extended period of time? If so, where and for how long?

_____

3) Which of the following varieties of English do you personally find most pleasant?

Received pronunciation General American Australian Canadian Scottish Irish Other: _____ I do not know

4) Do you strive to emulate any of the varieties above in your own speech? If so, which one?

_____

5) What variety of English do you teach (British English, American English, other)?

_____

6) What variety of English are the teaching materials that you use based on? (multiple answers are possible)

Textbooks (or titles):

_____

Other materials (books, magazines, films, songs...):

_____

7) What importance do you attribute to English pronunciation teaching?
(1 – not important, 5 – essential)
1 2 3 4 5

8) Please rate the importance of the following objectives for teaching English pronunciation:
(1 – not important, 5 – essential)

a) The learner will attain an accent close to the level of a native speaker.
1 2 3 4 5

b) The learner will understand the speech of native speakers easily.
1 2 3 4 5

c) The learner will be easily understood by native speakers.

1 2 3 4 5

d) The learner will easily communicate with other non-native speakers.
1 2 3 4 5

e) The learner is aware of the differences between varieties of English.
1 2 3 4 5

9) To what extent do you agree with the following statements?

a) A native speaker is a better pronunciation teacher than a non-native speaker.
definitely not – rather not – don't know – rather yes – definitely yes

b) Staying abroad, communication with native speakers or listening to authentic materials are more effective means of attaining English pronunciation than in-class pronunciation instruction.
definitely not – rather not – don't know – rather yes – definitely yes

c) Learners should be familiar with the symbols of the International Phonetic Alphabet (IPA) that are relevant to English. They should be able to associate them with the sounds that they represent.
definitely not – rather not – don't know – rather yes – definitely yes

d) Certain aspects of pronunciation are more important for successful communication than others.
definitely not – rather not – don't know – rather yes – definitely yes

e) British English is the most appropriate model for learning English as a foreign language.
definitely not – rather not – don't know – rather yes – definitely yes

f) It is possible to attain a native-like accent at any age.
definitely not – rather not – don't know – rather yes – definitely yes

e) American English is easier to understand for learners than British English.
definitely not – rather not – don't know – rather yes – definitely yes

About the authors

Lenka Čtvrtečková graduated from the Faculty of Education at Charles University with her bachelor thesis 'Perception of intrusive /r/ by Czech Learners of English' in the programme English Language Focused on Education. She completed a master's at the Institute of Phonetics, which is based at the Faculty of Arts, Charles University, with her diploma thesis on teacher cognition in pronunciation teaching.

Email: ctvrteckova.lenka@seznam.cz

Ondřej Fischer studied English and American Studies (BA) and English Language (MA) at the Faculty of Arts, Charles University. He also graduated from Jaroslav Ježek Conservatory where he studied composition. He currently works as a radio presenter and music director at a classical music radio station in Prague.

Email: ondrej.fis@gmail.com

Radek Skarnitzl is an Associate Professor at the Faculty of Arts, Charles University, Prague, and director of the Institute of Phonetics. He is interested in various aspects of speech communication, with expertise in speech prosody, acoustic phonetics, and second language (L2) acquisition. His research focuses on L2 pronunciation and on the effect of various pronunciation features on the socio-psychological evaluation of a speaker in both native and foreign languages. He is also interested in issues related to speaker identification and forensic phonetics, especially the effects of voice disguise on human and automatic speaker recognition performance.

Email: radek.skarnitzl@ff.cuni.cz

Duckinoska-Mihajlovska, I., & Kirkova-Naskova, A. (2023). A short teaching intervention on word-stress rules and pronunciation learning strategies: An exploratory study. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 46–60). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8174024

A short teaching intervention on word-stress rules and pronunciation learning strategies: An exploratory study

Ivana Duckinoska-Mihajlovska
Ss. Cyril and Methodius University, Skopje

Anastazija Kirkova-Naskova
Ss. Cyril and Methodius University, Skopje

Research has shown that word stress is important for improved intelligibility in an EFL context (e.g., Cutler, 2015, Levis, 2018). However, instruction on word stress is frequently avoided in the EFL classroom due to time limitation, which begs the question whether a shift of focus from classroom learning to autonomous learning by exploiting learning strategies is a viable option for overcoming time constraints. For instance, longer instruction in language learning strategy use has led to the improvement of general oral proficiency (Nakatani, 2005) or specific pronunciation features such as word stress, linking, and primary phrase stress among learners with different L1s (Sardegna, 2011, 2012; Sardegna & Dickerson, 2023), as well as greater learner autonomy.

This study investigates whether short word stress and strategy instruction yields improvement in learners with the same L1 in an EFL classroom setting. Forty Macedonian learners were assigned to a treatment and a control group (n = 20 each) and completed pre-, post-, and delayed post-tests. Only the treatment group received a four-week instruction which targeted stress placement in polysyllabic words based on four word-stress rules following the Covert Rehearsal Model (CRM) (Hahn & Dickerson, 1999). Learners were also taught to use pronunciation learning strategies (PLSs) for self-regulated practice out of class and completed a strategy diary. Results show that even a short teaching intervention on word stress and strategy use is beneficial for learners' ability to accurately apply word stress rules in production.

Keywords: pronunciation instruction, word-stress rules, pronunciation learning strategies, Covert Rehearsal Model (CRM)

1 Introduction

Word stress has received divided attention in pronunciation instruction over the years – from being prioritised as a feature leading to intelligible speech, to being considered a non-core feature in the Lingua Franca Core (Jenkins, 2000). Recently, its importance for intelligibility has been more widely acknowledged, as studies have shown that accurate word stress placement contributes to more intelligible non-native speech (Hahn, 2004) and that stress misplacement at the beginning of a conversation could affect word decoding thus hindering further message processing (Levis, 2018).

Unlike languages where stress is fixed to a particular syllable in a word, English word stress is lexically designated, i.e., its placement is governed by the word itself. Non-native learners coming from diverse linguistic backgrounds have difficulty predicting English word stress. They may mainly rely on their intuition and L1 stress patterns because descriptive stress rules seem complicated. Despite this, some authors argue that English word stress is not as random as it appears (Fritz & Kotzor, 2022) and it can be predicted in polysyllabic words by use of orthography-based rules (Dickerson, 2013, 2015). Word stress is also regarded as teachable and learnable; research shows improvement in word stress acquisition through teaching simplified rule-based strategies (Sardegna & Dickerson, 2023).

The current study explores the acquisition of English word stress by Macedonian EFL learners through the use of predictive stress-placement rules and learning strategies as described in the Covert Rehearsal Model (CRM) (Dickerson, 2013; Hahn & Dickerson, 1999). Such instruction assumes important roles for both teachers and learners; teachers are expected to select appropriate strategies and design suitable materials (Oxford, 1990), while learners are expected to use the strategies to guide them to more autonomous learning and to encourage further practice outside the classroom (Sardegna, 2011).

2 Previous research

Strategy instruction presupposes the identification of appropriate pronunciation learning strategies (PLSs), which have been defined as "steps taken by students to enhance their own pronunciation learning" (Peterson, 2000, p. 7). Although different studies employ diverse taxonomies that yield different results, strategy instruction with PLSs has proved to be effective in promoting learner autonomy and depends on various factors such as target feature, duration of teaching intervention, type of assessment, and strategy classification (Pawlak & Szyszka, 2018). Research evidence suggests that the success of strategy use in improving a particular pronunciation feature can be conditioned by the task type and learner's pronunciation knowledge. For instance, Szyszka (2021) had 58 first-year English majors (identified as high- or low-achievers) fill in a questionnaire that elicited PLSs they used while completing six tasks on vowels and diphthongs. She found that while some strategies were employed across all six tasks, there were differences in use between the high- and low-achievers, with the former group using a greater number and wider range of PLSs than the latter.

Strategy instruction has led to favourable results with various phonological phenomena. For instance, Haslam (2010) found that certain PLSs accounted for accuracy and improved comprehensibility, but not global foreign accent and fluency among EFL and ESL learners after a 10-week instruction on strategy use. Another instance is Ingels (2011) who investigated whether a self-monitoring program over a 16-week period improved participants' suprasegmental features. The results revealed that the self-monitoring strategies (critical listening, transcription, annotation, and rehearsal of corrections) led to improvement mainly in identifying message unit boundaries, linking, and vowel reduction in function words. Furthermore, Sardegna and MacGregor's results (2013) showed that scaffolded teaching with

PLSs led to the improvement of read-aloud accuracy of vowel reduction, linking, primary stress, and intonation during a 15-week intervention.

Research also shows that use of PLSs enhances learners' autonomous learning. Dickerson (2013, p. 5), for instance, suggests using the Covert Rehearsal Model (CRM), which aims to equip learners with a set of rules and learning strategies that guide them during their private practice of target structures. The CRM is a self-practising learning sequence consisting of six steps that centres around the concept of strategy instruction, where learners practice aloud in privacy out of class (steps 1 and 2), then they self-monitor and compare their performance with other models (steps 3 and 4), and finally they self-correct their performance and practice until satisfied and fluent (steps 5 and 6). The effectiveness of the CRM regarding different pronunciation features, including word stress, has been documented by a series of studies carried out by Sardegna (2011, 2012) and Sardegna and Dickerson (2023). In Sardegna (2011), the results indicate that a four-month intensive instruction with PLSs led to short- and long-term improvement in linking among learners with different L1s. In another study, Sardegna (2012) investigates the effect of individual learner differences when mastering linking and English word stress and found that these differences can predict individual progress over time. Sardegna and Dickerson (2023) tested the effect of three stress rules on the improvement of English word stress use by focusing on the extent to which ESL learners with different L1s practised in covert rehearsal and used PLSs after the instruction period. The findings reveal that the instruction led to improvement in the intervention group for all three stress rules. With regard to PLS use, the results showed a preference for perception strategies over prediction and production strategies.

Despite these promising results, there is insufficient research into strategy instruction focusing on word-stress rules with learners of a shared L1 and for a short instructional period. The current study aims to fill this gap by investigating a short strategy instruction in covert rehearsal for English word stress by Macedonian L1 learners. Given that Macedonian word stress is fixed and falls on the first syllable in disyllabic words or on the antepenultimate syllable in three or more syllable words (Koneski, 2004), Macedonian learners might regard English word stress placement as arbitrary and irregular. Hence, they could benefit from a specific instruction that equips them with a set of rules and learning strategies for practising English word stress placement.

3 Research methodology

3.1 Research questions

Considering the importance of word stress for intelligibility and the potential of the CRM for promoting learner autonomy, the present study aims to answer the following research questions:

RQ1: Does a four-week teaching intervention with the CRM approach lead to improvement of learners' word-stress placement in polysyllabic words? Which stress rule pattern (KSR, VSR, LSR, PSR) is the most effectively learnt?

RQ2: What learning strategies are used by learners in covert rehearsal?

3.2 Participants

Forty Macedonian EFL learners participated in the study. They were divided into a treatment group (n = 20; M = 3, F = 17; M_{age} = 20, age range 19–28) and a control group (n = 20; M = 3, F = 17; M_{age} = 19.45, age range 19–21). The participants were first-year English majors at Ss.

Cyril and Methodius University in Skopje, enrolled in the course Modern English 2 which targets language skills at B2 level according to the Common European Framework of Reference for Languages (Council of Europe, 2001). They had no prior formal knowledge or instruction in English pronunciation.

To determine whether the groups were similar or different in their performance at pre-test (T1), a *t*-test for independent groups was conducted. The results show that there was no statistically significant difference between the two groups at T1 with regard to their initial combined read-aloud accuracy scores ($t(38) = -0.18$, $p > 0.05$; $M_{TG} = 11.45$ vs. $M_{CG} = 11.6$). This indicated that the two groups were similar enough to be compared in further analyses (see §3.4).

3.3 Treatment procedure

Only the treatment group received formal instruction on four orthographic word-stress rules for polysyllabic words (Hahn & Dickerson, 1999): Key Stress Rule (KSR), Left Stress Rule (LSR), V/VC Stress Rule (VSR), and Prefix Stress Rule (PSR). In order to apply these rules, participants learnt how to identify parts of speech, affixes, syllable structure, and stressed/unstressed syllables. They focused on identifying the Key Syllable (positioned at the end of a word or left of an ending) and the Left Syllable (positioned to the left of the Key Syllable), either of which is always the main stress-carrier in the rule patterns presented in Table 1.

Table 1

Characteristics of KSR, LSR, VSR, and PSR Stress-Rule Patterns

Pattern	Word endings	Rule	Examples [a]
	Key Rule Endings:		
Key Stress Rule (KSR)	*-ia* (+V/C) *-io* (+ V/C) *-iu* (+ V/C) *-ienC*	Stress Key Syllable.	*rem**ed**(ial* *fall**ac**(ious* *cons**ort**(ium* *conv**en**(ienc(e*
	Other endings that follow Key Rule Endings:		
	-er, -ive *-al, -able, -ate* *-y, -ary, -ory* *-ize/ise, -ist, -ism* *-alise/alize, -alist, -alism*		*exec**ut**(ioner* *devi**at**(ional* *conc**il**(iatory* *cre**at**(ionism* *r**at**(ionalism*
Left Stress Rule (LSR)	*-y, -ies* (plural ending) *-fy, -fies, -fied,-fier, -fying* [b] *-ate, -ated, -ating,* *-ator* *-acy, -acies*	Stress Left Syllable.	*homogen**e**it(y* *common**al**it(ies* *f**ort**if(ied* *inc**in**er(ate* *p**erp**etr(ator* *deg**en**er(acy*

Pattern	Word endings	Rule	Examples [a]
V/VC Stress Rule (VSR)	*-al* (adj. only) *-ous* (adj.) *-ant* (adj. & n.) *-ance* (n.) *-ancy* (n.) *-ent* (adj. & n.) *-ence* (n.) *-ency* (n.) *-ic* [c] (adj. & n.)	1. Stress Left Syllable if Key Syllable is spelled with a V or VC. 2. Stress Key Syllable if spelled otherwise.	*in**aug**ur(al* ***om**in(ous* *c**ompl**im(ent* *c**omp**et(ency* *photogr**aph**ic* [c] *cub**oid**(al* *dis**astr**(ous* *adol**esc**(ent* *acc**ount**(ancy*
Prefix Stress Rule (PSR)	*-ary* *-ery* *-ory* *-ive* *-ative* *-atory* *-ature*	1. Stress Left Syllable if the prefix [d] is not part of the Left Syllable. 2. Stress Key Syllable if otherwise.	***arb**itr(ary* *inqu**is**it(ive* *c**ar**ic(ature* *inf**irm**(ary* *obj**ect**(ive* *cons**erv**(atory*

[a] The underlined syllable is the Key Syllable and the stressed syllable is in bold.
[b] The *-f-* in *certifies/certified* is not part of the left rule ending, but it helps identify the set of words to which the LSR applies.
[c] The VSR rule also applies to adjectives and nouns with a final *-ic*. Here, however, the final *-ic* is the Key Syllable not an ending, which is why there is no open parenthesis to mark the rule ending as in the examples *photogr**aph**ic* (adj.) and *econ**om**ics* (n.). Note that *-s* in *economics* is considered a neutral ending.
[d] Neutral prefixes are ignored when analysing words: *counter-/contra-, inter-/intro-, extra-, over-, retro-, super-*. Regular prefixes are relevant in the Left Syllable: *de-, re-, pre-, pro-, per-, ad-, ab-, ob-, sub-, in-, com-, con-, ex-, dis-*.

The teaching intervention also included training in strategy use. As a reference point for the PLS classification, we used the Prediction, Perception, Production (3Ps) Model (Dickerson, 2013) which emphasises the connection between orthography and prediction, thus facilitating perception and production. Once learners were equipped with predictive stress rule patterns, they could rely on orthography in covert rehearsal to practise and monitor their performance. They also used learning strategies to guide them when completing homework assignments (see Table 2). The intervention was conducted as eight 45-minute sessions over a four-week period. Lesson 1 introduced participants to the CRM, the PLSs, how to keep a strategy diary, and how to use online tools as speech models. Lessons 2–8 focused on syllable structure, identifying the Key and Left Syllables, and the four word-stress rules. Throughout the intervention, all participants in the treatment group attended lessons regularly, practised in class and at home, and completed seven homework assignments. The instructor gave explicit instructions about word-stress rules, suggested additional resources for the practice of the rules, provided opportunities for practice in/out of class, and supervised students' homework. They all kept a diary about their strategy use after each homework assignment. To facilitate their diary entry writing, they were given a list of questions as prompts and instructed to write either in their mother tongue or in English. They were required to send each diary entry to the course instructor who provided assistance when necessary.

Table 2

Pronunciation Learning Strategies

Strategy type	Code	Description
Prediction strategies	PRE1	I analyse the spelling to identify the syllables in a word.
	PRE2	I analyse word endings to identify the Key and Left Syllable in a word.
	PRE3	I use word endings to decide which syllable to stress in a word.
Production strategies	PRO4	I record myself saying polysyllabic words and then compare my own production against that of the model.
	PRO5	I listen to speech models and imitate their pronunciation of a word.
	PRO6	I read aloud a word several times and pay attention to which syllable is the loudest.
	PRO7	I read aloud sentences/passages with the target word.
	PRO8	I use the target word in a sentence.
Perception strategies	PER9	I listen to speech models (online tools/recorded material/native speakers).
	PER10	I listen to recorded material to identify the stressed syllables in words.
	PER11	I highlight or underline the stressed syllable in a word.

3.4 Data collection and analysis

This study uses a mixed-method design. For the quantitative part of the data, the participants completed pre-, post- and delayed post-tests. The test consisted of 20 target words, 5 per each stress rule and 11 distractors (see Appendix). Using the same test words in a shuffled order for all three tests allowed for a more objective measure of participants' progress over time. They were recorded saying each word aloud using the audio recording software Audacity[1]. The pre-test (T1) was conducted before the intervention, with a post-test (T2) after the intervention in week five, and a delayed post-test (T3) three weeks after the post-test. A total of 2400 tokens were analysed for accurate stress placement (1200 words per group). The benchmark against which the participants' pronunciation of the target words were assessed was the Cambridge Online Dictionary[2], i.e., the dictionary's audio recordings of the same stimuli words. Care was taken that the stimuli words had identical stress patterns in both British and American varieties. All tokens were marked by the instructor – value 1 was given for correct stress placement and value 0 for incorrect stress placement. The 0 category also included mispronounced words, e.g., **disorientantive* for *disorientate*, or truncation of syllables, e.g., **delination* for *delineation.*

A series of two-way mixed ANOVAs were performed to test the main effect of TIME (pre-test T1 vs. post-test T2 vs. delayed post-test T3) and GROUP (treatment vs. control), as well as their interaction effect (TIME x GROUP) on stress patterns (combined scores, as well as separate scores for KSR, VSR, LSR, and PSR). A follow-up one-way repeated measures ANOVA was applied to analyse differences in stress patterns as measured at T1, T2 and T3 among participants in the treatment and the control group separately. In all analyses, Mauchly's

[1] Audacity https://www.audacityteam.org/
[2] Cambridge Dictionary [Online] https://dictionary.cambridge.org/

test of sphericity was non-significant, implying that the assumption for using mixed ANOVA was met; therefore, mixed ANOVA and follow-up one-way repeated measures ANOVA could be applied. The Box's test of equality of covariance matrices was also not significant, indicating that the covariance matrices are equal. In addition, a *t*-test was used to analyse if there were differences in pre-test scores (baseline measures) on all stress patterns between the treatment and the control group; the results were not significant, indicating that the two groups did not differ at T1 (see §3.2).

As qualitative data, the participants' diaries were analysed to gain insight into their strategy use and to stimulate reflection on learning processes during assignment completion and private practice (Goh, 1997; Pawlak & Szyszka, 2018). Each participant from the treatment group completed seven diary entries, which were first analysed for the type of strategy used (identified using the strategy descriptions 1–11 in Table 2), and then the total number of responses for each reported strategy per participant was counted.

4 Results

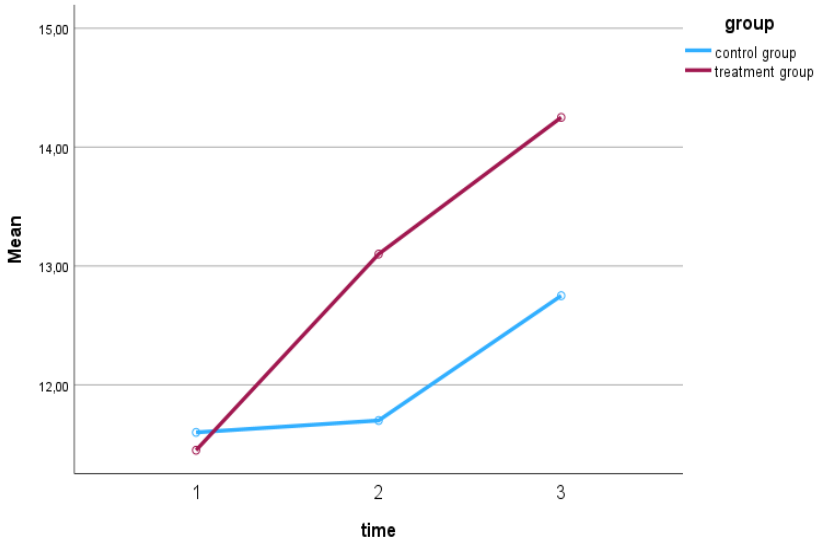
4.1 Effects of instruction

To test whether the treatment, i.e., the teaching intervention, yielded improvement of word-stress rules instruction in the participants' production of test words, subject variation from T1–T2 and from T1–T3 was calculated using a series of two-way mixed ANOVAs. The scores from all test words for all four stress rule patterns were combined and analysed to check for the effectiveness of the instruction as a whole. The scores from test words were further grouped by stress rule pattern (KSR, LSR, VSR, PSR) and analysed separately to check which pattern demonstrated best results.

The results for the overall instruction with combined scores for all stress rule patterns indicated no significant main effect of GROUP (treatment vs. control), $F(1, 38) = 0.99$, $p = .325$, $\eta^2 = .02$. However, they did indicate a significant main effect of TIME (T1, T2, T3) $F(2, 37) = 21.71$, $p = .00$, $\eta^2 = .54$, as well as a statistically significant interaction of TIME and GROUP $F(2, 37) = 4.04$, $p = .026$, $\eta^2 = .18$. Hence, while the two groups did not differ significantly because improvement was evident in both groups over time, the treatment group demonstrated greater improvement consistently from T1 to T3 compared to the control group (Figure 1, Table 3).

Figure 1

Treatment vs. Control Group: Overall Improvement



Note. TG n = 20; CG n = 20.

Table 3

Mean and Standard Deviation Values for the Treatment and Control Groups at T1, T2 and T3

Group	T1_all		T2_all		T3_all	
	M	SD	M	SD	M	SD
Treatment	11.4500	2.28208	13.1000	3.00701	14.2500	2.86310
Control	11.6000	3.15228	11.7000	3.98814	12.7500	3.16020

Note. TG n = 20; CG n = 20; all = combined accuracy scores of read-aloud polysyllabic words (KSR, VSR, LSR, and PSR).

A follow-up one-way repeated measures ANOVA revealed significant differences in mean scores across testing times for the treatment group ($F(2, 18) = 21.98$, $p = .000$, $\eta^2 = .71$) indicating that these participants demonstrated differences regarding the time they were tested. Pairwise comparisons based on Bonferroni adjustment at three testing times revealed that participants' test scores at T2 were significantly higher than T1 scores ($M_{T2} = 13.1$ vs. $M_{T1} = 11.45$; $p < 0.01$). Moreover, T3 scores were significantly higher than both T1 scores ($M_{T3} = 14.25$ vs. $M_{T1} = 11.45$; $p < 0.001$) and T2 scores ($M_{T3} = 14.25$ vs. $M_{T2} = 13.1$; $p < 0.05$), suggesting that participants' improvement was steady over time and their knowledge of word-stress rules was retained.

Significant differences in mean scores across testing times were registered for the control group as well ($F(2, 18) = 4.29$, $p = .030$, $\eta^2 = .32$). Pairwise comparisons at three testing times showed that the difference in participants' test scores at T3 in comparison to their test scores at T1 was significant ($M_{T3} = 12.75$ vs. $M_{T1} = 11.60$; $p < 0.05$), while no statistically significant

differences in the test scores at T1 vs. T2 (M_{T1} = 11.60 vs. M_{T2} = 11.70), and T2 vs. T3 (M_{T2} = 11.70 vs. M_{T3} = 12.75) were found.

To see which rule pattern was most successfully applied after the teaching intervention, participants' score variation was calculated for each stress rule category separately. Table 4 shows that participants' initial proficiency performance in the treatment and the control group was similar for all four categories, i.e., for KSR ($t(38) = 0.78$, $p > 0.05$; M_{TG} = 2.6 vs. M_{CG} = 2.4),VSR ($t(38) = -0.74$, $p > 0.05$; M_{TG} = 3.15 vs. M_{CG} = 3.4), LSR ($t(38)= -0.19$, $p > 0.05$; M_{TG} = 2.85 vs. M_{CG} = 2.9), and PCR ($t(38) = -0.78$, $p > 0.05$; M_{TG} = 2.8 vs. M_{CG} = 2.9).

The analyses of KSR, LSR, and VSR scores revealed a similar trend for these three stress rules (Figure 2, Table 4). The results for the KSR pattern indicated no significant main effect of GROUP, $F(1, 38) = 2.42$, $p = .128$, $\eta^2 = .02$, a significant main effect of TIME, $F(2, 37) = 4.21$, $p = .023$, $\eta^2 = .18$, and no statistically significant interaction of TIME and GROUP, $F(2, 37) = 1.01$, $p = .376$, $\eta^2 = .05$. The results for the LSR pattern indicated no significant main effect of GROUP, $F(1, 38) = 0.89$, $p = .351$, $\eta^2 = .02$, a significant main effect of TIME, $F(2, 37) = 10.92$, $p = .000$, $\eta^2 = .37$, and no statistically significant interaction of TIME and GROUP, $F(2, 37) = 1.86$, $p = .171$, $\eta^2 = .09$. The results for the VSR pattern indicated no significant main effect of GROUP, $F(1, 38) = 1.03$, $p = .317$, , $\eta^2 = .03$, a significant main effect of TIME, $F(2, 37) = 4.40$, $p = .019$, $\eta^2 = .19$, and no statistically significant interaction of TIME and GROUP, $F(2, 37) = .51$, $p = .606$, $\eta^2 = .03$. According to these results, both groups showed similar improvement from T1 to T3 (see Table 4); however, this improvement is not substantial enough to be attributed to the teaching intervention.

The results for the PSR pattern indicated no significant main effect (borderline) of GROUP, $F(1, 38) = 3.88$, $p = .056$, $\eta^2 = .02$, but a significant main effect of TIME, $F(2, 37) = 7.79$, $p = .002$, $\eta^2 = .30$, as well as a statistically significant interaction effect of TIME and GROUP, $F(2, 37) = 6.01$, $p = .005$, $\eta^2 = .25$. Such a trend for the PSR pattern can be observed in Figure 2. The results for the control group are similar at each testing time showing modest improvement, while the results for the treatment group demonstrate noticeable improvement from T1 to T3, indicating that the instruction was effective for this stress rule pattern.

Table 4

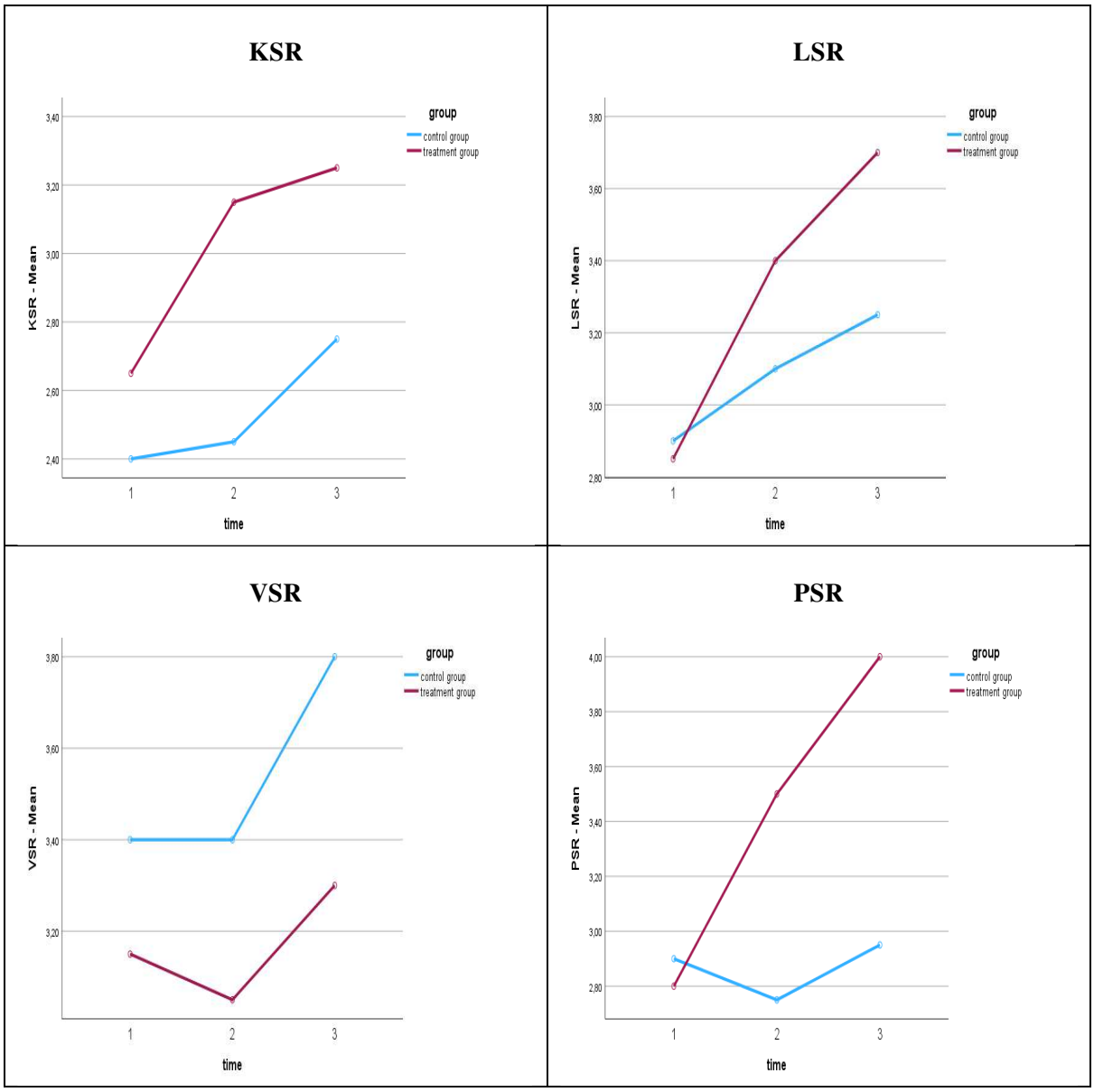
Mean and Standard Deviation Values for the Treatment and Control Groups at T1, T2, and T3 for KSR, LSR, VSR, and PSR

Group		KSR		LSR		VSR		PSR	
		M	SD	M	SD	M	SD	M	SD
Treatment	T1	2.6500	1.03999	2.8500	0.81273	3.1500	0.87509	2.8000	0.95145
	T2	3.1500	0.98809	3.4000	0.94032	3.0500	1.23438	3.5000	1.14708
	T3	3.2500	1.20852	3.7000	1.03110	3.3000	0.97872	4.0000	0.91766
Control	T1	2.4000	0.99472	2.9000	0.85224	3.4000	1.23117	2.9000	1.29371
	T2	2.4500	1.39454	3.1000	0.96791	3.4000	1.66702	2.7500	1.25132
	T3	2.7500	1.06992	3.2500	0.85070	3.8000	1.39925	2.9500	0.99868

Note. TG n = 20; CG n = 20.

Figure 2

Treatment vs. Control Group Progress for KSR, LSR, VSR, and PSR Stress-rule Patterns



Note. TG n = 20; CG n = 20.

4.2 Use of pronunciation learning strategies

Data from participants' diary entries was coded and analysed to identify which strategies were used during completion of homework assignments 1–7. The results in Table 5 show that all suggested strategies were used, with different combinations of strategy categories. The participants most frequently used prediction strategies either to identify syllables (PRE1 = 18), to analyse word endings for Key/Left Syllable identification (PRE2 = 59), or to use word endings to decide which syllable to stress (PRE3 = 46). Three out of five production strategies were also frequently used: the participants were recording themselves saying polysyllabic words and comparing their production against a model (PRO4 = 49); just listening to speech

models and imitating their pronunciation of the word (PRO5 = 36); or reading aloud a word and paying attention to the loudest syllable (PRO6 = 21). The remaining two production strategies were hardly ever used, i.e., participants did not choose to practise reading aloud sentences/passages with the target word (PRO7 = 4), or using the target word in a sentence (PRO8 = 2). The perception strategies were sporadically used compared to prediction and production strategies; when used, they included listening to speech models such as recordings of native speakers (PER9 = 18), listening to recordings to identify the stressed syllable (PER10 = 4), or highlighting/underlining the stressed syllable in a word (PER11 = 8).

Comparing the number of strategies reported in a single assignment, fewer strategies were used for homework 1–2, which focused on identification of syllable structure and Key/Left Syllable. The number of strategies used increased for homework 3–7, which focused on the four stress-placement rules, indicating that participants chose to apply the rules practised in class, and that during covert practice they experimented with a combination of strategies from all three categories.

Table 5

Frequency Count of Reported Pronunciation Learning Strategies

PLSs	Hw1		Hw2		Hw3		Hw4		Hw5		Hw6		Hw7		Total per PLS
	n	%	n	%	n	%	N	%	n	%	n	%	n	%	
PRE1	11	55	1	5	1	5	-	-	1	5	2	10	2	10	18
PRE2	-	-	14	70	13	65	6	30	12	60	6	30	8	40	59
PRE3	-	-	1	5	7	35	8	40	11	55	9	45	10	50	46
PRO4	7	35	-	-	8	40	10	50	8	40	8	40	8	40	49
PRO5	6	30	-	-	3	15	6	30	7	35	7	35	7	35	36
PRO6	5	25	1	5	2	10	4	20	3	15	2	10	4	20	21
PRO7	-	-	-	-	1	5	1	5	2	10	-	-	-	-	4
PRO8	-	-	-	-	-	-	2	10	-	-	-	-	-	-	2
PER9	2	10	-	-	2	10	2	10	4	20	5	25	3	15	18
PER10	-	-	-	-	3	15	-	-	-	-	-	-	1	5	4
PER11	1	5	1	5	1	5	2	10	-	-	1	5	2	10	8
Total per Hw	32		18		41		41		48		40		45		

Note. TG *n* = 20; PRE = prediction, PRO = production, PER = perception, Hw1–7 = homework 1–7.

5 Discussion

The first research question focused on whether instruction under the CRM for a short treatment period improved learners' word stress placement in polysyllabic words. It also addressed which stress rule pattern was most effectively learnt. Our findings revealed that although the treatment and the control group did not differ (the main effect of GROUP was statistically insignificant), the treatment group achieved better scores over time (the main effect of TIME was statistically significant as well as the interaction of TIME and GROUP). These results may indicate that the participants who received instruction tend to demonstrate enhanced ability to accurately

stress polysyllabic words compared to the participants who did not receive instruction. A possible explanation about the improvement demonstrated by the participants in the control group might be that during the treatment period they attended other English classes specialised in linguistics and literature, hence, such regular language exposure may have affected their progress from T1 to T3 positively. At the same time, the participants in the treatment group also attended the same classes and were exposed to the same language input. This might mean that for them the teaching intervention seems to make a difference on its own, because their improvement in comparison to the participants in the control group was noticeable over time (from T1 to T2 to T3). In addition, the separate analyses by stress rule pattern showed that the effect of instruction was indicative for KSR, LSR, VSR, and particularly evident for PSR. Learners appear to have benefited from orthographic rules for stress placement, which concurs with Sardegna and Dickerson (2023). The improvement of the treatment group was moderate but steady over testing times, so we may infer that the length of the intervention was too short. Nonetheless, even a short explicit teaching intervention of a particular pronunciation feature, such as word stress, might also prompt learners' awareness of that feature. As for the learnability of stress rule patterns, our results indicated that some rules seem easier to grasp (PSR) than others (KSR, LSR, VSR); although statistically insignificant, the results for KSR, LSR, and VSR appeared to indicate improvement, implying that the teaching intervention for these rules may have been helpful.

The second research question addressed the types of strategies used by learners during self-practice. The results indicated that learners, once trained how to employ strategies to their advantage, chose to combine them to complete assigned tasks. The two most commonly reported strategies were prediction strategies PRE2 and PRE3, i.e., the learners relied on analysing word endings to find the stressed syllable. This provides further support to the claim that, when orthographic stress-placement rules are practised, prediction facilitates learning. Strategy use seems to complement consolidation of knowledge acquired through explicit rule instruction and to help the learning process in general, which is consistent with results from research into strategy use (Ingels, 2011; Nakatani, 2005; Szyszka 2021).

The teaching intervention participants received was fully integrated in the course Modern English 2, which was mandatory in their curriculum. This aspect is relevant, as it shows that such integration is achievable and successful. Furthermore, time restrictions (often considered a major drawback for pronunciation instruction) can be overcome. It further supports the need for regular practice over time, for pronunciation rules to be internalised.

Effective application of rules and strategies has important implications for pronunciation pedagogy. The results of this study suggest that word-stress rules show great potential for integration into ESL/EFL teaching syllabi. Learners need explicit rules; hence, suggesting online resources for self-monitoring their performance, such as online dictionaries, speech models (e.g., Youglish, TED talks)[3] and voice recorders (e.g., Vocaroo, Voice spice)[4] might strengthen their motivation and willingness to improve. Given that learning strategies reinforce self-regulated practice (Pawlak & Oxford, 2018), teachers could select strategies to suit their learners' needs, provide training with those strategies, and monitor learners' progress with strategy diaries.

[3] YouGlish https://youglish.com/
TED talks https://www.ted.com/talks
[4] Vocaroo https://vocaroo.com/
Voice spice https://voicespice.com/

6 Conclusion

The aim of this study was to examine whether a short-period teaching intervention consisting of orthographic word-stress rules and pronunciation learning strategies under the CRM could help learners improve their word stress placement accuracy. It also took into consideration the learner profile (shared L1) and context (EFL classroom setting, integrated instruction in a general English language course). The results suggest that such instruction is beneficial for learners over time, as measured by a read-aloud task. The combination of rules and strategies tends to improve learners' ability to recognise and understand novel pronunciation structures, and, therefore, increases learner autonomy for self-practice outside the classroom.

Future research could investigate whether such combined rule and strategy instruction translates into improvement in word stress placement accuracy in spontaneous speech. Further research into the link between word stress placement and vowel quality change (for instance, nuances such as correct stress/correct vowel, or correct stress/incorrect vowel) is needed to better understand the acquisition process of word stress. From a methodological point of view, future studies might consider a larger participant sample to obtain more generalisable results, including a comparison group with implicit instruction, and finding ways of controlling for parallel language input.

Acknowledgements

We wish to express our heartfelt gratitude to the anonymous reviewers whose insights were most helpful. We are also grateful to Radek Skarnitzl and Biljana Blaževska-Stoilkovska, for their assistance with the choice of statistical analyses.

References

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, and assessment (CEFR).* Cambridge University Press.

Cutler, A. (2015). Lexical stress in English pronunciation. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 106–124). Wiley Blackwell.

Dickerson, W. B. (2013). Prediction in teaching pronunciation. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–7). Wiley Blackwell.

Dickerson, W. B. (2015). Using orthography to teach pronunciation. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 488–504). Wiley Blackwell.

Fritz, I., & Kotzor, S. (2022). 'Category and a'nalysis: Navigating English stress. *Speak Out!, 67*, 18–25.

Goh, C. (1997). Metacognitive awareness and second language listeners. *ELT Journal*, *51*(4), 361–369. https://doi.org/10.1093/elt/51.4.361

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*(2), 201. https://doi.org/10.2307/3588378

Hahn, L. D., & Dickerson, W. (1999). *Speechcraft: Discourse pronunciation for advanced learners*. University of Michigan Press.

Haslam, N. (2010). *The relationship of three L2 learning factors with pronunciation proficiency: Language aptitude, strategy use, and learning context*. [Master's thesis, Brigham Young University, Provo, UT]. Brigham Young University. Theses and Dissertations 2111. https://scholarsarchive.byu.edu/etd/2111/

Ingels, S. A. (2011). *The effects of self-monitoring strategy use on the pronunciation of learners of English.* [Doctoral dissertation, University of Illinois at Urbana-Champaign]. IDEALS, Graduate Dissertations and Theses. https://hdl.handle.net/2142/26259

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.
Koneski, B. (2004). *Gramatika na makedonskiot jazik* [A grammar of Macedonian]. Detska radost.
Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation.* Cambridge University Press.
Nakatani, Y. (2005). The effects of awareness-raising training on oral communication strategy use. *The Modern Language Journa*l *89*(1), 76–91. https://www.jstor.org/stable/3588552
Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Heinle & Heinle Publishers.
Pawlak, M., & Oxford, R. L. (2018). Conclusion: The future of research into language learning strategies. *Studies in Second Language Learning and Teaching*, *8*(2), 525–535. https://doi.org/10.14746/ssllt.2018.8.2.15
Pawlak, M., & Szyszka, M. (2018). Researching pronunciation learning strategies: An overview and a critical look. *Studies in Second Language Learning and Teaching*, *8*(2), 293–323. https://doi.org/10.14746/ssllt.2018.8.2.6
Peterson, S. S. (2000). Pronunciation learning strategies: A first look. Unpublished research report. (ERIC Document Reproduction Service No. ED450599). https://eric.ed.gov/?id=ED450599
Sardegna, V. G. (2011). Pronunciation learning strategies that improve ESL learners' linking. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference* (pp. 105–121). Iowa State University.
Sardegna, V. G. (2012). Learner differences in strategy use, self-efficacy beliefs, and pronunciation improvement. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 39-53). Iowa State University.
Sardegna, V. G., & Dickerson, W. B. (2023). Improving the pronunciation of English polysyllabic words through orthographic word-stress rules. In V. G. Sardegna & A. Jarosz (Eds.), *English pronunciation teaching: Theory, practice and research findings* (pp. 81–97). Multilingual Matters.
Sardegna, V. G., & McGregor, A. (2013). Scaffolding students' self-regulated efforts for effective pronunciation practice. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 182–193). Iowa State University.
Szyszka, M. (2021). Pronunciation learning strategies: A task-based perspective. In A. Kirkova-Naskova, A. Henderson, & J. Fouz-Gonzales (Eds.), *English pronunciation instruction: Research-based insights* (pp. 147–172). John Benjamins. https://doi.org/10.1075/aals.19.07szy

Appendix

Test Words (n = 20) and Distractors (n = 11)

Rule pattern	Test words	Word endings
KSR	rep**ud**iate	-iaC
	deline**at**ion	-ioC
	cons**ort**ium	-iuC
	dis**or**ientate	-ienC +nonbasic ending
	d**ev**iance	-ianC
VSR	anc**estr**al	-al
	stup**end**ous	-ous
	econ**om**ics	-ic = Key syllable
	adol**esc**ent	-ant/ent
	extr**av**agancy	-ancy/ency

Rule pattern	Test words	Word endings
LSR	heterogen**e**i<u>t</u>y	-y
	incr**im**i<u>n</u>ate	-ate
	appr**ox**i<u>m</u>ate	-ate
	obsti<u>n</u>acy	-acy
	in**acc**u<u>r</u>acy	-acy
PRS	**arb**i<u>t</u>rary	-ary
	sava<u>g</u>ery	-ery
	exp**os**itory	-ory
	ind**ic**a<u>t</u>ive	-ive
	impl**ic**ature	-ature
Distractors		
reading, matched, ended, fighting, sniffing, acted, breaking, seemed, blaming, running, glued		

Note. C = consonant letter.

About the authors

Ivana Duckinoska-Mihajlovska is a Language Instructor in the Department of English Language and Literature, Ss. Cyril and Methodius University in Skopje, North Macedonia, where she teaches Contemporary English courses and Academic Writing. Her main interests include English phonetics and phonology, learner autonomy, and teacher and student well-being. She serves as treasurer of IATEFL's PronSIG.

Email: iduckinoska@flf.ukim.edu.mk

Anastazija Kirkova-Naskova is an Associate Professor in English phonetics and morphology in the Department of English Language and Literature, Ss. Cyril and Methodius University in Skopje, North Macedonia. She also teaches research methodology (MA level). Her research interests include English pronunciation teaching and learning, foreign-accented speech, speech perception, pronunciation learning strategies, and teacher cognition. Together with Alice Henderson and Jonás Fouz-González, she co-edited *English pronunciation instruction: Research-based insights* (John Benjamins, 2021). She was an assistant editor-in-chief for *The Journal of Contemporary Philology*.

Email: akirkova@flf.ukim.edu.mk

Edensor-Costille, K. (2023). Englishville: A new way of practising prosody. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 61–69). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8173981

Englishville: A new way of practising prosody

Kizzi Edensor-Costille
University of Caen Normandy

Despite evidence that prosody plays an important role in the intelligibility, comprehensibility and accentedness of non-native discourse (Munro & Derwing, 1995, 1998), it is seen as difficult to teach (Setter et al., 2010). One way of making prosody easier to teach and understand is by using a real-time 3D spectrogram such as the one used on the website Englishville (Costille, 2020). Four groups of French students, enrolled in their third year of a BA in English, took part in this experiment. Thirty short sentences focusing on intonation were recorded by a female native British speaker. All participants read and recorded the same phrases as they appeared on the screen and groups 3 and 4 received specific explanations regarding the spectrogram and intonation contours. The first group simply read the phrases (limited input) and recorded their own productions. The other 3 groups received supplementary input: group 2 read the text and heard the corresponding audio recordings (audio input); group 3 read the text and saw the corresponding 3D spectrogram (visual input); and group 4 read the text, heard the audio and saw the corresponding 3D spectrogram (multi-sensorial input). The recordings were then compared in Englishville to the expected intonation pattern and given one point per matching pattern. The results do not show that seeing speech systematically improves students' intonation but did show that the students felt the tool was useful and easy to use.

Keywords: prosody, L2 learners of English, multi-sensorial input, Englishville

1 Introduction

Teaching and research in second language acquisition has long focused on auditory sources of input for both training sessions and experiments. More recently, linguists have started to use multi-sensorial modalities, looking for ways to enhance second language (L2) learning. For segmentals, teachers have found ways to render the theoretical aspects of phonemes more comprehensible by using multi-sensorial techniques – be it with their own mouth, videos of another's mouth or by demonstrating on sagittal sections. Learners can simultaneously see the different positions of jaw, lips, and tongue, and hear the phonemes. The representation of intonation or lexical stress is more abstract and tools such as Praat (Boersma & Weenink, 2001), however helpful for research, remain difficult to use if learners are not previously trained how to use them (Setter & Jenkins, 2005; Setter et al., 2010).

It might be that acquiring or improving prosody is complex partly because of the lack of physical or visual aids available to learners. It could even be argued that its abstractness dissuades teachers from teaching it. This led to the creation of Englishville (Costille, 2020) – a website dedicated to practising prosody by means of a 3D spectrogram. An experiment was set up to test its usability and usefulness regarding prosody.

This chapter gives a brief overview of prosody and research on L2 learners with a special focus on the prosodic elements in the experiment. It also discusses research that has used auditory and multi-sensorial tools. It then describes the methodology used, followed by the experiment and results. Lastly, the findings are discussed in light of the possible contribution of Englishville to the field.

2 L2 prosody

Prosody, also known as suprasegmentals (to be understood as all that is not segmental), includes elements such as: rhythm, intonation, stress, and pauses. Drawing on research studies investigating first language (L1) acquisition, Hirst and Di Cristo (1998) explain that prosody is likely to be acquired by a child before any other phonetic features, moreover it is likely to be the last feature lost when aphasia strikes or when another language is learnt. The fact that we learn the prosody of our L1 in the very early stages of development can explain why it is difficult to learn later in an L2. Previous studies have confirmed that acquiring L2 prosody is challenging even for advanced learners (Colantoni et al., 2014). However, it is generally accepted that using inaccurate intonation patterns, i.e., such that differ from native productions, can lead native listeners' to either misinterpret the intended meaning or show negative stereotyping towards the L2 speaker.

Studies focusing on L2 pronunciation instruction and the relevance of L2 speech in speaker interactions have investigated various prosodic features. Jenkins (2000) finds the most important suprasegmental features in NNS–NNS interactions to be contrastive stress, the direction of pitch movements, word stress placement, and stress-timed rhythm. Other authors share a similar point of view on prosody. For instance, Munro and Derwing (1995, 1998) and Hardison (2004, 2010), argue that prosody plays a significant role in L2 speech; learners who received instruction on prosodic features (intonation, rhythm, word stress, and sentence stress) showed significant improvement in comprehensibility and accentedness compared with those who had only received instruction on segments. Furthermore, Derwing and Rossiter (2003) have shown that L2 fluency and comprehensibility significantly improved after a 12-week instruction period on the pronunciation of prosodic features.

2.1 Research on visual and auditory tools

Since the mid-1970s, there has been an on-going stream of studies which have used computer-based methods in order to test and improve the perception and production of prosody (de Bot, 1983; James, 1976). De Bot (1983) concluded that visual feedback was more effective than auditory feedback – in other words, when the subject saw speech (in this experiment, the pitch contour in Praat was used) rather than just hearing it, the subject's intonation improved. Pitch visualisers (such as Praat and similar software) have been used in more recent research (Imber et al., 2017; Kartushina, et al., 2015; Offerman, & Olson, 2016; Olson, 2014, Setter et al., 2010). Gorjian et al. (2013) compared two methods of teaching stress and intonation: a) a traditional one that uses repetition and explanations about acoustic properties of speech; and b) a computer-assisted one that uses Praat software. The results showed that learning prosody with Praat was significantly more beneficial. As the authors point out, the first method is generally teacher-centred, leaving students passive in the classroom. In contrast, the use of multi-media tools and multi-sensorial software places the student at the centre of their acquisition.

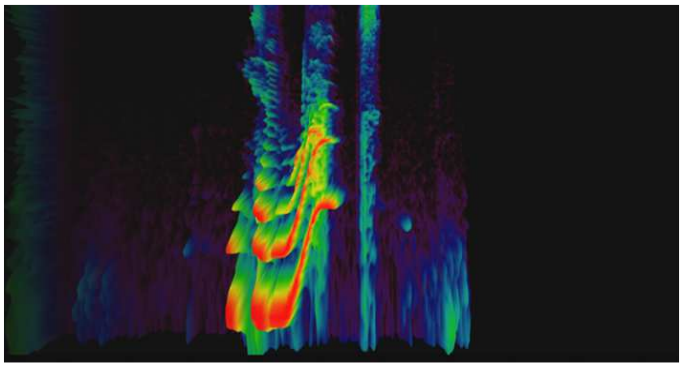
Software, such as Praat, require practice and training to use and some research has concluded that combining sound and image led to slightly more mixed results in learning prosody, often due to the complexity of the software used (Setter et al., 2010). Given the technical side of these tools, it can be difficult to motivate students to familiarise themselves with them and then work on intonation. The results of certain studies (Gorjian et al., 2013) showed that learners improved when using Praat, which could be explained simply by the additional time spent working on prosody to comprehend what they were seeing. Perhaps using real-time displays of intonation might prove to be even more comprehensible for the learners – they can see their speech appear as they speak, enabling them to test different intonation patterns more easily. We argue that the instantaneous effect of seeing speech makes intonation easier to perceive because not everyone can perceive it simply through their ears, some people need their eyes to validate or invalidate their aural perception. Therefore, depending on the pronunciation feature practised, the use of multi-sensorial tools and methods may be helpful, even having a global positive effect on L2 speech production in general.

3 Englishville: Methodological approach

The desire to create a free, user-friendly real-time tool for prosody in the domain of multi-sensorial learning of L2 English, and the idea that prosody should be at the centre of second language teaching motivated the creation of our website called "Englishville" (Costille, 2020). Englishville uses a 3D spectrogram and facilitates the capture of the audio stream so that it can be recorded on a server. These tools are then integrated in a website where it is possible to record a corpus, set up experiments and participate in them. One advantage of Englishville is that L2 learners can hear the audio, see the spectrogram and intonation contour and the corresponding text, then repeat and save their own productions. The spectrogram shows the direction of the tone of voice, making it possible to imitate a visual real-time model of an intonation pattern and to simultaneously compare it to one's own melodic pattern. The spectrogram used in Englishville can be seen in Figure 1 with an example of rising intonation.

Figure 1

A Real-time Display of Pitch Contour and Intensity in Englishville (Costille, 2020) of the Utterance "He said what?" Pronounced with a Rising Intonation



It is often considered difficult to perceive pitch movement for learners and even for teachers. Therefore, the main objective of Englishville and this experiment is to help learners see speech in order to improve their pronunciation of the different melodic patterns. The intonation patterns used in the experiment are simple. For example, it is generally acknowledged that the use of a falling tone on the nucleus indicates finality, and that a rising tone indicates non-finality (Wells, 2007). Wh-questions (open questions) are normally said with a falling tone on the last lexical item whereas Yes/No-questions (closed questions) are normally uttered with a rising tone on the nucleus. As for intonation in lists, Wells (2007) differentiates between 2 types: those that are finished and those that are not. To illustrate this point, he gives two versions of the same utterance and explains that "the fall on tea in (1) signals that there are no more options: you must choose either tea or coffee. The rise on tea in (2) signals that there may be other possibilities too, as yet unmentioned, e.g., or you could have an ↘orange juice" (p. 75).

(1) You can have ↗ coffee | or ↘ tea.
(2) You can have ↗ coffee | or ↗ tea.

The structure of the sentences used in the intonation task of the Englishville experiment correspond to the two main intonation patterns (fall, rise) and example (1) in the case of a closed list. French L2 speakers of English tend to struggle most with falling intonation and use rising intonation for all types of statements. This is typically noticeable when a speaker concludes an oral presentation with a rise, which can leave the listener frustrated and/or surprised when they realise that the presentation is, in fact, finished (the non-finality effect of rising intonation).

3.1 Research questions

The study aimed therefore to investigate the following research questions:

RQ1: Can seeing a real-time 3D spectrogram enable L2 learners of English to reproduce certain intonation patterns?
RQ2: What kind of input is more beneficial out of limited, audio, visual and multi-sensorial input?

3.2 Experiment

An experiment was set up with four groups of participants (see Table 1): Group 1 (control group) received limited input and read the sentences as they appeared on the computer screen; Group 2 had audio input, i.e., they heard the sentences as recorded by the female native speaker and saw the text for each sentence on the screen; Group 3 had visual input which corresponded to seeing the spectrogram and the melodic pattern (as in Figure 1) and read the text without hearing the audio recordings, and Group 4 had access to both aural and visual input. The aim of the experiment was to test if the participants who received multi-sensorial input (Group 4) produced better results, i.e., those closest to the expected intonation pattern.

Table 1

Experiment Conditions for each Testing Group: Type of Input and Procedure

Group	Type of input	Experiment procedure
1	limited	Records words and sentences.
2	auditory	Hears the utterances before recording them.
3	visual	Sees the spectrograms in Englishville.
4	multi-sensorial	Hears the utterances and sees the corresponding spectrograms on Englishville.

At the beginning of the experiment, the participants had one example to familiarise themselves with the user interface. Each participant read the same sentences in the same order as they appeared on the screen. They only heard or saw each item once, before recording their own production. They had to click the button *play* to hear or see the visual model recording and then had to click the microphone button on and off to record their own speech. They could not listen to the same sentence more than once, but they could pronounce it as many times as they liked while their microphone was activated. By pressing the microphone button a second time, their recording was saved, and the next utterance automatically appeared. At the beginning of the experiment, Groups 3 and 4 received supplementary information about what they were about to see. For example, they were informed that they would see the movement of the tone of voice in the spectrogram (downward or upward movement). The participants who received input, be it auditory or visual, were asked to imitate as closely as possible what they heard or saw.

In light of previous research, it was hypothesised that the combination of both audio and visual input would yield the best results. It was therefore expected that Group 4 would have better results than the other three groups because they would be able to see immediately if their spectrogram resembled or not the model and attempt to improve during the experiment.

3.3 Participants

Twenty French students (n = 5 in each Group 1–4; M = 8, F = 12) at the beginning of their third year of a BA in English Language and Literature at the University of Caen Normandy participated in this experiment. They were, on average, 20 years old and had been learning English for at least ten years. Their level of English was estimated to be B2+, based on their teacher's experience with the CEFR scales. They had all studied phonology and phonetics (including intonation). For these classes, the teaching model was British English. Prior to the

experiment, each participant completed an online questionnaire about their language and personal background.

The learner's evaluation of Englishville was also of high interest in this study, therefore, after the experiment they were asked to give feedback via an online questionnaire about Englishville and their impression of it and the experiment.

3.4 Stimuli

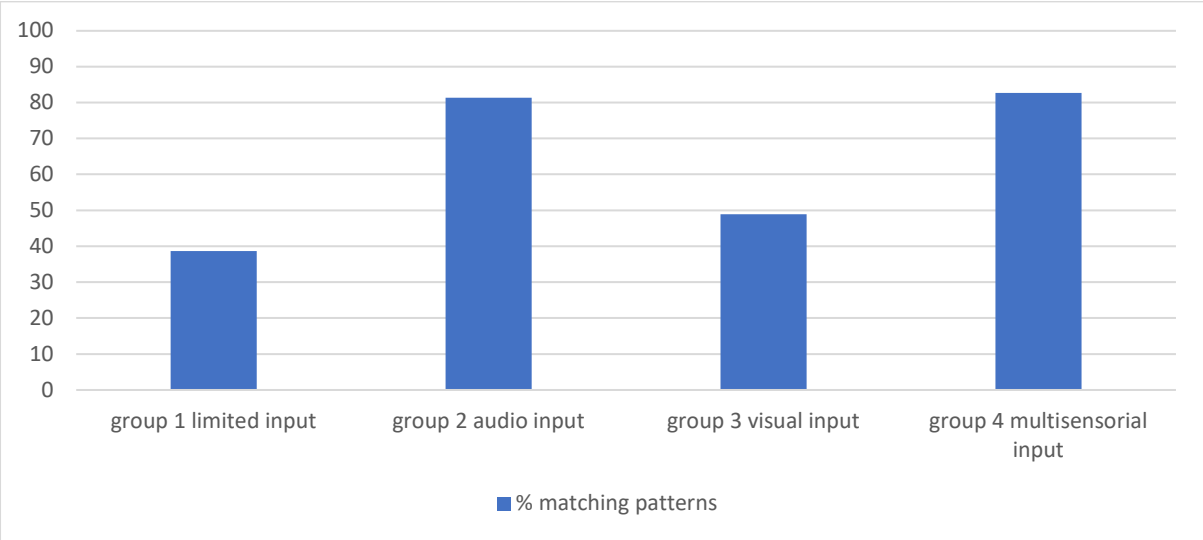
Thirty sentences consisting of simple patterns and short utterances were recorded by a female British native speaker, also the creator of Englishville. The sentences included ten statements and five of each of the following sentence types: Wh-questions, Yes/No-questions, echo questions and a two-element closed list. For example: *We live in London* (statement with falling intonation), *Where's the manual?* (Wh-question said with falling intonation), *May I lean on the railings?* (Yes/No-question said with rising intonation), *He is on the computer?* (echo-question with rising intonation), *Are you growing oranges or lemons?* (two-element closed list said with a rise followed by a fall).

4 Results

Recordings from the four groups were compared with the original recording and spectrogram to give an auditive and visual analysis. Intonation patterns found to match the model were awarded one point, for example, a final fall, rise or rise+fall (closed lists) whereas those that differed got zero. The number of matching realisations for intonation patterns were then calculated for each group, as can be seen in Figure 2.

Figure 2

Percentage of Matching Realisations of Intonation Pattern per Input Type (out of 30 sentences)



In general, the results show that having any kind of input is beneficial as Group 1, who received limited input, only matched 38.67% of the 30 sentences. For Group 3, seeing the spectrogram slightly improved their overall percentage of matching realisations (48.9%)

compared to Group 1 (marking an increase of 10.23% points). Group 2, who only heard the corresponding audio, had nearly the same percentage of matching realisations (81.33%) as Group 4 (82.67%), who benefitted from multi-sensorial input.

In the post-test questionnaire, which was completed by 10 participants from Groups 3 and 4, there was much positive feedback. Those who saw the spectrogram (Groups 3 and 4), found the tool useful (n = 7), very good (n = 7), easy both to comprehend (n = 7) and to use (n = 5). Only one participant found it hard to use (n = 1).

5 Discussion

The study aims first to see if there is a difference in the participants' oral productions depending on which type of input they received, and secondly to gather participants' views and impressions of Englishville as a teaching/learning tool. The results suggest that merely visualising the corresponding spectrogram is slightly more beneficial than having no input at all, but barely makes a difference when the participants had access to the audio recordings. The fact that Group 3 did better than Group 1 suggests that some participants were able to learn from what they were seeing and improve some of their productions.

The minor difference between the results of Group 2 and Group 4 begs the question of sensorial overload. It is possible that having to read the text, pay attention to the spectrogram and to the audio resulted in too much information for the participants and may have led them to ignore some or all the input. The sentences were simple and short, and the written text was probably superfluous. It would have been possible for this level of L2 learners to repeat the text simply from the audio recordings, without the written text. In hindsight, Groups 1 and 2 are said to have limited input or only audio input, where it could be argued that both also had visual input because they saw/read the written text for each sentence, no visual information about intonation accompanied the written forms.

Given the small sample and with only five participants per group, it is difficult to draw any firm conclusions due to speaker variation. For example, it is possible that the spectrogram helped certain speakers in Groups 3 and 4, but it is also possible that some participants paid no attention to it at all. In addition to speaker variation, the fact that the participants could only listen/see each item once and could only record their own speech once before moving on to the next phrase, left them with little possibility to improve. They could of course repeat their productions several times before validation, but few did this. The main and global objective of Englishville is to enable learners to improve their prosody because they can see and use the spectrogram. For future experiments, it would be important to provide learners with several opportunities to do the tasks.

6 Conclusion and further directions

This article addresses the difficulty of acquiring English prosody for L2 learners and the use of a multi-sensorial tool to improve it. Previous research has shown that using visual aids often yields better results in training sessions when learning prosody and even has lasting effects on speech production in general (Derwing & Rossiter, 2003). Englishville was designed to explore this issue. One of its advantages is its malleability, making it possible to add for example, words, phrases, speakers, accents, participants, or remove elements from it (for example, to only have one group) so that all participants have the same input.

Despite our inconclusive results regarding the benefit of visualising speech, the students' feedback leads us to believe that Englishville corresponds to their desire for technological teaching tools. Our overall objective is to provide a tool which is easy to use for both teachers and learners and is also a useful way for L2 learners to practice prosody and raise their

awareness of it. This is motivated by the observation that teaching and learning prosody seems challenging, and the fact that software such as Praat is too complicated to be used by untrained learners. Learners should, however, be at the centre of any modern pedagogical approach and Englishville makes this possible. Therefore, we recommend that future use of this tool – whether for research or teaching – should allow students to hear, see the model, and record their own productions as many times as they want.

Various factors had an impact on the participants, including the number of times they could listen to the recordings, speaker variation, and sensorial overload. While the overall results do not allow us to confirm that seeing the melodic pattern by means of a spectrogram improves learners' productions, positive feedback was given by those who saw and used the spectrogram. They found it especially useful when they were able to match their own spectrogram with the model provided. In this regard, Englishville can be considered successful, as it encourages learners to make autonomous, critical comparisons. Perhaps with more time, their actual productions would also show improvement. To that end, an eight-week training session is currently being carried out with first year university students specialising in English. We hope that the latter will provide a clearer picture on the potential of this multi-sensorial tool.

References

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Colantoni, L., Marasco, O., Steele, J., & Sunara, S. (2014). Learning to realize prosodic prominence in L2 French and Spanish. In R. T. Miller, K. I. Martin, C. M. Eddington, A. Henery, N. M. Miguel, A. M. Tseng, A. Tuninetti, & D. Walter (Eds.), *Selected Proceedings of the 2012 Second Language Research Forum* (pp. 15–29). Cascadilla Proceedings Project. http://www.lingref.com/cpp/slrf/2012/paper3082.pdf

Costille, K. (2020, May 6) Englishville. https://demo.englishville.ovh/

de Bot, K. (1983). Visual feedback of intonation: Effectiveness and induced practice behavior. *Language and Speech, 26*, 331–350. https://doi.org/10.1177/002383098302600402

Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning, 13*, 1–17.

Gorjian, B., Hayati, A., & Pourkhoni, P. (2013). Using Praat software in teaching prosodic Features to EFL learners. *Procedia - Social and Behavioral Sciences, 84*, 34–40.

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology, 8*, 34–52. http://llt.msu.edu/vol8num1/hardison/

Hardison, D. M. (2010). Visual and auditory input in second-language speech processing. *Language Teaching, 43*(1), 84–95. https://doi.org/10.1017/S0261444809990176

Hirst, D., & Di Cristo, A. (Eds). (1998). *Intonation systems: A survey of twenty languages*. Cambridge University Press. DOI: 10.1353/lan.2000.0088

James, E. (1976). The acquisition of prosodic features of speech using a speech visualizer. *International Review of Applied Linguistics in Language Teaching (IRAL)*, *14*(3), 227–243. https://doi.org/10.1515/iral.1976.14.3.227

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.

Imber, B., Maynard, C., & Parker, M. (2017) Using Praat to increase intelligibility through visual feedback. In M. O'Brien & J. Levis (Eds.), *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference* (pp. 195–213). Iowa State University..

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America, 138*(2), 817–32. https://doi.org/10.1121/1.4926561

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(2), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*(2), 393–410. https://doi.org/10.1017/S0261444811000103

Offerman, H. M., & Olson, D. J. (2016). Visual feedback and second language segmental production: The generalizability of pronunciation gains. *System, 59*, 45–60. https://doi.org/10.1016/j.system.2016.03.003.

Olson, D. J. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second language pronunciation instruction. *Hispania, 97*(1), 47–68. http://dx.doi.org/10.1353/hpn.2014.0030

Setter, J., & Jenkins, J. (2005). Pronunciation: State-of-the-art review article. *Language Teaching, 38*(1), 1–17. DOI: 10.1017/S026144480500251X

Setter, J., Stojanovik, V., & Martínez-Castilla, P. (2010). Evaluating the intonation of non-native speakers of English using a computerized test battery. *International Journal of Applied Linguistics*, 20(3): 368–385. http://dx.doi.org/10.1111/j.1473.

Wells, J. C. (2007). *English intonation: An introduction*. Cambridge University Press.

About the author

Kizzi Edensor-Costille is a Lecturer at the University of Caen, Normandy where she teaches in the Department of English and Language Science. She did her thesis in the Speech and Language Lab (LPL) at Aix-Marseille University and specialised in regional accents in the UK and Ireland. She has also developed an interest in the field of L2 English learning, particularly in terms of perception and pronunciation. More recently, she has been working on the acquisition of prosody by non-native speakers. For this purpose, she developed a website called Englishville where L2 learners can visualise a real-time 3D spectrogram to practice different intonation patterns by seeing both the model and their own intonation.

Email: Kizzi.edensor-costille@unicaen.fr

Galimberti, V., Mora, J. C., & Gilabert, R. (2023). Teaching EFL pronunciation with audio-synchronised textual enhancement and audiovisual activities: Examining questionnaire data. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 70–82). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8174014

Teaching EFL pronunciation with audio-synchronised textual enhancement and audiovisual activities: Examining questionnaire data

Valeria Galimberti
University of Barcelona

Joan C. Mora
University of Barcelona

Roger Gilabert
University of Barcelona

Synchronising the enhancement of target words to their auditory onset has been found to promote a focus on their phonetic form (Stenton, 2013). In the case of L2 subtitled video, post-viewing activities involving interpretation and repetition of speech from the video offer further opportunities for noticing target pronunciation features and incorporating them into the learners' developing L2 system. This study investigated three groups of high-school EFL learners. Two intervention groups watched TV series clips with or without audio-synchronised textual enhancement of words containing past tense <-ed> endings in the subtitles and performed pronunciation-focused audiovisual activities such as revoicing and subtitling, whereas the control group was not exposed to the learning materials and thus provided a baseline of past tense <-ed> pronunciation rule knowledge. Questionnaire data provided information on the participants' language learning profiles, their perception of the enhancement and their ability to describe the past tense <-ed> pronunciation rule, and their impressions of the intervention. The participants' perceptions of the intervention were favourable, and the enhancement seemed to positively impact self-reported noticing of the target verbs, although not the internalisation of the pronunciation rule. We outline ideas for future research involving the implementation of these pedagogical tools in the language classroom.

Keywords: input enhancement, multimodal input, audiovisual activities, pronunciation teaching, English regular past

1 Introduction

Watching TV in a second language is an activity with a strong pronunciation learning potential, as it provides exposure to large amounts of L2 speech even in instructional contexts where the L2 is not spoken outside the classroom. This extensive listening practice can be supported and further enhanced by the simultaneous processing of subtitles that contain a verbatim transcription of each utterance. While research has shown that exposure to video with L2 subtitles facilitates speech segmentation (Charles & Trenkic, 2015) and promotes the development of speech perception skills (Mitterer & McQueen, 2009), we know very little about its effects on the development of L2 pronunciation (see Wisniewska & Mora, 2020 for a pioneering study) and how the characteristics of subtitles in L2 videos can be manipulated to further promote pronunciation learning. In particular, it is of interest whether visually highlighting words that contain fossilised L1 sound/symbol correspondences would disrupt automatic reading behaviours and direct learners' attention to the soundtrack containing the target-like realisation of those words (Stenton, 2013). In this study, we audio-synchronised the highlighting of target words in the subtitles and combined them with video-based activities to teach the pronunciation of English regular past tense <-ed> ending. This paper analyses participants' perceptions of the intervention, their self-reported noticing of the enhanced target verbs and their acquisition of the past tense <-ed> pronunciation rule.

2 Literature review

Watching TV in a target language is not only a fun extracurricular activity but also an effective way to practise L2 reading and listening. TV series in particular tend to keep viewers engaged for many hours, increasing the total amount of exposure to the foreign language (Pujadas & Muñoz, 2019). Moreover, L2 video represents an accessible source of L2 auditory input even for learners at lower proficiency levels, thanks to the possibility to rewind the video and listen again as many times as needed, and to the widespread availability of subtitles providing a verbatim transcription of the speech (Vanderplank, 2015).

Watching subtitled video involves exposure to large amounts of visual and auditory information, including language-related visual cues such as facial expressions and gestures, written text, and natural monologic or conversational L2 speech. However, the extent to which various grammatical, lexical and phonological features of the input are attended to and effectively processed may vary greatly, depending on viewing context (leisure vs. classroom activity), as well as learners' proficiency and attitudes towards the use of recreational materials in language learning (Vanderplank, 2015). To improve learners' noticing of target vocabulary or grammatical constructions during the viewing, a number of studies have used textual enhancement by typographically enhancing (e.g., highlighting or underlining) target words in the subtitles (see for example, Lee & Révész, 2020; Montero Perez et al., 2015).

Previous research on subtitled video enhancement in L2 pronunciation teaching has found that synchronising the enhancement of target words with the corresponding auditory onset in the soundtrack may promote a focus on the pronunciation of those words (Galimberti et al., 2023). The timely noticing of target words' pronunciation may, in turn, enhance awareness of any differences between the phonetic form of words as perceived through target-like auditory input and the learners' stored representation of the word. Further support for the synchronised enhancement of auditory and written word forms comes from research on reading-while-listening,

where it has been used to promote the update of L2 lexical stress patterns (Stenton, 2013) and the development of L2 and L1 reading skills (Bailly & Barbour, 2011; Gerbier et al., 2018).

Input enhancement in L2 video can promote, in the context of a primarily meaning-focused activity, the noticing of language form, which is a necessary step in the conversion of L2 input into intake (Schmidt, 1990; Sharwood Smith, 1991). In the presence of sufficient depth of processing, noticing may result, over time, in the development of rule-based representations, which may be more or less generalisable and accessible for testing depending on the explicitness of the learning conditions (Robinson, 1997). In Leow's (2015) L2 processing model, the initial stages of learning involve moving from L2 input processing to intake by engaging in memory-based processing (item learning), and/or rule-based processing (system learning). While restructuring of the learner's L2 system can be triggered by the conceptually-driven processing required to formulate a rule, the cognitive effort required to process the data can be reduced by the automatisation of linguistic data through repeated exposure and meaningful practice (Leow, 2015). This assumption is in line with Han et al.'s (2008) recommendation, based on a meta-analysis of 21 studies on textual enhancement, to combine textual enhancement with other strategies such as explicit instruction and/or interactional tasks involving the target feature.

To test the efficacy of this recommendation, we designed a pronunciation teaching intervention that combined TV series clips containing audio-synchronised enhancement with audiovisual (AV) activities, such as revoicing a silent clip or writing the L2 subtitles for an unsubtitled clip (see the AV framework in Zabalbeascoa et al., 2012). These activities, traditionally used to train translators, have been recently introduced into the language classroom to promote the development of listening and speaking skills (Danan, 2010; Zhang, 2016). Of particular relevance to this study are the studies that implemented AV activities to teach pronunciation (e.g., Chiu, 2012), including those with a broader focus on fluency (Sanchez-Requena, 2018) and comprehensibility (Lima, 2020).

The pronunciation of the English regular past tense <-ed> ending was selected as a target structure since the choice among the three allomorphs /d/, /t/ and /əd/ or /ɪd/ is derived from a morphophonological rule which depends on the phonemic environment. The main aspects of this rule can be explained in terms of spelling (Brutten et al., 1986); verbs ending in <-t> and <-d> in their present form take the /əd/ or /ɪd/ pronunciation, while other spelling endings take either /d/ or /t/. Learning this was expected to reduce the most common mispronunciations involving the erroneous addition of epenthetic vowels (e.g., *worked* pronounced */wɜrkəd/). These mispronunciations are as critical as the deletion of inflectional endings because the addition of an unexpected extra syllable through epenthesis may affect comprehensibility, i.e., the listener's perception of how difficult it is to understand a message, and intelligibility, which is the listener's actual understanding of the message (Levis, 2018). Finally, targeting the accurate pronunciation of <-ed> endings was expected to be appropriate and beneficial for high school learners, because regular past tense <-ed> endings are hard to perceive and produce accurately even at advanced proficiency levels, due to their low perceptual salience and redundancy with time adverbials (Strachan & Trofimovich, 2019).

3 Research methodology

3.1 Research questions

The study aimed to answer the following research questions:

RQ1: After a teaching intervention based on audio-synchronised textual enhancement and audiovisual activities:
a) do learners report noticing the target L2 pronunciation feature?
b) can learners successfully describe the target L2 pronunciation rule?

RQ2: What are the learners' perceptions of:
a) videos with audio-synchronised textual enhancement in subtitles?
b) pronunciation-focused audiovisual activities?

3.2 Participants

The intervention was implemented with three intact classes of L1 Spanish and Catalan 15-year-old students learning English as a foreign language. Out of 78 students, 53 completed a survey after obtaining their parents' written consent. The students' English proficiency level was estimated to be intermediate, based on the textbook used in class and on the participants' vocabulary size (m = 2715.09, SD = 592.87) as assessed by the X-Lex test (Milton, 2010). The groups were not significantly different in terms of vocabulary size ($F(2, 50) = .52$, $p = .60$), time spent in an English-speaking country ($F(2, 50) = .09$, $p = .92$), total time spent on English extracurricular classes ($F(2, 50) = .90$, $p = .41$), and weekly exposure to L2 TV shows ($F(2, 50) = .04$, $p = .96$). In order to ensure participant anonymity, a unique identifier was generated using a combination of alphanumeric characters. For instance, participant 2 in intervention group A was assigned the code A02.

3.3 Intervention materials

Students watched five video clips in which, under the enhancement condition only, a selected number of target words (past tense regular verbs) were enhanced in the subtitles 500 ms before the corresponding auditory onset by highlighting the whole word in yellow and underlining the <-ed> ending together with the orthographical representation of its phonological context, i.e., the vowel or consonant preceding it. Low frequency words and words not clearly audible in the soundtrack were not enhanced, in order to avoid interference with comprehension. In each AV activity, the participants in the intervention groups (see Figure 1) re-watched a clip and were instructed to either: 1) complete subtitles in which some words, including the target words, were missing; 2) order and label excerpts of the clip containing the target words; 3) identify muted target words in shorter unsubtitled excerpts and repeat the whole sentence out loud; 4) revoice a muted clip with the help of the subtitles; and 5) order unsubtitled excerpts containing the target words and revoice the obtained sequence. Therefore, in the first session learners could self-test their perception of the target feature through subtitling, whereas in the second session they needed to pay close attention to L2 speech, although in the context of a meaning-focused comprehension task. Finally, the three sessions that involved revoicing aimed at the automatisation of accurate and fluent production, with the support of (part of) the target utterances spoken by native L2 speakers (the characters). After each AV activity, participants did an awareness raising activity in which they read or listened to a list of verbs and were asked, for example, to underline the letter preceding the <–ed> ending of some verbs and decide if the sound was voiced or voiceless; group the verbs based on how the <-ed> ending sounded; or decide whether the vowel representing letter <e> in the <–ed> ending

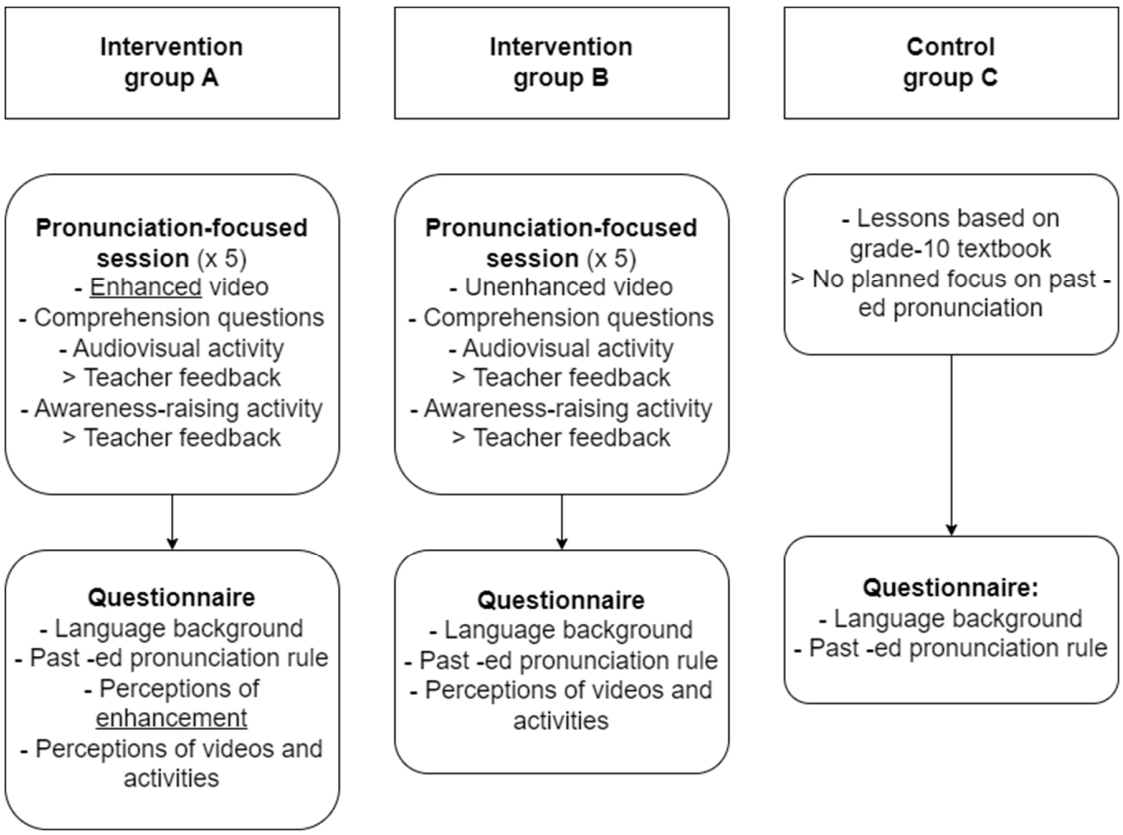
was pronounced or remained silent. These activities aimed at explicitly directing learners' attention to some aspects of regular past tense <-ed> pronunciation, such as the existence of different allomorphs, the difference between voiced and voiceless consonants, and the effects of the phonetic context preceding and following the <-ed> ending (Strachan & Trofimovich, 2019).

3.4 Procedure

The intervention lasted six weeks, with each group receiving fifty minutes of instruction per week (Figure 1). Intervention group A was exposed to audio-synchronised textual enhancement and carried out the AV activities; Intervention group B did the same activities but watched the clips without enhancement; Control group C followed their conventional textbook-based classes, and no planned or reactive focus on past tense <-ed> pronunciation was implemented by the teacher. The control group provided a baseline of past tense <-ed> pronunciation rule knowledge among learners who belonged to the same population as the intervention groups but had not received focused instruction.

Figure 1

Lesson Plan and Data Collection Procedure by Group



After a mock session, in each of the five sessions groups A and B watched a video containing enhanced and unenhanced target words respectively, worked on an AV activity in pairs while the teacher[1] walked around the classroom offering support, and they received feedback. The feedback phase involved having two or three pairs report on the activity in front of the whole class and asking other students if they agreed with the solution or performance proposed until the correct answer was provided. To conclude, each student answered ten comprehension questions and did an awareness-raising activity individually before receiving group feedback. A written questionnaire was administered in a quiet classroom the week after the intervention. Participants could choose between the Spanish and Catalan version and were asked to complete it within 20 minutes.

3.5 Questionnaire

After the language background section, which provided information on the participants' L1(s) and extracurricular exposure to English, all participants (groups A, B and C) were asked to describe the rule about the pronunciation of the <-ed> ending of past tense regular verbs, including examples if possible. Participants in group A and group B expressed their perceptions of the intervention by indicating to what extent they agreed with statements about the videos and the activities. The statements were mostly adapted from Sokoli's (2018) survey of learners' perceptions of AV activities, with a few novel items included. The item on peer collaboration was added as an initial (albeit very limited) measure of social interaction, a construct that has been related to active learning and to a higher focus on the task (Zabalbeascoa et al., 2012). In addition, participants were asked to indicate if they read the subtitles during the viewing, which provided a tentative measure of audiovisual processing under the circumstances (as collecting eye-tracking data during the implementation of whole-class activities was impossible). Sokoli's (2018) questions on participants' feeling of learning were adapted to assess whether the learners' general focus was primarily on grammar or pronunciation, since the intervention may have increased awareness of both the grammatical function and phonological form of the verbs. As a measure of reported noticing, participants in group A were also asked whether any letters were enhanced in the subtitles, what those letters had in common, and whether the participants believed that the enhancement was useful or distracting.

3.6 Data analysis

The data reported in this paper were collected via written questionnaire and analysed quantitatively. Yes/No questions resulted in binary variables (0 or 1), whereas the five-point Likert items on learner perceptions resulted in categorical variables with five levels from 1 (totally disagree) to 5 (totally agree). Knowledge of the past tense <-ed> rule was also operationalised as a categorical variable with four levels, with value range 0 (no response) to 3 (completely correct response). The rating of the responses was conducted by the author/teacher. Partially correct answers (value 1) mentioned some relevant elements but missed other important ones, e.g., "it is pronounced like a *t*" (A05). Answers were considered mostly correct (value 2) if they mentioned the existence of three allomorphs and/or the presence or absence of a vowel sound depending on the context, e.g., "in *walked* <e> makes no sound, in *provided* it sounds like /ed/ because the word ends in <e> (sic), other times it sounds like /t/ and others it makes no sound" (B46). An example

[1] The first author, Valeria Galimberti, was the teacher during the intervention.

of a response considered completely correct is: "There are verbs that in the past are pronounced as if they ended with /t/ (e.g., *walked*), others with /id/ (e.g., *waited*) or with /d/ (e.g., *turned*)" (A02).

Count data is presented for RQ1a, due to the small sample. To answer RQ1b, between group differences were also explored through Fisher's exact tests with Monte Carlo method (1e4 sampled tables), due to the low expected frequencies per variable level. When reporting in-text the participants' responses to the statements in RQ2, the response values 5 (agree) and 4 (somewhat agree) were collapsed into one category 5 (agree). Similarly, the response values 1 (disagree) and 2 (somewhat disagree) were collapsed into the category 1 (disagree). To offer a complementary picture of the data, the mean value and standard deviation on the original five-point scale of the responses to each statement were reported in Table 1 and Table 2 aggregated by group.

4 Results

4.1 Noticing and describing the target L2 pronunciation feature (RQ 1a, 1b)

All participants watching the videos with audio-synchronised enhancement (n = 18) reported noticing the enhanced words in the subtitles, with 16 finding the enhancement useful. Fourteen participants correctly identified that the enhanced words were regular past verbs and/or verbs ending in <-ed>, only one mentioned that the words had been enhanced because of their pronunciation, and three mentioned that the enhancement was related to pronunciation without further specification.

When asked to describe the rule about how to pronounce regular past tense <-ed> endings, 50% of the participants in group A, 12% of the participants in group B, and 22% of the participants in group C did not attempt to answer. The proportion of incorrect answers was 17% in group A, 71% in group B and 56% in group C. Of the twelve acceptable answers, only two for group A and one for group B were rated as mostly correct (11% and 6%, respectively), and two as completely correct (one in group A and one in group B). Fisher's exact tests with Monte Carlo method did not find significant differences between the responses of the three groups (two-tailed p = .66).

4.2 Learners' perceptions of videos with audio-synchronised textual enhancement and pronunciation-focused audiovisual activities (RQ 2a, 2b)

All participants reported understanding the videos, and around 80% in each group thought they were fun (Table 1). Two thirds of group A (videos with enhanced subtitles), but only half of group B (unenhanced subtitles), reported reading the subtitles. Around 70% of the participants in each intervention group believed that they had learned some English pronunciation from the video. While 65% also felt that they had learned some grammar or vocabulary from the videos in group B, only 50% of group A agreed.

Table 1

Responses (1–5) to Statements about the Enhanced Videos

	Intervention group A (Enhancement)			Intervention group B (No enhancement)		
	M	*SD*	95% CI	*M*	*SD*	95% CI
I understood the videos	4.83	.38	[4.64; 5.02]	4.53	.62	[4.21; 4.85]
The videos were fun	4.22	1.17	[3.64; 4.80]	4.12	.93	[3.64; 4.59]
I read the subtitles	3.72	1.02	[3.22; 4.23]	3.53	1.18	[2.92; 4.14]
I learned some English pronunciation from the videos	3.72	.57	[3.44; 4.01]	3.82	1.07	[3.27; 4.38]
I learned some English grammar or vocabulary from the videos	3.50	.71	[3.15; 3.85]	3.65	.99	[3.13; 4.16]

Note. 1 = totally disagree, 5 = totally agree

Almost all participants reported understanding the instructions of the AV activities and two thirds used the clues offered within each activity to complete them (Table 2). Eighty percent of the participants in group A and 60% of the participants in group B indicated that the activities were fun. Only one third of the participants in group A indicated that the activities were challenging, but in group B almost two thirds of participants found them challenging. Ninety percent of participants in group A and 75% in group B responded that both partners had contributed equally to the activity. Similar to the responses for the enhanced videos, around 70% of the participants in each group reported learning some pronunciation from the activities, but only half in each group reported learning some grammar and vocabulary.

Table 2

Responses (1-5) to Statements about the Audiovisual Activities

	Intervention group A (Enhancement)			Intervention group B (No enhancement)		
	M	*SD*	95% CI	*M*	*SD*	95% CI
I understood the instructions	4.39	.78	[4.00; 4.78]	4.29	.85	[3.86; 4.73]
We used the clues to do the activities	3.72	.89	[3.28; 4.17]	3.76	.90	[3.30; 4.23]
The activities were fun	3.89	1.23	[3.28; 4.50]	3.53	1.01	[3.01; 4.05]
The activities were challenging	3.00	1.19	[2.41; 3.59]	3.41	1.12	[2.84; 3.99]
My partner and I contributed equally to the activities	4.39	1.24	[3.77; 5.01]	4.06	1.20	[3.44; 4.67]
I learned some English pronunciation from the activities	3.89	.68	[3.55; 4.23]	4.06	.97	[3.56; 4.56]
I learned some English grammar or vocabulary from the activities	3.50	.71	[3.15; 3.85]	3.65	1.11	[3.07; 4.22]

Note. 1 = totally disagree, 5 = totally agree

5 Discussion

In relation to RQ1, participants who watched L2 videos with audio-synchronised, textually enhanced subtitles and did pronunciation-focused activities, reported noticing the enhanced words and connected the enhancement to the target feature. This suggests that they were able to move past the stage of *input processing* to that of *intake processing*, as the enhanced exemplars seemed to have been cognitively registered with some level of awareness (Leow, 2015, p. 17). However, even after five weeks of intervention, most participants were unable to describe the rule relative to regular past tense <-ed> pronunciation better than the control group. Despite adopting a sequential design in which learners were encouraged to process input for meaning first and then focus on form through subsequent activities, as recommended by Han et al. (2008), the intervention did not seem to promote the type of conceptually-driven processing necessary to extract abstract rules from the exemplars encountered in the input (Leow, 2015). One possible explanation is that the participants may have struggled to integrate visual and auditory input due to the low salience of the target phonological forms (Strachan & Trofimovich, 2019), and the salience created externally by highlighting the target words and creating activities that revolve around these words may not have aligned with the learners' internally created salience (Sharwood Smith, 1991). In line with this hypothesis, very few participants indicated the *pronunciation* of the regular past endings as the reason for their enhancement, which suggested that the processing of regular past verbs may have primarily focused on their grammatical or semantic properties rather than their phonological realisation. Other possible explanations for the null or negative findings associated with input enhancement typically involve the shortness and implicitness of the treatment (Han et al., 2008). However, in studies of similar length that assessed L2 speech production rather than rule acquisition, significant pronunciation gains have been observed from exposure to enhanced input (Stenton, 2013) as well as the implementation of AV activities (Sanchez-Requena, 2018). Considering that almost all participants in this study perceived the enhancement as useful and that the feeling of pronunciation learning was generally very high, exposure to audio-synchronised enhancement and AV activities may have benefited other, more implicit, dimensions of L2 pronunciation learning.

Regarding RQ2, learners' responses to the questionnaire seemed to indicate an overall positive perception of the videos and activities, in line with previous studies on AV activities (Danan, 2010; Sanchez-Requena, 2018; Sokoli, 2018). The learners indicated that they had understood the videos and the instructions of the activities, and that the videos were fun, confirming that the materials were appropriate for the target population. According to their responses to the questionnaire, participants in group A reported that they relied on subtitles more than participants in group B. If group A had been primarily processing the written input, the appearance of an enhanced word may have interrupted the automatic reading process and successfully redirected their attention to the corresponding auditory form (Stenton, 2013). However, this explanation is only tentative in the absence of online measures of attention allocation from eye-tracking and offline stimulated recall protocols. Moreover, participants in group A, who had already focused on the target words during the first exposure to the enhanced video and may have developed a stronger episodic memory of those auditory forms, were less likely to find the activities challenging than participants in group B. Almost all participants reported that both partners had contributed equally during the activities, suggesting that working in pairs may have helped learners overcome the challenges presented by the dual processing of meaning and form, fostering social interaction with positive effects on language learning (Zabalbeascoa, 2012). Finally, learners reported a higher feeling of learning for

pronunciation than for grammar and vocabulary, especially in relation to the AV activities. The lack of speech perception and production tests, which would have allowed us to draw more robust conclusions regarding pronunciation learning, is a major limitation of this paper, and will be addressed in future publications.

6 Conclusion

After an intervention featuring L2 video with audio-synchronised textual enhancement and video-based activities, our participants reported noticing the enhanced verbs in the subtitles but were unable to infer the past <-ed> pronunciation rule and describe it in writing. However, the intervention was well-received, and the participants' feeling of learning was high, in line with the hypothesis that incorporating these materials into the EFL classroom may foster active and collaborative learning (Zabalbeascoa, 2012). To ensure the successful implementation of AV activities, teachers should carefully select target features and video clips at the appropriate difficulty level and provide clear instructions before each activity. Prefacing the activities with explicit instruction may help direct learners' attention to the phonological properties of the target words, especially with a morphophonemic feature like the regular past <-ed>. To ensure active participation, the teacher should monitor the learners' execution of each stage, provide individualised feedback during pair work and foster a safe learning environment in which learners may be willing to perform the revoicing activities in front of the whole class.

This study focused on the participants' perceptions of audio-synchronised enhancement and AV activities and was therefore very limited in scope. The main limitation was the lack of objective measures of phonological development tapping into the learners' perception and production of the target feature. Recommendations for future research include the analysis of L2 pronunciation development through pre- and post-tests involving L2 speech production tasks, as well as the investigation of learners' attention allocation to audio-synchronised subtitle enhancement. Although collecting eye-tracking data in a classroom setting may not be feasible, a lab-based study featuring a comparable sample may provide valuable insights on the processing of audio-synchronised enhancement. Finally, due to our small and homogenous sample, further research is needed to assess how participants of different ages, proficiencies and mother tongues would respond to audio-synchronised enhancement and AV activities.

Acknowledgments

This work was supported by the Spanish Ministry of Science, Innovation and Universities [grant PID2019-107814GB-I00] and by the Secretary of Universities and Research of the Government of Catalonia and the European Social Fund [grant FI_2019].

References

Bailly, G., & Barbour, W. (2011). Synchronous reading: Learning French orthography by audiovisual training. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (pp. 1153–1156). ISCA. http://dx.doi.org/10.21437/interspeech.2011-342

Brutten, S. R., Mouw, J. T., & Perkins, K. (1986). The Effects of language group, proficiency level, and instruction on ESL subjects' control of the {D} and {Z} morphemes. *TESOL Quarterly*, *20*(3), 553–559. https://doi.org/10.2307/3586301

Charles, T., & Trenkic, D. (2015). The effect of bi-modal input presentation on second language listening: The focus on speech segmentation. In Y. Gambier, A. Caimi, & C. Mariotti (Eds.), *Subtitles and language learning* (pp. 173–197). Peter Lang.

Chiu, Y. (2012). Can film dubbing projects facilitate EFL learners' acquisition of English pronunciation? *British Journal of Educational Technology*, *43*(1), 24–27. https://doi.org/10.1111/j.1467-8535.2011.01252.x

Danan, M., (2010). Dubbing projects for the language learner: A framework for integrating audiovisual translation into task-based instruction. *Computer Assisted Language Learning*, *23*(5), 441–456. https://doi.org/10.1080/09588221.2010.522528

Galimberti, V., Mora, J. C., & Gilabert, R. (2023). Audio-synchronized textual enhancement in foreign language pronunciation learning from videos. *System, 116,* 103078. https://doi.org/10.1016/j.system.2023.103078

Gerbier, E., Bailly, G., & Bosse, M. L. (2018). Audio–visual synchronization in reading while listening to texts: Effects on visual behavior and verbal learning. *Computer Speech and Language, 47*, 74–92. https://doi.org/10.1016.j.csl.2017.07.003

Han, Z., Park, E. S., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics*, *29*(4), 597–618. https://doi.org/10.1093/applin/amn010

Lee, M., & Révész, A. (2020). Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks. *Studies in Second Language Acquisition*, *42*(3), 1–27. https://doi.org/10.1017/S0272263120000108

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach.* Routledge. https://doi.org/10.4324/9781315887074

Levis, J. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press. https://doi.org/10.1017/9781108241564

Lima, E. F. (2020). The Supra Tutor: Improving speaker comprehensibility through a fully online pronunciation course. *Journal of Second Language Pronunciation*, *6*(1), 39–67. https://doi.org/10.1075/jslp.18031.lim

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. In I. Vedder, I. Bartning, & M. Martin (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 211-232). Second Language Acquisition and Testing in Europe Monograph Series 1. EUROSLA.

Mitterer H., & McQueen J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE*, *4*(11), e7785. https://doi.org/10.1371/journal.pone.0007785

Montero Pérez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *Modern Language Journal*, *99*(2), 308–328. https://doi.org/10.1111/modl.12215

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, 47(4), 479–496. https://doi.org/10.1080/09571736.2019.1616806

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition, 19*, 223–47. https://www.jstor.org/stable/44488684

Sánchez-Requena, A. (2018). Intralingual dubbing as a tool for developing speaking skills. *Translation and Translanguaging in Multilingual Contexts*, *4*, 101–128. https://doi.org/10.1075/ttmc.00006.san

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Sharwood Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research, 7*, 118–132. https://www.jstor.org/stable/43104426

Sokoli, S. (2018). Exploring the possibilities of interactive audiovisual activities for language learning. *Translation and Translanguaging in Multilingual Contexts*, *4*(1), 77–100. https://doi.org/10.1075/ttmc.00005.sok

Stenton, A. (2013). The role of the syllable in foreign language learning: Improving oral production through dual-coded, sound-synchronized, typographic annotations. *Language Learning in Higher Education*, *2* (1), 145–161. https://doi.org/10.1515/cercles-2012-0009

Strachan, L., & Trofimovich, P. (2019). Now you hear it, now you don't: Perception of English regular past –ed in naturalistic input. *Canadian Modern Language Review*, *75*(1), 84–104. https://doi.org/10.3138/cmlr.2017-0082

Vanderplank, R. (2015). Thirty years of research into captions/same language subtitles and second/foreign language learning: Distinguishing between 'effects of' subtitles and 'effects with' subtitles for future research. In Y. Gambier, A. Caimi & C. Mariotti (Eds.), *Subtitles and language Learning* (pp. 19–40). Peter Lang.

Wisniewska, N., & Mora, J. C. (2020). Can captioned video benefit second language pronunciation? *Studies in Second Language Acquisition*, *42*(3), 599–624. https://doi.org/10.1017/S0272263120000029

Zabalbeascoa, P., Sokoli, S., & Torres, O. (2012). *Clipflair: Foreign language learning through interactive revoicing and captioning of clips* (Work Package 2, Lifelong Learning Programme 519085-LLP-1-2011-1-ES-KA2-KA2MP). Education and Culture DG, European Union. http://clipflair.net/wp-content/uploads/2014/06/D2.1ConceptualFramework.pdf

Zhang, S. (2016). Mobile English learning: An empirical study on an APP, English Fun Dubbing. *International Journal of Emerging Technologies in Learning, 11*, 4-8. https://doi.org/10.3991/ijet.v11i12.6314

About the authors

Valeria Galimberti is a PhD candidate in Applied Linguistics at the University of Barcelona, and a member of the GRAL (Language Acquisition Research Group). She has conducted research with high school and university students on L2 captioned video processing, vocabulary learning from TV series and production tasks in second language acquisition. Her PhD thesis explores the effects of input enhancement and video-based activities on the acquisition of L2 pronunciation.

Email: galimberti@ub.edu

Joan C. Mora is associate professor in the Department of Modern Languages and Literatures and English Studies and a member of GRAL (Language Acquisition Research Group) at the University of Barcelona (UB) in Spain. He is interested in understanding how contextual and individual factors shape L2 speech learning. His current research interests focus on the role of cognitive and emotional individual differences in the development of L2 pronunciation and speaking fluency, phonological learning in the mental lexicon, phonetic training methods, multimodal pronunciation training, and task-based pronunciation teaching and learning in instructed SLA.

Email: mora@ub.edu

Roger Gilabert is a lecturer and researcher at the University of Barcelona, and he is a member of the GRAL (Language Acquisition Research Group). His research has revolved around task and syllabus design for the last 25 years. The focus of his research has also been L2 oral and written production, multimodal input processing through caption videos (genres), game-based learning and reading skills. He is currently leading the taskGen project on the automation of second and foreign language task design.

Email: rogergilabert@ub.edu

Herment, S. (2023). From research to teaching: The case of English rising contours. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 83–96). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8174003

From research to teaching: The case of English rising contours

Sophie Herment
Aix Marseille University

It is quite common to hear incongruous rises in the speech of EFL learners, and yet, native English speakers produce rising contours in the same contexts. To understand where this discrepancy comes from and how it can be avoided, this paper focuses on the forms and functions of rising contours in English through examples of French learners of English and native English speakers. Depending on the variety of English, the type of speech, the conditions in which the interaction takes place, the relationship between the interlocutors, etc., rises can take different forms and have different functions. On the basis of two studies of native read speech, it is shown that rising contours are rare and that contrary to what is stated in the literature, their main function is not to indicate non-finality and continuation, but rather to convey attitudes. The pedagogical implications of these results are of importance: in read speech, learners should try and avoid rises, even in non-final tone units. In spontaneous speech, learners should be aware of the attitude conveyed by rising terminals. A few examples are given of how the visualisation of prosody can help to better understand the contours and to better hear and produce them.

Keywords: EFL, English rising contours, melodic forms, discursive functions, visualisation

1 Introduction

The aim of this paper is to show how research on intonation can inform L2 learning and teaching practice, illustrating this with the example of French learners of English.

For this study we adopted the theoretical framework of the British school of intonation (see, for example, Cruttenden, 1997, 2014; Roach, 2009; Wells, 2006). The British system is based on a configurational approach, which we have argued is better adapted to teaching than the American approach, based on a bitonal model (see Herment, 2018; Niebuhr et al., 2017). In the British tradition, speech is divided into Tone Units (TUs), which correspond roughly to syntax and semantics and are characterised by a coherent melodic contour. A TU is composed of at least one syllable, necessarily the nuclear syllable (or nucleus). The melodic movement starts on the nucleus and spreads on the post-nuclear syllables, if any. Fall (F), rise (R), rise-fall (RF) and fall-rise (FR) are the melodic contours commonly used to describe intonation (with a few variants according to Cruttenden, 1997, 2014; Roach, 2009; Wells, 2006).

French learners of English (hereafter referred to as learners) tend to produce (too) many rises when they speak English. This is why we focus on these contours, which are also found in the productions of native English speakers (hereafter referred to as natives). Rises in learners and natives are therefore examined, and their forms and functions discussed, as they can vary according to many factors, such as the variety of English, the type of speech, the conditions in which the interaction takes place, the relationship between the interlocutors, etc. The canonical functions of English intonation are then called into question based on two studies of native read speech, which are summarised. The following section deals with the pedagogical implications of the results and shows how visualisation can help improve the teaching and learning of prosody.

2 English rising contours

Several studies have shown that it is quite common to hear incongruous rises in the speech of learners of L2 English (or EFL learners), for questions (e.g., MacDonald, 2011; Pytlyk, 2008; Santiago-Vargas & Delais-Roussarie, 2012), but also at the end of declarative sentences (Contreras Roa, 2019; Horgues, 2010). Those rises sound inappropriate, and yet, native English speakers produce rising contours in the same contexts. Where does this discrepancy come from? How can it be avoided? How can learners be helped to integrate a more native-like prosody? To try and answer these questions, it is necessary to talk about variation and to focus on the forms and functions of rising contours, both in learners and natives.

2.1 Rising contours and learners

It is well known that it is difficult to master the prosody of a new language. One reason is that prosody is acquired in utero. Studies have shown that hearing is operational by the third trimester of pregnancy (Moore & Jeffery, 1994), but maternal tissues function as low-pass filters, removing the fine details of speech and allowing only the prosody (the melody and the rhythm) of the speech signal to pass through. Indeed, studies have shown that foetuses recognise their mother's voice. Kisilevsky et al. (2003) have shown that the heartbeat rate increases when foetuses hear their mother’s voice. It has also been shown that new-borns recognise their mother tongue: infants had longer sucking bursts on a nipple connected to a pressure transducer when played recordings in their native tongue than in a foreign language (Moon et al., 1993).

The influence that a speaker’s first language can exert on the production and perception of a second language is also well-known and variously referred to as *transfer* or *interference*. To

avoid the negative (and behaviourist) connotations of these terms, Sharwood-Smith (1983) suggested the theory-neutral term *cross-linguistic influence* (CLI) to refer to any kind of influence of a source language onto a target language.

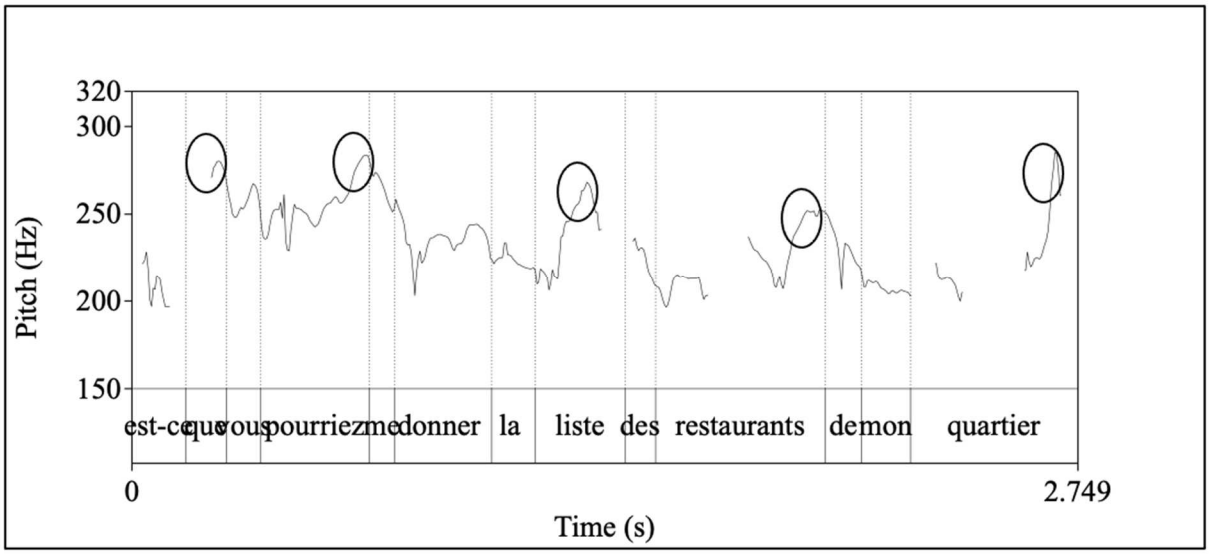
In the case of French learners of English, CLI from French to English is striking, particularly as far as rhythm is concerned. In French, the chunking (the division into rhythm units) is very different from English. An intonation phrase is divided into small groups of words (accentual phrase), which are realised with a rising contour on the very last syllable of the group, whatever the nature of the word. Figure 1 below, which is a screencapture from the PRAAT software (Boersma & Weenink, 2001), shows the production by a French native speaker of the following example sentence (1):

(1) Est-ce que ↗| vous pourriez ↗ | me donner la liste ↗ | des restaurants ↗ | de mon quartier ↗?|
Can you give me a list of the restaurants in the neighbourhood?

In the example above (and in the ones further on), the division into rhythmic groups is shown by the horizontal bars and the syllable bearing the melodic movement is underlined (the last syllable in each group in this example). The arrows show the melodic movement (rises here). In Figure 1, the rises are marked by the ellipses on the melodic curve.

Figure 1

French Sentence Produced by a Native French Speaker



Est-ce que vous pourriez-me donner la liste des restaurants de mon quartier ?

The production of the equivalent sentence in English by a French learner is shown in Figure 2 and exemplifies CLI. A native English speaker would have divided the sentence into two rhythmic groups, as in example (2):

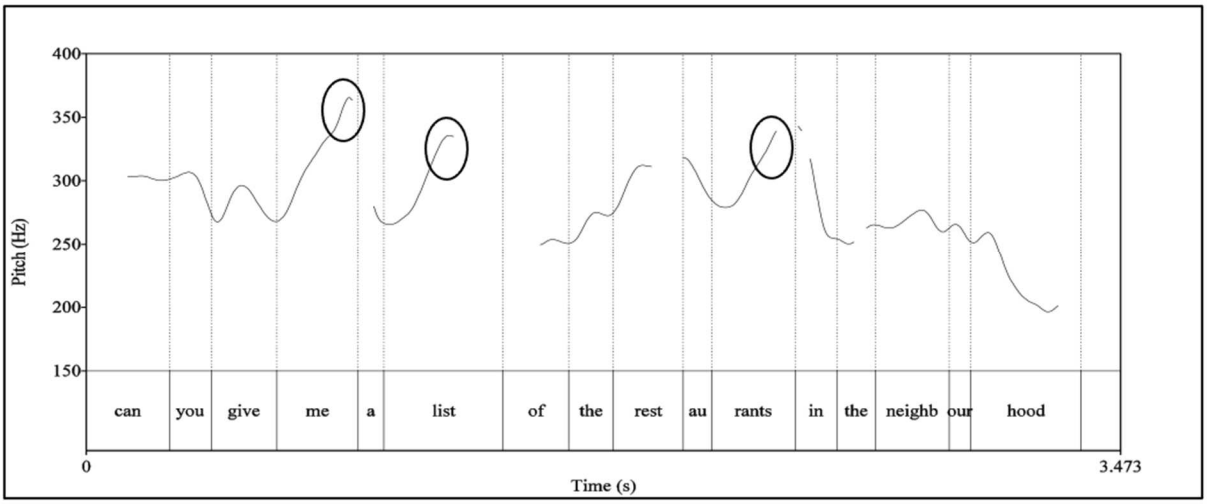
(2) Can you give me a list of the restaurants ↗ | in the neighbourhood? ↗ |

The French learner of English divides the sentence into four groups, as in example (3), and like in their mother tongue, pronounces rises on the last syllable of each group (except the last one), whatever the nature of the word and its stressed syllable:

(3) Can you give me ↗ | a list ↗ | of the restaurants ↗ | in the neighbourhood? ↘ |

Figure 2

English Sentence Produced by a French Learner of English



Can you give me a list of the restaurants in the neighbourhood?

Although this example is particularly striking as far as CLI is concerned, the rising contours in learners are probably not only due to language transfer. Santiago-Vargas and Delais-Roussarie (2015) studied the intonation of questions in Mexican learners of French. They show that some cases of erroneous productions can be attributed to the transfer of L1 contours to L2, but that not all errors can be explained in this way. The authors suggest that the acquisition process itself may be responsible for some of the erroneous contours produced by learners.

These numerous rises might also be linked to the pragmatic functions of intonation and specifically of the rising contours. An upward contour on declarative sentences, which English teachers are familiar with, is often referred to as tentative intonation. It sounds as if the learner integrates a question in their sentence, the pragmatics of which could be *Am I right? Is this the right answer?*. It denotes linguistic insecurity, which is also found in the production of native speakers (see example 4 below). An interesting example from a corpus of Canadian English is to be found in Rodrigues Da Mota and Herment (2016): the authors give an example (4) in which the speaker, who addresses a French person in English, pronounces the French word *Grenoble* with a high rising pattern, probably because she is not sure her pronunciation is right and expresses her insecurity in front of a native French person.

(4) and so now we don't have a car | so we can't get to Maastricht | so we're going to Grenoble ↗ | for a few days I think |

This suggests that the pragmatic functions of intonation should also be taken into account. When a learner uses rising intonation on a declarative, it often sounds tentative. Rising contours are also heard among natives on declaratives, with several pragmatic functions, as explained in the next section.

2.2 Rising contours and natives

Rising contours can take different forms and different functions among natives, according to the variety of English spoken, the context of speech or sociological factors. To allow comparison with learners, we focus here on rising contours at the end of declarative sentences.

2.2.1 Stylistic rises

The term *stylistic rises* is used in this paper (following Bongiorno, 2021) for the rising contours found at the end of declarative sentences in English for pragmatic and stylistic purposes. Other authors talk about *high rising terminals* (HRTs) (e.g., Wilhelm, 2015) or *uptalk* (e.g., Warren, 2016). We prefer stylistic rises because they do not necessarily appear at the end of a sentence and because we contrast them with systemic rises (see §2.2.2).

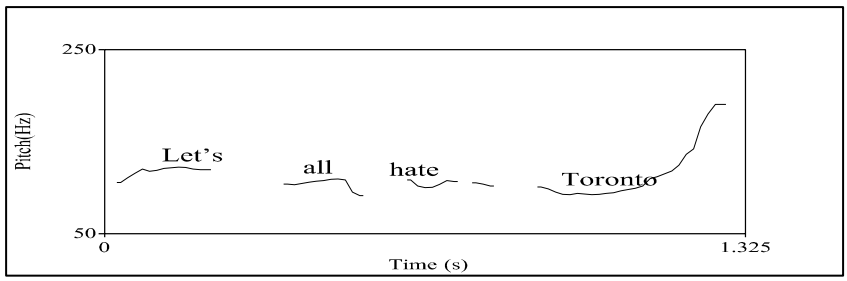
Stylistic rises have been widely documented (see Warren, 2016). First attested in Australia and New-Zealand (2nd half of the 19th century), then in North America (Pacific rim) at the end of the 19th century, they are now rapidly spreading in Great Britain. They can take different forms in different varieties of English, but they are most often characterised by a sharp rise and a flat prenuclear contour, as exemplified in Figures 3 and 4 below, in 2 different varieties of English.

Typically, these rises have interactional and discursive functions, among which and not exclusively:

- the integration of an underlying question, such as *Do you see what I mean? Do you know what I am talking about?*, to make sure the interlocutor follows what the speaker is saying;
- the sign of the speaker's uncertainty, for example, when they indicate an approximate distance;
- the mark of the speaker's linguistic insecurity;
- the speaker's wish to obtain feedback, e.g., *Is this what you're expecting from me?*.

Figure 3

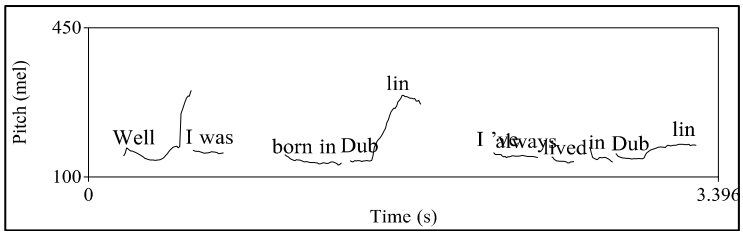
Canadian English



Let's all hate Toronto (Rodrigues da Mota & Herment, 2016)

Figure 4

Dublin English



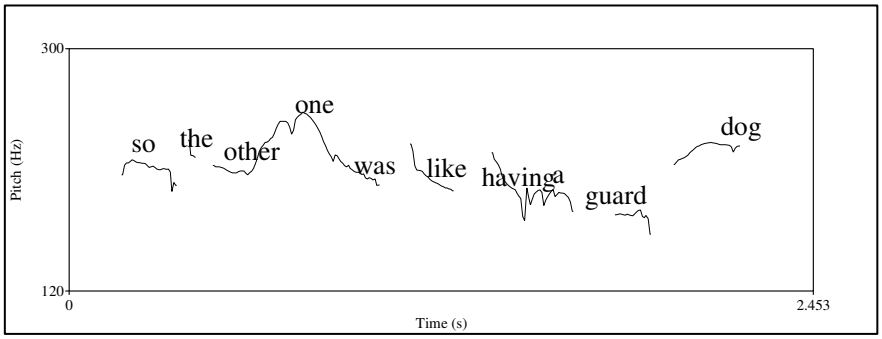
I was born in Dublin (Bongiorno & Herment, 2018)

2.2.2 Systemic rises

Another type of rises heard among natives is referred to as Urban Northern British Intonation (UNBI) and is found, as its name indicates, only in certain varieties of British English and more particularly in the large cities of the North: Belfast, Derry, Glasgow, Newcastle, Liverpool, Manchester, Leeds, Birmingham. In these places, the rising contour is the intonation by default on declaratives, hence the term *systemic* (as opposed to the stylistic rises mentioned above). According to Wilhelm (2015), five different types of rises are heard in UNBI: a) low rise, b) rise-slump, c) rise plateau, d) rise-plateau slump, and e) high plateau. The examples in Figures 5 and 6 exemplify c) and d) respectively.

Figure 5

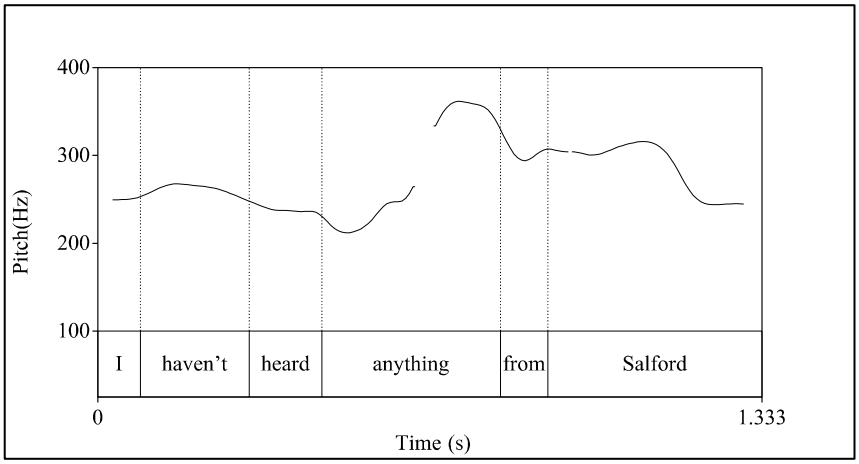
Rise-plateau on 'Guard Dog' in Donegal English



So the other one was like having a guard dog (Turcsan & Herment, 2015)

Figure 6

Rise-plateau-slump Starting on 'Anything' in Newcastle English



I haven't heard anything from Salford (Herment et al., 2020a)

2.2.3 Speaking styles

Intonation also varies according to speaking styles. In spontaneous conversations, speakers tend to use rising contours so as not to lose their speech turn. Rising contours with pragmatic functions (stylistic rises, see §2.2.1) are also more frequent in interactions (spontaneous conversations in particular). In non-prepared speech, hesitations are also more common and often trigger a rising contour. This is the case in sports commentaries, where journalists tend to use numerous rising contours to increase suspense (Samlowski et al., 2018). Falls are more frequent in formal speech. Teachers, for example, know that they will not be interrupted so they do not have to utter rising contours to keep the floor. In reading aloud, which is a rather codified exercise, a canonical intonation is expected, i.e., the intonation described in manuals or phonology books.

2.2.4 Canonical intonation

It is difficult to describe a canonical intonation, since so much variation exists – within speakers, between speakers and in different varieties of English. However, many authors like Cruttenden (2014), Tench (1996) or Wells (2006) agree on the following basic functions of intonation for standard British English:

- Rises signal non-finality, falls finality and completeness;
- Rises are produced in yes-no questions and falls in WH-questions;
- Falling-rising contours signal continuity, implication, contrast, the desire to draw the co-speaker's attention to what is being said (Brazil, 1997) or to what follows (Gussenhoven, 2004);
- Rise-falls are used for irony or surprise (Roach, 2009).

These principles are often taught to learners of English. However, when listening to recordings of British English natives, we noticed an amazingly high proportion of falling contours in read speech. We decided to analyse this speech style to determine how many rises

and falls occur, where and why. Our hypothesis is that incomplete statements (non-final tone units), usually described as displaying a non-fall contour, can be, and frequently are, pronounced with a falling tone in read speech, contrary to what is generally admitted.

The next section briefly describes two studies (Herment & Tortel, 2021 and Herment et al., 2020b) in which we tested this hypothesis.

3 Analysis of read speech by native English speakers

In order to study L2 English by French learners, two corpora were collected and analysed: the Anglish corpus (Tortel, 2008) and the AixOx corpus (Herment et al., 2014), which contain the same passages in English read by both native English speakers and French learners of English. The hypothesis mentioned above stems from listening to the native English speakers of these two corpora. Our results confirm our hypothesis (see Herment & Tortel, 2021): concerning non-final tone units, 78.5% are produced with a fall. Contrary to what is generally admitted, the rising tone is far from being the most common contour for incomplete statements as only 14.5% were found and 7% were falling-rising tones, which amounts to only 21.5% when combined. As far as final tone units are concerned (534 TUs), the most common tone is the fall with 85%, followed by 10.5% of fall-rises and 4.5% of rises. Falls are clearly the most common contours in read speech. We then conducted a qualitative analysis of the corpus and took a close look at the types of sentences, clauses, and phrases to see how the tones were distributed in relation to syntax and attitudinal functions (Herment et al., 2020b).

Tonal sequences (F/R, R/F, FR/F, etc.) were analysed, i.e., the intonation contours in sentences containing two or more TUs. The notion of dependency, which is important as far as syntax is concerned, was also considered: the elements which are not complete in themselves are dependent on some other structure. An independent element is an independent clause or main clause in a complex sentence or last clause in compound sentences.

The literature (see Cruttenden, 1997; Wells, 2006) reports that a fall is found on the main part of the utterance, and a non-fall on the subordinate or dependent part, whether final or initial (as in adverbials in trailing position). The sequence F/F (fall plus fall) is also commonly found: the information is then presented as two separate and potentially complete items, as is often the case with coordinate clauses, related by *and*, *but* and *or*.

The results of our qualitative analysis (248 tonal sequences) show that (see Herment et al., 2020b for details):

- the sequence F/F dominates in our data (especially in coordinate clauses with 95% and subordinate clauses with 82%);
- the rising contour can be described as exceptional in read speech, even in non-final clauses;
- the main function of rising contours is not to indicate non-finality and continuation, or dependency, but rather to convey the speaker's implication or emotional state.

These results run counter to what the literature reports as canonical intonation (see §2.2.4), at least in reading, and thus to what is generally taught. They therefore have important pedagogical implications.

4 Pedagogical perspectives

In this section, we develop two main pedagogical perspectives: a) the use of rises depending on the level of the learner, and b) the contribution of visualisation techniques in the teaching and learning of prosody.

4.1 Level of learners

Rises are very infrequent in the read speech of natives but they are quite frequent in interactional speech, where they take on particular pragmatic and discursive functions and convey emotions and attitudes. In the speech of French learners of English rises are very common, because of CLI and other reasons, but these rises tend to sound incongruous. Therefore, teaching practices should insist upon falling contours. We are aware that read speech is not necessarily a model for typical everyday speech communication. However, the fact that so many falls are found is intriguing, and shows that it is possible to use mainly falls to communicate, at least at the beginning of the learning process. Knowing this could counteract the overuse (and harmful effects) of rises among beginner learners. Therefore, the level of learners should be taken into account when teaching prosody.

The functions of rises correlate with other aspects such as the variety of English, the speech style, or being a learner. When a learner speaks, native speakers recognise that they are a learner because their pronunciation is not that of a native. This is very important because native interlocutors will adapt to the fact that they have a learner in front of them and will therefore probably understand a rise as a sign of insecurity or uncertainty. The lower the level of the learner, the more native speakers will be tempted to attribute rising contours to a feeling of insecurity. The more advanced the learner, the more native interlocutors will interpret a pragmatic function, as they would when listening to other native speakers.

Beginner or intermediate learners should therefore avoid rising contours, even in non-final contexts and in yes-no questions. Teachers should make them practise the realisation of falling contours as much as possible. In contrast, advanced learners should not ban rises from their speech, but they should be (made) aware of the pragmatic and discursive functions of rising contours, especially in final TUs, so that possible misunderstandings are avoided.

4.2 Visualising prosody

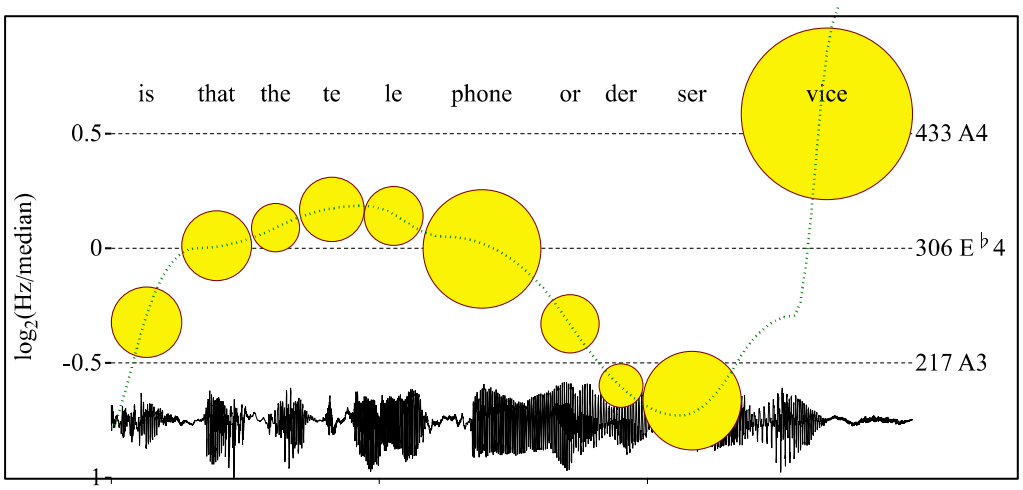
As already argued in Herment and Tortel (2021), learning (and teaching) prosody is facilitated by its visualisation. The representations of prosody shown in Figures 7–11 below were generated using the ProZed plugin (Hirst, 2005) implemented in PRAAT. In each Figure, from top to bottom, one can see the sentence read, the melodic curve (the green line), each syllable of the sentence (yellow balls on the curve), and the oscillogram. The green curve is a representation of the fundamental frequency, and thus provides a visualisation of the intonation contour. The size of the yellow balls corresponds to the duration of the syllable: the bigger the ball, the longer the syllable. The size of the ball is a good clue for lexical stress in English. If the ball is big, the syllable is very likely to be accented, at least not reduced (apart from the last syllable, which is longer than average because final and not necessarily accented). In Figures 7 and 8, the same sentence is read by a native and a learner respectively and the equivalent sentence in French is read by a native French speaker in Figure 9.

The intonation of the learner is clearly very different from that of the native, and strikingly modelled on that of French. Showing such representations to learners is probably very helpful because they can see what they produce, and it is well-known that seeing helps perceiving.

Another example of the difference between French and English contours is illustrated in Figures 10 and 11, where a WH-question is read, with a typical falling contour by the native, and a typical rising contour by the learner. The teacher could show such examples in class to help learners visualise the difference. The size of the yellow balls could also be commented upon by the teacher, to help learners with word stress and reductions (e.g., *for*).

Figure 7

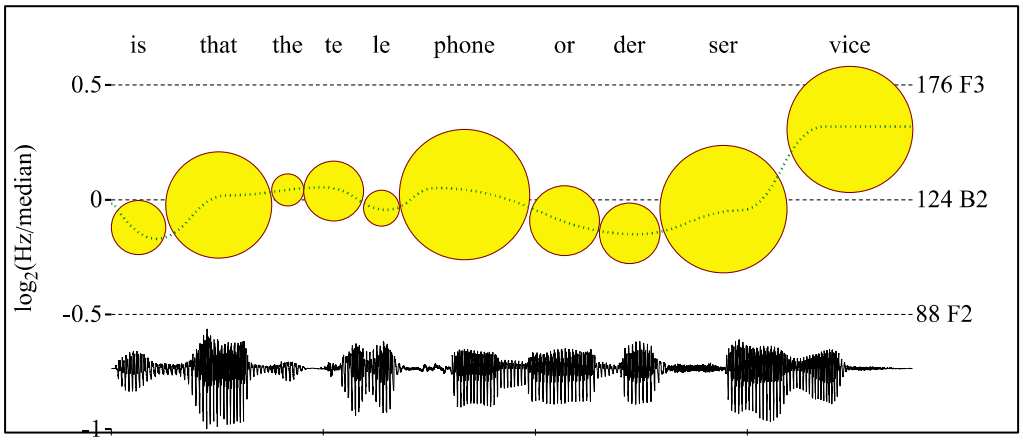
Native English Speaker



Is that the telephone order service?

Figure 8

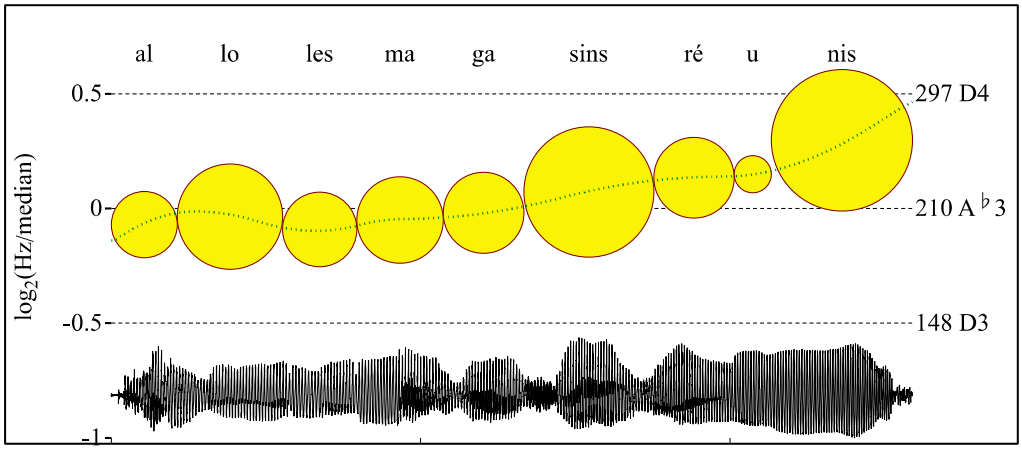
French Learner of English, Intermediate Level



Is that the telephone order service?

Figure 9

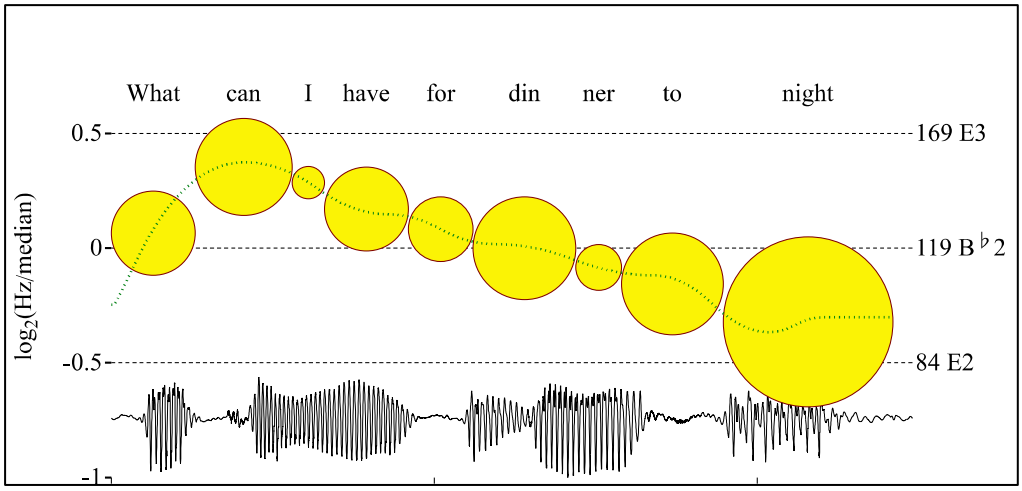
Native French Speaker



Allo, les Magasins Réunis?
Hello, is this the Magasins Réunis?

Figure 10

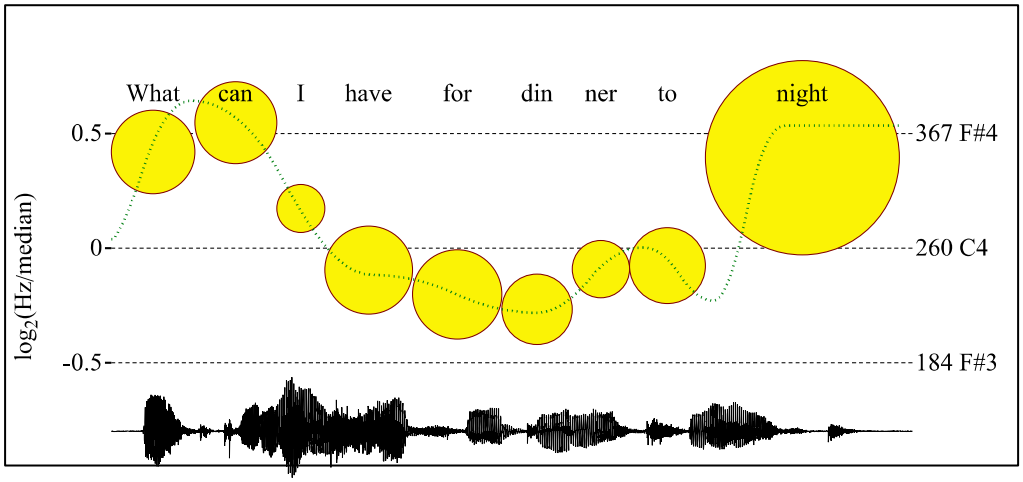
Native English Speaker



What can I have for dinner tonight?

Figure 11

French Learner of English, Intermediate Level



What can I have for dinner tonight?

5 Conclusion

Avenues of research and of teaching practice, derived from our experience of teaching English prosody to learners, have been proposed in this paper. The next step is to analyse the impact of visualisation on the learning of prosodic contours. A study testing two groups of learners, one learning prosody with the help of visualisation and the other without, could be set up. Other visualisation techniques could be used and developed. We believe that this is one way to effectively improve the learning and teaching of prosody, which is sorely lacking in English language classrooms.

References

Boersma, P., & Weenink, D. (2023). *PRAAT: doing phonetics by computer*. [Computer program]. Version 6.3.10, retrieved 3 May 2023 from http://www.praat.org

Bongiorno, J. (2021). *Étude du système intonatif de l'anglais parlé dans le sud de Dublin : focus sur les montées stylistiques*. [Unpublished doctoral thesis]. Aix-Marseille University, France.

Bongiorno, J., & Herment, S. (2018). A qualitative analysis of rising tones in Dublin English. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński & D. Śledziński (Eds.), *Proceedings of the 9th International Conference on Speech Prosody* (pp. 109-113), doi: 10.21437/SpeechProsody.2018-22.

Brazil, D. (1997). *The communicative value of intonation in English*. Birmingham Bleak House and ELR, 1985. Republished Cambridge University Press.

Contreras Roa, L. (2019). *Prosodie et apprentissage des langues. Étude contrastive de l'interlangue d'apprenants d'anglais francophones et hispanophones*. [Unpublished doctoral thesis]. University of Rennes 2, France.

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge University Press.

Cruttenden, A. (2014). *Gimson's pronunciation of English* (8th ed.). Routledge.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Herment, S. (2018). Apprentissage et enseignement de la prosodie : l'importance de la visualisation. *Revue Française de Linguistique Appliquée*, *8*(1), 73–88.

Herment, S., & Tortel, A. (2021). The intonation contour of non-finality revisited: Implications for EFL teaching. In A. Kirkova-Naskova, A. Henderson, & J. Fouz-González (Eds.), *English pronunciation instruction: Research-based insights* (pp. 175–195). John Benjamins.

Herment, S., Tortel, A., Bigi, B., Hirst, D., & Loukina, A. (2014). AixOx, a multi-layered learners corpus: Automatic annotation. In J. Díaz Pérez & A. Díaz Negrillo (Eds.), *Specialisation and variation in language corpora* (pp. 41–76). Peter Lang.

Herment, S., Leonarduzzi, L., & Bouzon, C. (2020a). The various rising tones in Newcastle English: A phonological distinction? *Anglophonia* [Online], *29*, http://journals.openedition.org/anglophonia/3297

Herment, S., Tortel, A., & Leonarduzzi, L. (2020b). The British English rising contour: An exception in read speech? *Proceedings of the 10th International Conference on Speech Prosody* (pp.196-200), doi: 10.21437/SpeechProsody.2020-40

Hirst, D. (2005). ProZed: A speech prosody editor for linguists, using analysis-by-synthesis. In K. Hirose & J. Tao (Eds.), *Speech prosody in speech synthesis: Modeling and generation of prosody for high quality and flexible speech synthesis* (pp. 3–17). Springer Verlag.

Horgues, C. (2010). *Prosodie de l'accent français en anglais et perception par des auditeurs anglophones natifs*. [Unpublished doctoral thesis]. University of Paris Diderot-Paris 7.

Kisilevsky, B.S., Hains, S.M., Lee, K., Xie, X., Huang, H., Ye, H.H., Zhang K., & Wang Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*(3), 220–224. https://doi.org/10.1111/1467-9280.02435. PMID: 12741744.

MacDonald, D. (2011). Second language acquisition of English question intonation by Koreans. In L. Armstrong (Ed.), *Proceedings of the 2011 Annual Conference of the Canadian Linguistic Association*. https://cla-acl.ca/pdfs/actes-2011/MacDonald_2011.pdf

Moon, C., Cooper, R. P., & Fifer W. P. (1993). Two-day-olds prefer their native language. *Infant Behaviour and Development*, *16*, 495–500.

Moore, D. R., & Jeffery, G. (1994). Development of auditory and visual systems in the foetus. In G. D. Thorburn & R. Harding (Eds.), *Textbook of fetal physiology* (pp. 278–286). Oxford University Press.

Niebuhr, O., Alm, M., Shümchen, N., & Fischer, K. (2017). Comparing visualization techniques for learning second language prosody: First results. *International Journal of Learner Corpus Research*, *3*(2), 250–277. https://doi.org/10.1075/ijlcr.3.2.07nie

Pytlyk, C. (2008). Interlanguage prosody: Native English speakers' production of Mandarin yes/no questions. In S. Jones (Ed.), *Proceedings of the 2008 Annual Conference of the Canadian Linguistic Association*. University of British Columbia. https://cla-acl.artsci.utoronto.ca/actes-2008-proceedings/

Roach, P. (2009). *English phonetics and phonology* (4th ed.). Cambridge University Press.

Rodrigues Da Mota, C., & Herment, S. (2016). The pragmatic functions of the final particle eh and of high rising terminals in Canadian English: Quite similar, eh!, *Proceedings of the 8th International Conference on Speech Prosody* (pp. 878-882), doi: 10.21437/SpeechProsody.2016-18.

Samlowski, B., Kern, F., & Trouvain, J. (2018). Perception of suspense in live football commentaries from German and British perspectives. *Proceedings of the 9th International Conference on Speech Prosody* (pp. 40-44), doi: 10.21437/SpeechProsody.2018-8.

Santiago-Vargas F., & Delais-Roussarie, E. (2012). La prosodie des énoncés interrogatifs en français L2. In L. Besacier, B. Lecouteux & G. Sérasset (Eds.), *Actes des Journées d'Études sur la Parole (JEP-TALN 2012)* (pp. 265–272). https://www.aclweb.org/anthology/F12-1034/

Santiago-Vargas, F., & Delais-Roussarie, E. (2015). The Acquisition of question intonation by Mexican Spanish learners of French. In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and languages in contact: L2 acquisition, attrition and languages in multilingual situations* (pp. 243–270). Springer Verlag.

Sharwood-Smith, M. (1983). On first language loss in the second language acquirer: Problems of transfer. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 222–231). Newbury House.

Tench, P. (1996). *The intonation systems of English*. Cassell.

Tortel, A. (2008). ANGLISH : base de données comparatives L1 et L2 de l'anglais lu, répété et parlé. *TIPA 27* [Online]. https://journals.openedition.org/tipa/321

Turcsan, G., & Herment, S. (2015). L'anglais d'Irlande du nord (Ulster). In I. Brulard, P. Carr & J.

Durand (Eds.), *La prononciation de l'anglais contemporain* (pp. 183–198). Presses Universitaires du Midi.
Warren, P. (2016). *Uptalk*. Cambridge University Press.
Wells, J. (2006). *English intonation, an introduction*. Cambridge University Press.
Wilhelm, S. (2015). Quand les paroles s'envolent : réflexions sur les caractéristiques et la forme phonétique du *High Rising Terminal* en anglais contemporain. *Anglophonia* 20 [Online]. https://journals.openedition.org/anglophonia/591

About the author

Sophie Herment is a full Professor in the department of English studies at Aix-Marseille University, France. She has taught English phonetics and phonology to French students at all levels of the university curriculum, including the teacher training programme, for 25 years. As a member of the Speech and Language Lab (LPL, co-funded by the university and the CNRS – the French National Centre for Research), of which she is deputy director, she works mainly on prosody. Her research interests concern L2 prosody, the varieties of English, the prosody/syntax interface and corpus prosody. She co-supervises the PAC programme (Phonology of Contemporary English).

Email: sophie.herment@univ-amu.fr

Horgues, C., & Scheuer, S. (2023). How L2 English pronunciation contributes to miscommunication in two French/English tandem conversation tasks. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 97–107). https://doi.org/10.5281/zenodo.8173999

How L2 English pronunciation contributes to miscommunication in two French/English tandem conversation tasks

Céline Horgues
Sorbonne Nouvelle-Paris 3 University

Sylwia Scheuer
Sorbonne Nouvelle-Paris 3 University

The paper is a contribution to the study of miscommunication in native–non-native speaker (NS–NNS) conversations. It analyses pronunciation-induced communication breakdowns (CBs) found in video-recorded face-to-face tandem conversations held in English by 21 pairs of students, each consisting of a native speaker of English and a native speaker of French (SITAF tandem corpus; Horgues & Scheuer, 2015). These NS–NNS interactions are shaped by specific linguistic, discursive, intercultural, and psycho-affective characteristics of the Tandem learning framework.

We draw on our previous research which showed that pronunciation was the single most important linguistic factor behind CBs arising from the speech of the NNS in a debating task. We now aim to look for possible task effects on the frequency and nature of the CBs by comparing two collaborative speaking tasks in English.

Our results confirm that the amount and types of CBs are indeed shaped by the tandem learning setting and tasks, and that pronunciation is the main impediment to the intelligibility of NNS English speech in the corpus. We hope our study contributes to a better understanding of the communicative impact of L2 pronunciation in authentic NS–NNS exchanges.

Keywords: L2 intelligibility, L2 pronunciation, tandem learning, NS–NNS communication, communication breakdowns

1 Introduction

Conversations between a native speaker (NS) and a non-native speaker (NNS) are generally described as presenting more opportunity for miscommunication than NS–NS exchanges (Varonis & Gass, 1985b), because the former setting presents a wider language proficiency and intercultural gap between interactants. *Miscommunication* has been used as an umbrella term to describe various kinds of "communicative turbulence" (Mauranen, 2006, p. 128), or instances of the flow of communication getting broken and therefore requiring some kind of reparation (repair negotiation in Nakahama et al., 2001). In the Second Language Acquisition (SLA) framework, miscommunication is often studied as the comprehension component of negotiation of meaning (NoM) (Bower & Kawaguchi, 2011; Gass, 2003; Varonis & Gass, 1985a, 1985b) or of language-related episodes (LREs). The latter are defined by Swain and Lapkin (1998) as moments in an interaction when participants "talk about the language they are producing, question their language use, or correct themselves or others" (p. 326), which involve the interactants attending to both mutual comprehension difficulties arising in conversation and non-target like language use through negative evidence (corrective feedback).

For the sake of clarity, we will use the term communication breakdown (CB) as a synonym for miscommunication or unintelligibility (taken in a broad sense, see §2), therefore including cases of non-understanding, misunderstanding or problematic comprehension. We will look at cases where communication breakdown is signalled (verbally, vocally or visually) by at least one of the interactants, and collaboratively attended to.

One particular context fostering NS–NNS communication is that of language tandem learning, which consists of non-formal spoken exchanges between two NSs of two different languages, who collaborate through regular, autonomous, spoken interactions to learn each other's language and culture (Brammerts & Calvert, 2003). Depending on which language is spoken in the tandem conversation, each participant takes on the role of the (relative) expert in the L1 language and culture, while the NNS partner is a (relative) novice or less proficient language user. What makes the tandem setting particularly fruitful is that the expert-novice relationship is fluid, dynamic, and reversible. The latter aspect refers to the role reversal resulting from the language switch, which allows for what Vassallo and Telles (2006) describe as a symmetrisation process (which they term *global symmetry*) reducing the local asymmetries in linguistic and cultural expertise between tandem participants (p. 95). The relationship between the two tandem interactants is, therefore, much more symmetrical, reciprocal, and non-hierarchical than in formal instructed language learning, or even in other NS–NNS conversations naturally occurring in daily life. These linguistic, intercultural, communicative and socio/psycho-affective characteristics of tandem learning will certainly play a key role in shaping the emergence and resolution of miscommunication in the course of tandem exchanges. Tandem partners may, on the one hand, be excessively charitable (Dascal, 1999; Varonis & Gass, 1985b) with their conversation partner and therefore downplay any communicative issues, but, on the other hand, such a non-threatening environment may encourage them to more readily signal miscommunication and engage in its management.

In this paper we focus on the emergence of CBs triggered (in part) by pronunciation issues (henceforth pronunciation-induced CBs) in the output produced by the NNS in the course of a tandem conversation in English with a NS partner. We aim to explore whether the speaking task may affect the quantity and quality of such miscommunication instances.

2 Previous research

Many studies that looked at how L2 speech may affect communication focus either on intelligibility or comprehensibility. With respect to the effects of L2 pronunciation in particular, the experimental method generally consists in submitting selected audio excerpts taken from L2 speech to (asynchronous) perceptive evaluation by NS listeners (see Wheeler & Saito, 2022, for a recent review of various intelligibility measurement techniques). This method tests how much verbal content these listeners can decode (word recognition level – narrow definition of intelligibility) or gauges how much processing effort is required of these listeners (definition of comprehensibility; see for instance, Levis, 2018; Munro & Derwing, 1995). It appears that not only segmental (Suzukida & Saito, 2019; Zielinski, 2008) but also – and sometimes predominantly – suprasegmental deviations in NNS speech (Henderson, 2008; Kang, 2010) hamper intelligibility and/or comprehensibility for NS interlocutors. These results need to be put into perspective, as the perceptual evaluation techniques and learners' L1 backgrounds are not directly comparable across studies. In line with Bamgbose (1998) who defined intelligibility as "a complex of factors comprising recognizing an expression, knowing its meaning, and knowing what that meaning signifies in the sociocultural context" (p. 11), we take a broader view on (un-)intelligibility, encompassing subconcepts such as intelligibility, comprehensibility, and interpretability because, although these levels can be teased apart in (quasi-)experimental perceptual designs, they are closely intertwined in situated, authentic interaction.

In experimental research, the evaluation of the predicted effects and gravity of L2 phonological deviations is often operationalised using the functional load principle, which takes a *theoretical* approach to the interpretation of lexical confusion (Munro & Derwing, 2006; Suzukida & Saito, 2019). However, the *actual* impact of such deviations on real-life interactions has been understudied and authors such as Henderson (2008) have also pointed out the necessity to take the sociolinguistic (not just functional) weighting of pronunciation errors into account as well. On the other hand, in research exploring NS–NNS miscommunication from an interactionist point of view (Bower & Kawaguchi, 2011; Nakahama et. al., 2001; Strawbridge, 2021; Varonis & Gass, 1985a, 1985b), pronunciation issues are generally given very little or no detailed attention. In his study of Spanish/English e-tandem exchanges, Strawbridge (2021) determined that 21% of the Reactive LRE triggers were phonetic (33% lexical, 33% global, 14% morphosyntactic). However, this proportion merged both L2 Spanish and L2 English speech results and his global category did not specify the contribution of pronunciation-related errors. Our previous study (Scheuer & Horgues, 2021) showed that in a selected part of our tandem corpus (Game 2 in English, see §3), L2 pronunciation represented the single most crucial CB trigger emerging from the speech of French learners of English addressed to NSs (ahead of lexical or morphosyntactic triggers). There were also more suprasegmental (especially word stress-related) than segmental triggers. We would now like to extend this study to explore potential task effects on the role of L2 pronunciation on NS/NNS communication.

In previous studies exploring these kinds of effects, task-types are often distinguished in terms of planning, structuration, control, and predictability of the speakers' output. These tasks vary from unplanned open-ended conversations resulting in basic "information exchange" (Strawbridge, 2021), to information-gap activities (e.g., picture description, map tasks, spot-the-difference) or academic lectures. Compared to controlled speech, less planned speech may be expected to present more comprehension difficulty (by virtue of being unpredictable and unstructured). On the other hand, it is also characterised by lower lexical density and fewer polysyllabic words (Henderson, 2008). Furthermore, it allows speakers to drop or avoid problematic topics (Nakahama et al., 2001). Nakahama et al. (2001) found that task-type not

only influenced the quantity of miscommunication events (fewer in uncontrolled conversations) but also their quality; there were more mechanical, local management and resolution events in the information gap-activity compared to the conversation task. Looking at pronunciation more specifically, contrary to Crowther et al., (2015), Suzukida and Saito (2019) found no task effect in their comparison of the contribution of L1 Japanese L2 English on NSs' comprehensibility ratings. High functional load consonant substitutions hampered comprehensibility equally in their two tasks (picture description task vs. unstructured IELTS[1] long-turn interview). Differences in the types of speech analysed (monologic vs. interactive), L1/L2 configurations, specifics of the learners' L1 phonology, but also in the methods used to compare speaking tasks or investigate L2 intelligibility/comprehensibility may explain the mixed results obtained in these studies.

Most previous research has focused on raters' indirect and *a posteriori* evaluation of the communicative impact of L2 pronunciation, where speaking materials and perceptual tasks are often not naturalistic or contextualised. It is, however, important to explore L2 intelligibility in action (i.e., in interaction) by focusing on its real-life *in-situ* effects. Therefore, we have adopted a methodology observing interactants' actual communicative behaviour in the course of authentic NS–NNS conversations. Exploiting a video-recorded corpus also makes it easier to examine the main speaker's and their interlocutor's non-verbal reactions, which are otherwise absent from any audio analysis. This was pointed out by Wheeler and Saito (2022) who regret that "the vast majority of L2 intelligibility research relies on audio-only stimuli" (p. 429).

2.1 Research questions

Our overarching research question is: does communicative task-type have an effect on how a NNS's output (especially L2 pronunciation) may generate miscommunication with a NS interlocutor? We will address it by raising the following sub-questions:

RQ1: Are CB episodes more, or less, prevalent in the narrative vs. the debating task in this tandem setting?
RQ2: What is the relative contribution of pronunciation issues compared to other types of triggers in these two tasks?
RQ3: What is the relative contribution of segmental vs. suprasegmental errors to the pronunciation-induced CBs?

3 Research methodology

The SITAF corpus (Horgues & Scheuer, 2015) consists of video-recorded face-to-face tandem conversations held in English and French by 21 pairs of students, each tandem being formed by a NS of English and a NS of French. All participants were undergraduate students at Sorbonne Nouvelle-Paris 3 University and their self-reported L2 proficiency level ranged from intermediate to advanced. The 21 Anglophones were speakers of different English varieties (British, American, Irish, Canadian, Australian) and the Francophones were French students majoring primarily in English studies. The tandem pairs met weekly for autonomous tandem conversations over one academic semester. They were recorded twice, performing the exact same speaking tasks in the two languages (two collaborative game-like activities and one monitored reading task) at the beginning of their tandem experience (session 1) and then three

[1] International English Language Testing System (IELTS)

months later (session 2). This study only concerns miscommunication arising in the semi-spontaneous activities held in English. Game 1 (story-telling) consists in the L2 speaker narrating a personal story integrating three lies which their interlocutor tries to elucidate by asking questions. The narrative performed by the NNS is monologic for the most part, albeit interactive at times. Game 2 (debating) consisted in the two interactants giving their opinion about a set of controversial topics, in order to then decide on the degree of like-mindedness between them. It is characterised overall by more symmetry (the NS and the NNS's speaking times and contributions to discourse) than Game 1.

We will look at cases where communication breakdown is signalled (verbally, vocally or visually) and resolved through interactional work. We will therefore exclude cases where breakdowns are prevented from happening through anticipation (Preemptive LREs; Strawbridge, 2021) or where a speaker self-resolves comprehension issues individually or avoids engaging in miscommunication management (let-it-pass strategy). To determine a case of CB, the two authors inspected all video-recorded sequences independently and followed the same perspective as Nakahama et al. (2001) in relying on the observation of the *interlocutor*'s verbal, vocal, and visual reactions to problematic NNS speech, i.e., when the recipient demonstrably had difficulty, or was incapable of, grasping the meaning of an utterance as seemingly intended by the speaker. We agreed in our identification and classification of 80% of CB cases and then discussed the remaining 20%, to reach a consensual decision after further and joint reviewing of the debatable multimedia sequences. We study CB sequences using Varonis and Gass (1985a)'s overall structure: Trigger (Speaker), Indicator or Signal (Hearer), Response (Speaker), Reaction to Response (Hearer). In this study, we are mostly interested in defining the type of trigger (pronunciation, morphosyntactic, lexical, pragmatic, cultural, or a mix of any of these) emerging from NNS output (i.e., the French learner of L2 English being the initiator). Example (1) is an instance of a CB sequence from Game 1, where the main trigger was L2 pronunciation:

(1) NNS: [There were] Hens (['ɛns]) *[TRIGGER]*
NS: *(silence at first)* Ants? *[SIGNAL]*
NNS: *(makes a clucking sound and a gesture of flapping wings)* *[RESPONSE]*
NS: Chickens? *[REACTION TO RESPONSE]*

We also quantified the amount of speech produced by the native and the non-native speakers by counting the number of words produced by each interactant during a collaborative speaking task. This was done through automatic extraction (Unix/Linux script) of word counts from the manual transcriptions prepared by the SITAF team members (Transcriber programme).

4 Data analysis and results

4.1 Communication breakdowns (CBs) in narrative vs. debating task in tandem setting (RQ1)

In total, the narrative Game 1 generated more CBs than the debating Game 2 (see Table 1 for a summary) – 39 tokens as opposed to 21 respectively, which represents an increase by a factor of nearly 1.9. The difference very nearly reaches statistical significance, although it needs to be considered in context. The two tasks differed significantly in length, understood as the number of words produced by the NNS; on average, 831 words in Game 1 and 501 in Game 2. The fact that, lengthwise, the narrative exceeded the debate by a factor of nearly 1.7 means that there were, proportionally, more communication breakdowns in Game 1 than in Game 2, although the difference between the CB-to-word ratios in the two tasks was not statistically

significant. It should therefore be concluded that no tangible task effect was found on the relative frequency of CB episodes generated by NNS discourse.

Table 1

Length and Communication Breakdowns Frequency: Game 1 vs. Game 2

	Game 1	**Game 2**	***p***
Total number of CBs in NNS discourse	39	21	.0537
Average number of words produced by NNS	831	501	***.00012**
CB-to-word ratio in NNS discourse	0.0024	0.0015	.36
Average number of words produced by both partners (NS + NNS)	1174	1063	.2

Note. All the word counts and ratios were calculated on a sample – deemed representative – of 15 pairs for which transcriptions were fully exploitable.

To further put our results in context, the relative contribution of the NS – i.e., the mis- or non-understander – to the conversation is worth mentioning. As expected, the nature of Game 1 gave more speaking time to the learner. This resulted in the NNS uttering on average 2.4 times more words than their NS partner (831 words vs. 343; very highly significant). Game 2, conversely, presented a fairly symmetrical picture, with the NS actually producing marginally more speech on average than the NNS (562 words vs. 501; non-significant). Game 1 was somewhat longer than Game 2 in terms of the average number of words produced by both tandem partners, i.e., the NS and the NNS: 1175 words in Game 1 vs. 1063 in Game 2 (non-significant). It should be clarified that, as per the task instructions, the two games were intended to be roughly of the same length (5 mins each).

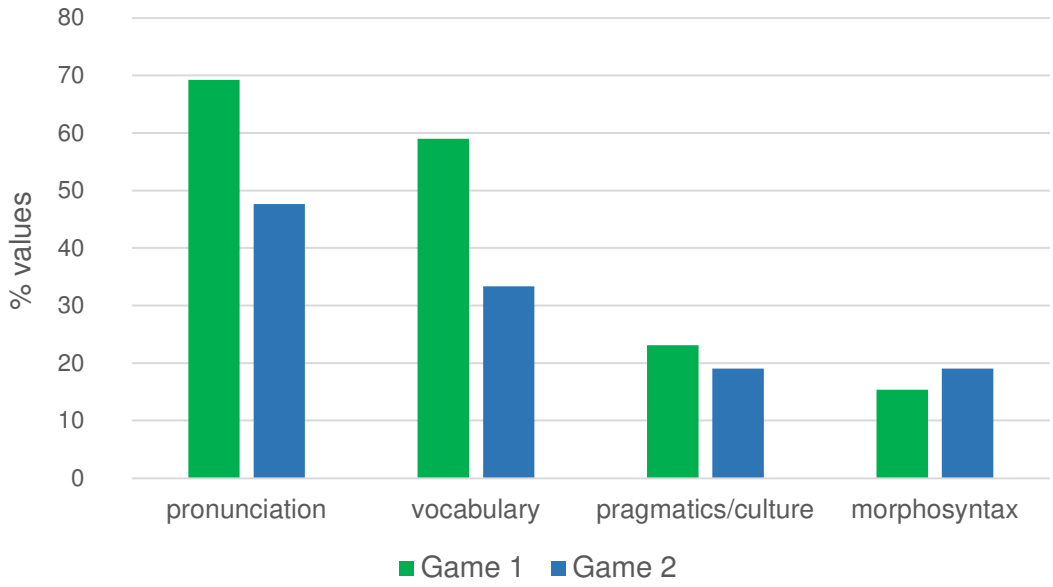
Finally, it should be noted that there were marked differences between individual pairs, both in terms of word counts and, especially, the number of signalled CBs arising during the two tasks. While some tandems consistently showed no CBs in either game or session, one (Pair 09) contributed 9 tokens, i.e., 15% of the total of 60 found in the corpus.

4.2 Relative contribution of pronunciation issues as triggers (RQ2)

As previously mentioned, pronunciation was identified as the most important trigger of CBs in Game 2, accounting, at least partially, for 10 out of the 21 CBs in the debating task (48%). There are, proportionally, even more pronunciation-induced CBs in Game 1: 27 out of 39 (69%). These numbers include the more complex mixed cases, where pronunciation was identified as one of the linguistic factors. Even though the difference is not statistically significant, pronunciation once again surpassed vocabulary, as well as the other triggers explored (see Figure 1).

Figure 1

Communication Breakdowns by Trigger Type



Note. The mixed cases are included in each applicable category. Therefore, some CB instances appear more than once in the above data, which accounts for % values which add up to more than 100.

Example (2) illustrates a mixed CB trigger, involving and counting as both pronunciation and vocabulary (wrong preposition):

(2) NNS: I went in Latvia [*'lad'vja]
NS: You went where?

4.3 Relative contribution of segmental vs. suprasegmental errors to the pronunciation-induced CBs (RQ3)

While the vast majority of the pronunciation-induced CBs in Game 2 had a suprasegmental overlay (9 out of 10), this was not the case in Game 1. In the narrative task, suprasegmentals were identified as playing a role in 14 out of the 27 instances, some of which were also compounded by segmental inaccuracies. Word stress emerged as the key issue among the suprasegmentals, featuring in 11 CB cases in Game 1. This is shown in the following exchange in example (3), where the stress problem was accompanied by a lexical one, at least from the American NS's perspective:

(3) NNS: I'm going to talk you about my last summer va... holiDAYS
NS: Summer…? *(forward movement of head / trunk)*
NNS: Summer vacation
NS: Vacation, yeah.

The remaining three suprasegmental items concerned incorrect syllable count (i.e., deletion or addition), as illustrated by example (4), where the NS's comprehension is compromised by a missing syllable:

(4) NNS: It was great. I was the sixtith [*ˈsɪkstiθ] err floor
NS: sixth floor *(nod)*?
NNS: yeah, err…sixteenth...err sixty *(hand gesture for height)*
NS: sixtieth floor?
NNS: yeah!

5 Discussion

The present analysis of the narrative Game 1 complements the existing picture of miscommunication in the English conversational data of the SITAF corpus. The tendency for this largely monologic task to generate more CBs than the dialogic Game 2 (RQ1), even after adjusting for length differences, is perhaps not altogether surprising. Game 1 revolved around the NNS's individual conceptualisation and encoding of their own personal story (life anecdotes unknown to the interlocutor). Since there was no common background set by the task instructions, most of the content was initiated solely by the NNS, who was therefore the only 'knower'. The NS, being largely the recipient of the story, may therefore have had a more difficult job trying to interpret their partner's L2 output than in Game 2, where there was a set, imposed topic, and therefore more room for collaborative structuring of the debate dynamics. Examples (2) and (3), which occurred at the beginning of Game 1, show how this 'out-of-the-blue' effect may have contributed to the lack of understanding: the NS was not primed to process *Latvia* or *summer holidays*, whose rendition was – additionally – imperfect. The task rules and goals of Game 1 may also have created a stronger need for the NS to understand the details of the NNS discourse than that of Game 2, and therefore a need to overtly signal communicative turbulence. Given that the NS was meant to identify the three lies incorporated into their partner's story, there was probably less room for the let-it-pass strategy, whereby the listener can choose to simply ignore certain processing difficulties. These possible task effects are not directly comparable to what previous studies found, as – although characterised by a convergent goal and outcome – the speaking tasks in our corpus are neither as open-ended and unplanned as free conversations (Strawbridge, 2021) nor as controlled and limiting as a map-task or spot-the-difference activity (Nakahama et al., 2001).

The findings relating to the linguistic triggers of communication breakdowns (RQ2) point to the powerful role played by L2 pronunciation problems in impeding intelligibility. This is true for both tasks, but the trend is even stronger in Game 1 than Game 2 (69% vs. 48%). The reasons may tie in with the interpretations previously offered: L2 mispronunciations have vast potential for blurring the word's identity, but this becomes an even more acute problem when the context is largely unknown and unpredictable, compared to a dialogue where the interlocutors usually build on each other's contributions. The fact that Game 1 provided more extensive data than Game 2 (both in terms of length and total number of CBs) testifies perhaps even more strongly to the true weight of pronunciation in maintaining smooth NS–NNS communication, relative to other linguistic domains such as vocabulary or morphosyntax.

Comparisons between the two tasks should, however, be drawn with caution, due to the sometimes-low number of tokens involved. This becomes particularly relevant when interpreting data from RQ3, with suprasegmental problems coming strongly to the fore in Game 2 but not as much in Game 1. Another complication stems from the fact that it is often difficult to disentangle suprasegmentals from segmentals. For example, stress and vowel quality are usually intertwined, to the extent that "nonstandard English word stress is largely

defined by the presence or absence of vowel errors" (Richards, 2016, p. 105). Despite such methodological issues, it is likely that the nature of the task played a role in shaping the results. The NNS, being in charge of the narrative in Game 1, may have been more successful in avoiding complex polysyllabic words than in Game 2, which resulted in the scope for suprasegmental errors being somewhat reduced in the former. Conversely, this may have enhanced the importance of segmental details in Game 1, as single segmental errors have a larger effect on reducing intelligibility in monosyllabic than in polysyllabic words (Wheeler & Saito, 2022).

6 Conclusion and implications

The SITAF miscommunication data, now complemented by the analysis of the narrative task, point to pronunciation as the principal trigger of CBs arising in connection with non-native English speech. Yet, pronunciation is often viewed as an optional component of L2 teaching, an "add it on if we have time" feature (Levis, 2018, p. 1). This is in contrast to morphosyntax, which tends to be a long-standing favourite in the EFL classroom, but which actually ranked last among the four linguistic triggers in our CB data. Importantly, our findings demonstrate the potential of seemingly minor (from the student's perspective) pronunciation errors for generating miscommunication, such as the missing /h/ in example (1), and the rightward stress shift in (3). The latter are characteristic of L1 French learners of English and their communicative impact should be brought to their attention. More general pedagogical implications suggest L2 learners should be taught how to develop strategies to manage these CBs effectively (rather than avoiding them altogether) in NS/NNS interactions. Our study also points towards some effect of task type and this should be investigated further to account for the emergence, type and resolution of pronunciation-induced CBs in our corpus. We hope our study contributes to a better understanding of the communicative impact of L2 pronunciation in authentic, real-life NS–NNS exchanges.

Acknowledgements

We are very grateful for the methodological and financial support received from Sorbonne Nouvelle-Paris 3 University, Labex EFL program (ANR-10-LABX-0083), Ircom/Ortolang, and M. Adda-Decker who helped with Unix/Linux scripts to export word counts. We thank all SITAF participants and team members.

References

Bambgose, A. (1998). Torn between the norms: innovations in world Englishes. *World Englishes*, *17*(1), 1–14. https://doi.org/10.1111/1467-971X.00078

Bower, J., & Kawaguchi, S. (2011). Negotiation of meaning and corrective feedback in Japanese/English etandem. *Language Learning & Technology*, *15*(1), 41–71. ISSN 1094-3501

Brammerts, H., & Calvert, M. (2003). Learning by communicating in tandem. In T. Lewis & L. Walker (Eds.), *Autonomous language learning in tandem* (pp. 45–59). Academy Electronic Publications.

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does a speaking task affect second language comprehensibility? *The Modern Language Journal, 99*(1), 80–95. https://doi.org/10.1111/modl.12185

Dascal, M. (1999). Introduction: Some questions about misunderstanding. *Journal of Pragmatics*, *31*(6), 753–762. https://doi.org/10.1016/S0378-2166(98)00059-9

Gass, S. M. (2003). Input and interaction. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 224–255). Blackwell. https://doi.org/10.1002/9780470756492.ch9

Henderson, A. (2008) Towards intelligibility: Designing short pronunciation courses for advanced field experts. *ASp*, *53-54*, 89–110. https://doi.org/10.4000/asp.369

Horgues, C., & Scheuer, S. (2015). Why some things are better done in tandem? In J. A. Mompeán & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 47–82). Palgrave Macmillan. https://doi.org/10.1057/9781137509437_3

Horgues, C., & Scheuer, S. (2018). "What are you not understanding?" Managing communication breakdowns in native/non-native English/French tandem interactions. In E. Babatsouli (Ed.), *Crosslinguistic research in monolingual and bilingual speech* (pp. 129–146). Institute of Monolingual and Bilingual Speech.

Levis, J. (2018). *Intelligibility, oral communication, and the teaching of pronunciation.* Cambridge University Press. https://doi.org/10.1017/9781108241564

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, *38*(2), 301–315. https://doi.org/10.1016/j.system.2010.01.005

Mauranen, A. (2006). Signaling and preventing misunderstanding in English as a lingua franca communication. *International Journal of the Sociology of Language*, *177*, 123–150. https://doi.org/10.1515/IJSL.2006.008

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–91. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, *34*, 520–531. https://doi.org/10.1016/j.system.2006.09.004

Nakahama, Y., Tyler, A., & van Lier, L. (2001). Negotiation of meaning in conversational and information gap activities: A comparative discourse analysis. *TESOL Quarterly*, *35*(3), 377–405. https://doi.org/10.2307/3588028

Richards, M. G. (2016). *Not all word stress errors are created equal: Validating an English word stress error gravity hierarchy* (Publication No. 16001) [Doctoral dissertation, Iowa State University]. Graduate Theses and Dissertations. https://doi.org/10.31274/etd-180810-5628

Scheuer, S., & Horgues, C. (2021). Corrective feedback and unintelligibility: Do they work in tandem during tandem interactions? In A. Kirkova-Naskova, A. Henderson & J. Fouz-González (Eds.), *English pronunciation instruction: Research-based insights* (pp. 224–251). John Benjamins. https://doi.org/10.1075/aals.19.10sch

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal*, *82*(3), 320–337. https://doi.org/10.1111/j.1540-4781.1998.tb01209.x

Strawbridge, T. (2021). Modern language: Interaction in conversational NS/NNS video SCMC eTandem exchanges. *Language Learning & Technology*, *25*(2), 94–110. https://doi.org/10125/73435

Suzukida, Y., & Saito, K. (2019). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Language Teaching Research, 25*(3), 431–450. https://doi.org/10.1177/13621688198582

Varonis, E. M., & Gass, S. (1985a). Non-native/non-native conversations: A model for negotiation of meaning. *Applied Linguistics*, *6*(1), 71–90. https://doi.org/10.1093/applin/6.1.71

Varonis, E. M., & Gass, S. (1985b). Miscommunication in native/nonnative conversation. *Language in Society*, *14*(3), 327–343. https://doi.org/10.1017/S0047404500011295

Vassallo, M. L., & Telles, J. A. (2006). Foreign language learning in-tandem: Theoretical principles and research perspectives. *The ESPecialist*, *27(*1), 83–118.

Wheeler, P., & Saito, K. (2022) Second language speech intelligibility revisited: Differential roles of phonological accuracy, visual speech, and iconic gesture. *The Modern Language Journal*, *106*(2), 429–448. https://doi.org/10.1111/modl.12779

Zielinski, B. W. (2008) The listener: No longer the silent partner in reduced intelligibility. *System*, *36*(1), 69–84. https://doi.org/10.1016/j.system.2007.11.004

About the authors

Céline Horgues is an Associate Professor in English phonetics and phonology at Sorbonne Nouvelle-Paris 3 University, France, where she teaches English phonetics/phonology, listening comprehension and acquisition/teaching of L2 pronunciation. In 2010 she defended a PhD on the prosody of French learners of L2 English. Her main research interests are L2 pronunciation teaching/learning, L2 prosody, foreign accentedness, native/non-native interactions and tandem learning. With Sylwia Scheuer, she coordinated the collection and analysis of a multimodal corpus of English-French tandem spoken interactions (the SITAF corpus, 2012-2014). Since 2012 she has been running a tandem course and a free tandem programme for the students of her department. In 2019 she co-edited a book on tandem language and culture learning in higher education with Claire Tardieu.

Email: celine.horgues@sorbonne-nouvelle.fr

Sylwia Scheuer is an Associate Professor in phonetics and phonology at the Department of English, Sorbonne Nouvelle-Paris 3 University, France. She got her PhD from the School of English, Adam Mickiewicz University in Poznań, Poland, where she worked as an Associate Professor until 2005. Her PhD dissertation provided a corpus-based analysis of interference-motivated pronunciation errors in Polish students of English. She was a visiting lecturer at the Department of Linguistics, University of Vienna, 1998-2001. Her teaching and research have focused on second language acquisition, phonetics and phonology, general linguistics, sociolinguistics, and English as an International Language. With Céline Horgues, she coordinated a research project for the collection and analysis of a multimodal corpus of English-French tandem spoken interactions (SITAF, 2012-2014).

Email: sylwia.scheuer-samson@sorbonne-nouvelle.fr

Kalvodová, L., & Skarnitzl, R. (2023). Production and perception aspects of weak forms in Czech-accented English. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 108–121). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8173995

Production and perception aspects of weak forms in Czech-accented English

Lenka Kalvodová
Charles University, Prague

Radek Skarnitzl
Charles University, Prague

Czech is mostly syllable-timed, while English is mostly stress-timed and partly constructs rhythm through connected speech processes (CSPs), which cause unstressed grammatical words to change into weak forms. Since these are particularly challenging for speakers with syllable-timed mother tongues, this study addresses the production and perception of weak forms in proficient Czech speakers. In the production part, we examined Czech-accented speakers producing material loaded with grammatical words and CSPs. Subsequent listening analysis confirmed that the more accented the speaker, the fewer weak forms and associated CSPs can be found in their production. In the perceptual part, we used two perception tests containing realisations where CSPs had been manipulated to create one more and one less native-like version of a sentence. Czech respondents were asked to assess comprehensibility and accentedness. The results suggest that comprehensibility is easier for them to assess than accentedness, regardless of whether they had received phonetics and phonology instruction prior to the experiment. Our results correspond with and expand upon previous findings, both for Czech speakers (e.g., Skarnitzl & Rumlová, 2019) and for non-native speakers in general (e.g., Barańska & Zając, 2014), showcasing non-native patterns in Czech speakers. Yet, despite their importance and frequency, CSPs have been given little space in research and teaching.

Keywords: connected speech processes, grammatical words, weak forms, Czech-accented English

1 Introduction

Given the differences between Czech and English, it is likely that Czech speakers will transfer their native language (L1) habits into their L2 English. Among others, this will affect rhythmic patterning, which includes, particularly in English, the pronunciation of weak forms (WFs) of grammatical words. The aim of our study is to establish whether there are any patterns in production and perception of WFs and associated connected speech processes (CSPs) in Czech-accented proficient learners of English. The importance of weak forms as part of connected speech (CS) in English language acquisition is highlighted.

2 Previous research

2.1 Speech rhythm and connected speech

Speech rhythm is treated today not as a regular occurrence of a speech unit in time, but rather as a sequence of contrasting units, i.e., those with greater and smaller prominence (Nolan & Jeon, 2014; Turk & Shattuck-Hufnagel, 2013). In line with this approach, the traditional binary division of languages into strict categories is replaced by a rhythm continuum whose theoretical ends would correspond to the stress- and syllable-based languages (Nolan & Jeon, 2014). Crucially, however, (strict) timing regularities are no longer part of the definition of speech rhythm (see also Cauldwell, 2002), and no language corresponds to either theoretical end. Languages' rhythm profiles lie on this continuum, meaning that, when compared, two languages can be relatively closer to or further away from each other, the latter being the case of Czech and English.

The mostly stress-based rhythm of English is largely constituted by prominence contrasts at stress group level (Brown & Kondo-Brown, 2006, p. 2), with stressed syllables functioning as peaks of prominence, as compared to unstressed syllables. In contrast, all syllables are of roughly the same prominence in Czech (Skarnitzl & Eriksson, 2017), a language traditionally described as syllable-timed.

Out of the four levels of stress in a unit of speech as defined by Volín & Johaníková (2018, p. 181), unstressed syllables are of interest for us, since they can be further divided into full unstress (containing a full vowel) or weak unstress (containing a reduced one).

A language's rhythm profile is facilitated by the (partially) language-specific CSPs; otherwise, rhythm would be disrupted, inhibiting perception (e.g., Barańska & Zając, 2014). CS helps maintain speech rhythm through several processes, with modifications taking place in some words when they occur in CS (Shockey, 2003). As there is still a lack of extensive research on CSPs, and the terminology is far from unified, we find it fit to include an overview of the core CSPs as referred to in this article (but see, e.g., Alameen & Levis, 2015 for a slightly different categorisation).

First of all, vowel reduction contributes strongly to the rhythmic profile of English (e.g., in the phrase *There was a man* [ðə wəz‿ə], the first three words will only have the mid central vowel schwa). Schwa is known to constitute approximately one quarter of all vowels occurring in natural English speech (e.g., Volín et al., 2013, pp. 32–33).

The second group of processes contains various types of assimilation: a) place of articulation (e.g., /n/ → /m/ in *in Prague* [ɪm˺praːg]); b) manner of articulation (e.g., /d/ → /n/ in *good night* [gʊn˺naɪt]); and c) coalescence (e.g., /d+j/ → /dʒ/ in *could you* [kʊdʒ‿u]). An instance of place assimilation is dentalisation, which occurs frequently in English (e.g., alveolar /t/ influenced by (inter)dental /ð/ in *get this* [geṯ˺ðɪs]). In native-like speech production, stop consonants are often unreleased (e.g., /t/ in final position in *not* in the phrase *not to* [nɒt˺tə]).

Linking is a key characteristic of English sound patterns, with words beginning with a vowel typically linked to the previous word (e.g., *can I* [kən‿aɪ], *see it* [siː(j)ɪt]). Finally, elision, or, the deletion of a sound, is also an integral part of English CS (e.g., /r/ in *from* in the phrase *from Paris* [fəm pærɪs], /d/ in *and* in the phrase *you and me* [juː(w)ən miː]). Naturally, in certain contexts, more CSPs may be combined, as is the case of an elision of word-initial [h] in *he* with linking to the preceding word (e.g., *does he* [dəz‿i]).

2.2 Weak form words in English as L1 and L2

Grammatical words constitute a major component of the rhythmic pattern of English: bearing little prominence and semantic load, they participate strongly in CSPs, where they are frequently realised as what is called their weak form (e.g., *must* realised as [məst] or [məs], *them* as [ðəm] or [əm]). These weak forms of grammatical words (see the comprehensive overview by Cruttenden, 2014, or the discussion of CSP typology by Shockey, 2003) almost invariably contain the reduced vowel schwa [ə], and they make up the "filling" between the prominent peaks of lexical words.

The absence of WFs and CSPs in the speech of L2 speakers may impact not only their accentedness (i.e., the strength of foreign accent), but also their comprehensibility (i.e., the subjective ease of processing of their speech). From the contrastive perspective, this may be partly caused by the absence of (systematic) vocalic reduction in the native language of L2 speakers. Alameen and Levis (2015, p. 160) point out that the way L2 speakers produce CS may pose a challenge to native speaker listeners. At the same time, comprehensibility issues may arise even for other L2 speakers who speak a different variety. However, native-like production of WFs by L2 learners may also cause problems in the perception of other L2 speakers of English, especially beginners.

Given the high frequency and the essential role of connected speech phenomena in English, they should certainly be targeted in pronunciation teaching. However, several publications, including Brown and Kondo-Brown (2006) or Alameen and Levis (2015), point out that CSPs and WFs are marginalised in EFL curricula. A number of studies (e.g., Alameen & Levis, 2015; Brown & Kondo-Brown, 2006; Izumi, 2003) show that WF and CSP instruction is beneficial for learners' production of the target language, as well as the perception of native speech. Greater attention should also be given to vowel quality instruction (see Barańska & Zając, 2014 or Volín & Johaníková, 2018).

2.3 Research questions and hypotheses

In this study, we are interested in how relatively advanced Czech speakers of English produce weak forms of grammatical words and associated connected speech processes in specially designed sentences which are rich in these phenomena. Our second objective is to find whether speech manipulated to feature near-native-like patterns of WFs and CSPs will be regarded by Czech listeners as less accented and more comprehensible.

Specifically, our hypotheses are as follows:

H1A: Czech-accented speakers of L2 English show deviations from native-like production of weak forms of grammatical words and the associated CSPs.

H1B: The deviations from native-like production are modulated by the speakers' overall accentedness.

H2A: Czech learners of English perceive WFs and CSPs in English through the qualities of accentedness and comprehensibility.

H2B: Comprehensibility is more difficult to assess for Czech-accented learners of English than accentedness.

H2C: The more advanced the student, the more their perceptual assessment of accentedness and comprehensibility will correspond to stimulus manipulations.

3 Experiment 1: Production

3.1 Research methodology

To map how advanced Czech learners of English produce weak forms, we recorded 34 volunteers reading out loud 24 sentences from paper. The sentences were designed to ensure the occurrence of all types of WFs. The expected default realisation of WFs and CSPs (see Appendix) was based on Standard Southern British English.

All speakers were L1 Czech female first-year BA students (aged 19–25) in English Studies and English Translation Studies at the Faculty of Arts in Prague, for which they had previously passed an entrance exam, ensuring proficiency at the B2/C1 CEFR level (Council of Europe, 2020). Although the speakers had not yet received any instruction concerning weak form words or CS at the time, to assure that the sentences are not perceived as unnatural, we used varied but simple lexical words to complement the large concentration of grammatical words. The recordings were obtained in a recording booth of the sound-treated studio at the Institute of Phonetics in Prague, using an AKG C4500 B-BC condenser microphone at a sampling rate of 32 kHz with 16-bit quantisation.

The 34 speakers were divided into two groups based on their overall accentedness, as determined by the authors. Subsequently, 12 speakers per group, representing the most and the least accented productions, were chosen for analysis by the authors. It should be noted that assessing the speakers' overall accentedness involves subconsciously taking into account WFs and associated CSPs, as these are likely to participate in the perception of accentedness. While it could therefore be argued that the group division is to a certain extent circular, this did not interfere with our research purpose.

Each realisation of a phenomenon of interest was analysed in Praat (Boersma & Weenink, 2021) and labelled, and all entries were extracted using a script. Processed in R (R Core Team, 2021), the refined data was visualised using the ggplot2 package (Wickham, 2016).

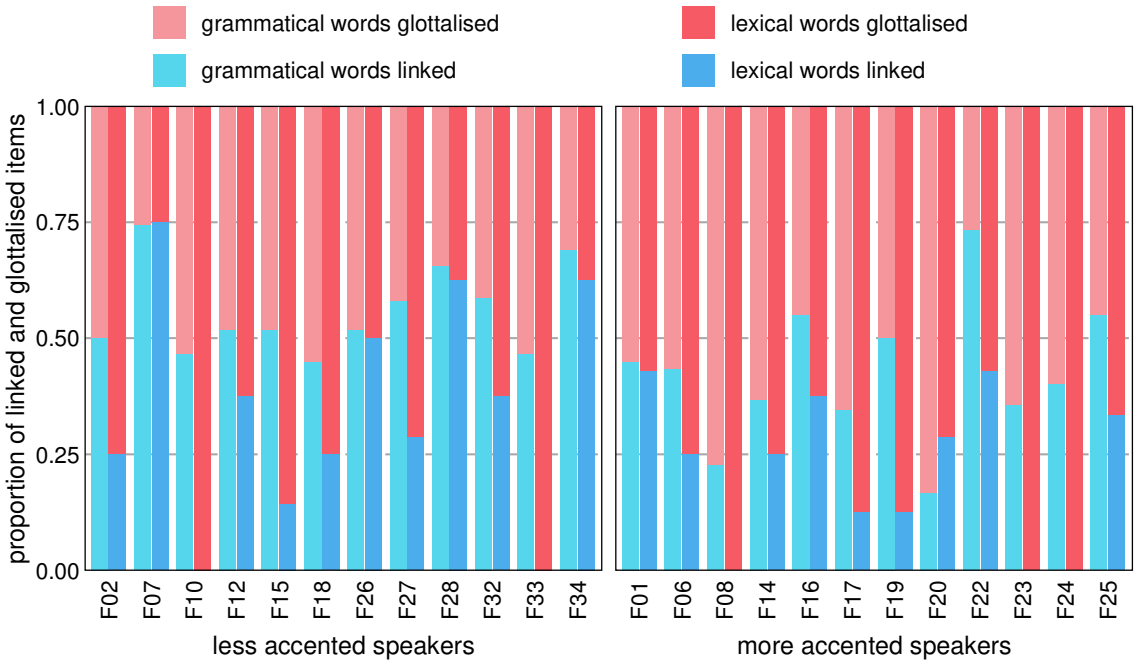
3.2 Data analysis and results

3.2.1 Linking and glottalisation

In Figure 1, the proportion of linked vs. glottalised (i.e., not linked) items is shown separately for linking towards lexical and towards grammatical words (for example, *who arrive* and *arrive at*, respectively). As expected, the less accented speaker group did link slightly more than the more accented one (51.4% for the former compared to 38.0% for the latter). Almost without exception, linking (marked in blue in the figure) was more frequent in grammatical words (in lighter shades, on the left for each speaker) than in lexical words. However, the results point to considerable between-speaker variability, even within the two groups. This is especially evident in the lexical words, with the less accented group having speakers who glottalised (as opposed to linking) in all instances, as well as some who linked most of the items.

Figure 1

Linking and Glottalisation in Less and More Accented Speakers (Separately for Grammatical and Lexical words)

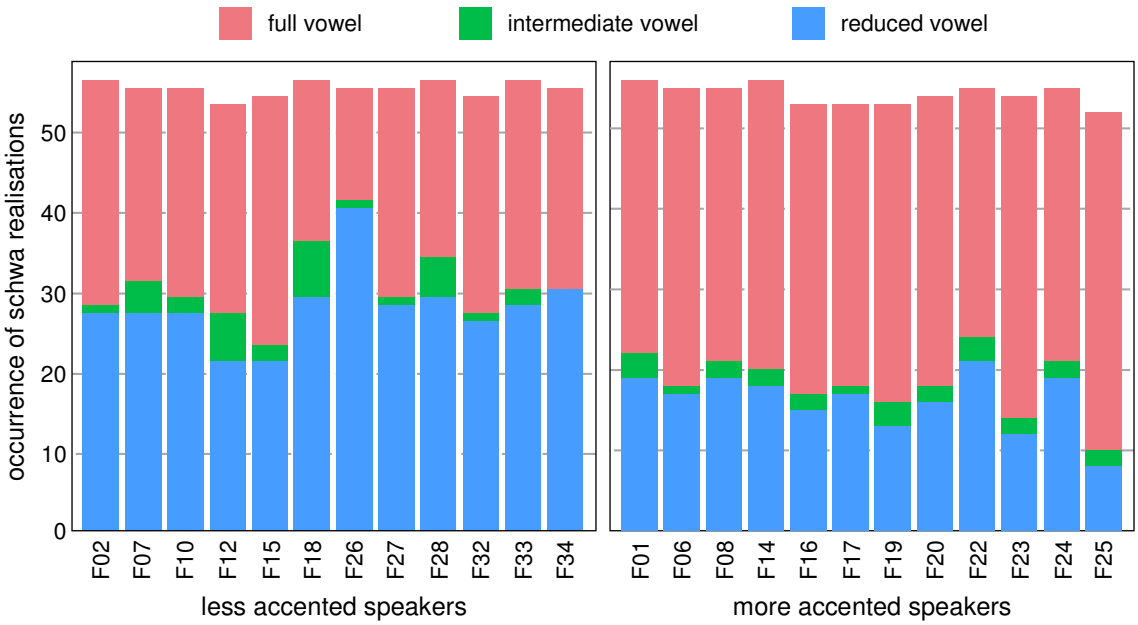


3.2.2 Vowel reduction

The difference in the degree of accentedness is more salient in vowel reduction: as shown in Figure 2, the more accented group of speakers reduced the quality of canonical schwa vowels markedly less frequently, and the difference amounts to approximately 20 percentage points. An 'intermediate' realisation category was introduced, since 4.4% of all vowel realisations were neither full, nor reduced, but elsewhere on the spectrum. For instance, the first vowel in *to you* was realised as not only [uː] or [ə], but also [u], [uˑ], or [ʊ], thus lying on the continuum from full to reduced. Finally, in contrast with linking, within-group variability is much lower here.

Figure 2

Reduction of Canonical Schwa Vowels in Less and More Accented Speakers

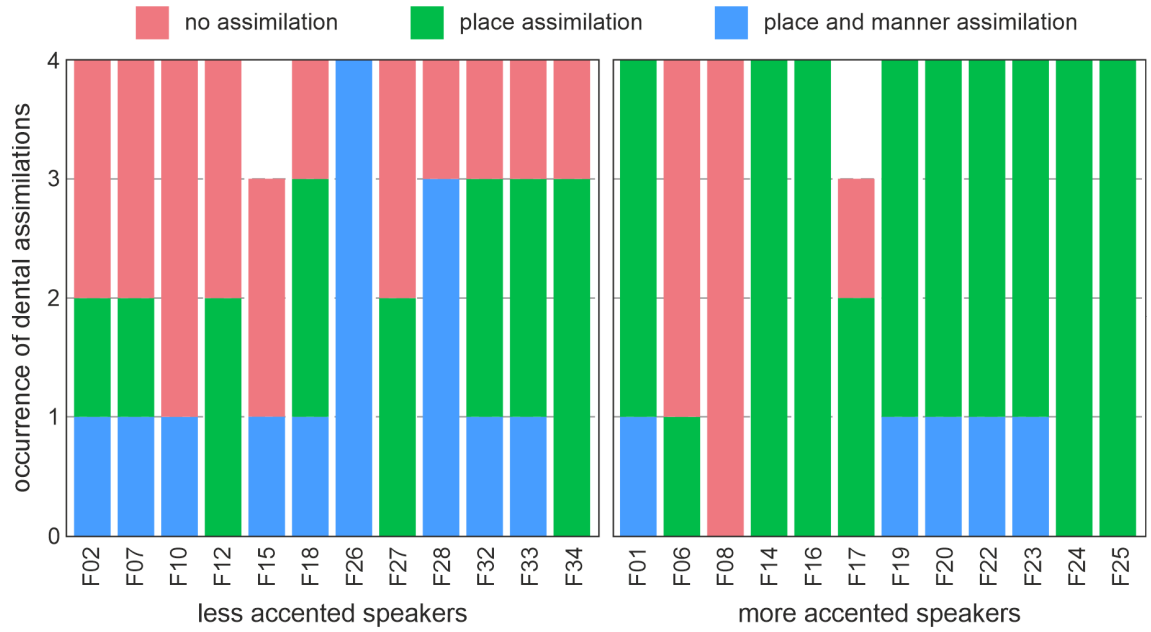


3.2.3 Assimilation

As for assimilation, the task yielded seven possible contexts of coalescence; of these, pronunciation with coalescence (e.g., /z+j/ → /ʒ/ in *as your* [æʒɔː] or /d+j/ → /dʒ/ in *did you* [dɪdʒə]) appeared in 55% of the items of the less accented group, and in 37% of the items of the more accented group – a result in line with our expectations. A more interesting tendency is revealed in the context of [n] before the dental [ð]; this occurred four times in the text. As shown in Figure 3, there were three possible realisations: a) with no assimilation (e.g., [ɪn ðə] or, more frequently, [ɪn̻ də]; b) with partial assimilation, which involved the change in articulation place [ɪn̪ ðə]; and c) complete assimilation, which involved regressive place and progressive manner assimilation [ɪn̪ n̪ə]. Items where no assimilation was realised were much more frequent in the less accented speakers; the same applies, however, for the complete assimilation where there is, phonetically speaking, a long dental nasal sound. On the other hand, intermediate realisations were dominant in the more accented speaker group.

Figure 3

Assimilation of Place and Manner of [n] and [ð]



3.2.4 Consonant elision

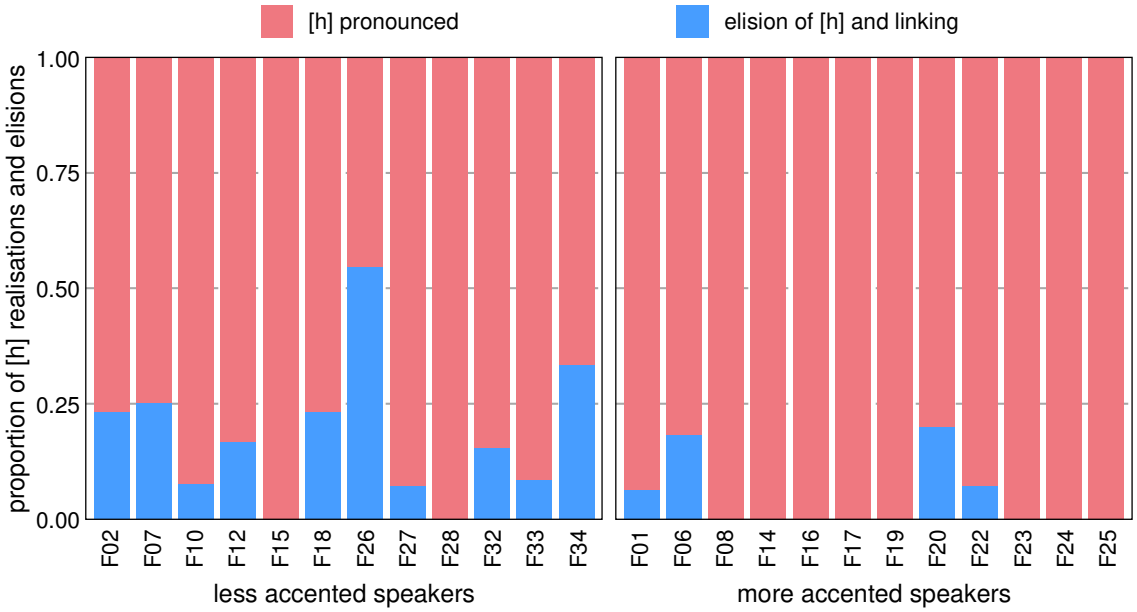
The results for elision show that speakers from both groups deleted the word-final [d] in *and* [t] in *must* before another consonant in approximately one half of the instances, and almost always pronounced stops as unreleased before another stop (e.g., *some men*; *what does*).

3.2.5 Elision (h-dropping) followed by linking

The results for the combined phenomenon of [h] elision and subsequent linking, as in *does he* (see the end of §2.1) can be observed in Figure 4 which shows that linked pronunciation without [h] was quite rare, occurring in only 10% of the items. However, there is a difference between the less and more accented speaker groups, with productions like *should have* [ʃʊd‿əv] or *for his* [fər‿ɪz] appearing more often in the former group.

Figure 4

Elision of [h] Accompanied by Linking



4 Experiment 2: Perception

4.1 Research methodology

We used two perception tests to examine whether and, if so, then how advanced Czech speakers of English perceive accentedness and comprehensibility of the recordings described in the previous section. A total of 24 sentences were read by 10 of the 34 speakers. Recordings from these 10 speakers were selected based on the accentedness level, so as to facilitate the subsequent manipulations. Each sentence was manipulated towards a more native-like version of the original sentence (containing uses of WFs and CSPs more natural for spoken English), and towards a less native-like version (where WFs and CSPs were suppressed).

Adobe Audition and Praat were used to perform the manipulations. We simulated full and reduced vowels by changing their duration (deleting or adding pitch periods) and amplitude. To eliminate linking, we spliced in a short period of silence; it was sometimes necessary to make the release of the word-final consonant more robust. To simulate lack of assimilation, release, and absence of elision in consonants, we have extended the duration of plosive closure or the noise component of fricatives and plosives; vice versa to achieve native-like consonant segments. Another method, used for both vowels and consonants, was splicing the target segment taken from the same speaker's data into the word being manipulated. Finally, fundamental frequency (f_0) was altered, transitions smoothed out, and amplitude adjusted where necessary.

Because of the social distancing measures associated with the COVID-19 epidemic, two perception tests were created using the MFCExperiment tool in Praat and sent to potential respondents via email. One test targeted accentedness, the other comprehensibility, and the respondents' task was to choose the recording (one more and one less native-like, as described above) which they regarded as less accented and as easier to understand, respectively. The order of the stimuli was randomised for each listener. The test set was accompanied by detailed instructions, which included saving the results into a binary file and sending it back to the

experimenter by email. Based on the Praat reaction time statistics, the whole test took 15 minutes, on average.

Another aim was to explore whether proficiency plays a role; for that reason, we had two groups of respondents, all between 19 and 25 years old. Group 1 consisted of 15 first-year students (F = 10, M = 5) with little formal knowledge of the sound patterns of English. Group 2 were eight third-year students (F = 7, M = 1), who had already completed a year-long course in English phonetics, including an overview of WFs and CSPs. None were familiar with the topic or the specifics of the study.

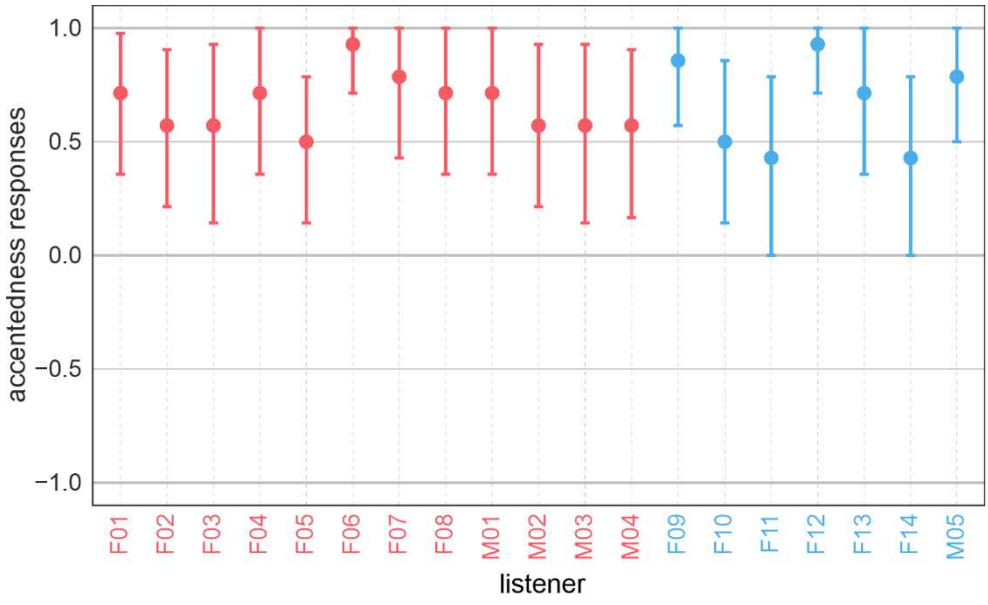
4.2 Data analysis and results

To assess the statistical significance of the perceptual comparisons, we employed the bootstrap method; this method is suitable for estimating the confidence interval of the mean value of a relatively small number of binary responses which are not normally distributed. Instances where our respondents were able to correctly assess accentedness (i.e., pick the more accented sentence realisation from a pair) were coded as 1, the opposite (i.e., they were not able to recognise the more accented version) as –1.

Figure 5 shows that, at the alpha level of 0.05, listeners in both groups were mostly able to correctly recognise the more accented sentence realisation. Interestingly, two listeners in the third-year group were slightly more ambivalent, with their responses approaching the significance boundary (the confidence interval nearly including zero).

Figure 5

Estimation of the Listeners' Ability to Assess Accentedness



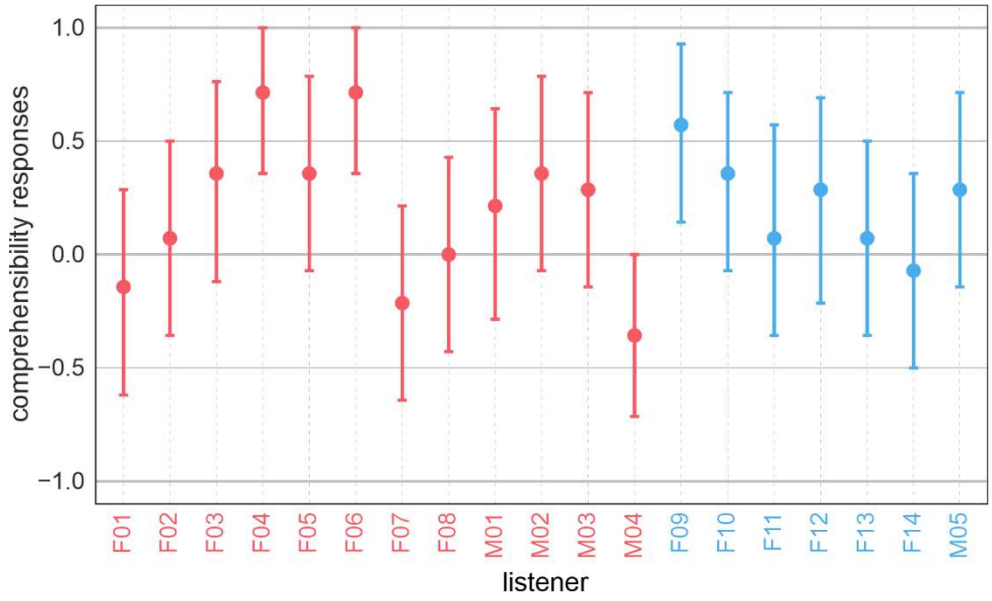
Note. First-year respondents in red, third-year respondents in blue.

The results are much less unambiguous for comprehensibility, as shown in Figure 6. Most listeners' intervals intersect the zero value, which means that they were mostly unable to assess comprehensibility levels correctly. In total, only three listeners displayed significantly higher

awareness of the phenomenon (F04, F06, F09), while one listener's responses may be considered as marginally significant in the opposite direction (M04).

Figure 6

Estimation of the Listeners' Ability to Assess Comprehensibility



Note. First-year respondents in red, third-year respondents in blue.

5 Discussion

In line with hypothesis H1A, Czech-accented EFL speakers showed deviations from the native-like production of WFs and associated CSPs. No single CSP was produced in a native-like way in all contexts by a single speaker. As is characteristic of non-native speech, the percentage of native-like CSP use ranged widely, from 10% (elision of [h] with linking) to 90% (unreleased consonants). Intermediate realisations on the CSP continuum (Alameen & Levis, 2015) are also typical of L2 speakers, represented in our results by various "degrees" of vowel reduction, consonant elision lacking subsequent linking, and dentalisation instead of assimilation. Similarly to Barańska and Zając (2014, pp. 281–282), vowel reduction was found to be problematic, a characteristic which Slavic speakers of English seem to share. In all, this shows that there are, indeed, more problematic aspects of CS that deserve attention in EFL teaching.

Our results also supported H1B, according to which more accented speakers will employ fewer WFs and CSPs, with the less accented group typically producing by 15-20 percentage points more CSPs. However, the overall difference between the groups was not substantial, which is in line with the findings by Barańska and Zając (2014).

The results of the perception test show that Czech listeners are at least somewhat aware of WFs and CSPs in English speech (H2A). The hypothesis claiming that comprehensibility is more difficult for Czech speakers of English to assess than accentedness (H2B) was also confirmed, with accentedness assessed correctly in 80-85% of cases, and comprehensibility in 60%.

Kukačka (2018) found that having taken phonetics and phonology classes had no effect on the results in WF production (p. 51). We reached the same conclusion, failing to support H2C (the more experienced the respondent, the more successful they are in assessing accentedness and comprehensibility), as the two groups' results were quite similar.

6 Conclusion and implications

Connected speech phenomena and weak-form realisation of grammatical words constitute an important component of the sound patterns of English, and specifically of its rhythm. This study has shown that even advanced, university-level students of English do not pronounce these in a native-like way, which may impact their comprehensibility (Anderson-Hsieh, 1990; Brown & Kondo-Brown, 2006; Ito, 2006). Our research has also presented evidence for the claim that none of the phenomena analysed here should be regarded as binary categories; rather, the realisation of WFs and CSPs typically occurs along a continuum. Similarly, in perception the awareness of these phenomena also ranges widely on a spectrum with the (theoretical) ends of 'no awareness whatsoever' and 'native-like perception'.

References

Alameen, G., & Levis, J. (2015). Connected speech. In M. Reed & J. Levis (Eds.), *Handbook of English pronunciation* (pp. 159–174). Wiley Blackwell.

Anderson-Hsieh, J. (1990). Teaching suprasegmentals to international teaching assistants using field-specific materials. *English for Specific Purposes, 9*, 195–214.

Barańska, A., & Zając, M. (2014). The influence of attention to language form on the production of weak forms by Polish learners of English. *Research in Language, 12*(3), 279–290. https://doi.org/10.2478/rela-2014-0017

Boersma, P., & Weenink, D. (2021). *Praat: doing phonetics by computer* (Version 6.1.38) [Computer software]. www.praat.org

Brown, J. D., & Kondo-Brown, K. (2006). Introducing connected speech. In J. D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 1–15). National Foreign Language Resource Center, University of Hawaii at Manoa.

Cauldwell, R. (2002). The functional irrhythmicality of spontaneous speech: A discourse view of speech rhythms. *Apples – Journal of Applied Language Studies, 2*(1), 1–24.

Council of Europe. (2020). *Common European framework of reference for languages*. Council of Europe. Cruttenden, A. (2014). *Gimson's pronunciation of English*. Routledge.

Ito, Y. (2006). The significance of reduced forms in L2 pedagogy. In J. D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 17–26). National Foreign Language Resource Center, University of Hawaii at Manoa.

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics, 24*(2), 168–196. https://doi.org/10.1093/applin/24.2.168

Kukačka, A. (2018). *The production of weak forms of function words by non-native English speakers with a special focus on prepositions*. [Doctoral dissertation, Masaryk University]. https://is.muni.cz/th/oq2jj/

Nolan, F., & Jeon, H.-S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society B*, *369*(1658), 20130396. https://doi.org/10.1098/rstb.2013.0396

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Shockey, L. (2003). *Sound patterns of spoken English*. Blackwell Publishing.

Skarnitzl, R., & Eriksson, A. (2017). The acoustics of word stress in Czech as a function of speaking style. In F. Lacerda, D. House, M. Heldner, J. Gustafson, S. Strömbergsson, & M. Włodarczak

(Eds.), *Proceedings of Interspeech 2017* (pp. 3221–3225). International Speech Communication Association. https://doi.org/10.21437/Interspeech.2017-417

Skarnitzl, R., & Rumlová, J. (2019). Phonetic aspects of strongly-accented Czech speakers of English. *Acta Universitatis Carolinae Philologica, 2*, 109–128. https://doi.org/10.14712/24646830.2019.21

Turk, A., & Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology*, *4*(1), 93–118. https://doi.org/10.1515/lp-2013-0005

Volín, J., & Johaníková, T. (2018). Weak structural words in British and Czech English. In J. Volín & R. Skarnitzl (Eds.), *The pronunciation of English by speakers of other languages*, 181–195. Cambridge Scholars Publishing.

Volín, J., Weingartová, L., & Skarnitzl, R. (2013). Spectral characteristics of schwa in Czech accented English. *Research in Language, 11*, 31–39. https://doi.org/10.2478/v10015-012-0008-6

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Appendix

The predicted distribution of weak word forms and connected speech processes in the 24 sentences used as stimuli in the production experiment.

Schwa	Assimilation - Assimilation of place or manner - Coalescence (= coalescent/fusional assimilation)
Consonant elision	Unreleased consonants
Linking	*Elision (h-dropping) + linking* (where the linking can only occur together with elision)

1. I would have told you about it.
[aɪ wəd‿əv ˈtəʊldʒ‿u(w)əbaʊt‿ɪt]

2. Can I talk to you about something?
[kən‿aɪ ˈtɔːk tə ju(w)əbaʊt sʌmθɪŋ]

3. When did you want to meet and discuss it?
[ˈwen dɪdʒ‿ə wɒnt̚ tə ˈmiːt‿ən̚ dɪˈskʌs‿ɪt]

4. There was a young man there.
[ðə wəz‿ə ˈjʌŋ ˈmæn ðɛː]

5. We must look in the locker and in the drawer.
[wi məs̚ ˈlʊk‿ɪŋ̚ n̪ə ˈlɒkər‿| ən̚ ‿ɪŋ̚ n̪ə ˈdrɔːə]

6. What have you been doing?
[ˈwɒt‿əv jə bɪn ˈduːɪŋ]

7. The people who arrive at five PM are too late.
[ðə ˈpiːpl̩ hu(w)əˈraɪv‿| ət faɪv piː(j)ˈem‿| ɑː tuː ˈleɪt]

8. When does he arrive from Paris?
[ˈwen dəz‿i(j)əˈraɪv f̚əm ˈpærɪs]

9. She should have asked for his permission.
[ʃi ʃəd‿əv‿ˈɑːskt fər‿ɪz pəˈmɪʃn̩]

10. You shall be on the list.
[ju ʃəl ˈbiːˍ(j)ɒn̪ n̪ə ˈlɪst]

11. He can't have gone behind your back!
[hi ˈkɑːnt‿əv ˈgɒn | bəˈhaɪndʒə ˈbæk]

12. He got me her number and email.
[hi ˈgɒʔ miˍ(j)ə ˈnʌmbər‿ən‿ˈiːmeɪl]

13. There is a lack of answers.
[ðər‿ɪz‿ə ˈlæk‿əv‿ˈɑːnsəz]

14. Could you have been there?
[kədʒuˍ(w) əv ˈbiːn n̪ɛː]

15. There were some men who knew them.
[ðə wə səm ˈmen‿ u ˈnjuː ðəm]

16. His university was as good as your college.
[hɪʒ‿uːnɪˈvɜːsətɪ | wəz‿əz ˈgʊd‿əʒ‿ə ˈkɒlɪdʒ]

17. No, but there is an umbrella.
[ˈnəʊ | bʌt n̪ˈdər‿ɪz‿ən‿ʌmˈbrelə]

18. Do you like her more than Jane?
[dʒə ˈlaɪk‿ ɜː mɔː ðən ˈdʒeɪn]

19. It's a gift from us for him and his wife.
[ɪts‿ə ˈgɪft fəm‿ˈʌs | fə ˈhɪm‿ən‿ɪz ˈwaɪf]

20. The parents were nice to them.
[ðə ˈpɛːrənts wə ˈnaɪs tə ðəm]

21. To be, or not to be?
[tə ˈbiːˍ(j)| ə ˈnɒt tə bi]

22. Out of everyone here, I am the best.
[aʊt‿əv‿ˈevrɪwʌn ˈhɪər‿| ˈaɪˍ(j)æm ðə ˈbest]

23. What does it mean to us?
[ˈwɒt‿dəz‿ɪʔ ˈmiːn tʊˍ(w)əs]

24. He said that his brother was an artist.
[hi ˈsed | d̪ət‿ɪz ˈbrʌðə wəz‿ən‿ˈɑːtɪst]

About the authors

Lenka Kalvodová is an alumna of the BA in English studies at the Faculty of Arts, Charles University, Prague. Having previously focussed on phonetics and phonology of English, she is currently pursuing an MA in Phonetics at the University of Helsinki, Finland. Her background includes experience in EFL (English as a foreign language) pedagogy and in technology, and her interest is in exploring the intersection of these. She is particularly interested in how combined knowledge from these fields can be applied in various domains of our lives (e.g., the use of speech technology in language acquisition or accessibility).

Email: lenka.kalvoda@gmail.com

Radek Skarnitzl is an Associate Professor at the Faculty of Arts, Charles University, Prague, and director of the Institute of Phonetics. He is interested in various aspects of speech communication, with expertise in speech prosody, acoustic phonetics, and second language (L2) acquisition. His research focuses on L2 pronunciation and on the effect of various pronunciation features on the socio-psychological evaluation of a speaker in both native and foreign languages. He is also interested in issues related to speaker identification and forensic phonetics, especially the effects of voice disguise on human and automatic speaker recognition performance.

Email: radek.skarnitzl@ff.cuni.cz

Kirkova-Naskova, A. (2023). Engaging research: Empowering ESL/EFL teachers to teach pronunciation. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 122–137). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8232364

Engaging research: Empowering ESL/EFL teachers to teach pronunciation

Anastazija Kirkova-Naskova
Ss. Cyril and Methodius University, Skopje

The ongoing debate about the research-practice link has raised the point that research advances are far more progressive and teaching practice is failing to keep up. An important aspect that has received little attention is understanding teachers' classroom reality – a reality that for many involves an ever-increasing workload, limited resources, and a lack of autonomy and self-confidence. It is evident that teachers should be encouraged to address pronunciation in their teaching more frequently. However, they also need support to navigate the personal and institutional challenges.

This paper proposes a conceptual agenda which empowers teachers to learn how to reflect on their pronunciation teaching practices. Current issues in pronunciation research relevant for the teaching context are discussed, followed by an overview of the most effective research findings that can be successfully applied in the classroom. These insights are then contrasted with preliminary results from a qualitative study that investigates teachers' experience with pronunciation teaching. In the last section, ways of providing support for teachers' learning and professional development are explored.

Keywords: pronunciation research, pronunciation teaching, research-practice link, classroom practices, professional development

1 Introduction

Pronunciation is without doubt as important as any other language skill. Yet, with the rise of communicative language teaching (CLT), its equal presence in coursebooks and in classrooms has been overlooked, as traditional pronunciation activities did not fit ideally within the CLT framework and new ones were not adequately implemented (Celce-Murcia et al., 2010). One particular reason why L2 pronunciation will always matter to L2 learners is the fact that it is not only a cognitive skill but it is also a motor skill. As such, it has a physical reality that results in the speech produced by learners, which is often L1-accented. From a linguistic perspective, having accented speech results in pronunciation that may lead to communication breakdowns. From a social perspective, such speech may have social consequences on the learners, who may experience stereotyping, bias, and discrimination in personal communication or in the workplace (Lippi-Green, 1997).

Traditionally, foreign accent (FA) was treated as a by-product of L2 pronunciation and its elimination was the end goal of pronunciation learning. However, three aspects lurked over this conceptualisation, making it difficult for teachers to know what to do. Firstly, the complexity of the FA phenomenon was difficult to grasp despite authors investigating the predictors of FA (Piper & Cansin, 1988; Purcell & Suter, 1980) and factors affecting the degree of FA (see overview in Piske et al., 2001). Secondly, globalisation processes led more and more people to communicate in English as a lingua franca, which made it clear that L2 pronunciation does not always prevent successful communication. Thirdly, the diversity of English varieties posed a serious problem when choosing a native variety as a reference model for L2 learners. A conceptual breakthrough came with the work of Munro and Derwing (1995) when they deconstructed the phenomenon of foreign accent and demonstrated that it consists of three perceptual phenomena: accentedness (i.e., how different a pattern of speech sounds when it is compared to the local variety), comprehensibility (i.e., how easy/difficult a listener finds it to understand someone's speech), and intelligibility (i.e., to what degree a listener actually understands an utterance). Munro and Derwing's (1995) oft-cited study was followed by two influential developments, in 2000 and 2005. Jenkins (2000) observed interactions between non-native speakers (NNS) and concluded that the assumption that L2 learners of English communicate solely or even mostly with native speakers (NS) is not accurate; in fact, NNSs are more likely to communicate with other NNSs than with NSs. She proposed a Lingua Franca Core (LFC), i.e., a set of pronunciation features that are likely to cause breakdowns in intelligibility (p. 158–160), hence, they could be given precedence in pedagogy. Furthermore, Levis (2005) addressed the issue of pronunciation models and goals and made a distinction between the Nativeness vs. Intelligibility Principle, suggesting that intelligibility is the way forward; while it is possible to achieve native-like pronunciation in the foreign language classroom, learners more importantly need to be understandable and to communicate successfully despite a noticeable or a strong foreign accent.

Such a change in viewpoints has prompted a re-examination of traditional teaching approaches. Pronunciation research has flourished with many studies investigating the development of L2 pronunciation, the potential of various teaching methods and activities, as well as various social and individual factors that affect successful speech acquisition. New avenues in the interplay of speech phenomena have been explored to provide answers to relevant questions such as:

- Is pronunciation instruction effective?
- Should the focus of pronunciation instruction be on the auditory mode (perception) or the articulatory mode (production)?

- Which is more important, teaching segmental features or teaching suprasegmental features?
- What kind of pronunciation techniques do teachers use in the classroom?
- What types of pronunciation tasks are most effective: controlled or spontaneous?
- How do teachers address pronunciation errors?
- Is pronunciation learned more effectively if taught separately or combined with the other language skills?
- Is pronunciation integrated in coursebooks and national curricula?
- Do teachers teach pronunciation regularly? Do they feel competent to teach pronunciation (Foote et al., 2011; Henderson et al., 2012; Macdonald, 2002)?

All in all, the results overwhelmingly show that pronunciation instruction is effective (see Lee et al., 2015; Saito & Plonsky, 2019; Thomson & Derwing, 2015). However, are these promising research findings applicable in the classroom?

2 Effective pronunciation teaching: Research findings and recommendations

In their comprehensive overview of 150 years of pronunciation teaching, Murphy and Baker (2019) define emerging trends in empirical research about pronunciation teaching. They classify the following macro-level themes: a) explorations on *what* features of English phonology are necessary to teach; b) explorations on *how* to teach L2 pronunciation effectively; and c) teachers' cognitions (beliefs and knowledge) and learners' views on pronunciation instruction (p. 56–58). This paper only focuses on the second theme with its specific micro-level subthemes (priorities, instruction, strategies), presenting positive research findings (from a selection of empirical and review studies) and analysing their implications for the L2 classroom in §2.1–§2.4.

2.1 Priorities in pronunciation teaching

Establishing priorities in pronunciation teaching is the first logical step in helping teachers make research-informed decisions. This includes shedding light on issues such as foreign-accented speech, gravity of pronunciation errors, effectiveness of pronunciation instruction, and L2 identity (teacher and learner).

Results from research studies investigating foreign-accented speech show that L2 speech can be accented, yet remain intelligible and comprehensible (Munro 2008). This is an important insight because it supports the view that accent reduction should not be the goal of pronunciation teaching and learning. Various linguistic aspects also play different roles when it comes to processing L2 speech. For instance, Trofimovich and Isaacs (2012) identified that accent is related to aspects of phonology (rhythm, segmental accuracy, and syllable structure accuracy), while comprehensibility is related to grammatical accuracy and lexical richness.

Research has also pointed out that some but not all phonological errors cause communication breakdowns, especially when L2 speakers use English as a lingua franca (ELF) in NNS–NNS interactions. Jenkins (2000), for instance, defined the core phonological errors as LFC (here presented in a simplified way): all consonants except /r/, /t/, /θ/, /ð/, [ɬ], aspiration, fortis/lenis consonant distinction due to their effect on vowel length, consonant clusters (initial not simplified; medial/final simplified), vowel length contrasts, nuclear stress placement, and division of speech. Many authors immediately recognised the practical potential of LFC and explored ways applying it in the classroom (e.g., Walker 2010). Even so, more important was the conclusion that in such interactions intelligibility should be prioritised; hence the

recommendation that the goal of pronunciation teaching and learning is acquiring intelligible speech (Levis, 2018; McAndrews & Thompson, 2017).

Understanding the impact of pronunciation instruction was also given meticulous attention in research studies. Results show that pronunciation instruction is indeed effective and should be integrated in coursebooks and in teacher training programs (see Darcy, 2018; Derwing & Munro, 2005; Derwing & Rossiter, 2003; Jones, 2016; Lee et al., 2015; McGregor & Reed, 2018; Saito & Plonsky, 2019; Thomson & Derwing, 2015)

As for developing acceptable L2 identity, studies show that raising awareness of native and non-native varieties should also be addressed in the classroom, preferably through comparison and discussion. Teachers still see native accents as necessary reference models; on the positive side, there is a wider awareness and acceptance of non-native accent diversity (Červinková Poesová & Lancová, 2021).

To summarise, these findings have the following classroom implications:

- teachers should accept intelligible L1-accented pronunciation;
- teachers should create opportunities for NNS–NNS interaction practice;
- teachers should promote intelligibility and positive attitude to non-native accents;
- teachers should teach pronunciation as frequently as all other language skills in any educational context regardless of learners' age and proficiency level.

As simplified as they seem, implementing these aspects in the teaching practice might require a change in teachers' mentality, especially in terms of diverging from the nativeness principle, but also in prioritising pronunciation when necessary over other language skills. Valuable pronunciation practice time should not be systematically sacrificed for other language skills to be practised.

2.2 Choices to make in pronunciation instruction

Giving priority to pronunciation instruction in the classroom implies that teachers make instantaneous choices and decisions about: 1) the type of phonological feature their students struggle with and which needs to be taught and practised; 2) the most appropriate type of approach they should adopt; and 3) the types of techniques/activities they should choose in order to practise specific pronunciation features. The implications of research findings related to these aspects are discussed in §2.2.1, §2.2.2, and §2.2.3.

2.2.1 Type of phonological feature

Addressing learners' pronunciation difficulties is a daunting task – teachers cannot attend to every potential mispronunciation but have to decide which aspects of pronunciation are most important and which ones should be tackled at a later stage. This challenge has been acknowledged by research, especially when it comes to which should be taught first, segmental or suprasegmental features. Research findings show that neither should be neglected or given precedence. In fact, the overall conclusion is that it is important to address both segmental and suprasegmental features at once (broad framework suggested), with prosodic features (suprasegmentals) given priority over segmental features when aiming for general improvement in oral communication (Derwing et al, 1998; Levis & Muller Levis, 2018). Moreover, teachers should select phonological features for instruction and practice based on their learners' goals, L1 backgrounds, and proficiency levels (Saito, 2012).

A concept that has been revisited in empirical research with regard to its effects on listeners is the Functional Load Principle (Brown, 1988; Catford, 1987), or ranking phonemic contrasts

according to their frequency and importance in English pronunciation. High functional load errors largely affect comprehensibility and accentedness, whereas low functional load errors have only minimal impact on comprehensibility (Munro & Derwing, 2006). In addition, errors with high communicative value help listeners distinguish between low- and mid-level proficiency learners, while errors with low communicative value help listeners distinguish between mid- and high-level proficiency learners (Suzukida & Saito, 2022).

An aspect of pronunciation that has also been given prominence in research is the relevance of connected speech. For example, Cauldwell (2013, 2018) advocates teaching learners how to decode rapid informal spontaneous speech, emphasising that it should be the goal of learning listening and that it can also help to improve oral fluency.

From a teacher's perspective, these findings indicate the following classroom implications:

- teachers should address segmental and suprasegmental features in a balanced way;
- teachers should assess learners' pronunciation and identify learners' goals;
- teachers should focus on errors that impede communication more frequently;
- teachers should explain connected speech phenomena as an example of where good listening skills can go hand-in-hand with pronunciation work.

To make effective use of these implications, teachers are expected to have a certain level of phonological competence and knowledge of key concepts. It is assumed that they are sufficiently trained and skilled to teach pronunciation.

2.2.2 Type of pronunciation instruction

In terms of what type of instruction is most effective, many attempts have been made to test different approaches. For example, Saito and Plonsky (2019) conclude that explicit pronunciation instruction at a controlled level allows teachers to explain detailed phonetic information, which enables learners to notice and practise the accurate production of segments, syllables, prosodic, and temporal features of speech in a careful manner. Saito (2012) shows that Focus on Form (FonF) instruction in meaning-oriented communicative contexts enables learners to improve both during controlled and spontaneous practice (meaning that these learners can more easily generalise their knowledge), while Focus on Forms (FonFs) instruction yields improvement only in controlled contexts (with focus on accuracy via mechanical drills and choral repetition) and does not allow learners to transfer what they learnt in the classroom to outside of the classroom. Such approaches develop perceptual and noticing skills in learners, and these seem to be as necessary as practising L2 productive skills, i.e., perception-based pronunciation instruction is equally important as production-based instruction (Lee et al., 2020).

Concerning the communicative focus of L2 pronunciation, it has been shown, for instance, that regardless of whether task-based pronunciation teaching directs attention either to form or to meaning (depending on the task instructions), it leads to better accuracy in the long run (e.g., Mora & Levkina, 2017). Within the discourse-pragmatic approach (Pickering, 2018), findings show that intonation (pauses, prominence, pitch) is easier to understand through discourse contexts as opposed to isolated or partial utterances. Levis (2018) argues that in both ESL and EFL contexts the intelligibility-based teaching approach (i.e., focus on pronunciation features that affect intelligibility) is more appropriate.

These findings are particularly relevant for guiding teachers' choices as they reflect on the following classroom implications:

- teachers should be able to explain the sound system in detail but with simple language appropriate to their learners' age and proficiency level;
- teachers should vary activity types between controlled and communicative;
- teachers should work on both speech perception and speech production;
- teachers should integrate pronunciation through task- and project-based activities;
- teachers should work on pronunciation at both utterance and discourse level.

Perhaps these implications, if regularly implemented into the teaching practice, will seem to impose a heavy burden onto teachers' otherwise busy professional lives. They might need to develop skills such as being resourceful, being creative with developing materials, and becoming efficient time managers.

2.2.3 Type of pronunciation teaching technique

The effectiveness of various pronunciation teaching techniques has also been investigated, because not all activities which are often recommended in coursebooks and practice resources give favourable results and thus valuable classroom time may be wasted on them. One such activity that has proven most beneficial is High Variability Phonetic[1] Training (HVPT). Research has shown that HVPT used with nonwords and real words is effective and results in long-lasting improvement in learners' pronunciation (Ortega et al., 2021; Thomson, 2018). Another activity frequently employed in the classroom is the use of phonemic symbols or keywords to focus on segmental differences. Both serve as effective reference labels for developing and consolidating perceptual sound categories, as shown by Fouz-González and Mompeán (2021).

The usefulness of critical listening has been confirmed, in terms of raising learners' awareness of acceptable vs. unacceptable L2 speech (Couper, 2011). Its benefits are especially enhanced if oral corrective feedback is provided in the form of reformulations (recasts and explicit correction) and prompts for self-repair (elicitation, metalinguistic clues, clarification requests, repetitions) – strategies which are favoured by learners (Lyster et al., 2013).

The potential use of technology for teaching and learning pronunciation has also been examined. Research shows that technology-based activities such as shadowing (Foote & McDonough, 2017) or Automated Speech Recognition (ASR) dictation (McCrocklin et al., 2019), to name but a few, improve learners' comprehensibility and fluency and promote authentic language use.

In sum, such findings lead to the following classroom implications:

- teachers should use HVPT to enhance L2 sound discrimination;
- teachers should teach phonemic symbols to develop sound-to-symbol connections;
- teachers should record and analyse learners' speech and give corrective feedback frequently;
- teachers should integrate the use of online tools and resources, both in the classroom and for self-learning.

To use these activities with learners , teachers are expected to first understand their benefits, then familiarise themselves with step-by-step procedures, and finally employ them regularly in L2 lessons.

[1] The term 'phonetic' in HVPT is sometimes optionally replaced by 'pronunciation' (see Thomson, 2018) or 'perceptual' (see Qian et al., 2018).

2.3 Pronunciation strategies

Another avenue of research has considered the effectiveness of pronunciation strategy use. For instance, Osbourne (2003) concluded that advanced ESOL learners employ self-monitoring strategies to repair their mispronunciations by using imitation, paralanguage (speed, volume, clarity), voice quality settings, and by focusing on individual sounds/clusters or syllable/words, and on prosodic structure. Other studies show that high- and low-achieving learners use different pronunciation learning strategies (PLSs) and these differ depending on the task (Szyszka, 2021). Szyszka (2021) noted that the most frequently used PLSs by both groups of learners were checking pronunciation in the dictionary and reading words and texts aloud. Sardegna (2022) conducted a strategy-based instruction and tested its efficacy. She found out that such instruction promotes learner autonomy and self-regulated learning – learners trained to use pronunciation strategy protocols demonstrated: a) greater success when they practised frequently; b) higher motivation to continue practising after instruction ended; and c) a high sense of self-efficacy.

With obvious learning potential, these findings suggest the following classroom implications:

- teachers should train learners to use various pronunciation learning strategies;
- teachers should teach their learners to use self-monitoring strategies;
- teachers should encourage autonomous learning.

Familiarising their learners with pronunciation learning and self-monitoring strategies may be far from a simple task – it might, in fact, require devoting class time to train learners in the appropriate use of strategies.

2.4 The reality clash: Teacher and learner challenges

Recent calls for more research-informed pronunciation practice (e.g., Derwing & Munro, 2015; Pennington & Rogerson-Revell, 2019) and the ensuing re-evaluation of the research-practice link have brought to light a frustrating state of affairs: even with such a wealth of findings, the connections between L2 pronunciation research and L2 pronunciation teaching have become more blurred, with "many interesting studies today [that] do not have clear implications for teaching, and many practically oriented publications [that] show minor grounding in research" (Levis, 2021, p. 18). In addition, research into L2 pronunciation has continued to advance with impressive strides, while at the same time, there are few observable improvements in developing and adapting teaching materials (in the form of publications, e.g., Jones, 2016; or in the form of online resources, e.g., English Accent Coach 3.0 by Thomson, 2012–2023[2]). In other words, pedagogy is failing to keep pace with research progress.

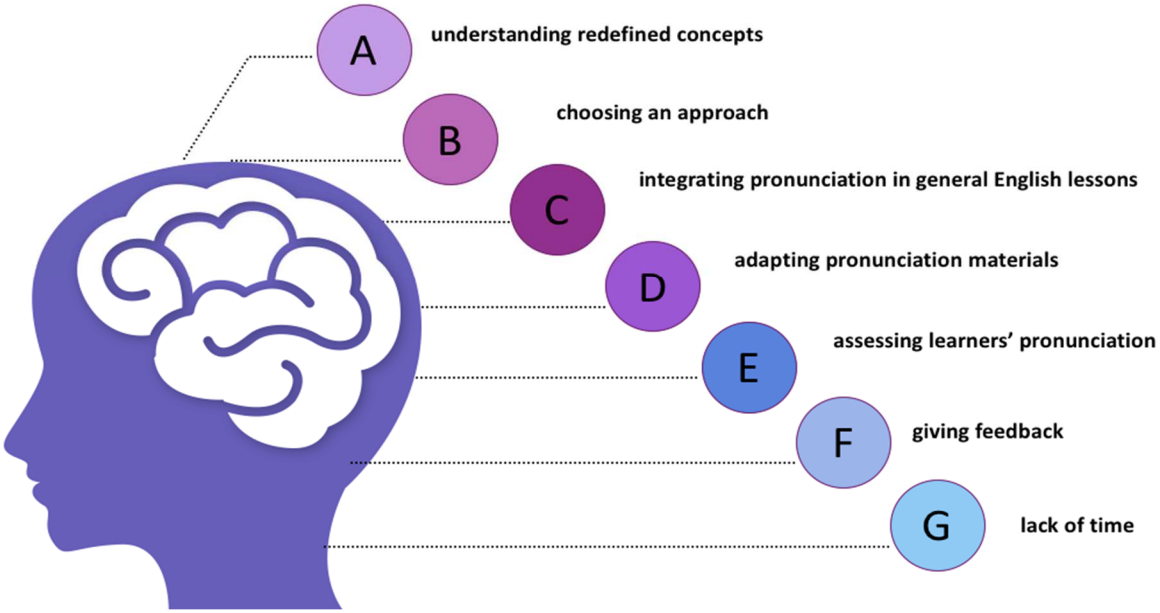
Despite the latest attempts to narrow this gap by promoting publications that devote special sections to practical applications of key research findings (e.g., Kirkova-Naskova et al, 2021; Levis et al., 2022; Sardegna & Jarosz, 2023), pronunciation is still marginalised in the classroom and teachers are left to rely on their own intuitions and experiences as L2 learners. They struggle with real challenges (see Figure 1), including: a) difficulty grasping *re-defined concepts* such as intelligibility, comprehensibility, and accent; b) dilemmas on *which approach* to adopt given that there are so many (conflicting) ideas and techniques; c) a lack of competence on *how to integrate* pronunciation features in general English lessons; d) not knowing *how to adapt materials* to learners of different proficiency levels and ages; e) a lack

[2] https://www.englishaccentcoach.com/

of self-assurance on *how to assess* their learners' pronunciation; f) being uncertain about *how to give feedback* on their learners' pronunciation goals, given their varied linguistic experience and developing L2 identities; and g) inability to cope with a *lack of time*.

Figure 1

Challenges Teachers Face



Another angle that needs to be considered is that applying research findings in the teaching practice is not always straightforward, for example, the LFC and intelligibility. From teachers' perspective, LFC was interpreted as a set of features that should be *the only* focus of pronunciation instruction and was widely accepted by teachers as the much-needed tool that would provide that quick fix to all problems. In fact, the LFC only lists the phonological errors that are most likely to occur and cause misunderstanding when NNS of various L1 backgrounds interact. It does not mean that *all* features are problematic for a given group of learners that share the same L1. From learners' perspective, the LFC diminished the importance of practising pronunciation. Unless properly trained, learners do not easily comprehend the complexity of the intelligibility concept and thus interpret the LFC and being intelligible as a free pass to use 'whatever' pronunciation. As a result of such extremes, neither improvement nor successful implementation are achieved, leaving teachers and learners struggling with L2 pronunciation issues. It seems that putting research-based classroom implications into effect, as suggested in our analysis in §2.1–§2.3, is not a simple undertaking and demands teachers devote substantial effort to familiarising themselves with current research findings and, maybe more vitally, being creative and motivated to try out new ideas.

3 A case in point: Teachers' views on their pronunciation teaching practices

3.1 Aims and a research question

In a larger on-going study, I am investigating teachers' views on their formal pronunciation instruction and current teaching practices. More specifically, the study explores: a) their reflections on the relevance of pronunciation vis-à-vis nativeness/intelligibility principle; b) their knowledge of the English sound system prior to university; c) their views on the formal pronunciation instruction they received during their English Phonetics and Phonology undergraduate course; d) the type of teacher training they received for teaching pronunciation; and e) their current pronunciation teaching practices. The analysis presented below focuses on the last section only, i.e., their practical experience with teaching pronunciation, addressing the following research question:

> **RQ**: What kind of practices do teachers employ when they teach pronunciation with reference to: frequency of teaching; type of approach used; pronunciation features taught; pronunciation activities used; and type of corrective feedback given?

3.2 Participants, instruments and data analysis

Twenty-nine Macedonian teachers of English as a foreign language (F = 23; M = 6) had participated in the study by April 2022. Their age ranges from 24–39 years old and their teaching experience from 2–16 years. They work in varied teaching contexts: a) in state schools (elementary and high); b) in private language schools (all levels); and c) a private international school (elementary, middle, high). Qualitative data was elicited through semi-structured interviews and then analysed with thematic category analysis. The participants were coded MK01–MK29 for anonymity.

3.3 Preliminary results

At the beginning of the interviews, the participants were asked about their impressionistic view on teaching pronunciation, in particular, whether they find it necessary to teach and whether they find it teachable at all. All participants (n = 29) gave favourable responses to both questions, indicating that they regarded pronunciation as a relevant language skill.

However, when asked how much class time on average they spent on teaching pronunciation features, most of them replied that they sporadically address pronunciation in class (n = 13) and a few responded that they regularly address pronunciation, especially with young learners (n = 4). The rest of the participants specified the approximate time they usually spend on pronunciation issues (n = 11), with 10 minutes on average. Those who mentioned that they do not pay enough attention to pronunciation gave the following reasons: lack of time (n = 5) and pronunciation not being addressed in the national curriculum (just one mention), therefore, not being required to assess it (n = 3). The following remark by one participant best exemplifies this viewpoint:

> … Pronunciation is one of the most important elements of language learning but it's been really neglected by us as teachers. First of all, because we don't want our students to feel down if we correct them all the time about their pronunciation. Second, our pronunciation is not as good as it should be I think, and third, the curriculum does not give us enough space

> to teach pronunciation. We teach grammar, we teach reading, we teach writing, we teach listening, but, teaching pronunciation is not accentuated in the programs. (MK05)

When pronunciation is addressed in the classroom, it is rarely addressed separately, e.g., a whole lesson being devoted to explaining and practising pronunciation features (n = 5); it is typically integrated in the L2 lesson (n = 24). This is mostly done when new vocabulary is introduced (accurate pronunciation presented), when speaking is practised (mispronunciations corrected), when reading is practised (mispronunciations corrected), or when listening is practised (the focus being more on listening comprehension exercises in the coursebook than discriminating sounds in speech). Apparently, they do not systematically teach pronunciation and barely mention pronunciation aspects; the presentation of various pronunciation features is rather unplanned. Only a selection of features are taught, predominantly word stress, then vowels (long vs. short, /æ – e/), diphthongs, consonants (mainly /θ, ð, t, d/, i.e., consonants that are difficult for Macedonian learners of English), intonation patterns (e.g., in tag questions or the meaning of rising and/or falling tones), and the pronunciation of <-s> and <-ed>. Phonemic symbols and letter-to-sound connections are seldom taught, as well as syllable division. It was interesting to notice from their responses that they frequently used the term *understandable* to mean intelligible and/or comprehensible.

The type of pronunciation activities that are practised is varied. When the focus is on specific speech sounds, the activities are more controlled and include elicited imitation (listen and repeat), individual sound identification/categorisation, minimal pairs for sound discrimination, reading-aloud, dictation with phonemic symbols, and tongue twisters. When the aim is to make the lesson more interactive and fun, then communicative activities are practised, such as games, quizzes, and dialogues in the form of role-plays (not necessarily with a strong focus on pronunciation practice). One participant mentioned organising a debate club and mock trial courts (MK06), and another a drama studio (MK24) – these teachers use such activities to address pronunciation issues in context. Several teachers sometimes use activities from a coursebook (n = 13), few create materials or use other resources in addition to the coursebook activities (n = 7), and some do not even have coursebooks, as in some schools teachers are required to create their own materials and do not use coursebooks at all (n = 9). Those teachers who use coursebooks but do not practise pronunciation activities in the coursebook, reported that they do so because it is not required in the curriculum. Several teachers (n = 10) make use of resources that are incidentally found online and the online resources they select are chosen in relation to their learners' age, including short stories, songs, listening activities, interactive minimal pair activities, etc. Many teachers (n = 16) consult specific sites, for instance, Kahoot![3], online dictionaries, Live Worksheets[4], slam poetry websites – to name a few. One teacher (MK23) reported that when she noticed that her learners struggled with a particular pronunciation aspect, she would consult specialised teacher development sites; it is important to point out here that this teacher enjoys institutional support in this respect (the institution covers the subscription cost but this is only the case in the private international schools). Three teachers reported not using online resources at all.

With respect to the ways they correct their learners' pronunciation, the responses were mixed. In general, the teachers use reformulations (predominantly explicit correction, occasionally recasts) and prompts for self-repair (mainly repetition). Corrective feedback is given mostly individually and immediately when the mistake is made. Some teachers prefer giving delayed feedback (they wait for the learners to finish the task and then they clarify the

[3] Kahoot! https://kahoot.com/
[4] Life Worksheets https://www.liveworksheets.com/

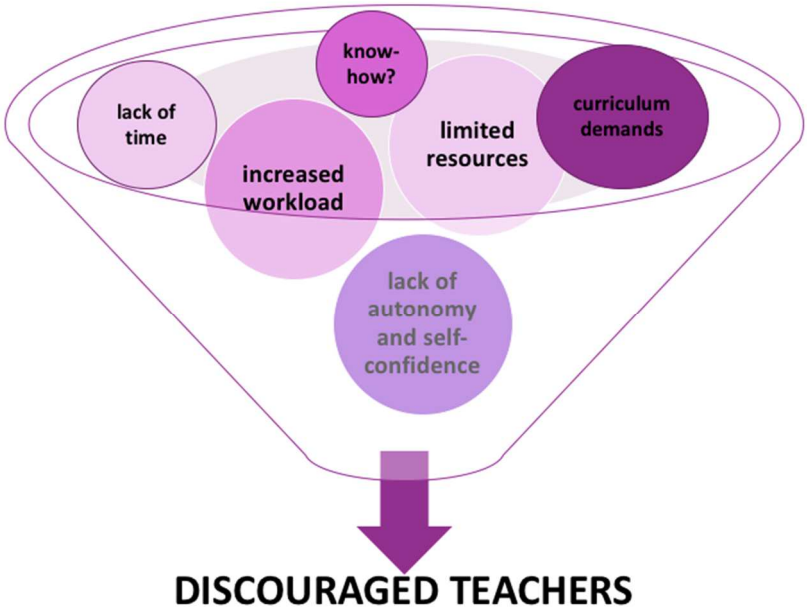
mispronunciation either to the learner who made the mistake or as a general correction in front of the whole group).

During the interviews teachers made unsolicited remarks: the majority reported forgetting core pronunciation knowledge (phonetic/phonological rules). With this weakness in mind, they observed that they would benefit from specialised courses for teaching pronunciation. Such incidental findings ring alarm bells about teachers' self-identified needs for systemic guidance and opportunities for continuing professional development.

These preliminary results reveal some of what actually goes on in classrooms and provide insight into an important aspect that has received little attention in research: genuinely understanding classroom reality – a reality that assumes increased workload, limited resources, and a lack of autonomy and self-confidence (see Figure 2). Such a constraining professional context inevitably leads to teachers being discouraged to engage in teaching pronunciation and who, in truth, need support to navigate the personal and institutional challenges. They also need to be encouraged to address pronunciation issues in their teaching more frequently.

Figure 2

Professional Contextual Constraints Discouraging Teachers from Engaging to Teach Pronunciation



4 Discussion

The findings presented in this analysis at least hint that perhaps one way of addressing teachers' discouragement with regard to teaching pronunciation is to have researchers understand their harsh reality and make their research more accessible and comprehensive – research should empower teachers to develop a proactive approach to teaching pronunciation rather than a reactive one. The ideal teacher profile is a competent teacher who is trained to teach pronunciation and informed about the latest research evidence of what is effective. Equally important, teachers should be skilled and confident to make more spontaneous decisions in the

classroom. This is, of course, easier said than done, but it *is* in truth achievable. Instead of overwhelming teachers with a profusion of theoretical information, recommendations could be made on how to take small steps of action. For instance, it is paramount that teachers be equipped with a reasonable phonological know-how on which they can build their teaching practice. Table 1 shows initial recommendations on how a teacher who is assumed to know little about pronunciation teaching or feels discouraged about it might enhance their competence and classroom routines over time.

Table 1

Approaching Pronunciation Teaching in the Classroom

Having the know-how	Noticing the problem	Addressing the problem
• Inform yourself about the L2 sound system. • Do a diagnostic assessment of your students' speech.	• Observe your classroom context and start with one pronunciation activity (research-grounded). • See what works best for your students and build on that experience.	• Foster intelligible pronunciation for communication. • Encourage autonomous learning and use of online resources.

It is important to highlight that providing support, collaborating, and partnering are key elements for nurturing a successful approach to pronunciation teaching.

In terms of the circulation of information, new knowledge needs to be regularly disseminated and shared through various fora and publications. A perfect example of a successful connection between research and practice are teachers' associations. In particular, the following deserve a special mention: the IATEFL Pronunciation Special Interest Group (PronSIG), the TESOL Speech Pronunciation Listening Interest Section (SPLIS), and the CATESOL Teaching of Pronunciation Interest group (TOP-IG).[5] These associations boast a community of enthusiasts who are active researchers and practitioners and who regularly organise conferences and online webinars (both theoretical and practical), share professional development content, publish their own journals[6], as well as share pronunciation-related teaching tips via podcasts, blogs, and social media reels. Furthermore, other good practices are promoted via various platforms where resources for teachers and learners are available such as (among others): English Accent Coach[7] for practising L2 perception through HVPT exercises; task banks, e.g., TBLT Language Learning Task Bank[8]; SLA Speech Tools[9] website, which

[5] IATEFL Pronunciation Special Interest Group – PronSIG https://pronsig.iatefl.org/
TESOL Speech Pronunciation Listening Interest Section – SPLIS https://www.tesol.org/
CATESOL TOP-IG: Teaching of Pronunciation Interest group https://www.catesol.org/

[6] Speak Out! https://pronsig.iatefl.org/journal/
TESOL Quarterly https://onlinelibrary.wiley.com/journal/15457249)
The CATESOL Journal http://www.catesoljournal.org/

[7] https://www.englishaccentcoach.com/

[8] https://tblt.indiana.edu/tasks/details.html?id=73

[9] http://sla-speech-tools.com/

offers a variety of tools for research and teaching; PhoTransEdit[10] application for practising phonemic transcription, etc. Recently, publications with a specific focus on enhancing teachers' pronunciation knowledge – Liu et al. (2023) being a notable example – have captured publishers' attention. Practically-oriented publications with detailed descriptions of courses with an integrated pronunciation component (Murphy, 2017) have also been available on the market. These are examples of already successful practices that should continue to be encouraged.

These achievements – in creating communities and resources – could be combined in an all-inclusive multi-content platform that would serve as a reference point for researchers, teachers, and learners. Such a platform – call it a Pronunciation Core - could serve as a medium for teaching and learning English pronunciation. It would aim to provide professional help to English language teachers by joining the two worlds of research and practice and narrowing the existing gap, where ideas and practical suggestions based on research findings could be offered, thus assisting teachers in making research-informed decisions and improving their teaching practices. These recommendations could be written in non-technical, simple language so that they address pre-service and in-service teachers' immediate needs. The platform would be an ideal spot where various points of interests could intersect: a) concepts explained; b) research updates and practical implications shared; c) useful links and resources for the classroom linked; d) typical features of diverse native and non-native accents described; e) professional development instructional videos uploaded; f) social media networks built; etc.

Finally, one aspect that begs for urgent action is opening a conversation between researchers and teachers, on the one hand, and coursework developers on the other. It is crucial that pronunciation as a language skill is given equal treatment in published teaching materials. After all, pronunciation should not be treated as an optional element in the classroom – it needs to be part of every lesson so that teachers can accommodate to their learners' needs.

5 Conclusion

This chapter examines aspects of the on-going debate on how to link pronunciation research and practice. It evaluates effective research findings related to important pronunciation issues, which pedagogy needs to take into consideration. It also considers their implications in the L2 classroom, especially from teachers' point of view, and discusses the challenges they face in their professional lives. In doing so, the paper has a two-fold aim: to invite researchers to reflect on how they see pedagogy and to redefine how teachers approach research.

Future endeavours will reveal whether value-added relations are developed between these two communities, who do not always have overlapping roles. Just as much as teachers need support, in the form of highlighted research results and their incorporation into hands-on materials, researchers also need directions from teachers as to which pedagogical issues are more pressing and should be brought to their attention. Undoubtedly, more research should be focused on innovation and practical application in designing actual output resources that can be applied in the classroom. In addition, teachers should reflect on their teaching practice and context in light of current research findings, so that they can inform themselves, build on their existing knowledge, and strengthen their confidence to teach pronunciation. Such an approach is useful in the long run and will empower teachers to make more spontaneous decisions in the classroom about what and how to teach. In this way, they will be better able to help their learners reach their goals. By bringing in a reasonable amount of structure, and affording teachers enough opportunities for autonomous action, feelings of discouragement about

[10] http://www.photransedit.com/

teaching pronunciation could be minimised and better practices could be promoted and developed.

References

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly, 22*(4), 593–606.

Catford, J. C. (1987). Phonetics and the teaching of pronunciation: A systematic description of English phonology. In J. Morley (Ed.), *Current perspectives on pronunciation: Practices anchored in theory* (pp. 87–100). TESOL.

Cauldwell, R. (2013). *Phonology for listening: Teaching the stream of speech*. Speech in Action.

Cauldwell, R. (2018). *A syllabus for listening: Decoding*. Speech in Action.

Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.

Couper, G. (2011). What makes pronunciation teaching work? Testing for the effect of two variables: Socially constructed metalanguage and critical listening. *Language Awareness, 20*(3), 159–182. https://doi.org/10.1080/09658416.2011.570347

Červinková Poesová, K., & Lancová, K. (2021). In A. Kirkova-Naskova, A. Henderson, & J. Fouz-González (Eds.), *English pronunciation instruction: Research-based insights* (pp. 91–117). John Benjamins. https://doi.org/10.1075/aals.19.05cer

Darcy, I. (2018). Powerful and effective pronunciation instruction: How can we achieve it? *The CATESOL Journal, 30*(1), 13–45. http://www.catesoljournal.org/wp-content/uploads/2018/03/CJ30.1_darcy.pdf

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), 379–397. https://doi.org/10.2307/3588486

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.

Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency and complexity of L2 accented speech. *Applied Language Learning*, *13*(1), 1–17.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favour of a broad framework for pronunciation instruction. *Language Learning*, *48*(3), 393–410. https://doi.org/10.1111/0023-8333.00047

Foote, J. A, & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation, 3*(1), 34–56. https://doi.org/10.1075/jslp.3.1.02foo

Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal*, *29*(1), 1–22. https://doi.org/10.18806/tesl.v29i1.1086

Fouz-González, J., & Mompeán, J. A. (2021). Exploring the potential of phonetic symbols and keywords as labels for perceptual training. *Studies in Second Language Acquisition, 43*(2), 297–328. https://doi.org/10.1017/S0272263120000455

Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkova-Naskova, A., Waniek-Klimczak, E., Levey, D., Cunningham, U., & Curnick, L. (2012). English pronunciation teaching in Europe survey: Selected results. *Research in Language*, *10*, 5–27. https://doi.org/10.2478/v10015-011-0047-4

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.

Jones, T. (2016). *Pronunciation in the classroom: The overlooked essential*. TESOL Press.

Kirkova-Naskova, A., Henderson, A., & Fouz-González, J. (Eds.). (2021). *English pronunciation instruction: Research-based insights*. John Benjamins.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics, 36*(3), 345–366. https://doi.org/10.1093/applin/amu040

Lee, B., Plonsky, L., & Saito, K. (2020). The effects of perception- vs. production-based pronunciation instruction. *System, 88,* 102185. https://doi.org/10.1016/j.system.2019.102185

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369–377.

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press.

Levis, J. M. (2021). Connecting the dots between pronunciation research and practice. In A. Kirkova-Naskova, A. Henderson, & J. Fouz-González (Eds.), *English pronunciation instruction: Research-based insights* (pp. 17–37). John Benjamins. https://doi.org/10.1075/aals.19.02lev

Levis, J. M., & Muller Levis, G. (2018). Teaching high-value pronunciation features: Contrastive stress for intermediate learners. *CATESOL Journal, 30*(1), 139–160. http://www.catesoljournal.org/wp-content/uploads/2018/03/CJ30.1_levis2.pdf

Levis, J. M., Derwing, T. M., & Sonsaat-Hegelheimer, S. (Eds.). (2022). *Second language pronunciation: Bridging the gap between research and teaching*. Wiley Blackwell.

Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. Routledge.

Liu, D., Jones, T., & Reed, M. (2023). *Phonetics in language teaching*. Cambridge University Press.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching, 46*(1), 1–40. https://doi.org/10.1017/S0261444812000365

Macdonald, S. (2002). Pronunciation – Views and practices of reluctant teachers. *Prospect*, *17*(3), 3–18. http://www.ameprc.mq.edu.au/__data/assets/pdf_file/0012/230061/17_3_1_MacDonald.pdf

McAndrews, M. M., & Thompson, R. (2017). Establishing an empirical basis for priorities in pronunciation teaching. *Journal of Second Language Pronunciation, 3*(2), 267–287. https://doi.org/10.1075/jslp.3.2.05mca

McCrocklin, S., Humaidan, A., & Edalatishams, E. (2019). ASR dictation program accuracy: Have current programs improved? In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference* (pp. 191–200). Iowa State University. https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2019/07/10th-Annual-PSLLT-Proceedings-FINAL.pdf

McGregor, A., & Reed, M. (2018). Integrating pronunciation into the English language curriculum: A framework for teachers. *The CATESOL Journal*, *30*(1), 69–93. http://www.catesoljournal.org/wp-content/uploads/2018/03/CJ30.1_mcgregor.pdf

Mora, J. C., & Levkina, M. (2017). Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition, 39*(2), 381–399. https://doi.org/10.1017/S0272263117000183

Munro, M. (2008). Foreign accent and speech intelligibility. In J. Hansen & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 193–218). John Benjamins.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning*, *45*, 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. System, *34*(4), 520–531. https://doi.org/10.1016/j.system.2006.09.004

Murphy, J. (2017). *Teaching the pronunciation of English: Focus on whole courses*. University of Michigan Press.

Murphy, J., & Baker, A. (2019). History of ESL pronunciation teaching. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 36–65). Wiley Blackwell.

Ortega, M., Mora-Plaza, I., & Mora, J. C. (2021). Differential effects of lexical and non-lexical high-variability phonetic training on the production of L2 vowels. In A. Kirkova-Naskova, A. Henderson, & J. Fouz-González (Eds.), *English pronunciation instruction: Research-based insights* (pp. 327–355). John Benjamins. https://doi.org/10.1075/aals.19.14ort

Osburne, A. G. (2003). Pronunciation strategies of advanced ESOL learners. *International Review of Applied Linguistics*, *41*, 131–143. https://doi.org/10.1515/iral.2003.005

Pennington, M. C., & Rogerson-Revell, P. (2019). *English pronunciation teaching and research: Contemporary perspectives*. Palgrave Macmillan.

Pickering, L. (2018). *Discourse intonation: A discourse-pragmatic approach to teaching the pronunciation of English*. University of Michigan Press.

Piper, T., & Cansin, D. (1988). Factors influencing the foreign accent. *The Canadian Modern Language Review*, *44*(2), 334–342. https://doi.org/10.3138/cmlr.44.2.334

Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191–215. https://doi.org/10.1006/jpho.2001.0134

Purcell, E. T., & Suter, R.W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning, 30*, 271–287. https://doi.org/10.1111/j.1467-1770.1980.tb00319.x

Quian, M., Chukharev-Hudilainen, E., & Levis, J. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *Language Learning & Technology, 22*(1), 69–96. https://dx.doi.org/10125/44582

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, *46*(4), 842–854. http://www.jstor.org/stable/43267893

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning, 69*, 652–708. https://doi.org/10.1111/lang.12345

Sardegna, V. G. (2022). Evidence in favour of a strategy-based model for pronunciation instruction. *Language Teaching*, *55*(3), 363–378. https://doi.org/10.1017/S0261444821000380

Sardegna, V., & Jarosz, A. (Eds.) (2023). *English pronunciation teaching: Theory, practice and research findings*. Multilingual Matters.

Suzukida, Y., & Saito, K. (2022). What is second language pronunciation proficiency? An empirical study. *System*, *106*, 102754. https://doi.org/10.1016/j.system.2022.102754

Szyszka, M. (2021). Pronunciation-learning strategies: A task-based perspective. In A. Kirkova-Naskova, A. Henderson, & J. Fouz-González (Eds.), *English pronunciation instruction: Research-based insights* (pp. 147–171). John Benjamins. https://doi.org/10.1075/aals.19.07szy

Thomson, R. (2018). High Variability [Pronunciation] Training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation, 4*(2), 208–231. https://doi.org/10.1075/jslp.17038.tho

Thomson, R., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326–344. https://doi.org/10.1093/applin/amu076

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition, 15*(4), 905–916. https://doi.org/10.1017/S1366728912000168

Walker, R. (2010). *Teaching the pronunciation of English as a lingua franca.* Oxford University Press.

About the author

Anastazija Kirkova-Naskova is an Associate Professor in English phonetics and morphology in the Department of English Language and Literature, Ss. Cyril and Methodius University in Skopje, North Macedonia. She also teaches research methodology (MA level). Her research interests include English pronunciation teaching and learning, foreign-accented speech, speech perception, pronunciation learning strategies, and teacher cognition. Together with Alice Henderson and Jonás Fouz-González, she co-edited *English pronunciation instruction: Research-based insights* (John Benjamins, 2021). She was an assistant editor-in-chief for *The Journal of Contemporary Philology*.

Email: akirkova@flf.ukim.edu.mk

Kivistö de Souza, H. & Lintunen, P. (2023). "Thinking about your pronunciation": Examining phonological self-awareness with a novel task. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 138–148). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225603

"Thinking about your pronunciation": Examining phonological self-awareness with a novel task

Hanna Kivistö de Souza
Federal University of Santa Catarina, Florianópolis and University of Turku

Pekka Lintunen
University of Turku

Noticing the gap in one's pronunciation is notoriously demanding (Piske, 2008), and yet becoming aware of pronunciation challenges is beneficial for overall pronunciation competence (Kivistö de Souza, 2017). Previous studies on phonological self-awareness have employed global tasks such as journaling (e.g., Kennedy & Blanchet, 2014), and have pointed out that more explicit learning conditions lead to more noticing (White & Ranta, 2002). The objective of this paper is to present an instrument that examines second language phonological awareness by bringing the phonetic detail explicitly into the learners' attention. The participants were 33 L1 Finnish advanced university learners of English attending an undergraduate course on English phonetics and phonology. At the beginning of the semester, the participants provided a speech sample targeting tricky English sounds. At the end of the semester, a "Thinking about your pronunciation" task was administered in which the samples were played back to the participants. They were asked to indicate any pronunciation deviations they could perceive and to elaborate on how they perceived their own intelligibility and their abilities in recognising phonetic and phonological phenomena in their own and others' speech. Our observations with the task indicate that the instrument can be a helpful and reliable tool in tapping into phonological self-awareness.

Keywords: phonological awareness, phonological self-awareness, pronunciation instruction, noticing, language awareness

1 Introduction: Research on phonological self-awareness

Language users possess vast amounts of knowledge about the phonological systems of the languages they speak, known as phonological awareness, as evidenced by accurate production and perception of target language sounds, syllables, stress, and intonational patterns. Furthermore, they are able to recognise phonological deviations in the form of a foreign accent, even when the speech sample is extremely short and played backwards (Munro, Derwing & Burgess 2003). Adult language users' phonological awareness thus entails knowledge about the target language phonological system at the subphonemic, segmental, and suprasegmental levels.

Phonological awareness also involves knowledge about one's own phonological competence (Kivistö de Souza, 2015), and includes noticing the gap (Schmidt & Frota, 1986) between one's own production and the target production. Such awareness is also evidenced by language users' ability not only to recognise foreign accented speech but also to provide accuracy judgments on others' speech. Noticing the gap has also been referred to as phonological self-awareness (Kivistö de Souza, 2015) or as phonological self-assessment or self-perception (Isbell & Lee, 2022; O'Brien, 2019). Nevertheless, noticing (i.e., becoming aware of a specific stimulus) does not necessarily entail understanding (i.e., verbalisation of the underlying rules), as these are seen as two distinct levels of language awareness (Schmidt, 1990).

Previous research on phonological self-awareness indicates that speakers' assessment of their pronunciation abilities correlates moderately with the actual performance (e.g., Saito et al., 2020; Trofimovich et al., 2016). However, speakers often tend to either over- or under-estimate their pronunciation skills rather than to provide accurate self-assessments. Language learners whose phonological self-awareness is accurate have been shown to have more accurate segmental (e.g., Saito, 2019) and suprasegmental (e.g., O'Brien, 2019) pronunciation. Noticing phonetic detail in regular classroom interactions can be challenging, and research suggests that drawing learners' attention to phonetic detail and explicitly focusing on L2 pronunciation features is beneficial for L2 pronunciation development (Saito, 2021).

Many of the previous studies about L2 phonological awareness have presented the participants with global tasks such as free journaling (Kennedy & Blanchet, 2014; Kennedy & Trofimovich, 2010), stimulated recall (Wrembel, 2011, 2013) or imitating L2 accented speech (Mora, Rochdi & Kivistö-de Souza, 2014). The objective of the present study was to develop an instrument that would bring phonetic detail into the learners' attention in a more focused and controlled manner by prompting learners to engage in in-depth reflection about their pronunciation.

2 Methodology

2.1 Thinking about your pronunciation: Task format

An instrument we called a "Thinking about your pronunciation" task was created to examine the phonological self-awareness of advanced English speakers taking a course in English phonetics and phonology. The objective of this task was to encourage the students to engage in self-reflection in relation to their English pronunciation. The task was carried out towards the end of the semester and thus also served as an opportunity to revisit the course contents. Our previous experience with similar tasks suggested that students would have difficulties in noticing phonological challenges in their own pronunciation (Lintunen, 2013). For this reason, we tried to design a task consisting of four parts (see Appendix) that would be as explicit as possible and would offer ample opportunities for noticing.

In Part 1, the participants listened to a speech sample they had provided at the beginning of the semester and indicated any pronunciation deviations they could notice. The speech sample was a wordlist recording of 12 words containing phonemes and features known to be challenging for L1 Finnish speakers of English (e.g., Lintunen, 2004), namely: aspiration, voiced plosives in word initial and final position as well as reinterpretation of distinctions in vowel contrasts. All the target items were high frequency monosyllabic CVC words. When recording the speech sample, the participants were not aware of the target sounds nor that they would analyse the pronunciation later in the semester. To increase noticing, each word was played three times, and with each time, the participant was asked to focus on one target sound (the initial consonant, the vowel or the final consonant). The participants were asked to mark whether they had pronounced the target sound correctly or not. Optionally, they could also explain why they thought their pronunciation was inaccurate. We chose to use the term 'correct' to describe phonologically accurate, non-deviant productions to facilitate the participants' comprehension. The definition of 'correct' was not given in the instructions, but we assumed that the participants' perceptions of correctness would adhere to a nativeness norm and vary from native production to near-native-like.

Once the participants had focused on each segment in each of the 12 target words separately, they were asked in Part 2 to indicate if they thought that certain words were entirely unintelligible for other English speakers. The objective of Part 2 was to allow the participants to elaborate in more detail on the items they considered especially challenging.

Part 3 consisted of the participants' self-assessment of their overall English pronunciation. They were asked to indicate on a 7-point Likert scale how comprehensible they thought their speech is for native English speakers. This question aimed to guide the participants' attention to their pronunciation as a whole, in comparison to focusing on specific pronunciation instances as in the first two sections. Moreover, whereas Part 1 focused on perceptions of accuracy (i.e., how much the pronunciation deviates from the target) and Part 2 on the perception of intelligibility (i.e., how well the productions are understood by other English speakers), Part 3 focused on self-perceived comprehensibility (i.e., how much effort the listener needs to understand the speech) (Munro & Derwing, 1995).

The final section, Part 4, was adapted from Kivistö de Souza (2015), and it focused on the participants' self-assessment of their phonological abilities. The section consisted of a set of phonological self-awareness questions the participants were asked to provide their opinion on. These questions asked, for instance, how easy it was for them to: notice pronunciation mistakes in other non-native English speakers' and their own speech, identify English spoken with different accents and explain mistakes of phonological nature. The objective of Part 4 was to explore the participants' self-reports on the two levels of language awareness, i.e., noticing and understanding (Schmidt, 1990).

2.2 Participants and raters

The participants were 33 L1 Finnish speakers doing an undergraduate degree in English language. Their mean age was 21.2 years, and their English proficiency was estimated to correspond to CEFR level C1, as indicated by their LexTALE scores (M = 86.70, SD = 7.35) (Lemhöfer & Broersma, 2012, p. 341). Twelve participants were male, 18 female and three did not disclose gender information. At the time of data collection, the participants were enrolled in an obligatory first-year practical course on English phonetics and phonology which aimed at improving students' pronunciation through practical exercises, as well as description and transcription of English phonemes. The raters were three university professors specialising in English Phonetics and Phonology with extensive experience in rating speech samples. Two raters were L1 Finnish speakers and one rater was an L1 Portuguese speaker.

2.3 Procedures and analyses

The speech samples (word list readings) for the first part of the task were recorded at the beginning of the semester as part of a larger project. First, the recordings' sound quality was improved by removing noise and normalising the speech samples with Audacity© (Audacity Team, 2021). Then, for each of the 33 participants individual recordings were created in Praat (Boersma & Weenik, 2022). The target words were isolated from a larger set of stimuli set and combined into a new sound file together with silent pauses. Each word was copied into the file three times with a one second pause between the words. The final repetition of the word was followed by a three second pause to indicate the change of a set. The recordings were approximately 2.5 minutes long.

The recordings and the instructions were made available through the course's virtual learning platform Moodle, to which the researchers were granted access by the course instructors. Additionally, printed copies of the answer sheets were handed out to the participants in class. Each participant had access to their own recording only. The participants completed the self-paced task at home as an additional non-graded homework assignment. The instructions were given to carry out the task in a calm environment with headphones, if available. Repeated listening of the recordings was allowed. Once the participants had completed the entire task, they submitted the answer sheets through Moodle. Instances of noticing of pronunciation deviations (instances marked as 'inaccurate' in the answer sheet) were computed for each participant and target feature, and then converted into percentages.

To determine whether the participants were noticing actual pronunciation deviations or indicating deviations in accurate productions (i.e., being overly critical), the participants' scores were compared to performance rating scores given by three expert raters. Two raters assessed all samples for accuracy and their overall agreement was 85.7%. When agreement could not be reached, the samples were additionally assessed by a third rater.

3 Results

This section describes our observations on the use of the instrument. (For detailed results see Kivistö de Souza and Lintunen (forthcoming)).

Very few data were missing from Part 1 (11 out of 1188 instances = 0.01%) in which participants listened to their speech samples and assessed their correctness, suggesting that they understood the task and put some effort into completing it, despite not being in the presence of a researcher. Participants differed greatly in their self-assessments and reported noticing on average 9.12% (range: 0–30%) of the pronunciation deviations (i.e., indicated as 'not correct' on the answer sheet). Most of the deviations noticed were in initial consonants (M = 12.13%, range: 0–62%), followed by final consonants (M = 7.81%, range: 0–34%) and vowels (M = 7.41%, range: 0–33%). Among the phonetic phenomena analysed (aspiration, voiced stops, vowel quality), the most frequent problems were in pronouncing initial voiceless plosives (inadequate VOTs; 14.7%, range: 0–100%). Deviations in voiced stops (devoicing) were slightly less frequent (M = 10.58%, range: 0–40%).

When the participants' reported noticing is compared with their actual performance, a slightly different image appears. By looking at the raters' assessments of the participants' pronunciation accuracy and comparing them with the participants' reported noticing, the participants' overall pronunciation accuracy was very high (M = 97.90%, range: 37.5–100%), and the participants noticed on average 25.9% of the pronunciation deviations present. However, the range of noticing varied from 0 to 100%, indicating a large individual variation in participants' phonological self-awareness.

On very few occasions, the participants (P) chose to elaborate on the deviations they had noticed (93 out of 1188 instances = 0.07%), most likely because this was presented as an optional activity. Short comments were more frequently used, for instance: “not clear enough” (P28), “too Finnish” (P57), “not aspirated” (P73). Longer and more detailed comments were also present in the data though less frequently, for example: “I emphasise the sound too much in order to not make it sound like the consonant "p", which makes it sounds weird” (P80).

Examining the responses to Part 2 where the participants could elaborate on specific items they considered unintelligible, three participants did not answer the question (10%), nine (27%) reported that they did not think any of their productions would be unintelligible, and the remaining 21 (63%) identified at least one word they considered to be unintelligible. The word that was mentioned as unintelligible most frequently (n = 9) was *pub*, followed by *buck* (n = 5). Both were reported to be easily mistaken for *pup* and *puck*, hence, the participants most likely perceived problems in consonant voicing. One participant commented that it would be difficult to know how other speakers would perceive their pronunciation.

In Part 3 where the participants rated their self-perceived comprehensibility, there was no missing data. The participants favoured the upper range of the scale (M = 5.6, range: 4–7), indicating that they did not expect native English speakers to experience problems in understanding them. This self-perception seems accurate as participants had a high English proficiency and were rated as highly accurate by expert raters.

The reliability of the set of phonological self-awareness questions was examined in section four. An earlier version of these questions was found to be an acceptable measure (n = 71, α = .75) for phonological self-awareness for Brazilian Portuguese learners of English (Kivistö de Souza, 2015). In the version used in the present study, Cronbach’s Alpha of .84 indicated that the items had a relatively high internal consistency and thus could be seen to tap into the same underlying construct (phonological self-awareness). There were three missing data points (out of a total of 396 instances) indicating that even though the participants might have considered some of the questions difficult, they still tried to answer.

Looking at the individual questions that made up the scale, the participants as a group reported the highest ability in recognising Finnish-accented English (quite easy M = 4.3) as well as noticing pronunciation mistakes in their own pronunciation (quite easy M = 4.1). The lowest ability was reported in explaining why a heard sound combination is possible or impossible in English (very difficult M = 2.6) and why the heard intonation and rhythm patterns are correct or incorrect (very difficult M = 2.7).

4 Discussion and conclusions

Phonological awareness has been shown to be positively related to pronunciation accuracy (e.g., O’Brien, 2019; Saito, 2019), and consequently, methods and instruments to increase language learners’ phonological awareness are highly relevant for L2 classrooms. In this paper we have described one such instrument that could be useful in drawing learners’ attention to their L2 pronunciation. We tested the instrument with 33 advanced Finnish learners of English who were attending a course in English phonetics and phonology. Our objective was to create an instrument that would reliably tap into phonological self-awareness and encourage students to engage in self-reflection about their L2 pronunciation in a practical manner. Our preliminary observations suggest that the instrument can be useful in examining language learners’ phonological self-awareness. Nevertheless, some issues arose from Part 1 of the instrument which researchers should address if they are interested in employing the tool.

We were aware that noticing the gap might be challenging, even for advanced language learners, which is why we tried to make Part 1 as explicit as possible. The participants received explicit instructions about which aspects to focus on and the task was self-paced, so the sound

files could be played several times. The words were all carefully selected to be monosyllabic and known to the learners, and they were presented in a controlled order in which minimal pair words followed each other (e.g., *pub* was followed by *pup*). This methodology contrasts with think-aloud protocols (Wrembel, 2013, 2015; Zuengler, 1988) and journaling (Kennedy, 2012; Kennedy & Blanchet, 2014; Kennedy & Trofimovich, 2010) which have been used previously to examine phonological awareness, as these methods allow for a wider attentional focus. Despite these efforts, the participants seemed to have a hard time identifying their segmental deviations, as testified by an average of 25.9% of noticed deviations, even though the participants possessed great amounts of metaphonological knowledge due to the course they were attending. Unsurprisingly, the participants found the syllabic and suprasegmental features difficult to explain, as the course they were attending was more focused on segmental phonology and had less emphasis on developing metalinguistic knowledge about English suprasegmentals.

Another possible reason for the low degree of noticing could be the participants' advanced proficiency level: when the production is accurate, there are fewer deviations to notice. However, phonological awareness is positively related to language proficiency (Kivistö de Souza, 2015), so the issue might not be as straightforward. It would be interesting to test participants with intermediate proficiency levels to determine how much noticing takes place.

Another possible reason for the participants' lack of noticing might be the task structure: as the task was carried out at home without researcher's supervision, the participants might have gotten distracted and might not have paid as much attention to the speech samples as they could have. However, as each participant received a different speech sample, working in a language lab was not a possibility for us.

A serious concern for researchers working with phonological self-awareness in this manner is the time required to create and administer the task. First, the researchers have to record the participants, create individualised sound files for each and finally individually send them through a cloud service or learning platform, for example. Second, when learners are asked to notice deviations in their own speech, the researchers have to make sure that those deviations are actually present in the speech samples. This requires either acoustic analysis or perceptual assessment. In the present study, we opted for the perceptual assessment of three teachers, but we are aware that ideally the speech samples should be presented to a larger number of raters.

A point should also be made about what is being noticed. Although we treat segments separately, speech is continuous and segments are affected by their surrounding context and speech rate, among others. When analysing the accuracy of the participants' productions, we had the specific difficulty of deciding, for instance, whether final inaccurate devoicing should be understood as a problem in the preceding vowel (that was too short) or in the lack of voicing of the consonant. These are issues that the researchers have to discuss in advance and while coding the data, in order to maintain consistency.

Finally, we would like to make a recommendation about combined measures of phonological self-awareness. Verbalising language awareness is difficult and verbalising phonological awareness appears to be even more so (Kivistö de Souza, 2015, p. 105; Schmidt, 1990, p. 132). It might be a good idea to complement tasks tapping into explicit phonological awareness, like the present one, with tasks that tap into non-verbalisable phonological self-awareness (e.g., perception, imitation, non-word recognition) for a comprehensive picture of the learners' phonological self-awareness (see Kivistö de Souza & Lintunen, forthcoming).

To conclude, despite the limitations, this instrument is a useful tool that can be used to address the complex aspects of phonological awareness, and it may help shed more light on how language users perceive their own pronunciation.

References

Audacity Team (2021). Audacity: Free Audio Editor and Recorder (Version 3.0.0) [Computer application]. https://audacityteam.org/

Boersma, P., & Weenik, D. (2022). Praat: doing phonetics by computer (Version 6.2.06) [Computer program]. htpps://www.praat.org

Isbell, D. R., & Lee, J. (2022). Self-assessment of comprehensibility and accentedness in second language Korean. *Language Learning, 72*(3), 806–852. https://doi.org/10.1111/lang.12497

Kennedy, S. (2012). Exploring the relationship between language awareness and second language use. *TESOL Quarterly, 46*, 398–408.

Kennedy, S., & Blanchet, J. (2014). Language awareness and perception of connected speech in second language. *Language Awareness, 23*, 92–106. https://doi.org/10.1080/09658416.2013.863904

Kennedy, S., & Trofimovich, P. (2010). Language awareness and second language pronunciation: a classroom study. *Language Awareness, 19*, 171–185. https://doi.org/10.1080/09658416.2010.486439

Kivistö de Souza, H. (2015). *Phonological awareness and pronunciation in a second language* [Unpublished doctoral dissertation]. University of Barcelona. https://fonapli.paginas.ufsc.br/files/2017/03/Kivist%C3%B6-de-Souza_PhD-dissertation1.pdf

Kivistö de Souza, H. (2017). Brazilian EFL learners' awareness about L2 phones: Is "mall" pronounced as "mal"? *Trabalhos Em Linguística Aplicada*, *56*(1), 235–258. https://doi.org/10.1590/010318135071200021

Kivistö de Souza, H., & Lintunen, P. (forthcoming).Did I say 'pup' or 'pub'?- An analysis of foreign language learners' phonological self-awareness.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343. http://dx.doi.org/10.3758/s13428-011-0146-0

Lintunen, P. (2004). *Pronunciation and phonemic transcription: A study of advanced Finnish learners of English.* University of Turku.

Lintunen, P. (2013). The effect of phonetic knowledge on evaluated pronunciation problems. In J. Przedlacka, J. Maidment, & M. Ashby (Eds). *Proceedings of the Phonetics Teaching and Learning Conference* (pp. 55–58). University College London. https://www.ucl.ac.uk/pals/proceedings-2013

Mora, J. C., Rochdi, Y., & Kivistö-de Souza, H. (2014). Mimicking accented speech as L2 phonological awareness. *Language Awareness, 23*(1-2), 57–75. https://doi.org/10.1080/09658416.2013.863898

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning, 45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., Derwing, T. M., & Burgess, C. S. (2003). The detection of foreign accent in backwards speech. In M. J. Sol´e, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 15)* (pp. 535–538). Futurgraphic.

O'Brien, M. G. (2019). Attending to second language lexical stress: Exploring the roles of metalinguistic awareness and self-assessment. *Language Awareness, 28*(4), 310–328. https://doi.org/10.1080/09658416.2019.1625912

Piske, T. (2008). Phonetic awareness, phonetic sensitivity and the second language learner. In J. Cenoz & N. H. Hornberger (Eds.), *Encyclopedia of language and education Vol. 6: Knowledge about language* (pp. 155–166). Springer Science.

Saito, K. (2019). Individual differences in second language speech learning in classroom settings: Roles of awareness in the longitudinal development of Japanese learners' English /ɹ/ pronunciation. *Second Language Research*, *35*(2), 149–172. https://doi.org/10.1177/0267658318768342

Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, *55*(3), 1–35. https://doi.org/10.1002/tesq.3027

Saito, K., Trofimovich, P., Abe, M., & In'nami, Y. (2020). Dunning-Kruger effect in second language speech learning: How does self-perception align with other perception over time? *Learning and Individual Differences, 79*, Article 101849. https://doi.org/10.1016/j.lindif.2020.101849

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237–326). Newbury House.

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition, 19*, 1–19. https://doi.org/10.1017/S1366728914000832

White, J., & Ranta, L. (2002). Examining the interface between metalinguistic task performance and oral production in a second language. *Language Awareness, 11,* 259–290.

Wrembel, M. (2011). Metaphonetic awareness in the production of speech. In M. Pawlak, E. Waniek-Klimczak & J. Majer (Eds.), *Speaking and instructed foreign language acquisition* (pp. 169-182). Clevedon: Multilingual Matters.

Wrembel, M. (2013). Metalinguistic awareness in third language phonological acquisition. In K. Roehr & G. A. Gánem-Gutiérrez (Eds.), *The metalinguistic dimension in instructed second language learning* (pp. 119–144). London: Bloomsbury.

Wrembel, M. (2015). Metaphonological awareness in multilinguals: A case of L3 Polish. *Language Awareness, 24*, 60–83. https://doi.org/10.1080/09658416.2014.890209

Zuengler, J. (1988). Identity markers and L2 pronunciation. *Studies in Second Language Acquisition, 10*, 33–49. https://doi.org/10.1017/S027226310000694X

Appendix

Thinking About Your Pronunciation Answering Sheet

Instructions in a nutshell:

1. Download your recording from Moodle ("Thinking about your pronunciation homework")
2. Make sure that you are in a quiet place where you can concentrate. If you have headphones, please use them.
3. Keep this answering sheet and a pen ready and play the recording. You will hear each word three times and each time you should pay attention to different parts of the word.
4. Listen carefully and tick the answer that applies. If you're unsure, you can listen to the word again. You can also stop the recording to have more time to answer.
5. Once you have listened to all words, answer the questions at the end.
6. Return the answering sheet to your teacher on the class 16th November. Alternatively, you can take pictures or scan your answers and upload them on Moodle.

Part 1. Did you pronounce the indicated part of the word correctly? If you didn't, you can explain shortly why. *There is a longer pause and a beep before the next word is presented. If you need more time before the next word, pause the recording when you hear the beep.*					
1. Pup					
The initial consonant **p**up	☐	Yes	☐	No	(Why not?)
The vowel p**u**p	☐	Yes	☐	No	(Why not?)
The final consonant pu**p**	☐	Yes	☐	No	(Why not?)
2. Pub					
The initial consonant **p**ub	☐	Yes	☐	No	(Why not?)

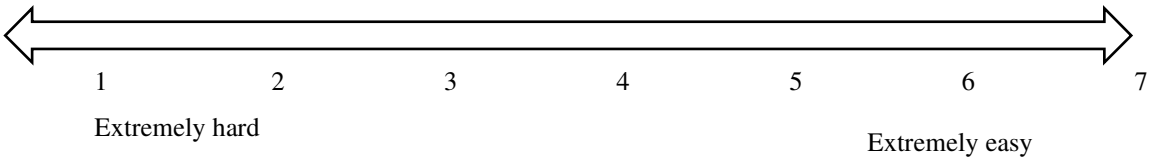
The vowel p**u**b	☐ Yes	☐ No	(Why not?)
The final consonant pu**b**	☐ Yes	☐ No	(Why not?)
3. Bet			
The initial consonant **b**et	☐ Yes	☐ No	(Why not?)
The vowel b**e**t	☐ Yes	☐ No	(Why not?)
The final consonant be**t**	☐ Yes	☐ No	(Why not?)
4. Bed			
The initial consonant **b**ed	☐ Yes	☐ No	(Why not?)
The vowel b**e**d	☐ Yes	☐ No	(Why not?)
The final consonant be**d**	☐ Yes	☐ No	(Why not?)
5. Buck			
The initial consonant **b**uck	☐ Yes	☐ No	(Why not?)
The vowel b**u**ck	☐ Yes	☐ No	(Why not?)
The final consonant bu**ck**	☐ Yes	☐ No	(Why not?)
6. Bug			
The initial consonant **b**ug	☐ Yes	☐ No	(Why not?)
The vowel b**u**g	☐ Yes	☐ No	(Why not?)
The final consonant bu**g**	☐ Yes	☐ N	(Why not?)
7. Deed			
The initial consonant **d**eed	☐ Yes	☐ No	(Why not?)
The vowel d**ee**d	☐ Yes	☐ No	(Why not?)
The final consonant dee**d**	☐ Yes	☐ No	(Why not?)
8. Did			
The initial consonant **d**id	☐ Yes	☐ No	(Why not?)
The vowel d**i**d	☐ Yes	☐ No	(Why not?)
The final consonant di**d**	☐ Yes	☐ No	(Why not?)
9. Beat			
The initial consonant **b**eat	☐ Yes	☐ No	(Why not?)
The vowel b**ea**t	☐ Yes	☐ No	(Why not?)
The final consonant bea**t**	☐ Yes	☐ No	(Why not?)
10. Bit			
The initial consonant **b**it	☐ Yes	☐ No	(Why not?)
The vowel b**i**t	☐ Yes	☐ No	(Why not?)
The final consonant bi**t**	☐ Yes	☐ No	(Why not?)
11. Peak			
The initial consonant **p**eak	☐ Yes	☐ No	(Why not?)
The vowel p**ea**k	☐ Yes	☐ No	(Why not?)
The final consonant pea**k**	☐ Yes	☐ No	(Why not?)
12. Pick			
The initial consonant **p**ick	☐ Yes	☐ No	(Why not?)
The vowel p**i**ck	☐ Yes	☐ No	(Why not?)
The final consonant pi**ck**	☐ Yes	☐ No	(Why not?)

Part 2. Do you think that there is any word that you pronounced that might not be understood by other speakers of English (native or non-native)? Which one(s)?

_____

_____

Part 3. How easy do you think it is for native speakers of English to understand your pronunciation when you speak in English? ***Circle the corresponding number on the scale.***



Part 4. How easy it is for you to...
Write ***X*** *on the corresponding box.*

	5 Very easy	**4 Quite easy**	**3 Quite difficult**	**2 Very difficult**	**1 I can't do this at all**
notice pronunciation mistakes in the production of *individual sounds* in other non-native English speakers' speech?					
notice pronunciation mistakes in *intonation and rhythm* in other non-native English speakers' speech?					
notice pronunciation mistakes in *your own* English speech?					
tell where a *native* speaker of English comes from based on their accent?					
tell whether a non-native speaker of English is *Finnish* based on their English accent?					
tell where a *non-native* speaker of English (not Finnish) comes from based on their English accent?					
notice whether a sound combination you hear is possible in English or not?					
notice whether the intonation and rhythm you hear in an English sentence are possible or not?					
notice whether an individual sound you hear is pronounced correctly in English or not?					
explain why a sound combination you hear is possible or impossible in English?					
explain why the intonation and rhythm you hear are correct or incorrect in English?					
explain why an individual sound you hear isn't pronounced correctly in English?					

About the authors

Hanna Kivistö de Souza is an Associate Professor at the Federal University of Santa Catarina (UFSC) in Brazil where she teaches and advises at the undergraduate and graduate levels. Her main research interests are L2 speech acquisition and cognitive- and psycholinguistics applied to SLA. She is especially fascinated by the role of phonological (self)awareness in L2 speech acquisition and has conducted research mostly on explicit and implicit phonological awareness as well as phonological self-awareness. In 2021, she was a visiting researcher at the University of Turku.

Email: hanna.kivistodesouza@gmail.com

Pekka Lintunen is a Professor of English at the University of Turku, Finland, where he heads the Second Language Acquisition track. Lintunen's main research field is the development of spoken L2 skills. Recently, his research has focused on the fluency and complexity or learner language, the learning and teaching of L2 pronunciation, learner perceptions and beliefs and informal language learning. Currently, he coordinates research projects for his FlowLang research group. Lintunen has also co-edited several publications, including *Fluency in L2 Learning and Use* (Multilingual Matters, 2020).

Email: pekka.lintunen@utu.fi

Martin-Rubió, X. (2023). English-medium-instruction lectures: Tackling the issue of (dis)fluency and accuracy measures. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 149–159). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225405

English-medium-instruction lectures: Tackling the issue of (dis)fluency and accuracy measures

Xavier Martin-Rubió
University of Lleida

More and more higher education institutions require certification of language proficiency in order to teach content in English, but countries assess proficiency differently. Moreover, fluency and pronunciation accuracy are just two aspects in determining proficiency.

A group of researchers from five European universities collected data from interviews and classroom recordings of six lecturers per university in 2018-2019, to compare the realities of English Medium Instruction in these contexts and, later, to improve training for EMI lecturers. A sub-group of researchers conducted a CEFR classification of the 30 lecturers. Another sub-group carried out a (dis)fluency analysis of two, 3-minute monologic stretches of 10 lectures, from 2 different lecturers in each country. To measure fluency, Mean Syllables per Run, Rate of Speech Time and Time Ratios were considered, while for disfluency the number of repetitions and false-starts/self-corrections per 100 syllables was determined.

This chapter reflects upon the complexity that surrounds the concept of pronunciation accuracy by narrowing in on a specific case: one Italian lecturer's use of epenthetic vowels after final consonants. This case triggered a lengthy discussion among the researchers as to whether this was a pronunciation error, a feature of an Italian accent, or a filled pause. Assessing lecturers' proficiency is revealed as a complex process which is subject to broader norms.

Keywords: EMI, fluency, pronunciation accuracy, accent, higher education

1 Introduction

English Medium Instruction is defined by Macaro (2018) as "the use of the English language to teach academic subjects (other than English itself) in countries or jurisdictions where the first language of the majority of the population is not English" (p. 19). Normally coupled with the idea of internationalisation, this phenomenon has been gaining ground across European universities. Wächter and Maiworm (2014) differentiated six regions within Europe and argued that there is a north-south divide when it comes to the number of English-Taught programs. According to the authors, the Nordic and Central-West regions have the highest number of such programs, even though recent increases are below the average because they have a high percentage of existing programs in English (e.g., 30% in the Netherlands and 38% in Denmark). Conversely, the South-West and South-East regions have seen impressive growth rates, although the overall numbers are still low (1.2% for Croatia, 2.3% in Spain, and 2.9% in Italy).

A group of researchers from five European universities received an Erasmus+ grant in order to compare the EMI realities in their five countries, to determine whether this north-south divide was confirmed, and to develop tools for improving EMI lectures. These countries cover four of the six regions identified by Wächter and Maiworm (2014): Denmark represents the Nordic region, the Netherlands the Central-West, Spain and Italy represent the South-West, and the South-East is represented by Croatia. The project, entitled "Transnational Alignment of English Competences for University Lectures" (TAEC henceforth), ran from September 2017 to April 2021.

2 The TAEC project: Sub-groups and actions

There were seven transnational meetings in TAEC, and the objective to collect data was adopted at the first one held in Copenhagen, October 2017. The TAEC research group consisted of researchers from the following universities: University of Copenhagen, in Denmark (UCPH), Universitat de Lleida, in Spain (UdL), Maastricht University, in the Netherlands (UM), Università degli Studi di Torino, in Italy (UNITO) and the Faculty of Humanities and Social Sciences of the University of Rijeka, in Croatia (FHSS). Six EMI lectures were collected and six interviews were conducted per university. Table 1 provides a visual summary of the sub-groups and their actions.

Table 1

TAEC Project Sub-groups and Actions

Sub-group	Members	Actions
A	6 TOEPAS experts 6 CEFR experts	• aligning TOEPAS (test used in UCPH) with CEFR
B	8 raters (TAEC members)	• rating 10–15 minutes of the 30 lectures using the CEFR scale
C	2 TAEC members	• calculating (dis)fluency measures of two, 3-minute stretches for 10 of the lectures • identifying mismatches between CEFR ratings and measures
D	7 TAEC members	• setting-up of transcription guidelines and corpus annotation criteria

Sub-group	Members	Actions
E	all members	• organising the revision of the transcriptions • transcribing and revising the transcriptions of the lectures • deciding how to annotate (mis)pronunciations • deciding whether an epenthetic vowel constitutes a filled pause rather than an error

In the third meeting (Maastricht, May 2018), it was agreed that each university would audio and video-record six lectures by six different lecturers from different disciplines and academic levels, and that the lectures would be transcribed. The lecturers were also to be interviewed. Thus, a total of 30 interviews and 30 lectures would be gathered altogether. It was also agreed that insights from the analyses of the videos would be used to create a handbook that could then be used to train EMI lecturers.

The research team was organised into different overlapping sub-groups. One of them was in charge of rating the English level of the lecturers. Eight raters were asked to pay attention to eight aspects of oral production and to provide a holistic assessment as well. Each rater was asked to rate the videos independently, without discussing the results with other raters and/or colleagues. Ten to 15 minutes of each lecture that included both monologue and (when possible) interaction were to be selected by each evaluator. A small sub-group of just two researchers carried out a (dis)fluency analysis of two, 3-minute monologic stretches of 10 lectures, from 2 different lectures in each country. The chosen fluency measures were: Mean Syllables per Run; Rate of Speech Time; and Time Ratios. For disfluency, the number of repetitions and false-starts/self-corrections per 100 syllables was determined. Finally, another sub-group of seven members managed the transcription of the lectures, tagging grammatical and pronunciation errors in the transcription. Initially, this was done with the intention of calculating (pronunciation) accuracy measures of the lecturers, but the validation process for the transcriptions revealed great discrepancies in relation to what constituted an error, so the transcriptions were finally cleared of these tags.

3 Lecturer proficiency levels: TOEPAS and TAEC ratings

A content lecturer in higher education needs an adequate level of the language of instruction to effectively implement the different tasks they must carry out. However, what this adequate level should be remains a controversial matter.

In the last two decades, a number of tests have appeared, specifically to address this issue in relation to EMI lecturers. The Test of Oral English Proficiency for Academic Staff[1] (TOEPAS henceforth) is the test used by the Center for Internationalisation and Parallel Language Use, University of Copenhagen. TOEPAS only focuses on oral production. The test, which is taken by three lecturers in a single session, is structured into three sections: a warm-up section, a 20-minute lecture simulation, and a final part for questions. As part of TAEC (see above), the TOEPAS test was aligned to CEFR. The alignment took place over a three-day standardisation event in October 2018 (see Dimova, 2018). There were 12 participants in the event, six of whom were familiar with TOEPAS, whereas the other six had expertise in rating using the CEFR (Council of Europe, 2018) scale.

The TOEPAS is a performance-based speaking test, and the maximum number of points is 60. A 30-point rating corresponds to a pass: "the lecturer has demonstrated sufficient English

[1] https://cip.ku.dk/english/documents/TOEPAS_2.0_A4_final.pdf

language proficiency for university teaching. No training is required but is strongly recommended". The alignment established that the B2+ level corresponded to this 30-point score, i.e., the minimum level that guaranteed an adequate level for teaching in English. Crucially, the B2 range is the widest one in the CEFR scale. In these discussions, B2 thus covered both 20-point and 30-point scores, and it was the difference between B2- and B2+ levels that determined whether the minimum level had been attained. A minimum acceptable level of B2+ agreed during the alignment discussions reflects the tendency of European universities, as data from O'Dowd (2018) shows. In his study, 44% of the European university representatives surveyed required their lecturers to demonstrate a C1 level, 43% a B2, and only 13% a C2. Spain has one of the most lenient language requirements, as more than 50% of universities required a B2 level (or less) to teach in English-Taught Programs (Halbach & Lázaro, 2015). The overall ratings by the TAEC sub-group in charge of this endeavour rated only four lecturers as having a B2 level (L14 and L15 from Spain, and L25 and L30 from Italy). The remaining 26 lecturers obtained ratings of B2+ or higher.

4 (Dis)fluency and (pronunciation) accuracy annotation

A sub-group of seven TAEC members was constituted in the fourth meeting (Rijeka, September 2018) to set up the transcription guidelines and corpus annotation criteria to be used in the transcription process. These guidelines were piloted on a lecture and revised in the fifth (Copenhagen, October 2018) and sixth (Lleida, February 2019) meetings. One of the most contested aspects in the guidelines was the annotation of mispronunciations, which this section explains.

The transcription guidelines (TAEC, 2020) which were adopted had proposed using: "standard orthography, even when words are *pronounced with a foreign accent* [emphasis added]" (p. 14). To tag what was referred to as mispronounced words, these words were supposed to be "transcribed in the standard form (when it is possible to understand the intended word)" (p. 15) with the <PRON> tag added before the mispronounced words and the </PRON> tag after it. Some TAEC members had proposed transcribing in a way that more accurately reflected the actual pronunciation, thus leaving the transcription reader to judge the accuracy of that pronunciation. If the lecturer produced a vowel before <s-> in *students*, for example, pronouncing three rather than two syllables, one could transcribe it as **estudents*, thus signalling the non-standard pronunciation, but without judging whether it is accented or inaccurate English. Another possibility was to transcribe it as <PRON> students </PRON>, tagging this as a mispronunciation. The final sub-group decision was to tag mispronunciations with <PRON> and ignore accented speech. However, there was no alignment training to better identify and agree on what falls within the realm of accent or within that of mispronunciation. The individual transcribers were given the responsibility to decide upon these aspects, and due to this lack of criteria unification, the results were inconsistent.

This issue was discussed at length in the seventh meeting (Torino, October 2019), using specific examples from the transcriptions of some lectures, and it was decided that the transcriptions would have to be revised by two more TAEC members, so these discrepancies could be reduced to the minimum. Three sections of each transcript were revised by these two TAEC members, who modified the transcript when they disagreed with what had been transcribed. In some cases, this entailed changing what had been transcribed, and on other occasions it meant adding new content to the transcript. It turned out that whereas there were few changes and additions when it came to content in general or pauses, there were numerous changes and additions when it came to the tagging of grammatical and pronunciation errors. The TAEC Corpus Report (TAEC, 2020) suggests that a "more fine-grained definition of 'pronunciation mistake'" (p. 6) would have made it easier for the taggers to carry out their task.

Although the amount of changes and additions did not go over the minimum standards established, it was finally agreed to remove the tags from the transcriptions.

In relation to the rating of the lecturers, it was stated in the instructions that each rater would rate all the videos individually without discussing the results with other raters and/or colleagues. Raters were asked to individually select 10–15 minutes of each lecture, contemplating both monologic and dialogic parts of the lecture, and to provide a holistic assessment, as well as to look at the following items, as defined by the sub-group: range, accuracy, fluency, interaction, addressing audiences, coherence, phonology, and mediation.

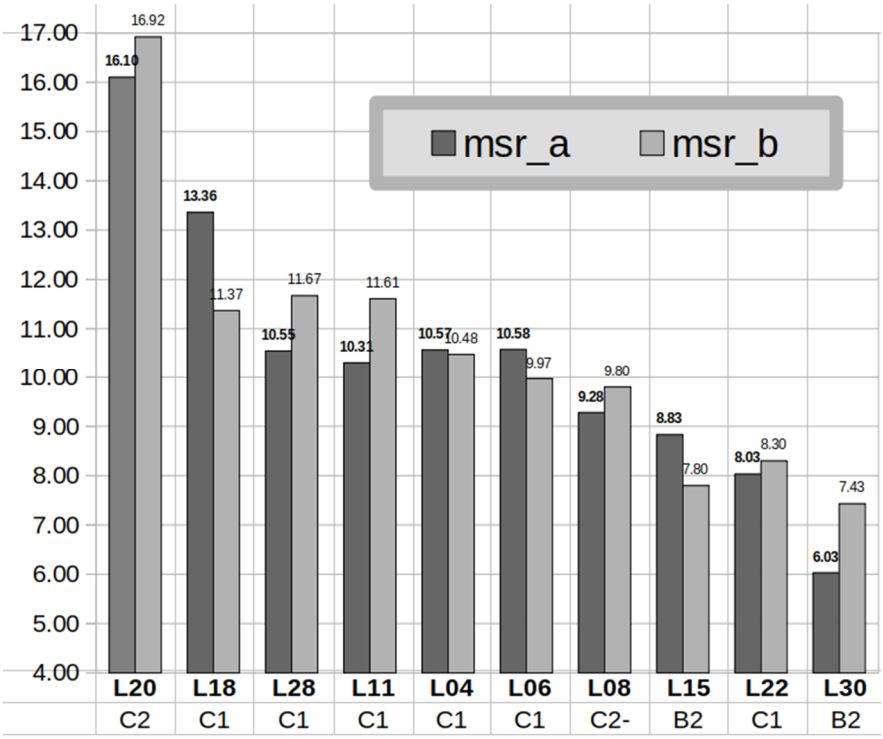
5 (Dis)fluency measures

As stated above, one sub-group of researchers selected portions of ten lectures on which to carry out (dis)fluency analyses. The speech samples used for analysis were taken from two parts of each lecture and each sample lasted 180 seconds. The first part (A moments) occurred within the first 10 minutes of the lecture, whereas the second part (B moments) occurred between minute 25 and minute 45 of the lecture. Only moments when lecturers were in lecturing mode were selected, defined as moments when lecturers were providing explanations, examples or definitions, rather than interacting with students.

In line with Martin-Rubió (2021), the (dis)fluency measures were calculated as follows. The two 180-second audio files from each lecture were opened in the software Audacity, and three kinds of chunks were identified: 1) between-pauses units (bp-units), i.e., stretches of syllables between pauses often referred to as runs in the literature; 2) filled pauses, i.e., sounds employed as hesitation elements irrespective of length; and 3) silent pauses, i.e., silent gaps of 0.25 seconds or longer.

Figure 1

MSR in A Moments and B Moments of the Ten Lectures

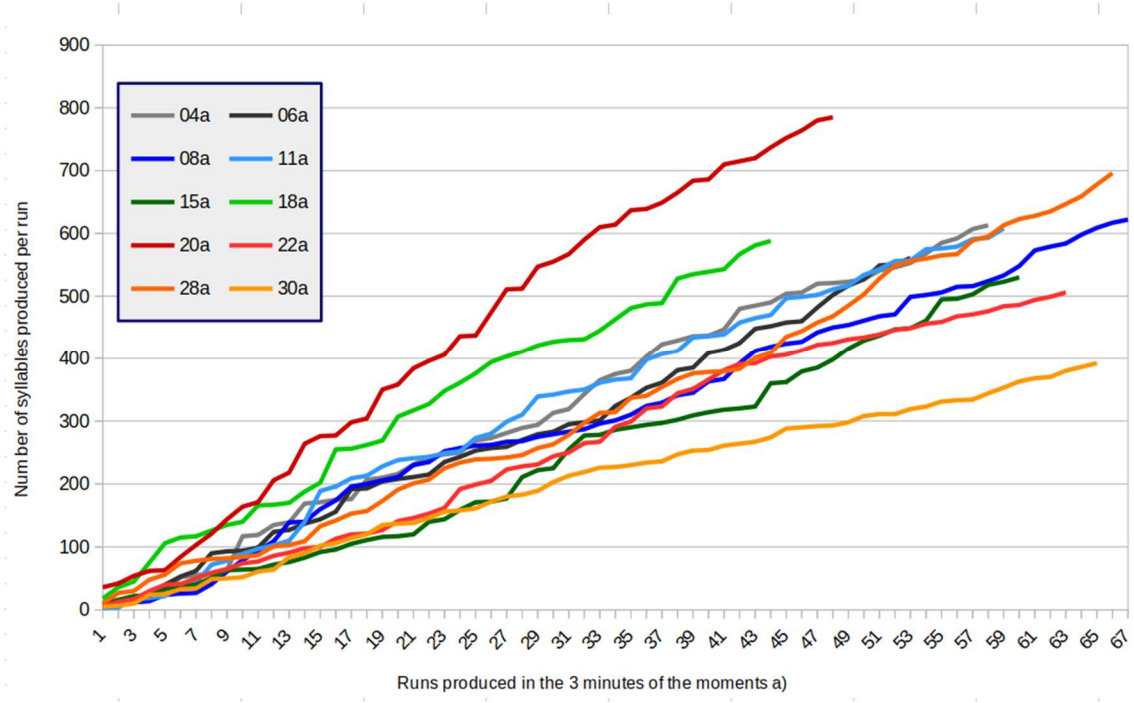


The measures used in the current analysis were taken from Ginther et al. (2010), and fall within three categories. The first category is Total Response Time, which is normally an important quantity measure. It is always 180 seconds in this case, so the focus is rather on the total number of syllables produced in those 180 seconds (see Appendix). The second category of measures includes three fluency measures: 1) Mean Syllables per Run (MSR) represents the average numbers of syllables produced between pauses (both filled and silent) and is calculated by dividing the total number of syllables by the number of runs; 2) Rate of Speech Time (ROST) measures the speed at which syllables are delivered and is calculated by dividing the number of syllables by the Speech Time; and 3) fluency ratio measures: Speech Time Ratio (STR), Silent Pause Time Ratio (SPTR), and Filled Pause Time Ratio (FPTR) indicate the proportion of time (in percentages) which the lecturer spends delivering meaningful syllables, silent-pausing or filled-pausing. Finally, the third category of measures includes two disfluency measures, x/100 syllables: 1) repetitions; and 2) false-starts and self-corrections (see Appendix for the main fluency and disfluency measures, together with the CEFR level and total number of syllables). This paper will now focus on the MSR measures, to show how one fluency measure is not sufficient to determine proficiency level.

The distribution of MSR in moments A and B, based on the CEFR level of the 10 lecturers, is shown in Figure 1 above. Lecturer 20 (L20), who is at C2 level according to the TAEC group's rating, has the highest MSR score (stretches of more than 16 syllables on average). He is followed by five lecturers at C1 level, and L15 and L30, who are at B2 level. Only L08 and L22 seem to diverge from the general alignment because, although L08 is at C2- level, his MSRs are lower than most lecturers at C1 level. L22's MSRs, on the other hand, are lower than one of the B2 level performances although he is at a C1 level.

Figure 2

Syllables and Runs for A Moments in the Lecture

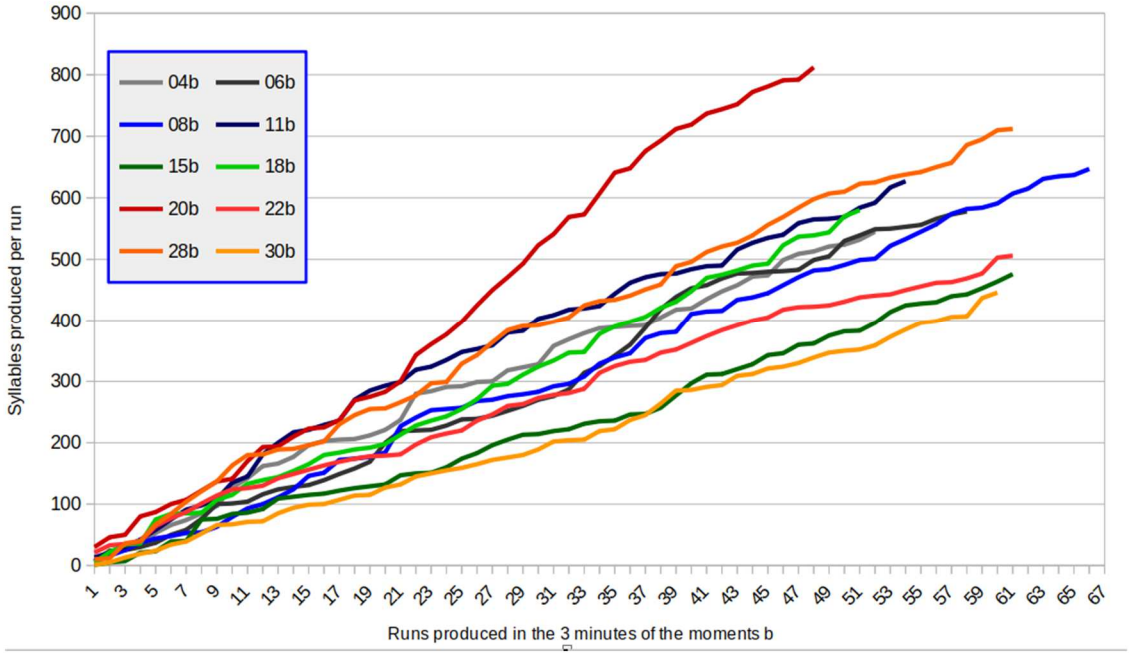


Figures 2 (above) and 3 (blow) represent the progression of the production of syllables throughout the 180 seconds in each of the two speech samples (A and B moments). In these figures, the cumulative number of syllables appear on the *y*-axis, whereas the runs appear on the *x*-axis. For instance, the A moments of lecturers L15 and L30 (Figure 2) start with a similar pattern. However, between runs 27 and 32, L15's production of syllable increases more rapidly (see Figure 2; the dark green line 15a and the light orange line 30a). The production flattens out for a few runs, but it goes up again in runs 44–45 and in runs 56–57.

In the end, L15 produces 530 syllables in 60 runs (MSR of 8.83 syllables/run), whereas L30 maintains the same pace and ends up producing 392 syllables in 65 runs (MSR of 6.03 syllables per run). In the B moments of these two lecturers (see Figure 3), though, both follow a very similar pace and in fact end up producing almost the same number of syllables in the same number of runs, L15 produces 476 syllables in 61 runs (7.80 syllables per run) and L30 produces 446 syllables in 60 runs (7.43 syllables per run).

Figure 3

Syllables and Runs for B Moments in the Lecture



Overall, the three fluency measures align well with the CEFR ratings. The results suggest that one fluency measure may not be sufficient to understand the role of fluency in lecturers' proficiency. For example, MSR is considered an important indicator of proficiency because of the assumption that lower proficiency level speakers cannot produce many syllables between two pauses, because their speech production is not automatised. In other words, they need to pause often in order to retrieve and articulate the necessary linguistic structures. If only MSR is taken into consideration, then L08's assessment at C2 level seems inconsistent with his MSR, which seems lower than most lecturers rated at C1 level. However, when ROST is considered, then it becomes apparent that he is able to produce more utterances within the same time slot than the lecturers at C1 level. The speed of his production suggests that his speech is automatised, and he does not need time to retrieve the linguistic structures he needs to express himself. One may hypothesise that he uses frequent pauses to allow the listeners to process

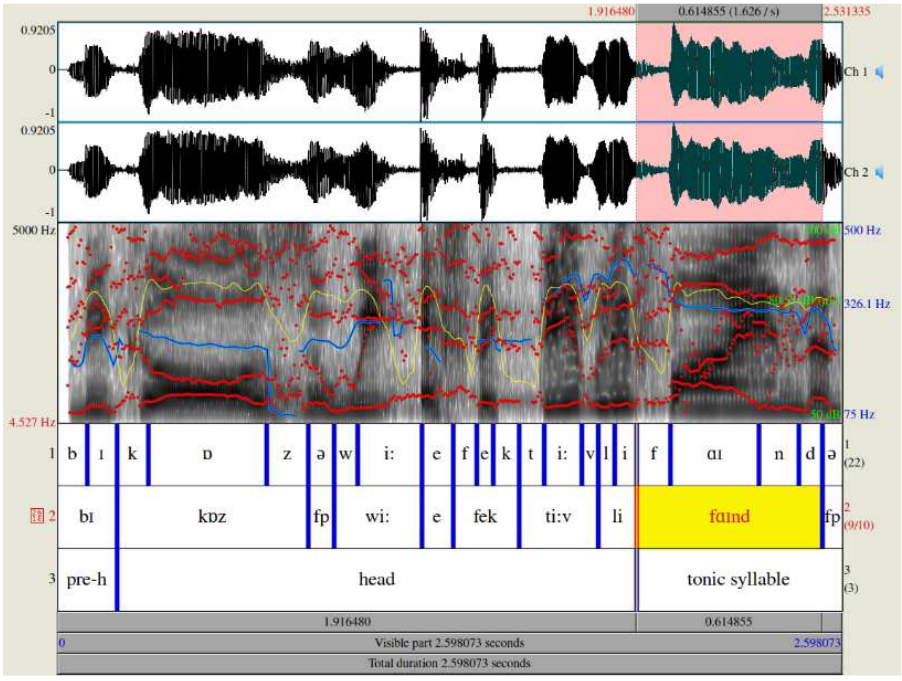
information before he proceeds. Therefore, it is recommended that fluency analyses of EMI lecturers include more than one fluency variable.

6 Lecturer 28 and epenthetic vowels

In this section, the A and B moments of L28 are looked at in greater detail. Excerpts from this transcription were discussed in the seventh TAEC meeting. It had been agreed that the TAEC members of one institution would transcribe the lectures from the same university, which meant that researchers would transcribe lecturers with accents that are familiar to them. Lecturer 28 was rated with a C1 level, and she had the second highest ROST, only behind the lecturer rated C2. However, there was a feature of her speech that had an impact on how these measures were to be calculated. She very often produced epenthetic vowels after final consonants. Figure 6 shows these two extra sounds after the consonants /z/ and /d/ in the words /bɪ'kɒz.ə/ and /'faɪn.də/. As Duguid (2001) explains, final consonants are rare in Italian, so they are often given "a following vowel, usually a *schwa*" (p. 76). More generally, "the stress-timed patterns of English cause great difficulty to Italian learners"; as a result, "[Italian] learners will expect full value to be given to all syllables" (p. 77).

Figure 4

Extra Vocalic Sounds in L28's Lecture



When this feature was brought up in the seventh TAEC meeting, and it was suggested that this could be considered a mispronunciation, the Italian TAEC members objected. This was simply a feature of the Italian accent when speaking English, they argued. However, not all Italians add vocalic sounds in this fashion when speaking English, for example the Italian TAEC members present in the meeting. These extra vowels, irrespective of the reason why the Italian speaker was resorting to them, were de facto acting as filled pauses. Once it was agreed this could never be considered an error, it was decided that they could, however, be regarded

as filled pauses. The measures provided in the Appendix were calculated with this criterion in mind — filled pauses, not errors. To summarise, data coding was impacted by different ideas of how to categorise this epenthetic vowel — whether as an error or as a feature of accented speech. This in turn reflects the influence of listener background (see Cutler, 2000; Kang et al., 2019).

7 Discussion

The goal of this study was to analyse the use and interpretation of (dis)fluency measures for ten EMI lecturers across Europe, as well as to examine the pronunciation accuracy of one lecturer. The (dis)fluency analysis shows that, overall, all three fluency measures (MSR, Ratios and ROST) align only partially with the CEFR proficiency levels assigned by the raters, who assessed the EMI lecturers' performance based on ten items, including fluency and language accuracy. Accent was not supposed to be an item in the rating, but L28 is categorised as B2 level in the phonology item, whereas she is rated C1 or C2 for all the other items. She shows some deviations in vowel and consonant pronunciation, and many filled-pauses. Her pronunciation is a crucial aspect of her teaching, given that 20% of her students were international, (from Eastern European, Asian and African countries); her Italian-accented English will need to be intelligible to these students, not just to her Italian students.

Jensen and Thøgersen (2017) found that although accent has been shown to have little impact on intelligibility in simple tasks, accented-speech could pose difficulties for students attending an EMI lecture, as the cognitive load of lectures can already be quite high. Valcke and Pavón (2015) state that "the ability to pass on academic content effectively greatly depends on the use of proper communication strategies, among which adequate pronunciation for the achievement of comprehension" (p. 336). They support this claim by giving examples of students' comments who point out that they find it difficult to understand their lecturer and cannot focus on understanding the content of the lecture. Similar results were obtained by Yildiz et al. (2017), as students reported difficulty understanding their lecturer because of pronunciation issues.

The findings from this study highlight the tension between requiring pronunciation to be either intelligible or at least minimally accented — which could be interpreted as requiring nativelike pronunciation (see Levis, 2005). When deciding whether intelligibility is an appropriate goal in a specific EMI context, the circumstances surrounding interactions should be taken into account. As Hynninen and Solin (2018) explained, scenarios in which English is used as a lingua franca differ. Tourists ordering a meal in a restaurant abroad differs significantly from an EMI context: in the latter, English is being employed by a content lecturer specialised in a field who is providing students with complex explanations, definitions and using terminology that is new for the students, who sometimes have different L1s and English levels. In this regard, Mauranen (2011) distinguished 'users' from 'learners', and although it is obvious that all learners are users and that learning is a lifelong process, the distinction can be helpful to separate these two realities. If one's priority is to successfully complete a transaction (say, agree on what to have for dinner), then using a variety of ways to refer to a type of food might suffice/be effective. However, when one's goal is to learn a new disciplinary concept for future use, then the type of pronunciation through which one accesses the concept might be highly significant. The insights provided here about how lecturers' pronunciation is assessed, should inform training courses, always bearing in mind that this is just one of the many elements lecturers must bring to their teaching.

References

Council of Europe (2018). *Common European Framework of Reference for Languages: Learning, teaching, assessment companion volume with new descriptors*. Strasbourg: Language Policy Division.

Cutler, A. (2000). Listening to a second language through the ears of a first. *Interpreting*, *5*(1), 1–23. https://doi.org/10.1075/intp.5.1.02cut

Dimova, S. (2018). Linking the TOEPAS with the CEFR: Technical report. https://cip.ku.dk/english/documents/Alignment_of_TOEPAS_with_the_CEFR.pdf

Duguid, A. (2001). Italian speakers. In M. Swan & B. Smith (Eds.) *Learner English: A teacher's guide to interference and other problems* (pp. 73–99). Cambridge University Press.

Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing, 27*(3), 379–399. https://doi.org/10.1177/0265532210364407

Halbach, A., & Lázaro, A. (2015). *La acreditación del nivel de lengua inglesa en las universidades españolas: Actualización 2015*. British Council.

Hynninen, N., & Solin, A. (2018). Language norms in ELF. In J. Jenkins, W. Baker, & and M. Dewey (Eds.) *The Routledge handbook of English as a lingua franca* (pp. 267–278). Routledge.

Jensen, C., & Thøgersen, J. (2017). Foreign accent, cognitive load and intelligibility of EMI lectures. *Nordic Journal of English Studies*, *16*(3), 107–137. https://doi.org/10.35360/njes.414

Kang, O., Thomson, R., & Moran, M. (2019). The effects of international accents and shared first language on listening comprehension tests. *TESOL Quarterly, 53*(1), 56-81. https://doi.org/10.1002/tesq.463

Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching, *TESOL Quarterly*, *39*(3), 369–377.

Macaro, E. (2018). *English medium instruction*. Oxford University Press.

Martin-Rubió, X. (2021). Audacity and Praat as pedagogical tools: Analysing fluency and pronunciation accuracy, *Philologia*, *19*(1), 37–50. https://doi.org/10.18485/philologia.2021.19.19.4

Mauranen, A. (2011). Learners and users - Who do we want data from?. In F. Meunier, S. de Cock, G. Gilquin, & M. Paquot (Eds.), *A taste for corpora: In honor of Sylviane Granger* (pp. 155–171). John Benjamins.

O'Dowd, R. (2018). The training and accreditation of teachers for English medium instruction: An overview of practice in European universities. *International Journal of Bilingual Education and Bilingualism, 21*(5), 553–563. https://doi.org/10.1080/13670050.2018.1491945

TAEC – Transnational Alignment of English Competences for University Lecturers (2020). TAEC Corpus Report. http://taec.ffri.uniri.hr/assets/files/TAEC%20Corpus%20Report_FINAL.pdf

Valcke, J., and Pavón, V. (2015). Use of pronunciation strategies. In R. Wilkinson & M. L. Walsh (Eds.), *Integrating content and language in higher education: From theory to practice* (pp. 323–341). Peter Lang.

Wächter, B., & Maiworm, F. (2014). *English-taught programmes in European higher education: The state of play in 2014*. Lemmens.

Yildiz, M., Soruç, A., & Griffiths, C. (2017). Challenges and needs of students in the EMI (English as a medium of instruction) classroom. *Konin Language Studies, 5*(4), 387–402. https://doi.org/10.30438/ksj.2017.5.4.1

Appendix

The Main (Dis)fluency Measures per Lecturer, at A Moments and B Moments

L	syllables		level	MSR		ROST		STR		repetitions		FS/SCs	
	A	B		A	B	A	B	A	B	A	B	A	B
L04	613	545	C1	10.57	10.48	4.40	3.93	77.10	76.21	1.14	2.02	1.47	2.02
L06	561	578	C1	10.58	9.97	3.75	3.94	82.30	79.82	0.18	0.35	0.53	0.52
L08	622	647	C2-	9.28	9.80	4.41	4.71	78.14	81.88	0.80	0.93	0.48	0.46
L11	608	627	C1	10.31	11.61	4.12	4.39	81.23	78.23	0.66	0.48	1.32	1.12
L15	530	476	B2	8.83	7.80	3.81	3.52	76.21	73.71	2.26	2.73	2.08	1.47
L18	588	580	C1	13.36	11.37	3.88	4.13	82.98	77.96	1.87	1.55	0.85	1.72
L20	773	812	C2	16.10	16.92	4.96	5.14	86.40	87.37	0.78	0.99	0.52	0.37
L22	506	506	C1	8.03	8.30	4.28	4.13	65.65	67.64	0.40	0	1.13	0.40
L28	696	712	C1	10.55	11.67	4.57	4.66	83.34	84.20	1.01	0.28	0.86	0.98
L30	392	446	B2	6.03	7.43	2.92	3.31	74.12	74.64	1.28	1.12	1.79	2.02

Note. L = Lecturer

About the author

Xavier Martin-Rubió holds bachelor degrees in English Philology and Audiovisual Communication from Universitat de Lleida, and a Master's degree in European Studies from Maastricht Universiteit and University of Portsmouth. He defended his doctoral dissertation in April 2011. He worked as an English teacher in a high school, two official language academies and two private universities before moving back to Universitat de Lleida. He started working full-time at UdL in December 2014. He has been involved in four competitive projects as a member of CLA, and after a gap of several years, he is publishing again on EMI in higher education, beliefs and emotions in additional language learning, and fluency and accuracy measurements across tasks.

Email: xavier.martinrubio@udl.cat

Messum, P & Young, R. (2023). L2 pronunciation: Seven learning/teaching paradigms found in instructed learning. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 160–171). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225633

L2 pronunciation: Seven learning/teaching paradigms found in instructed learning

Piers Messum
Pronunciation Science Ltd.

Roslyn Young
Pronunciation Science Ltd.

For good production in L2 pronunciation, learners have to be able to form new sounds in many syllabic contexts and at speed, so pronouncing an L2 requires learning new motor skills. The basic unit of motor skill development is the action-perception cycle, within which there is a need for the learners' attempts at a target to be evaluated, by themselves or by an expert. We present a theoretical framework based upon how learning/teaching paradigms meet the need for such evaluation. This yields a taxonomy of seven paradigms.

Keywords: L2 pronunciation learning, taxonomy, L2 speech sound production, motor skill

1 Introduction

For many years, we have been concerned that classroom-based pronunciation teaching has been developed, practised, and researched without sufficient prior investigation into how learners learn to pronounce an L2. Others have made similar points. For example, Foote and Trofimovich (2018) describe a lack of theory to guide L2 pronunciation research as being one of the most acute problems in the field. They explore how linguistic, psychological, interactionist, sociocultural, identity and sociocognitive perspectives can be useful for research, but acknowledge firstly that these have little to contribute to L2 pedagogy and secondly that this is a significant shortcoming, because, "research in L2 pronunciation should ultimately inform pronunciation teaching" (p. 85). For theory to support pronunciation teaching, it seems clear that it should address how learners learn to pronounce an L2.

This article starts to address one aspect of this issue: learning to produce L2 speech sounds which are not present in the L1 inventory (e.g., the production of English /l/ and /r/ by Japanese speakers, or French /y/ by English speakers). When we refer to learning new L2 speech sounds, we include learning new sounds in different contexts, not just in isolation (i.e., the learner producing sounds, sounds within clusters, sounds in words and then in phrases). Pronouncing an L2 sound, therefore, requires developing new motor skills.

It is obvious that one should try to understand how something is learnt before teaching it, so why is the mechanism used in learning to pronounce L2 so rarely discussed by researchers? Presumably the issue has seemed to be unproblematic: teachers and researchers know that some forms of vocal learning in speech and singing can be done by imitation (by which they mean a self-evaluated auditory matching-to-target process) and assume that this is possible for L2 speech sounds. They may think that children learn L1 pronunciation in this way, and that we can invoke the same mechanism in older learners, albeit with teaching practices that are adapted to take account of differences between the two age groups (summarised by Strevens, 1974).

The field in general, both researchers and teachers, has thus assumed that learners will develop L2 pronunciation if they are presented with exemplars to be copied. In so-called Intuitive-Imitative (I-I) approaches (Celce-Murcia et al., 2010; Kelly, 1969), this is imagined to be sufficient. In Analytic-Linguistic (A-L) approaches, this stage is preceded by training to improve how learners hear L2 and/or with phonetic information and rules which, it is believed, will become know-hows (automatised procedural knowledge) through repeated application.

We have identified six other ways in which it has been or could be imagined that L2 pronunciation is learnt/taught. To create this taxonomy, we approached the problem from two directions. First, we asked, "Since L2 pronunciation is a socially transmitted motor skill, what are the implications for how it is learnt?", and then, "L2 pronunciation is taught in different ways; what are the learning mechanisms that can be inferred to underlie each of them?".

We start by outlining some conceptual points drawn from psychology and other disciplines which inform our approach (see §2). We describe the basis of our taxonomy (see §3), and then we describe the different learning mechanisms it identifies and the teaching practices which are based upon them (see §4–6).

2 Conceptual points for understanding pronunciation learning

Messum and Young (2021) gave reasons why learning to pronounce an L2 is "a highly unusual activity" (p. 170) among socially transmitted skills, including the fact that the actions involved are largely invisible and that production and perception skills developed for L1 pronunciation interfere with the process. Furthermore, we argue that five conceptual points need to be taken into consideration if one is to think clearly about learning to pronounce an L2. These are presented one by one in the following sub-sections.

2.1 Pronunciation learning is not imitation

When teachers provide a spoken model for their learners in pronunciation classes, they are presenting them with the results of their actions and not the actions themselves, since most of the actions involved are hidden inside the mouth. In this way, teaching pronunciation is unlike teaching most skills, where a model shows the learner what the demonstrator is doing as well as the result. Metaphorically speaking, if a winner of the Masters were teaching golfers how to drive a ball by simply striking 300-yard shots off the tee, we do not think they would learn as much as if he gave them advice on how to improve their own swings. But at least they might pick up something from watching him in action. Now imagine if he were hitting those 300-yard drives while standing behind a tarpaulin so that the golfers could not see what he was doing and could only see the result: a ball sailing down the fairway every time he produced a 'model'. Then they would get very little, if anything, from the experience.

In technical discussions, *imitation* refers to copying of actions as well as reproducing the demonstrator's results. For this reason, the colloquial use of the word imitation for the result of a self-evaluated matching-to-target process in L2 pronunciation is inappropriate. The field should be using the word *emulation*, or more strictly, *goal emulation*: the adoption of the goals of the demonstrator, the reproduction of his/her results but not the copying of his/her actions (Call & Carpenter, 2002; Whiten & Ham, 1992). It is important for us to use terminology that acknowledges the complexity of the process of learning L2 speech sounds. In imitation, there are two sources of information that the learner can attend to at will: the actions of the model and the results obtained. In emulation, there is only one source, the results, and for L2 speech these are an acoustic signal whose interpretation by most learners is distorted by the mechanism of their L1 perception.

2.2 Two ways of listening: Autocentric and allocentric

There has been a longstanding scientific interest in the two products of our senses, sensation and perception, and in their relation. Humphrey (1992) opens his discussion of this by quoting Reid (1785): "The external senses have a double province – to make us feel, and to make us perceive" (p. 46).

In the auditory domain, Öhman (1975) gives the following example: we might experience a household event either as the sound of a refrigerator door shutting in another room or as meaningless "concrete music", that is not about the state of the world but is "an immediate awareness of the developing states of [our] auditory sense" (p. 42). Similarly, when listening to a speech sound, it is possible to attend to it in two different ways: as something meaningful or as noise. For instance, we might experience an event as a linguistic /p/ or as the meaningless noise of an explosion of air from the mouth. Since the first way of attending to sounds gives us potentially useful information about the outside world, it is the attentional set that we use almost all the time. Indeed, when meaning is available from a signal, it is often difficult to attend to it as noise.

In research on speech perception, various distinctions have been drawn between the attentional sets of listening, and a variety of terms have been used to describe them. There have even been proposals for three-factor models of speech perception, with auditory, phonetic, and phonemic components (Werker & Logan, 1985). However, to describe the particular distinction that Öhman (1975) drew and to avoid any possible confusion with other distinctions that have been made, we will use the terms proposed by Schachtel (1959), i.e., *autocentric* for awareness of sensations and *allocentric* for meaningful perception. These terms capture the point that the experience of mere noise is an awareness of the state of our own auditory sense

(autocentric mode) while experiencing meaning is an awareness of the state of the outside world (allocentric mode).

It is necessary to make this distinction when considering pronunciation because speech sounds (instances of phonemes, or strings of phonemes) are noises to which we give linguistic significance in speech. When we learn a new L2 sound we are learning to produce a noise (or set of noises) that will be recognised as that particular sound by other speakers of the L2.

2.3 Two forms of memory: ASM and ALTM

Current models of memory distinguish short-term and long-term memory, and posit at least two components of short-term memory that deal with sound (Scott & Mishkin, 2016). The first is a phonological store that can be supplemented with subvocal rehearsal to sustain speech representations. However, this is not of any obvious use in learning to produce new L2 sounds.

The second is auditory sensory memory (ASM), which Nees (2016) describes as "a set of acoustic features organized in time that can be consulted to complete behavioural tasks, including comparing sounds to one another" (p. 1). There is a panoply of other names in use: the sensory register, echoic memory, acoustic short-term memory, passive short-term memory, pre-perceptual auditory memory, and so on. ASM is reported as decaying rapidly, usually within a few seconds (Nees, 2016).

Auditory long-term memory (ALTM) may also be relevant to learning new L2 sounds. As it is usually conceived, it does not store raw, sensory intake (noises). However, since selective attention is considered to isolate information in perception and then store it in long-term memory (Barsalou 1999), there does not seem to be a principled reason against any detail at any level being captured in ALTM (e.g., a novel feature of an L2 sound), provided the learner has noticed it.

2.4 Perceptual representations are not production representations

One determiner for what is captured in long-term memory is the current task, since a task is a driver for selective attention. If the task is to learn to identify or discriminate any type of object, then the abstraction will be optimised for that purpose, but not necessarily structured for recall or to guide subsequent production. Perceptual training in L2 pronunciation is usually evaluated by the success of learners in identifying or discriminating new L2 sounds. However, this cannot be taken to mean that they have developed a type of representation that can directly inform production.

To better appreciate this point, we suggest you stop reading and, from memory, draw the Ford Motor Company logo, trying in particular to get the style of the lettering right. Despite your undoubted ability to recognise this logo (which you have probably seen thousands of times), and to discriminate it from similarly coloured or shaped logos, you would be unusual if your production of the <f> or <r> matches the detail of the original. In any perceptual modality, to be able to recognise something does not mean one knows how to produce it.

2.5 Mirroring: Learning through others

Before video recording technology was available, springboard divers had to rely upon a coach to evaluate their performance and hence to improve. The coach acted as a form of mirror to the learner, and psychology has demonstrated many ways in which we learn about ourselves through mirroring interactions. Moreover, it is very possible that children's L1 pronunciation develops this way, during vocal exchanges in which their caregivers imitate their babbling, reformulating it into well-formed L1 speech sounds (Messum & Howard, 2015). Munhall et

al. (2021) lend support to this hypothesis, considering it to be consistent with current thinking on how speech perception develops in children.

Maas et al. (2008) use the terminology of Schmidt's (1975) schema theory (discussed below) to describe this mechanism:

> Before the recognition schema can be used to judge the accuracy of the movement, the system must first learn which sensory consequences are to be considered "correct." There is often a clear reference of correctness (e.g., a golf ball must end up in the hole), but there are cases in which the reference of correctness is not directly available or interpretable to the learner but instead depends on feedback from an instructor, such as when learning to perform a somersault in diving. In such cases, the learner must calibrate the expected sensory consequences with an externally provided reference of correctness, so that the internal error signal may serve to correct errors on future trials without external feedback. (p. 279)

The reference of correctness for L2 speech sounds may be neither directly available to learners nor interpretable by them: *unavailable* because of the perceptual opacity of speech sounds (Heyes & Ray, 2000), with the learners hearing their own voice mixed in with bone-conducted sound and hearing what they expect themselves to be producing rather than the actual output (Munhall et al., 2021); and *uninterpretable* because of the effect of their L1 listening expertise on how they hear L2. In these circumstances, an externally provided reference of correctness becomes necessary.

3 A taxonomy of L2 pronunciation learning/teaching paradigms

Now that these conceptual points have been made, we present seven mechanisms for how the production of L2 speech sounds might be developed through conventional learning, with a note in most cases of an associated teaching practice or practices (Figure 1). A more comprehensive taxonomy will include learning mechanisms proposed in the literature that invoke innate, neural capacities (Messum & Young, in preparation).

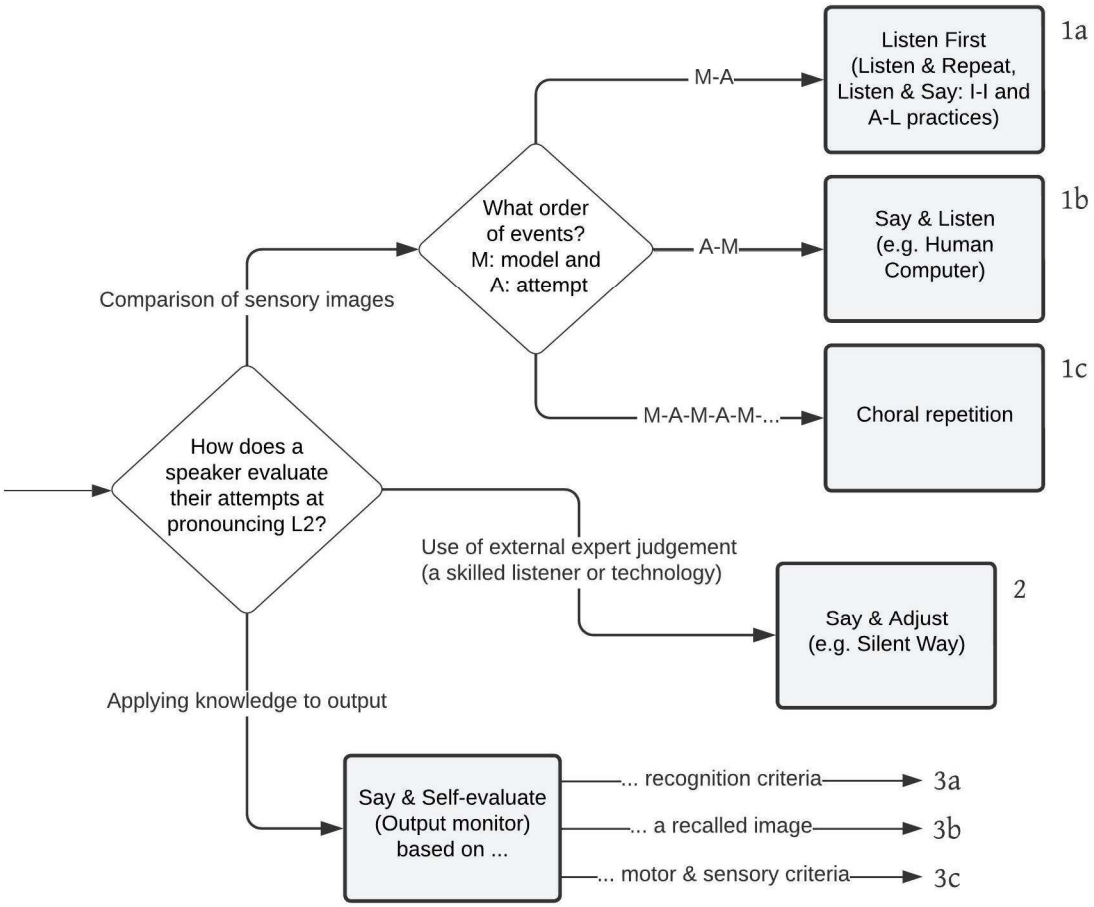
Before discussing each mechanism in turn, we first explain why our main taxonomic question is how a speaker evaluates their attempts at L2 pronunciation.

While the learning and teaching of L2 pronunciation can be considered from various perspectives, there should be a common aim: that learners pronounce L2 proficiently. Learners have to develop new willed actions — do something different with themselves (i.e., with their articulatory apparatus) — if they are to pronounce L2 differently from how they pronounce L1. Pronouncing is a motor skill. With this starting point, we can broaden our understanding of L2 pronunciation learning by reference to the established body of research on motor skill development.

The foundational theories in this field — Closed-loop Motor Learning (Adams, 1971, 1987) and Schema Theory (Schmidt, 1975; Schmidt & Lee, 2005) — agree that action and evaluation are two key elements in motor learning. These inform the memory trace and perceptual trace respectively in Adams' theory, and the recall schema and recognition schema in Schmidt's. In repeated action-perception (A-P) cycles (Cutsuridis et al., 2011), learners develop sensorimotor contingencies between their actions and the results these produce. The action part of the cycle relies on the learner having some control over what they do: the ability to do the same thing again or to do something new. In the perception part of the cycle the learner evaluates the results of the action judged against the target. This evaluation can be performed either by the learner, or by a competent other person who communicates their evaluation.

Figure 1

A Taxonomy of Learning/Teaching Paradigms in Instructed L2 Pronunciation



Note. The labels within the numbered boxes describe what a learner does in each case, with the name of an associated teaching practice noted if appropriate. Abbreviations: I-I, Intuitive-Imitative; A-L, Analytic-Linguistic. M, Model; A, Attempt. (Thus 'M-A' means 'model followed by learner attempt'.)

Before the first attempt is made, the learner has a (perhaps unspoken) question, "How can I do this?". After the attempt, they need to answer the question, "How successful was that?". If they continue, they then have to ask, "What am I going to do now?" and, after the next attempt, "Was that better or worse than last time?" and "What difference did the new thing I tried make?".

To establish the possible learning mechanisms for L2 pronunciation, we focus on the perception/evaluation side of the A-P cycle. Thus, our taxonomic question on the left of Figure 1 asks how the learner evaluates their attempts at pronouncing an L2. The possible ways in which this could be done are documented on the three branches which lead away from the central question and which describe the learner's mental activity:

1. The learner compares *sensory* images: the image they retrieve from their own attempt that they compare to a model they hear. One (or perhaps both) of these would be held in ASM.
2. The learner is informed about their attempt from another's evaluation of it — be it from a teacher, another speaker of the L2, or some form of technology.
3. The learner evaluates their attempt using *knowledge* previously acquired. This knowledge might be: a) auditory criteria for correctness for the L2 sound; b) a fine-grained exemplar of the sound in ALTM that they use for comparison; or c) motor and sensory (proprioceptive) criteria for correctness. The knowledge needed for a) and b) would have been acquired through listening; that for c) would have been acquired from previous cycles of A-P learning.

Note that any actual classroom interaction might give the learner more than one source of information about their performance, but the sources will be drawn from these three basic categories.

The final termination points for the three branches from the central question in Figure 1 are seven learning/teaching paradigms for L2 pronunciation, labelled 1a to 3c. The boxes describe what the learner does in each case, with the name of an associated teaching practice noted if appropriate. The next sections (§4–6) explain the paradigms resulting from each of the three branches.

4 Paradigms 1a, 1b, and 1c: Comparison of sensory images

4.1 Listen First: Intuitive-Imitative and Analytic-Linguistic approaches (Paradigm 1a)

If the learner evaluates their performance based on a comparison of sensory images (the aural image created by their attempt at a sound and an aural image they hear as a model) then the most familiar classroom order of such events (and the usual order of such events in language learning apps on mobile phones) is that of Listen & Repeat: a model followed by a learner attempt. This is the exercise that best characterises Intuitive-Imitative (I-I) pronunciation teaching.

However, learners complain that they cannot *hear* the model, i.e., that they are not aware that the model L2 token is different from some similar sound in L1. Nothing can be achieved by asking them to copy what they do hear in these circumstances, so teachers seek ways to remedy this. Hence Analytic-Linguistic (A-L) approaches supplement the supposed natural mechanism of imitation with phonetic information and/or preparatory perceptual training. A-L practices are considered to complement I-I practices rather than replace them (Celce-Murcia et al., 2010). However, to acknowledge that the learner's response after A-L preparation is likely to be more carefully considered than what they can attempt in simple Listen & Repeat, we call it Listen & Say.

Whether or not there is a preparatory phase, notice that the principal learning move for production in both I-I and A-L approaches is for the learners to match an L2 sound which has been presented to them. These two Listen First approaches aim at a learner's ASM, through autocentric mode listening. Most current classroom practice is therefore based within this single paradigm in our taxonomy.

4.2 Say & Listen (Paradigm 1b)

It is possible to reverse the order of events in Listen First approaches, i.e., for the model to follow the attempt. For example, in the Human Computer technique from Community Language Learning (Curran, 1976), the teacher repeats the learner's sound, word or phrase in their own, correct L2. The learner is then not allowed to repeat the phrase after the teacher.

This prohibition on repetition prevents reflexive copying where the learner is not aware of their speech motor activity, i.e., what they themselves are doing to pronounce. Instead, there is a period of silence, in which the learner can compare the image they have of their own production to what they have just heard. This gives them the opportunity to evaluate the former and consider what they might change about it. When enough time has passed for the teacher's utterance to have faded from the learner's ASM, the learner is allowed to say the phrase again, informed by what they learned from the previous cycle. The teacher will repeat the phrase again in good quality L2, and the learner has the opportunity to compare and contrast the two renditions.

4.3 Choral repetition (Paradigm 1c)

In choral repetition, the teacher's model comes neither before nor after the learner's attempt, but rather in both places. In repeated, rapid sequences of model – attempt – model – attempt, it becomes unclear which of the mechanisms in boxes 1a and 1b describes the learning episode. Choral repetition of this type could involve either of them, depending upon how the learner chooses to direct their attention, potentially switching between the two more than once during any session.

Kjellin (2004) describes his rationale for choral practice led by the teacher, and his own classroom methodology that implements this. Based upon extensive classroom experience, he insists that 10, 20 or 30 repetitions of a phrase are inadequate and counterproductive, but that a block of 50-100 repetitions leads to success. Jones (2018) describes her own rationale and gives further advice on the use of choral repetition.

Immonem et al. (2022) report success training 7-year-old children to produce new L2 vowels using this learning/teaching paradigm: automatic, alternating presentation of two words containing the vowels with a fixed interstimulus interval of three seconds, over four training sessions of 30 pairs of repetitions. With similar training regimes, older children and adults have also been successful, but more slowly. It might be that younger children are better at attending to sounds as sensory objects and to be less deeply in the *grip* of L1 (Underhill, 2013).

5 Paradigm 2: Use of external expert judgement

In Paradigm 2, Say & Adjust, learners rely on the evaluation of their performance by an expert other: a metaphorical *mirror*. Within a mirror learning mechanism, it is acknowledged (and unproblematic) that the learners do not yet know the target: either how to create it, what it feels like or even what it sounds like. They only have some approximate idea that enables them to start production. With feedback from an expert, they can modify what they do, all the while getting to know the target better.

Sixty years of classroom practice has demonstrated the efficacy of such a process in L2 pronunciation teaching, starting with Gattegno's first version of the Silent Way and developed by him and other teachers since (Messum & Young, 2021). The paradigm need not be associated with the Silent Way. The empirical evidence presented by Warsi (2001) suggests that he was considerably more successful than Bradlow et al. (1997) in teaching the production of English /l/ and /r/ to Japanese learners through Say & Adjust. Furthermore, the source of

evaluation need not be human; technological evaluation and feedback can also be used (e.g., Kartushina et al., 2015; Sakai, 2016).

While the Say & Adjust paradigm is consciously deployed in Silent Way classes, it may also appear in conventional classes which seemingly employ a Listen First approach. When a teacher says, "Make a […]" and the learner takes what they hear to be a *cue* rather than a *model* (and this distinction is key), they ignore the possibility of copying what the teacher has said and produce their current best attempt at the sound. Then, when the teacher gives feedback on how well the learner has done, informing a new attempt, this is a Say & Adjust exchange.

6 Paradigms 3a, 3b, and 3c: Applying knowledge to output

Three other paradigms involve self-evaluation based upon knowledge acquired previously. This notion can also be found in the field of Child Phonology. For example, based on developmental data, Kuhl (2000) described one way in which young children may learn to produce L1 speech sounds: "… early in life, perceptual representations of speech are stored in memory. Subsequently, these representations guide the development of motor speech" (p. 11854).

An analogue of this for L2 was described by Bradlow et al. (1997): "… perceptual learning leads to more accurate internal acoustic representations of the target speech sounds, and these improved representations function as acoustic templates that play an important role in monitoring the articulatory output" (p. 2307). They called such templates *output monitors*.

A monitor of some type, operating to evaluate the speaker's own output, could rely upon any of three distinct types of previously acquired knowledge: recognition criteria, a recalled image, or motor and sensory criteria.

6.1 Self-evaluation based on recognition criteria (Paradigm 3a)

The speaker might have developed criteria to identify and/or discriminate L2 sounds produced by others (in isolation or in words) which they now apply to evaluate their own output. To return to the visual analogue involving the Ford logo, a novice illustrator trying to reproduce it without a model to copy might evaluate their attempt by wondering whether it looks right.

6.2 Self-evaluation based on a recalled image (Paradigm 3b)

The speaker might have developed an image of the L2 sound which is now stored within ALTM. They could compare their output against a version of this that they recall. With respect to the Ford logo, the novice in this case would be able to evoke (or recall) a detailed mental image they had created to guide their attempt.

6.3 Self-evaluation based on motor and sensory criteria (Paradigm 3c)

During previous work on the L2 sound, the speaker might have developed some or all of three different types of criteria that now help them during independent practice. They will know more about: 1) how (gesturally) to make the sound (i.e., what actions they need to perform); 2) how it feels to make the sound; and 3) how it sounds to them to make the sound.

These criteria alone do not allow them to exceed their best previous level of production accuracy, but they do allow them to consolidate their learning: producing the sound in different contexts, at different rates, with greater automaticity, etc.

Returning to the Ford logo, in this case the novice would have already drawn the logo several times, checking their attempts against an original, and hence developing facility and accuracy in its production. Now they are making a further attempt.

7 Conclusion

Our taxonomy is based on the characteristics of the perception side of the A-P cycle; specifically, the mechanism by which a learner evaluates their successive attempts. We have identified a number of possible evaluation mechanisms and the teaching approaches that rely upon them which, considered together, we call learning/teaching paradigms. This taxonomy will be expanded (Messum & Young, in preparation) to include paradigms that do not rely upon instruction, and we will also discuss the likely effectiveness of each of them.

As suggested by Pennington (2021): "… approaches that challenge standard practices on theoretical grounds deserve to be systematically investigated and their effects and effectiveness compared to those of the standard practices" (p. 17). The taxonomy presented here shows that there are plausible learning/teaching paradigms other than Listen First and technology-driven approaches. We support Pennington's call for researchers to investigate the effectiveness of these currently non-standard practices.

Acknowledgements

We would like to thank our two anonymous reviewers from EPIP 7, as well as Marc-Georges Nowicki, Alice Henderson, Dan Frost, Anastazija Kirkova-Naskova, Martha Pennington, Jonathan Marks, and Adrian Underhill.

References

Adams, J. A. (1971). A closed-loop theory of motor learning. *Journal of Motor Behavior*, *3*(2), 111–150. https://doi.org/10.1080/00222895.1971.10734898

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention and transfer of human motor skills. *Psychological Bulletin*, *101*(1), 41–74. https://doi.org/10.1037/0033-2909.101.1.41

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Bradlow, A. R., Pisoni, D., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*(4), 2299–2310. https://doi.org/10.1121/1.418276

Call, J., & Carpenter, M. (2002). Three sources of information in social learning. In K. Dautenhahn & C. L. Nehaniv (Eds.), *Imitation in animals and artifacts* (pp. 211–228). MIT Press.

Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.

Curran, C. (1976). *Counseling-learning in second languages*. Apple River Press.

Cutsuridis, V., Hussain, A., & Taylor, J. G. (Eds.). (2011). *Perception-action cycle: Models, architectures and hardware*. Springer.

Foote, J. A., & Trofimovich, P. (2018). Second language pronunciation learning. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation.* Routledge.

Heyes, C. M., & Ray, E. D. (2000). What is the significance of imitation in animals? *Advances in the Study of Behavior*, *29*, 215–245. https://doi.org/10.1016/S0065-3454(08)60106-0

Humphrey, N. (1992). *A History of the mind: Evolution and the birth of consciousness*. Simon & Schuster.

Immonen, K., Alku, P., & Peltola, M. S. (2022). Phonetic listen-and-repeat training alters 6–7-year-old children's non-native vowel contrast production after one training session. *Journal of Second Language Pronunciation*, *8*(1), 95–115. https://doi.org/10.1075/jslp.21005.imm

Jones, T. (2018). Repeat after me. *As We Speak (TESOL SPLIS Newsletter)*, *March*. http://newsmanager.commpartners.com/tesolsplis/print/2018-03-15/3.html

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, *138*(2), 817–832. https://doi.org/10.1121/1.4926561

Kelly, L. G. (1969). *25 Centuries of language teaching*. Newbury House.

Kjellin, O. (2004). *Choral Practice – The Neurophysiological Opportunist's Way*. https://www.academia.edu/2184625/Choral_Practice-the_Neurophysiological_Opportunists_Way

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences USA*, *97*(22), 11850–11857. https://doi.org/10.1073/pnas.97.22.11850

Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology*, *17*(3), 277–298. https://doi.org/10.1044/1058-0360(2008/025)

Messum, P. R., & Howard, I. S. (2015). Creating the cognitive form of phonological units: The speech sound correspondence problem in infancy could be solved by mirrored vocal interactions rather than by imitation. *Journal of Phonetics*, *53*, 125–140. https://doi.org/10.1016/j.wocn.2015.08.005

Messum, P., & Young, R. (in preparation). L2 pronunciation: a taxonomy and assessment of nine learning/teaching paradigms.

Messum, P., & Young, R. (2021). Teaching students to pronounce English: A motor skill approach in the classroom. *RELC Journal*, *52*(1), 169–178. https://doi.org/10.1177/0033688220964107

Munhall, K. G., Cui, A., O'Donoghue, E., Lamontagne, S., & Lutes, D. (2021). Perceptual control of speech. In J. S. Pardo, L. C. Nygaard, R. E. Remez, & D. B. Pisoni (Eds.), *The handbook of speech perception* (pp. 97–121). Wiley. https://doi.org/10.1002/9781119184096.ch4

Nees, M. A. (2016). Have we forgotten auditory sensory memory? Retention intervals in studies of nonverbal auditory working memory. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01892

Öhman, S. E. G. (1975). What is it that we perceive when we perceive speech? In A. Cohen & S. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 36–47). Springer.

Pennington, M. C. (2021). Teaching pronunciation: The state of the art 2021. *RELC Journal*, *52*(1), 3–21. https://doi.org/10.1177/00336882211002283

Reid, T. (1785). *Essays on the intellectual powers of man*. Longman.

Sakai, M. (2016). (Dis)connecting perception and production: Training adult native speakers of Spanish on the English /i/–/ɪ/ distinction [Unpublished doctoral dissertation]. Georgetown University.

Schachtel, E. (1959). *Metamorphosis: On the development of affect, perception, attention, and memory*. Basic Books.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, *82*(4), 225–260. https://doi.org/10.1037/h0076770

Schmidt, R. A., & Lee, T. D. (2005). *Motor control and learning: A behavioral emphasis* (4th ed.). Human Kinetics.

Scott, B. H., & Mishkin, M. (2016). Auditory short-term memory in the primate auditory cortex. *Brain Research*, *1640*, 264–277. https://doi.org/101016/j.brainres.2015.10.048

Strevens, P. (1974). A Rationale for Teaching pronunciation: The rival virtues of innocence and sophistication. *ELT Journal*, *28*(3), 182–189. https://doi.org/10.1093/elt/XXVIII.3.182

Underhill, A. (2013). Cinderella, integration and the pronunciation turn. *Speak Out!*, *49*, 4–8.

Warsi, J. (2001). Effects of visual instruction on second language productive phonology [Unpublished doctoral dissertation]. Boston University.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35–44. https://doi.org/10.3758/BF03207136

Whiten, A., & Ham, R. (1992). On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research. In P. J. B. Slater, C. Beer, M. Milinski, & J. S. Rosenblatt (Eds.), *Advances in the study of behavior,* Vol. 21 (pp. 239–283). Academic Press.

About the authors

Piers Messum has taught English in Japan, France, and the UK. His PhD in Phonetics from University College London investigated how children learn L1 pronunciation. He co-authored *How We Learn and How We Should be Taught* (2011) with Roslyn Young. He now teaches in London, and does teacher training and materials development for Pronunciation Science Ltd.

Email: p.messum@gmail.com

Roslyn Young worked at the University of Franche-Comté in the Applied Linguistics Centre from 1968 until she retired, teaching spoken English and sometimes spoken French in intensive courses. She ran her first 'pronunciation clinic' in 1978 and has worked continuously since then on ways of improving the teaching of pronunciation. She wrote her doctoral thesis on Caleb Gattegno's model of learning and its relevance to his work in language teaching. She has published numerous articles on the teaching of pronunciation and on how people learn.

Email: roslynyoung@gmail.com

Mora-Plaza, I. (2023). Task-based pronunciation teaching helps to improve L2 vowel production: Generalisation effects. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 172–187). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225325

Task-based pronunciation teaching helps to improve L2 vowel production: Generalisation effects

Ingrid Mora-Plaza
University of Barcelona

A number of task-based pronunciation teaching interventions have been shown to raise learners' awareness of relevant properties of L2 speech input during interaction (e.g., Solon et al., 2017). However, it still remains unclear whether pronunciation gains are generalisable to diverse lexical contexts (unfamiliar tokens) and elicitation modes (words vs. sentences). This study investigates whether a focus on phonetic form improves learners' production of English /iː – ɪ/ and /æ – ʌ/ and leads to generalisation effects and retention. Sixty-three L1 Catalan/Spanish EFL learners carried out 20 dyadic, problem-solving tasks over 7 weeks. Task completion required the distinction of the target lexical items (e.g., *bean – bin, cat – cut*). Gains in production and generalisation effects were assessed through delayed word and sentence repetition tasks, and Mahalanobis distances were measured between confusable vowels and between learners' and native speakers' productions. Results showed that not only did learners increase the qualitative distance between the target confusable vowels, but their L2 vowel productions also approximated those of native speakers at post-test in words produced in isolation and in sentences. In addition, gains generalised to untaught tokens and improvement was retained after 11 weeks.

Keywords: pronunciation instruction, task-based language teaching, English vowel production, generalisation effects, retention

1 Introduction

Learning foreign language (FL) phonology is an arduous task because input in formal FL contexts is generally insufficient, for instance, typically limited to a few hours of language-focused instruction per week (Muñoz, 2014) and often L1-accented. Consequently, L2 categorical perception, speech segmentation, and lexical activation and retrieval processes are inevitably affected by the already-automated L1 perceptual system. According to the Perceptual Assimilation Model (PAM-L2; Best & Tyler, 2007), L1-based perception causes difficulties in phonetic learning, especially when phonetically similar L2 sounds are perceptually mapped onto single L1 sound categories, making L2 sound contrasts confusable.

At the phonetic processing level, the English vowel contrast /iː – ɪ/ is very challenging for Catalan/Spanish speakers because they only have one single high vowel /i/ with no comparable tense–lax distinction. In PAM-L2 terms, Catalan/Spanish speakers assimilate English /iː/ and /ɪ/ (e.g., *sheep – ship*) to Catalan/Spanish /i/ via a category-goodness assimilation pattern. Thus, whilst the English vowel /iː/ has been found to be highly similar to the Catalan /i/, the English vowel /ɪ/ has been perceived as a poorer fit for the same L1 sound, being identified as Catalan /i/ or /e/ (Cebrian, 2021). The English vowel contrast /æ – ʌ/ is also difficult to perceive and produce for Catalan/Spanish speakers due to the presence of the Catalan low central vowel /a/ (Cebrian, 2021). PAM-L2 would classify it as a single-category assimilation, as English /æ/ and /ʌ/ are assimilated to Catalan /a/.

Despite the difficulty in developing learners' L2 phonological awareness in a FL context, directing learners' attention to phonetic form through various training and instructional techniques has proved effective in developing the speech perception and production of L2 learners with different proficiency levels and L2 experiences (Lee et al., 2015). Few studies have explored the role of tasks in generating a focus on phonetic form during interaction (Gurzynski-Weiss et al., 2017), as well as the extent to which L2 pronunciation gains generalise to different contexts and different speakers/voices and remain over time.

1.1 Attention to phonetic form

In order to acquire new speech sounds, it is indispensable for learners to notice and pay attention (Schmidt, 1990) to cross-linguistic differences between L1 and L2 phonologies. One way to achieve this is by explicitly instructing learners to attend to specific aspects of the speech input (Guion & Pederson, 2007) while ignoring others. Lab-based high-variability phonetic training (HVPT) can be used to raise learners' awareness of L2 phonology and its gains have been shown to be robust, generalising to new lexical items and speakers (Thomson, 2018).

Drawing attention to phonological form through explicit pronunciation instruction has also been found to be effective (Lee et al., 2015), as it helps learners notice the difference between their own productions and those of more proficient L2 speakers. However, several studies have found Focus on Form (FonF) instruction to be more effective than Focus on FormS (FonFS) instruction[1] in developing intelligibility, comprehensibility and L2 pronunciation accuracy (e.g., Darcy et al., 2021). Saito's (2012) synthesis of 15 quasi-experimental studies showed that, whereas FonFS interventions resulted in improvement only at a controlled level, FonF interventions enabled learners to improve at both controlled and spontaneous speech levels.

[1] Saito (2012) refers to FonF instruction when learners practise pronunciation form while being involved in contextualised meaning-oriented communicative activities, and FonFS instruction when learners are asked to practise the accurate use of pronunciation form via mechanical drills and choral repetition.

1.2 Task-based instruction and pronunciation

Task-based language teaching (TBLT), also known as task-based instruction (TBI), is a teaching approach that focuses on having learners complete meaningful tasks using authentic language input. It aims to draw their attention to a particular linguistic structure during interaction, with tasks specifically designed to offer opportunities for practising the target structure. It is believed that real-world interaction encourages learners to refine and restructure their inter-language by drawing their attention to linguistic code features during negotiation for meaning (Long, 2015). Following Ellis (2009), tasks may direct learners' attention to meaning while predisposing them to focus on challenging L2 phonological forms through task-essentialness (Loschky & Bley-Vroman, 1993).

Whereas most research examining the facilitative role of tasks has focused on grammar, lexical structures or pragmatics, only a few TBLT studies have investigated the effectiveness of tasks in drawing learners' attention to phonetic form during communicative task performance, by testing learners' L2 pronunciation improvement (Gurzynski-Weiss et al., 2017). Although task-based pronunciation teaching (TBPT) has been found to be beneficial for L2 pronunciation development (Mora-Plaza et al., 2018), it remains unclear to what extent TBPT helps L2 pronunciation development over time and whether gains may be transferred to new lexical contexts. Furthermore, little is known about how L2 vowel production accuracy may vary depending on the context where vowels are embedded, namely, in isolated words or in sentences (but see Mora et al., 2022).

1.3 Research questions

The present study extends this line of research by assessing the effects of form-focused instruction on the production of English high /iː, ɪ/ and low /æ, ʌ/ vowels in communicative decision-making tasks during a longitudinal intervention. Improvement in L2 vowel production was assessed in terms of elicitation mode and generalisation to untaught tokens. The present study is therefore guided by the following research questions:

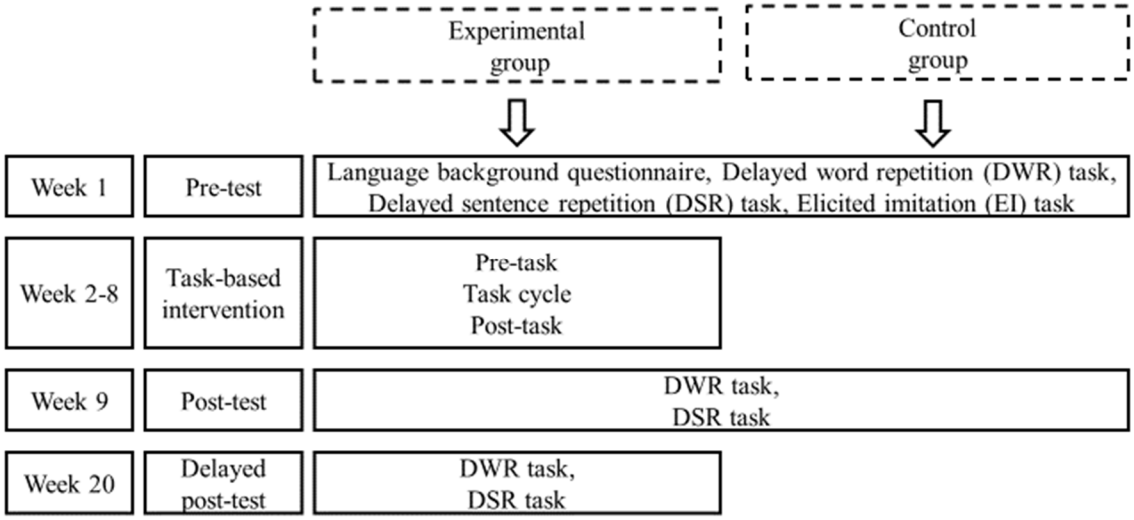
RQ1: Does TBPT improve learners' production of L2 vowels embedded in words elicited in isolation?
RQ2: Does TBPT improve learners' production of L2 vowels embedded in words elicited in sentences?
RQ3: Are gains in L2 vowel production generalised to untaught tokens elicited in isolation and in sentences?

2 Methods

This study followed a pre-/post-test design, with a delayed post-test that learners performed 11 weeks after the task-based intervention. All participants carried out a battery of perceptual (discrimination, lexical sensitivity) and production (delayed word, sentence) tasks targeting the English vowel contrasts /iː – ɪ/ and /æ – ʌ/ individually before and after 7 weeks. A language background questionnaire and an elicited imitation (EI) task were administered before the first testing session, in order to obtain learners' biographical information and L2 proficiency level (Figure 1). In the EI task, learners were instructed to repeat 30 sentences after a beep signal. The sentences increased in grammatical and lexical complexity. Learners' productions were recorded and assessed for accuracy following Ortega et al.'s (2002) rubrics, where each sentence received a score from 0 to 4. Individual scores could thus range from 0 to 120 points.

Figure 1

Research Design



The experimental group was exposed to a series of task-based lessons three times per week. Learners dedicated the first 20–30 minutes of their English class to practising L2 reading, listening, writing and/or use of English and the last 30–40 minutes were devoted to doing pronunciation-focused oral tasks in pairs. The control group completed the pre- and post-test, and continued with their regular English classes, without taking part in any task-based intervention.

2.1 Participants

Ninety-two Catalan/Spanish English as a Foreign Language (EFL) learners from a public high school participated in the study. They belonged to three intact classes selected for convenience, as it would have been logistically impossible to randomly assign participants to different groups. The number of males (M) and females (F) was balanced across groups (Experimental group: M = 33, F = 30; Control group: M = 14, F = 15) and their age ranged from 16 to 17 years old. Their self-estimated English proficiency level ranged from intermediate to upper-intermediate (see Table 1). The experimental and control groups were not significantly different in terms of demographic and linguistic variables ($p > .05$).

Table 1

Participants' Demographic and Linguistic Information

	M	*SD*	Range	95% Confidence Interval (CI)	
				Lower	Upper
Experimental group (*n* = 63)					
Age (years)	16.0	0.2	16-17	16.0	16.1
Age of onset (years)	5.6	1.9	3-9	5.2	6.1
L2 instruction (years)	10.3	1.9	7-13	9.8	10.7
L2 use (hours/week)	3.5	3.1	0-14	2.7	4.1
Self-estimated proficiency [a]	6.0	1.6	1-9	5.3	6.2
L2 proficiency [b]	71.2	20.0	32-113	67.6	79.1
Control group (*n* = 29)					
Age (years)	16.0	0.3	16-17	15.9	16.1
Age of onset (years)	6.1	1.6	3-10	5.5	6.7
L2 instruction (years)	9.9	1.6	6-13	9.2	10.4
L2 use (hours/week)	2.7	1.5	0-6	2.1	3.2
Self-estimated proficiency [a]	5.7	1.7	1-9	5.0	6.3
L2 proficiency [b]	71.7	21.2	35-116	63.3	80.1

[a] Averaged self-estimated ability to speak spontaneously, understand, read, write, and pronounce in English.
[b] Obtained through an elicited imitation (EI) task (Ortega et al., 2002).

2.2 Task-based intervention: Stimulus materials and procedure

Over 7 weeks, learners in the experimental group performed a sequence of 20 tasks that simulated an end-of-the-course trip to London. Tasks contained 80 consonant-vowel-consonant (CVC) words, coming from 24 minimal pairs containing the two target contrasts /iː – ɪ/ and /æ – ʌ/ as well as other words containing the four target vowels without being minimal pairs. Half of the words were monosyllabic and half were disyllabic (Appendix A). Each contrast was presented once in every task, and the stimuli consisted of six minimal pairs and eight extra words containing the target vowels. Four Southern Standard British English (SSBE) speakers recorded the aural input of the listening comprehension activity in the pre-task phase. Tasks were designed following Willis' (1996) framework for task-based learning, namely, a three-phase framework corresponding to a pre-task, task cycle, and language focus stage (see examples in Mora-Plaza, 2021; Mora-Plaza et al., 2022).

The pre-task lasted for 10 minutes. First, the teacher presented the topic area of the session through an illustration of a real-life activity (e.g., packing a suitcase) and elicited students' experiences. Relevant words and expressions were written on the whiteboard and noted down by students. Then, students listened to a conversation which replicated the task each pair was going to carry out during the task cycle. Students listened for overall comprehension, then they had to pay attention to certain words which were inserted, in order to trigger a focus on phonetic form (e.g., *cap* /kæp/ and *cup* /kʌp/). Finally, the teacher showed a picture of the target object

on screen and elicited the word. Students had to guess the word and listen to it. The teacher set the goal of the main task and gave planning time to prepare for it.

The task cycle lasted approximately 15–20 minutes and consisted of 20 problem-solving tasks which were two-way, close, and convergent. Additionally, tasks were "designed to provide opportunities for communicating using some specific linguistic feature" (Ellis, 2009, p. 223); namely, learners had to be able to distinguish L2 vowel contrasts (/iː – ɪ/ or /æ – ʌ/) in order to perform the task successfully (i.e., task essentialness; Loschky & Bley-Vroman, 1993). The task cycle consisted of three different phases:

1. Task – Students performed the task in dyads and the teacher monitored students' performance. The teacher made sure students were using the L2 and promoted spontaneous talk and confidence building.
2. Planning – Students rehearsed the outcome of the task and organised their discourse before presenting it in front of their classmates.
3. Report – Students presented their reports of the task in front of their classmates. The teacher noted down any inaccuracies related to the target vowels, to comment on them during the language focus stage.

The language focus stage lasted for 5–10 minutes. During this stage, the teacher prepared some language-focused tasks to consolidate the phonological contrasts encountered through communication during the task cycle. In the analysis stage, students did consciousness-raising activities and, in the practice stage, they consolidated the target pronunciation features through communicative tasks.

2.3 Assessment

Learners' L2 phonological knowledge was assessed through perception and production tests. In this paper, we will only report the results of L2 vowel production. Learners produced the L2 target vowels through delayed word repetition (DWR) and delayed sentence repetition (DSR) tasks, which were administered in DMDX[2] (Forster & Forster, 2003) on a laptop computer.

In the DWR task, participants heard the word (e.g., /kæt/) followed by a 1500ms pause, before a tone signal prompted them to repeat it. In order to test for generalisation effects, the testing stimuli comprised 24 taught and 24 untaught words, and were produced by two speakers that participants had not been exposed to during the intervention. The test consisted of a total of 68 trials (64 test and 4 practice trials). The testing stimuli were 32 monosyllabic and 32 disyllabic words: 48 were words from minimal pairs /iː – ɪ/ and /æ – ʌ/, and 16 words had no contrasting counterpart but contained the 4 target vowels and also had appeared during the intervention (Appendix B).

In the DSR task, learners were asked to: 1) read the sentence appearing in standard orthography on the computer screen for 3000ms; 2) listen to the sentence over the headphones; and 3) repeat the sentence from memory after a sound signal occurring 1500ms after the offset of the sentence stimulus. The DSR stimuli were identical to the DWR ones. Learners were exposed to 64 test sentences and four practice sentences, which were four words long. They were always formed by the determiner/pronoun THE/THEY + TARGET WORD containing /iː, ɪ, æ, ʌ/ + VERB + OBJECT (e.g., *The <u>bin</u> is empty*). All test words (DWR) and sentences

[2] DMDX (Forster & Forster, 2003) is a Win32-based software display system used in psychological and linguistic experiments for stimulus presentation (providing fast-action, dynamic gaming experience) and for measuring reaction times to visual and auditory stimuli.

(DSR) were distributed into two separate randomised blocks (1st block /iː/, /ɪ/; 2nd block /æ/, /ʌ/) with 32 stimuli each and a short break in between.

2.4 Data analysis

To obtain an L2 vowel production measure, vowel frequencies were extracted in Praat. A Bark-distance normalisation procedure was used to provide speaker-independent estimates of vowel quality. The difference in Bark between F1 and *f*0 (B1-B0) estimated vowel height, whereas the difference between F2 and F1 (B2-B1) estimated vowel frontness.

Mahalanobis distances were used to calculate a measure of vowel distinctiveness (i.e., every token of vowel /æ/ and the centroid of the distribution of the tokens of the other contrasting vowel /ʌ/, and vice versa). A larger distance meant less of an overlap between the two vowels (Melnik-Leroy & Peperkamp, 2021).[3] In addition, to measure vowel nativelikeness, we calculated the distance between native speakers' and learners' productions of each target vowel produced in the same phonetic context, so a smaller distance meant a more target-like production (Kartushina et al., 2015).

In order to answer RQ1 and RQ2, linear mixed-effects models were performed in SPSS 27 with GROUP (Experimental vs. Control), TIME (pre-test vs. post-test), CONTRAST (/iː – ɪ/ vs. /æ – ʌ/) or VOWEL (/iː, ɪ, æ, ʌ/), and their interactions as fixed effects, and SUBJECT and ITEM as random intercepts. As for RQ3, we calculated gains from pre-test to delayed post-test scores for the experimental group only. In a linear-mixed effects model, we included TOKEN TYPE (taught vs. untaught) as fixed effects and a random intercept for SUBJECT. The parameter estimates are given in Appendix C.

3 Results

RQ1 queried the effects of TBPT on L2 vowel production in words elicited in isolation. On the one hand, mixed effects models revealed significant main effects of TIME (F[1, 8680] = 7.08, p = .008) and CONTRAST (F[1, 8680] = 9.17, p = .002) on Mahalanobis distances between vowels /iː – ɪ/, /æ – ʌ/, i.e., distinctiveness measure. A significant GROUP × TIME interaction (F[1, 8680] = 6.90, p = .009) revealed that, at post-test, the experimental group significantly produced a larger distance between the vowels in the contrasts (pre-test: M = 10.24; SD = .37; post-test: M = 12.96; SD = .32; p < .001) than the control group (pre-test: M = 10.18; SD = .61; post-test: M = 10.16; SD = .80; p = .984). Bonferroni-adjusted pairwise contrasts from the GROUP × TIME × CONTRAST interaction (F[1, 8680] = 1.16, p = .282) showed that, while the task-based intervention seemed to have helped learners produce a larger distance between /iː/ and /ɪ/ (t[8680] = 2.20, p = .027), and /æ/ and /ʌ/ (t[8680] = 5.00, p < .001), none of the control group's contrasting vowels distinguished significantly at post-test (p > .05) (see Figure 2, left panel).

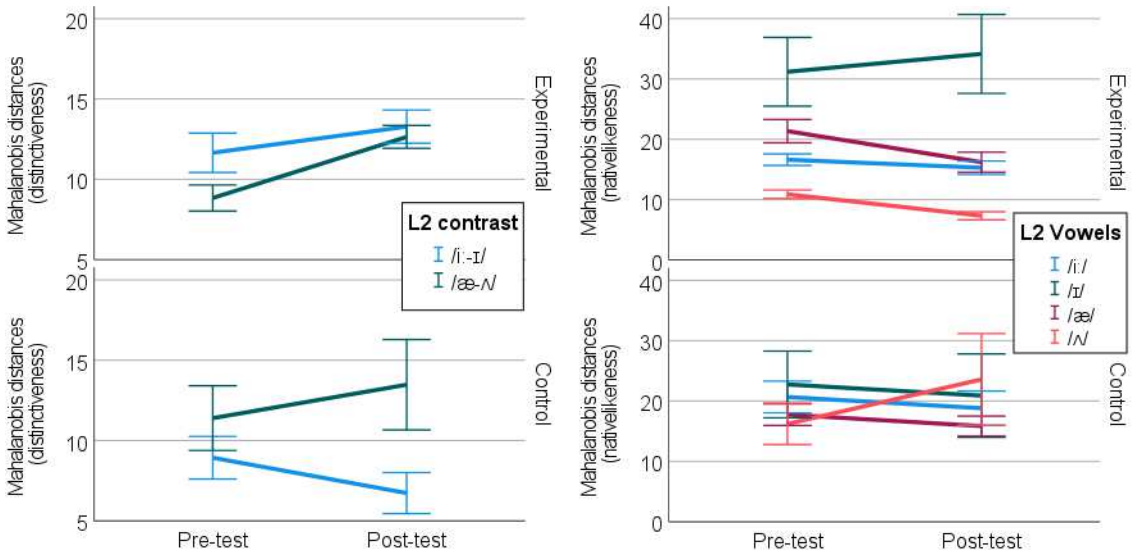
On the other hand, mixed effects models showed non-significant main effects of TIME (F[1, 8672] = .17, p = .679) and significant effects of VOWEL (F[3, 8672] = 29.37, p < .001) on Mahalanobis distances between learners' and native speakers' vowels /iː, ɪ, æ, ʌ/, i.e., nativelikeness measure. Despite the non-significant GROUP × TIME interaction (F[1, 8672] = 2.59, p = .107), Bonferroni-adjusted pairwise contrasts showed that overall the experimental group (but not the control group) had significantly shortened Mahalanobis distances compared to native speakers (i.e., the participants became more accurate) between testing times (pre-test: M = 20.03; SD = .79; post-test: M = 18.26; SD = .89; p = .05). As illustrated in Figure 2 (right

[3] Mahalanobis distances are defined as the distance in standard deviations between a point and the centroid of a distribution (Melnik-Leroy & Peperkamp, 2021).

panel), only learners' productions of /æ/ and /ʌ/ significantly approximated the values of native speakers' vowel productions. Finally, learners in the experimental group kept separating the confusing vowels in the contrasts (/i: – ɪ/ and /æ - ʌ/) as demonstrated by the vowel distinctiveness measures at the delayed post-test (post-test: $M = 12.96$, $SD = .32$; delayed post-test: $M = 13.98$, $SD = .39$; $p = .05$) and maintained similar Mahalanobis nativelikeness distances (post-test: $M = 18.26$, $SD = .89$; delayed post-test: $M = 17.34$, $SD = .83$; $p = .47$), suggesting that learning was retained 11 weeks after the treatment.

Figure 2

Mahalanobis Distances for Distinctiveness (on the Left) and for Nativelikeness (on the Right) Produced in Words in Isolation



Note. The graphs are organised by GROUP (Experimental vs. Control) and TIME (pre-test vs. post-test) in the X-axis. Error bars show 95% CI.

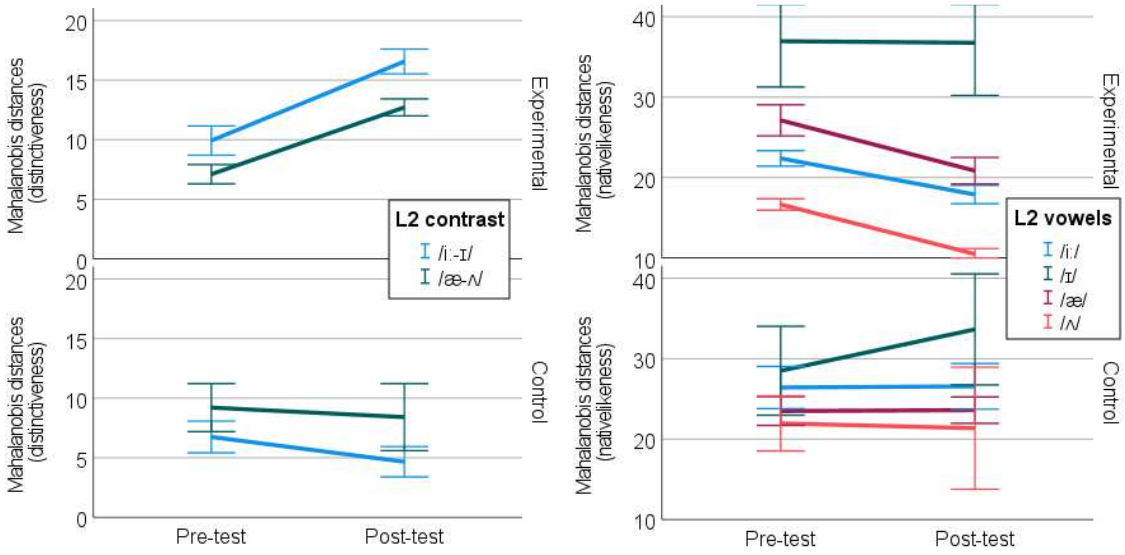
RQ2 also asked about the effects of TBPT on L2 vowel production but this time in words produced in sentences. In terms of Mahalanobis distances of vowel distinctiveness, mixed effects models revealed significant main effects of TIME ($F[1, 8680] = 21.31$, $p < .001$) and CONTRAST ($F[1, 8680] = 20.55$, $p < .001$). In addition, results from the GROUP × TIME interaction ($F[1, 8680] = 49.41$, $p < .001$) indicated that the experimental group produced significantly greater distances between the confusing vowels for the two contrasts after the task-based intervention (pre-test: $M = 8.52$; $SD = .37$; post-test: $M = 14.64$; $SD = .32$; $p < .001$), whereas the control group did not make any significant distinction between any of the target vowels in the contrasts (pre-test: $M = 8.00$; $SD = .61$; post-test: $M = 6.57$; $SD = .79$; $p = .157$) (see Figure 3, left panel).

Concerning Mahalanobis nativelikeness distances, linear mixed models revealed non-significant effects of TIME ($F[1, 8675] = .17$, $p = .241$) and significant effects of VOWEL ($F[3, 8675] = 45.20$, $p < .001$). Furthermore, a significant GROUP × TIME interaction ($F[1, 8675] = .17$, $p = .241$) arose because, according to Bonferroni-adjusted pairwise contrasts, experimental learners' vowels became more target-like (i.e., the distance to native speaker

values was reduced) from pre- (M = 25.78, SD = .79) to post-test (M = 21.50, SD = .89; p < .001), but the control group's vowel productions did not change significantly across times (pre-test: M = 25.05, SD = .91, post-test: M = 26.23, SD = 1.37; p = .268). As seen in Figure 3 (right panel), learners' vowel qualities became significantly more accurate after the intervention, except for vowel /ɪ/ (p = .798). Lastly, evidence for retention effects was found as learners in the experimental group still distinguished the confusing vowels in the contrasts at the delayed post-test (post-test: M = 14.64, SD = .32; delayed post-test: M = 13.75, SD = .39; p = .06) and did so in a native-like direction (post-test: M = 21.50, SD = .89; delayed post-test: M = 20.17, SD = .83; p = .29).

Figure 3

Mahalanobis Distances for Distinctiveness (on the Left) and for Nativelikeness (on the Right) Produced in Words in Sentences



Note. The graphs are organised by GROUP (Experimental vs. Control) and TIME (pre-test vs. post-test) in the X-axis. Error bars show 95% CI.

Interestingly, gains in distinctiveness were moderately correlated with gains in nativelikeness when vowels were embedded in words produced in isolation (r = .330, p = .008) and in sentences (r = .429, p = .001). This suggests that, overall, learners who produced more distinct vowel qualities, also produced vowels that were more target-like.

Finally, RQ3 looked into the comparison between vowel production gains for taught and untaught tokens in two different contexts: words and sentences. On the one hand, learners obtained similar gains in taught (M = 4.06; SD = .94) and untaught (M = 3.40; SD = .78) tokens produced in isolation with regards to vowel distinctiveness (F[1, 501] = .28, p = .593). In contrast, gains for vowel nativelikeness were greater (albeit non-significantly; F[1, 501] = 1.11, p = .291) for those words that had not been taught and were unfamiliar to learners than taught tokens. Likewise, Mahalanobis distance gains between vowels produced in sentences, were similar when they appeared in taught (M = 5.55; SD = .96) and untaught (M = 4.90; SD = .79) tokens (F[1, 501] = .26, p = .606). Gains in how much learners approximated native speakers'

vowel qualities were larger in untaught than taught tokens, but the effects of TOKEN TYPE did not reach significance (F[1, 499] = 1.12, p = .293).

4 Discussion

Results from this investigation suggest that carefully designing and manipulating tasks induces a focus on phonetic form during meaningful interaction, and generating a linguistic focus through task-essential language raises learners' awareness about challenging L2 pronunciation features, eventually leading to more accurate vowel production (see Solon et al., 2017). Learners who took part in the task-based intervention produced L2 vowels more contrastively and more accurately than learners who did not. In word-elicitation contexts (RQ1), the overlap between confusable vowels became significantly smaller for both contrasts and L2 vowels became more target-like. Similarly, in sentence-elicitation contexts (RQ2), learners produced L2 vowels more distinctively and accurately after the task-based intervention. This was not the case for the control group, whose L2 vowel qualities remained stable. Whereas producing words in isolation may have led to more conscious reflections on form (thus emphasising the distinctiveness of vowels in terms of spectral distances), producing the target words embedded in sentences may have mirrored the occurrence of such forms during the interactive tasks, in a more realistic context, where vowel differences relied mainly on quality (Mora et al., 2022). Changes in vowel distinctiveness were significantly associated with changes in nativelikeness, meaning that TBPT helped learners align their initially unstable vowel productions with those of native speakers of English. While the goal of the TBPT intervention was not to achieve a native-like accent, approximating native vowels may have helped learners to become more intelligible, hence, to produce L2 vowels that were sufficiently distinct in order not to confuse interlocutors during communication. In addition, L2 vowel learning seems to be robust, as L2 pronunciation gains were retained 11 weeks after the intervention. Finally, as found by HVPT studies (Thomson, 2018), gains in L2 pronunciation accuracy generalised to untaught tokens in isolation, as well as to tokens embedded in sentences (RQ3) and those spoken by unfamiliar speakers/voices.

Overall, findings from this study suggest that pronunciation instruction can be easily integrated in communicative tasks (in line with current pedagogical principles) by making L2 pronunciation features salient through task manipulation (e.g., making L2 vowel contrasts essential for task completion). Instead of teaching pronunciation in an explicit, often decontextualised manner, TBPT thus advocates for an analytic approach where learners deal with challenging L2 pronunciation features as they are communicating. In line with previous form-focused communicative studies (e.g., Darcy et al., 2021), exposing learners to L2 pronunciation features repetitively and in meaningful contexts results in L2 pronunciation gains; it also prepares learners for out-of-class conversations. Hypothetically, other L2 oral skills (e.g., fluency, prosody) may also develop along with segmentals while learners interact. Finally, L2 pronunciation improvement may be assessed in terms of successful task completion as well as objective and subjective pronunciation proficiency measures.

5 Conclusion

This TBPT study has shown that, despite the time constraints that teachers often suffer, L2 pronunciation can be part of a TBLT curriculum. Form-focused communicative instruction, which is based on tasks that are inherently repetitive yet genuinely communicative (see Sardegna, 2022), may enhance L2 pronunciation learning and lead to generalisation effects in diverse lexical contexts and elicitation modes. Further research should investigate how many segmental features can be addressed in a given task, as well as how often learners should

practise the same minimal pairs for acquisition to occur. More broadly, future work should explore: the effects of task design and manipulation on L2 intelligibility and comprehensibility in face-to-face and online settings; how TBPT may apply to the teaching of suprasegmentals; and which learner factors (experiential, affective, and/or cognitive) should be considered when designing tasks that promote second language pronunciation learning.

Acknowledgements

This research was funded through grant PID2019-107814GB-I00 from the Spanish Ministry of Science, Innovation and Universities. I would like to thank Miren Adrian, Daniela Avello and Valeria Galimberti for their help in data collection, and Dr. Joan C. Mora and Dr. Roger Gilabert for their constant support.

References

Best, C., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Cebrian, J. (2021). Perception of English and Catalan vowels by English and Catalan listeners: A study of reciprocal cross-linguistic similarity. *The Journal of the Acoustical Society of America*, *149*(4), 2671. https://doi.org/10.1121/10.0004257

Darcy, I., Rocca, B., & Hancock, Z. (2021). A window into the classroom: How teachers integrate pronunciation instruction. *RELC Journal*, *52*(1), 110–127. https://doi.org/10.1177/0033688220964269

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221–246. https://doi.org/10.1111/j.1473-4192.2009.00231.x

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods*, *35*(1), 116–124. https://doi.org/10.3758/BF03195503

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 57–77). John Benjamins. https://doi.org/10.1075/lllt.17.09gui

Gurzynski-Weiss, L., Long A.Y., & Solon M. (2017). TBLT and L2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis? *Studies in Second Language Acquisition*, *39*(2), 213–224. https://doi.org/10.1017/S0272263117000080

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, *138*(2), 817–832. https://doi.org/10.1121/1.4926561

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345–366. https://doi.org/10.1093/applin/amu040

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based learning. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123–167). Multilingual Matters.

Melnik-Leroy, G. A., & Peperkamp, S. (2021). High-variability phonetic training enhances second language lexical processing: Evidence from online training of French learners of English. *Bilingualism: Language and Cognition, 24*(3), 497–506. https://doi.org/10.1017/S1366728920000644

Mora, J. C., Ortega, M., Mora-Plaza, I., & Aliaga-García, C. (2022). Training the pronunciation of L2 vowels under different training conditions: The use of non-lexical materials and masking noise. *Phonetica, 79*(1), 1–43. https://doi.org/10.1515/phon-2022-2018

Mora-Plaza, I. (2021). Off to the library! TBLT Language Learning Task Bank. https://tblt.indiana.edu/tasks/details.html?id=73

Mora-Plaza, I., Mora, J. C., & Gilabert, R. (2018). Learning L2 pronunciation through communicative tasks. In J. Levis, C. Nagle, & E. Today (Eds.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference* (pp. 174–184). Iowa State University.

Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J.-M., & Tierney, A. (2022). Tools for second language speech research and teaching. http://sla-speech-tools.com

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics, 35*(4), 463–482. https://doi.org/10.1093/applin/amu024

Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002, October). *An investigation of elicited imitation tasks in crosslinguistic SLA research*. Paper presented at the Second Language Research Forum, Toronto, Canada.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly, 46*(4), 842–854. https://doi.org/10.1002/tesq.67

Sardegna, V. (2022). Evidence in favor of a strategy-based model for English pronunciation instruction. *Language Teaching, 55*(3), 363–378. https://doi.org/10.1017/S0261444821000380

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2017). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition*, *39*(2), 347–380. https://doi.org/10.1017/S0272263116000425

Thomson, R. (2018). High Variability [Pronunciation] Training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208–231. https://doi.org/10.1075/jslp.17038.tho

Willis, J. (1996). *A framework for task-based learning*. Longman.

Appendix A

Intervention Stimuli by Vowel (/iː/, /ɪ/, /æ/, /ʌ/), Type (Minimal Pairs vs. Extra Words also Containing the Target L2 Vowels) and Syllable (One- vs. Two-Syllable Words)

	/iː/	/ɪ/	/æ/	/ʌ/
Minimal pairs				
1 syllable	bean	bin	bag	bug
	cheek	chick	bat	butt
	feast	fist	cap	cup
	peel	pill	cat	cut
	sheep	ship	mag	mug
	teen	tin	ram	rum
2 syllables	heating	hitting	amber	umber
	keeper	kipper	ankle	uncle
	lever	liver	babble	bubble
	sleeper	slipper	batter	butter
	sneakers	Snickers	carry	curry
	weeping	whipping	natty	nutty
Extra words				
1 syllable	leave	kill	act	run
	weed	fish	hat	drum
	tea	chips	ham	bun
	jeans	pin	jam	gun
2 syllables	illegal	bitter	jacket	public
	kiwi	whiskey	baggy	nugget
	Peter	Jimmy	Patrick	Luster
	Sheila	Lily	Cathy	Sunset

Appendix B

Testing Stimuli by Vowel (/iː/, /ɪ/, /æ/, /ʌ/), Token Type (Taught vs. Untaught, and Practice Items) and Syllable (One- vs. Two-Syllable Words)

	/iː/	/ɪ/	/æ/	/ʌ/
Taught				
1 syllable	bean	bin	bag	bug
	cheek	chick	cat	cut
	sheep	ship	ram	rum
	leave	fish	hat	drum
	tea	chips	jam	bun
2 syllables	keeper	kipper	amber	umber
	lever	liver	batter	butter
	sneakers	Snickers	carry	curry
	illegal	bitter	jacket	public

	/i:/	/ɪ/	/æ/	/ʌ/
	kiwi	whiskey	baggy	nugget
Untaught				
1 syllable	beef	biff	crash	crush
	feel	fill	lag	lug
	seal	sill	stab	stub
2 syllables	greeting	gritting	attar	utter
	litre	litter	bagger	bugger
	weaner	winner	clatter	clutter
Practice items				
	feet	hill	rat	sun

Appendix C

Parameter estimates of linear mixed-effects models for the measures of distinctiveness and of nativelikeness

RQ1: Words in Isolation

	β	***SE***	***t***	***Sig.***	***95% CI Lower***	***Upper***
Intercept	13.633	1.1654	11.698	0.000	11.348	15.917
Group	-0.858	1.4082	-0.609	0.542	-3.619	1.902
Time	-2.157	1.2021	-1.795	0.073	-4.514	0.199
Contrast	-6.838	1.1556	-5.917	0.000	-9.103	-4.573
Group x Time	-1.741	1.4328	-1.215	0.224	-4.550	1.068
Group x Time x Contrast	-2.102	1.9523	-1.077	0.282	-5.929	1.725

	β	***SE***	***t***	***Sig.***	***95% CI Lower***	***Upper***
Intercept	24.079	3.1393	7.670	0.000	17.925	30.232
Group	-17.648	3.7929	-4.653	0.000	-25.082	-10.213
Time	-8.243	3.4130	-2.415	0.016	-14.934	-1.553
Vowel	-4.969	3.3487	-1.484	0.138	-11.533	1.595
Group x Time	12.267	4.0817	3.005	0.003	4.265	20.268
Group x Time x Vowel	-11.446	5.6656	-2.020	0.043	-22.552	-0.340

RQ2: Sentences

					95% CI	
	β	***SE***	***t***	***Sig.***	***Lower***	***Upper***
Intercept	15.810	1.9139	8.261	0.000	12.059	19.562
Group	1.978	2.3216	0.852	0.394	-2.573	6.529
Time	0.527	1.2041	0.437	0.662	-1.834	2.887
Contrast	-6.830	1.1772	-5.802	0.000	-9.137	-4.522
Group x Time	-6.187	1.4285	-4.331	0.000	-8.987	-3.387
Group x Time x Contrast	-2.564	1.9207	-1.335	0.182	-6.329	1.201

					95% CI	
	β	***SE***	***t***	***Sig.***	***Lower***	***Upper***
Intercept	21.844	3.1414	6.953	0.000	15.686	28.002
Group	-12.259	3.7954	-3.230	0.001	-19.698	-4.819
Time	-0.259	3.4145	-0.076	0.939	-6.953	6.434
Vowel	5.016	3.3501	1.497	0.134	-1.551	11.583
Group x Time	6.804	4.0835	1.666	0.096	-1.201	14.808
Group x Time x Vowel	-0.905	5.6680	-0.160	0.873	-12.016	10.205

RQ3: Words in Isolation

						95% CI	
	β	***SE***	**df**	***t***	***Sig.***	***Lower***	***Upper***
Intercept	3.566313	1.064224	501	3.351	0.001	1.475421	5.657205
Token Type	0.657299	1.228860	501	0.535	0.593	-1.757055	3.071653
Contrast	-0.320170	1.228860	501	-0.261	0.795	-2.734525	2.094184

						95% CI	
	β	***SE***	**df**	***t***	***Sig.***	***Lower***	***Upper***
Intercept	5.244276	4.671135	499	1.123	0.262	-3.933241	14.421792
Token Type	-4.414901	4.177990	499	-1.057	0.291	-12.623522	3.793720
Vowel	-1.363169	5.908571	499	-0.231	0.818	-12.971912	10.245573

RQ3: Sentences

						95% CI	
	β	***SE***	**df**	***t***	***Sig.***	***Lower***	***Upper***
Intercept	4.307979	1.083649	501	3.975	0.000	2.178924	6.437035
Token Type	0.645394	1.251290	501	0.516	0.606	-1.813027	3.103816
Contrast	1.198348	1.251290	501	0.958	0.339	-1.260073	3.656770

						95% CI	
	β	***SE***	**df**	***t***	***Sig.***	***Lower***	***Upper***
Intercept	8.799831	4.684930	499	1.878	0.061	-0.404789	18.004451
Token Type	-4.414901	4.190329	499	-1.054	0.293	-12.647763	3.817961
Vowel	-1.363169	5.926020	499	-0.230	0.818	-13.006195	10.279856

About the author

Ingrid Mora-Plaza is a Lecturer in the Department of Modern Languages and Literatures and English Studies at the University of Barcelona, Spain, where she teaches English phonetics and phonology courses. She is part of the Barcelona L2 speech research group and GRAL research group on the acquisition of languages. Her research focuses on L2 pronunciation learning and teaching in foreign language contexts and the effects of individual differences on L2 speech acquisition. She is especially interested in the potential benefits of task-based pronunciation teaching and high-variability phonetic training on L2 speech development in instructional settings.

E-mail: imoraplaza@ub.edu

Mouquet, M., and Mairano, P. (2023). Effects of silent letters on the L2 English pronunciation of L1 French learners. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 188–198). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225081

Effects of silent letters on the L2 English pronunciation of L1 French learners

Marine Mouquet
Aix Marseille University

Paolo Mairano
University of Lille

This paper investigates the effects of orthography on second language (L2) phonology. We replicate a study by Bassetti and Atkinson (2015) investigating the intrusion of silent consonants in production; while the original study was conducted on L1 Italian learners of L2 English, we focus on L1 French learners of L2 English. We recruited 110 French undergraduate students specialising in English in various French universities and we ran two tests to elicit the production of 7x2 target words containing silent letters. In the first task (reading aloud task) participants saw the spelling of target words, while in the second task (word repetition task), participants initially saw their spelling, but then the spelling vanished before beginning the production. The results show that: 1) participants were significantly affected by spelling in both tasks and produced a high proportion of intrusive consonants prompted by silent letters; and 2) when orthographic input was present, more intrusive consonants corresponding to silent letters were produced. These results extend the findings of the original study on Italian participants to French participants. Furthermore, since French (contrary to Italian) has a relatively opaque spelling system, these results suggest that silent letters can have an effect on L2 pronunciation not only for learners with a transparent L1 orthographic system, but also for those who have an opaque L1 orthographic system.

Keywords: L2 phonology, orthography, silent letters, speech production, second language acquisition

1 Background

Research in second language (L2) phonology has long focused on the cross-linguistic influence of the native language (L1) on the L2. Among the numerous models formulated to explain the difficulties faced by L2 learners in acquiring L2 phonology through their L1 phonological systems, the most influential are The Perceptual Assimilation Model (Best, 1994) and the Speech Learning Model (Flege, 1995). However, when an L2 is acquired via instructed learning, the acquisition process is achieved not only via spoken input, but also and often mainly via written input (as opposed to L1s, which are acquired through spoken input). It is therefore understandable that many recent studies have turned to the effects of orthography on L2 production and perception.

Research has found that orthography can either be a facilitating or a misleading factor in second language pronunciation, especially at the early stages of language acquisition. Among the beneficial effects of orthography, it has been demonstrated that providing orthographic input to L2 learners enhances the memorisation and retrieval of vocabulary (Bürki et al., 2019) and assists L2 learners in perceiving and distinguishing L2 sounds (Escudero et al., 2008; Escudero et al., 2014). In contrast, among the undesirable effects of orthography, researchers have shown that orthography can negatively affect L2 phonological representations, resulting in non-target-like productions (Bürki et al., 2019), notably segmental additions, omissions and substitutions (Bassetti, 2008). This can even lead to a phonological category from the L1 being imported; notably, Bassetti et al. (2018) reveal that Italian learners produce a singleton-geminate contrast in English reflecting spelling, e.g., *finish* vs. *Finnish*.

L2 learning may vary according to different factors, and the degree of transparency of the orthographic system in the L1 and the L2 is one of them. The orthographic system of some languages can be considered transparent: for instance, Finnish, Dutch, and Italian are acknowledged to have transparent orthographic systems, characterised mostly by one-to-one grapheme-phoneme and phoneme-grapheme correspondences (Aro, 2013). On the contrary, the orthographic systems of English, Irish, and French are considered to be opaque (or deep) since they often deviate from such correspondences (Erdener & Burnham, 2005). Therefore, it seems reasonable to claim that the phonological patterns of L2 languages with a transparent orthography can be easier to predict, because orthographic input can give direct information about the pronunciation of words (Seymour et al., 2003). Furthermore, researchers have suggested that learners whose L1 has a transparent orthographic system may tend to rely more heavily on orthography than learners whose L1 has an opaque orthographic system (Erdener & Burnham, 2005). Clearly, this may then be problematical for learners whose L1 has a transparent orthographic system targeting an L2 with an opaque orthographic system. For instance, Bassetti and Atkinson (2015) investigated the effects of silent letters (i.e., orthographic cues that are never produced) in reading aloud and word repetition tasks, revealing that L1 Italian learners of L2 English tended to pronounce intrusive consonants for silent letters, and more so when orthographic input was directly displayed.

In this study, we sought to replicate the two tests carried out by Bassetti and Atkinson (2015) originally on Italian learners, by running similar tasks with L1 French learners of L2 English. Contrary to Italian, the orthographic system of French is opaque and includes many one-to-many and many-to-one correspondences, including many cases of silent letters. This may lead us to expect, based on Erdener and Burnham (2005), that French learners of English may have fewer issues with silent letters in English than Italian learners: since their L1 orthographic system is less transparent and includes silent letters, L1 French learners may rely less than L1 Italian learners on orthography to predict phonological patterns of the L2. However, this prediction contradicts the experience of many teachers of English in France, who regularly witness the challenges posed by silent letters to French learners. Moreover, the impact of such

letters on production is not well-documented. Thus, our study not only replicates the original one, but extends it to a different population (L1 French learners), thereby providing a testbed for the claims of Erdener and Burnham (2005).

2 Research methodology

2.1 Research questions and hypotheses

The study aims to investigate the following research questions:

RQ1: To which degrees will French learners be affected by silent letters resulting in additional segmental units in their output?

RQ2: Do learner variables: a) living abroad and length of stay abroad; b) onset of English acquisition; c) L2 proficiency level; and d) variety comprehension, interact with orthographic effects?

We hypothesise that even though there are many silent letters in French, orthographic forms will still affect French learners' pronunciation of silent letters in English. Following Bassetti and Atkinson (2015), orthographic input is provided in the reading aloud task, but only partially in the word repetition task (i.e., orthographic input is provided briefly but vanishes as soon as the audio starts playing). Thus, we predict that more intrusive consonants will be produced for silent letters in the former task than in the latter. Following Erdener and Burnham (2005), we predict that French learners will be misled by silent letters to a lesser extent than Italian learners, since French is orthographically less transparent than Italian and includes many silent letters.

We also expect some learner variables, such as L2 proficiency, to counterbalance those effects. For instance, second-year undergraduate students at the English Department, University of Lille, are given lectures on grapho-phonemic rules and particularly on silent letters as part of their course in English phonetics in the first semester. In order to evaluate the impact of such lectures, the current study compares the production of first-year vs. second-year and third-year students. It is predicted that second-year students will produce fewer silent letters than the two other groups. This expectation arises from the fact that first-year students have not received any instructions on silent letters and that second-year students have more recently received lectures on silent letters than third-year students. Therefore, third-year students are also expected to produce fewer silent letters than first-year students.

2.2 Participants, task stimuli, and experiment procedure

Due to the COVID-19 lockdown, the two tests were hosted on a website, run remotely, and conducted via computers, without being available on smartphones. Therefore, unlike Bassetti and Atkinson (2015), participants used their own equipment. They were advised to perform the experiment in a quiet room and to use headphones for the repetition task. The website was coded in Javascript/JQuery for the dynamic part and in HTLM/CSS for the static part. More precisely, the website also used Boostrap for the layout. For the recordings, we used the library Recorder.js, published under the MIT licence by Matt Diamond. All the recordings and data were gathered on a MySQL database thanks to PHP.

The tasks were completed by 110 L1 French participants of L2 English with no reported language or reading impairments: 38 first-year, 24 second-year and 48 third-year students. All of them were undergraduate students specialising in English in various French universities (F = 92, M = 18, a typical gender bias among students in language departments in France). The

median self-reported level of English was B2 and ranged from B1 to C2. Twenty respondents had lived in an English-speaking country (mainly the UK and the US but also Canada, South Africa, Singapore, and New Zealand) with an average duration of 20 months (*MD* = 10.5, *SD* = 18.77). Participation was voluntary and unpaid.

Written instructions were given in French before each task. All participants were exposed to 37 stimuli appearing in the same order determined randomly for both tasks (see Table 1): seven target words included the silent letters <b>, <d>, and <l> in different positions, which were the focus of our analysis; 24 fillers (homophones[1]); three words including optionally silent letters[2]; and three filler words. The word *landscape* was included in the list of target words in Bassetti and Atkinson (2015) but it was discarded from our study because the silent <d> is frequently produced by native speakers[3].

Table 1

Stimuli Words

Target words	**Homophones**		**Words with optional silent letters**	**Filler words**
clim**b**	aloud	allowed	san**d**wich	okay
com**b**	caught	court	gran**d**son	like
de**b**t	flour	flower	lan**d**scape	um
lam**b**	higher	hire		
We**d**nesday	one	won		
sa**l**mon	principal	principle		
wa**l**k	right	write		
	sauce	source		
	seas	seize		
	son	sun		
	which	witch		
	wood	would		

Homophones were notably included in the stimuli because Bassetti and Atkinson (2015) include four different studies investigating several aspects of orthographic effects. We did not test orthographic effects on the pronunciation of homophones. However, the same stimuli were used in order to: 1) do a valid replication and be able to compare results, and 2) to distract participants from silent letters with different fillers. Stimuli were sorted randomly and presented to participants, although it was impossible to apply the exact same order as this information was not provided in Bassetti and Atkinson (2015).

In the first task, each word appeared on the screen and participants had to click to start recording, and then read the written word aloud within a timeframe of three seconds. No acoustic input was provided in this task. As the experiment was conducted at home and not in a supervised laboratory environment, the test screen displayed a 'Retry' button, in case

[1] The homophones were used for a different study by Bassetti and Atkinson (2015).
[2] These words have letters that are claimed to be pronounced by British English natives and were therefore discarded from the analysis (see Bassetti & Atkinson, 2015).
[3] As already shown with *sandwich* and *grandson* in Bassetti and Atkinson (2015).

participants encountered recording issues or other technical problems. However, participants were encouraged to be spontaneous and record each word just once. Only the last version was kept for the analysis. A bar at the top tracked their progress for each task. Once they had recorded the 37 words, they could move on to the second task.

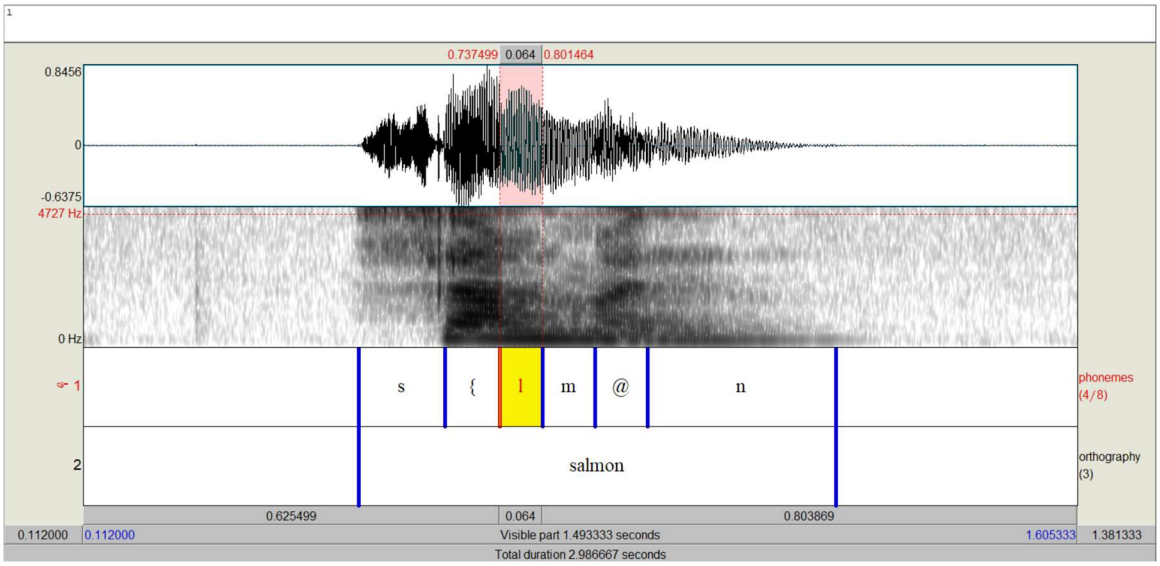
In the word repetition task, participants had to repeat words on the basis of a recording. The orthographic version of each word was initially shown on the screen, but vanished when participants listened to the audio. Only one listening was possible. The audio was extracted from the Cambridge Dictionary[4] (online version) and was based on Southern British English (SBE). As in the previous task, participants had to record themselves repeating the target stimuli. Participants were advised to record themselves just once, but they could retry if they encountered any technical issues.

3 Data analysis and results

The recordings containing the seven target words were examined in isolation for each student twice (once for each task). Based on an auditory and spectrographic analysis with Praat (Boersma & Weenink, 2019), each target silent letter was coded by the authors as intrusive or non-intrusive. A sample of the recordings is presented in Figure 1.

Figure 1

Spectrogram Analysis of the Word 'Salmon' with a Pronounced /l/ by a First-year Participant

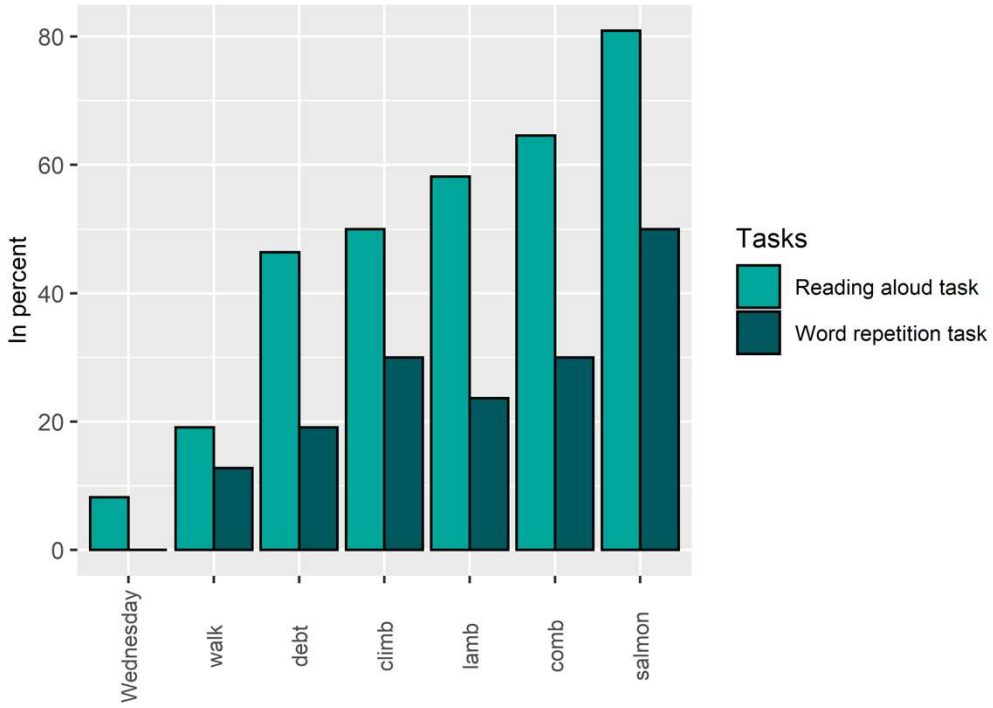


In general, the results show that on average 47% of the target words were produced with an added phone during the reading aloud task vs. 24% during the word repetition task. A paired *t*-test assuming unequal variances confirmed that there were significantly more cases of added silent consonants in the reading aloud task than in the word repetition task ((t (209) = 7.25, p < .001; M_{RAT} = 3.27 vs. M_{WRT} = 1.7), thus suggesting that acoustic input reduces the intrusion of silent letters.

[4] Cambridge Dictionary https://dictionary.cambridge.org

Figure 2

Proportion of Pronounced Silent Letters in Target Words: Reading Aloud Task vs. Word Repetition Task



Note. Data is presented in percentage and classified by significance

More detailed results of the tasks for each word are indicated in Figure 2. The <l> in the word *salmon* was the most problematic for French participants in both tasks: they produced 81% of <l> occurrences in the reading aloud task vs. 50% in the word repetition task. In addition, French participants encountered overall difficulties with the <b> in final position in *comb, lamb*, and *climb* and in middle position in *debt* (55% of the participants produced them in the word reading task vs. 26% in the word repetition task). Conversely, participants did not seem to face many difficulties with the silent <d> in *Wednesday* as only 8% of them pronounced it during the first task and none of them did during the second one. Most students who produced a /d/ also mispronounced other phonemes and/or stress (e.g., pronouncing it as */wedˈnesdeɪ/ and */wedˈnezdeɪ/).

In addition, written and spoken frequencies were extracted from the British National Corpus (see Table 2). In order to determine whether spoken and written lexical frequencies affected the percentage of intrusive consonants corresponding to silent letters, we examined their correlation (see Figure 3).

Table 2

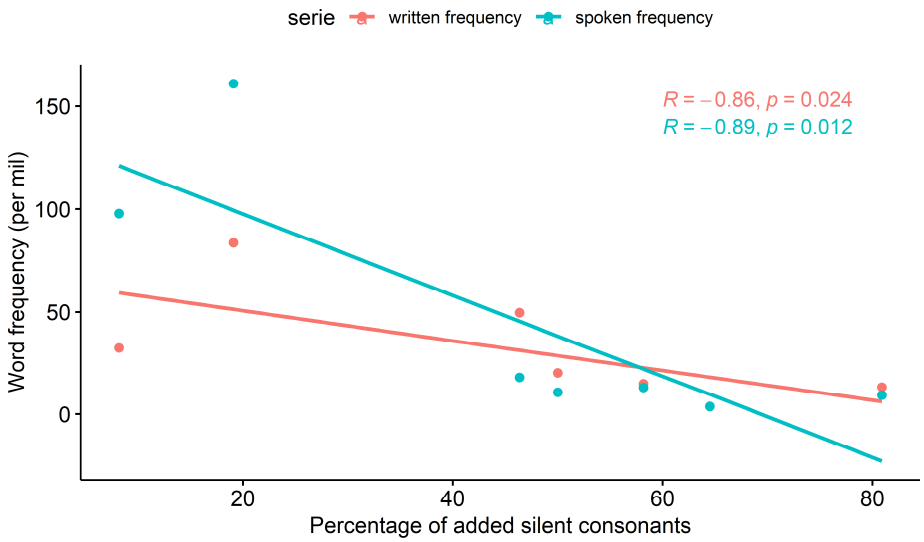
Occurrences of Written and Spoken Target Words in the British National Corpus (in Lemma, per Mil)

Words	Written frequency	Spoken frequency
walk	83.59	160.88
Wednesday	32.54	97.76
debt	49.64	17.97
climb	20.16	10.54
lamb	14.8	12.95
salmon	13.11	9.23
comb	3.62	3.61

Note. Data retrieved from the British National Corpus in May 2021 (Davies, 2004)

Figure 3

Number of Pronounced Silent Letters in the Reading Aloud Task: Spoken-to-written Frequency



Words with lower frequencies were more challenging for participants. A Spearman correlation test was used to check the distribution between the number of intrusive silent consonants for each word and its spoken and written frequencies. The correlation yielded significant results for both spoken (r = -0.89, p = 0.012) and written (r = -0.86, p = 0.024) frequencies. Spoken frequency correlates even more significantly than the written one; this is of course expected, as it is reasonable to assume that the pronunciation of words more

frequently encountered in speech is better known to learners. However, more samples are needed to confirm the validity of this finding, as only seven words were involved in this task.

Additionally, no significant correlation was found between orthographic effects and learner-level variables: living abroad and length of stay abroad, onset of English acquisition, proficiency level in English, and English variety comprehension. Surprisingly, a *t*-test assuming unequal variance revealed that second-year and third-year participants from the English Department in Lille (who had received grapho-phonemic lectures on silent letters (n = 16)) did not significantly diverge from first-year participants (n = 11) from the same Department (t (25) = 1.76; p = 0.09; M_{FIRST} = 4.45 vs. M_{OTHERS} = 3.25) for the reading aloud task and (t (20) = -0.08; p = 0.9; M_{FIRST} = 2.27 vs. M_{OTHERS} = 2.31) for the word repetition task. This may suggest that lectures on grapho-phonemic rules were not only ineffective, but also that there does not seem to be a change from the first year to the second and third year in terms of pronunciation of silent letters.

4 Discussion

In line with Bassetti and Atkinson's (2015) study, these findings confirm the significant effects of orthography on L2 production: French participants produced intrusive consonants corresponding to silent letters in 47% of words with silent letters in the reading aloud task and 24% in the word repetition task. The effects that we found for French learners of L2 English are smaller than those found by Bassetti and Atkinson (2015) for Italian participants. This result seems to support Erdener and Burnham's (2005) claim that the transparency of L1 orthography has an effect on the extent to which learners rely on spelling to predict pronunciation patterns in an L2. However, this interpretation should be tempered, because comparisons of L2 learners from different L1s may be misleading, and many factors may influence these findings, e.g., our French participants had less L2 instruction than the Italian participants in the original study, having had on average, 10 years vs. 11 years. However, French participants were older and were university students learning English, thus frequently speaking, listening, and writing English for a minimum of 20 hours per week, whereas Italian participants in the original study were high-school students from Rome. In addition, although French participants' self-reported proficiency levels were examined, we did not include a control group with experienced learners who did not receive lectures on grapho-phonemic rules. It may be that more experience with the L2 could lower orthographic effects.

Nonetheless, beyond differences potentially due to the L1 and other factors, the difference in results between the two tasks remains significant: when provided with orthography, participants were decoding graphemes, and therefore the effects of spelling were more noticeable. Yet, when investigating L2 transfer, phonological models such as PAM (Best, 1994) and SLM (Flege, 1995) do not account for orthographic effects, suggesting that more research is needed to attempt to capture learners' difficulties in acquiring a second language.

The impact of cognates was also investigated. In the original study, Italian learners did not encounter any difficulties with the cognate word *salmon* (in Italian *salmone*) even though the <l> is produced in the L1. Conversely, the French *saumon* does not include any <l> but *salmon* was the word with which French learners struggled most (i.e., 80% of them produced an intrusive consonant corresponding to silent letters in the reading aloud task and 50% in the word repetition task). This suggests that cognate words did not have any evident effects on the production of intrusive consonants prompted by silent letters, confirming Bassetti and Atkinson's (2015) findings.

The present study also yielded some unexpected results. The final <-mb> cluster (with silent <b>) seems to be challenging for French learners of L2 English, resulting in /mb/ realisations in 52% of the cases in the reading aloud task and 30% in the word repetition task. Yet, this

consonant cluster does not exist in French, suggesting that the effects of orthography can override L1 phonotactic restrictions. Bassetti et al. (2018) had similar results, with geminate consonants produced by L1 Italian learners and late bilinguals in contexts where they would not be present in the L1. Additionally, we observed an interaction of orthographic and phonotactic effect in the reading aloud task, where many participants had the tendency to nasalise the preceding vowel in *lamb* and *comb* leading to non-target-like realisations such as */ˈlɑ̃b/ (12%) and */ˈkɔ̃b/ (43%). This effect was barely noticeable in the repetition task (1% for the former vs. 8% for the latter), thereby suggesting that phonotactic constraints and orthography interact with one another, resulting in the nasalisation of the vowel (since <-am-> and <-om-> are realised as /ɑ̃/ and /ɔ̃/ in French) and the realisation of an intrusive consonant prompted by the silent letter.

5 Conclusion and perspectives

To the best of our knowledge, our study is the first to analyse the effects of English silent letters on French learners of English. The results replicated those found by the original study (Bassetti & Atkinson, 2015): orthographic forms can affect production, even for experienced learners, all of whom produced at least some intrusive consonants for silent letters. Particularly, the effect was stronger in the reading aloud task than in the word repetition task, as in the original study. This confirms that providing orthographic input to learners can increase orthographic effects, even in a language with a phonologically opaque orthography.

From a teaching perspective, one solution involves withholding written input. Winitz and Yanes (2002) suggested that at the early ages of acquisition of second language learning, L2 learners should not be provided with written input until they acquire a correct pronunciation of these words; yet, preventing access to written material in an L2 is obviously very impractical for many reasons, if at all possible. Thornbury (1999, p. 86) attempted to delay students' exposure to the orthographic forms until the correct pronunciation of silent-letter words was completely acquired; however, results were not conclusive, as learners tended to revert to orthographic input once exposed to it.

However, it is important to stress that orthography should not be banished from L2 learning. Another way to minimise the influence of spelling-pronunciation habits is to encourage the learning of phonetic notation (Mompeán & Fouz-González, 2021). Not only can phonetic symbols provide all learners with the correct pronunciation of vocabulary, but as they are based on visual displays, they can also facilitate learners' categorisation, and conceptualisation of mental representations for target sounds – and might even correct fossilised mispronunciations.

An extension of this study (Mouquet & Mairano, 2023) will expand the limits of the present study, which in itself does not address all the effects of silent letters. Particularly, the limited number of target words, the limited variety of silent letters (five items for <b> and merely two for <d> and <l>), as well as the limited number of contexts (three items for <-mb>, one for <-lb>, one for <-lk>, one for <-lm>, and one for <-dn->) do not give us a full picture. Therefore our findings do not provide an adequate basis for drawing generalisable conclusions. This study will therefore be expanded through an alternative design incorporating a greater number of letters, balanced contexts, and a larger set of items. Additionally, by conducting two tasks, a picture naming task and an ABX task in which a phoneme was added or deleted to the recording of each stimulus (e.g.,*/ˈsælmən/ for *salmon*) via Mbrola resynthesis, direct orthographic exposure may be discarded. This will be the opportunity to investigate whether L2 orthography, and more particularly silent letters, is entrenched in memory. Future research will attempt to provide further evidence for orthographic effects on L2 phonology.

References

Aro, M. (2013). Learning to read: The effect of orthography. In R. Malatesha Joshi, & P. G., Aaron (Eds.), *Handbook of orthography and literacy* (pp. 545–564). Routledge.

Bassetti, B. (2008). Orthographic input and second language phonology. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 191–206). https://doi.org/10.21832/9781847691118-013

Bassetti, B., Sokolovic-Perovic, M., Mairano, P., & Cerni, T. (2018) Orthography-induced length contrasts in the second language phonological systems of L2 speakers of English: Evidence from minimal pairs. *Language and Speech*, *61*(4), 577–598. https://doi.org/10.1177/0023830918780141

Bassetti, B., & Atkinson, N. (2015). Effects of orthographic forms on pronunciation in experienced instructed second language learners. *Applied Psycholinguistics*, *36*(1), 67–91. https://doi.org/10.1017/S0142716414000435

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). MIT Press.

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer* [Computer program]. Version 6.0.49, from http://www.praat.org/

Bürki, A., Welby, P., Clément, M., & Spinelli, E. (2019). Orthography and second language word learning: Moving beyond "friend or foe?". *The Journal of the Acoustical Society of America*, *145*(4), EL265-EL271. https://doi.org/10.1121/1.5094923

Davies, M. (2004). British National Corpus. Oxford University Press. https://www.english-corpora.org/bnc/

Erdener, V. D., & Burnham, D. K. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning*, *55*(2), 191–228. https://doi.org/10.1111/j.0023-8333.2005.00303.x

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*(2), 345–360. https://doi.org/10.1016/j.wocn.2007.11.002

Escudero, P., Simon, E., & Mulak, K. E. (2014). Learning words in a new language: Orthography doesn't always help. *Bilingualism: Language and Cognition*, *17*(2), 384–395. https://doi.org/10.1017/S1366728913000436

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Mompeán, J. A., & Fouz-González, J. (2021). Phonetic symbols in contemporary pronunciation instruction. *RELC Journal*, *52*(1), 155–168. https://doi.org/10.1177/003368822094343

Mouquet, M. & Mairano, P. (2023). Orthography and the mental lexicon: The effects of English silent letters on French learners. *Proc. of ICPhS 2023*, Prague (Czech Republic), 7-11 August 2023.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174. https://doi.org/10.1348/000712603321661859

Thornbury, S. (1999). *How to teach grammar*. Longman Harlow.

Winitz, H., & Yanes, J. (2002). The development of first year, self-instructional university courses in Spanish and German. In P. Burmeister, T. Piske, & A. Pohde (Eds.), *An integrated view oflLanguage development: Papers in honor of Henning Wode*. Wissenschaftlicher Verlag Trier.

About the authors

Marine Mouquet is currently a PhD student at the University of Aix-Marseille, France, in the LPL research group (*Laboratoire Parole et Langage*, UMR 7309) under the supervision of Sophie Herment (LPL) and Paolo Mairano (STL). She received a Master's degree of English Linguistics and Didactics at the University of Lille, France during which she wrote two MA theses supervised by Paolo Mairano investigating the effects of L2 English silent letters on L1 French learners' pronunciation and perception and its impact on the mental lexicon. Her main

research interests lie in English phonetics and phonology, second language acquisition, the effects of orthography and psycholinguistics.

Email: marine.mouquet@univ-amu.fr

Paolo Mairano is currently Associate Professor in English Linguistics at the University of Lille, and has previously worked in various academic and industrial institutions (University of Warwick, University of Rouen, University of Turin, Nuance Inc.). His research focus lies in the acquisition of L2 sounds as well as other areas of phonetics, and his studies adopt a quantitative approach.

Email: paolo.mairano@univ-lille.fr

Nowacka, M. (2023). Lyrics as a means of raising phonetic awareness: Transcription, pronunciation, and descriptive phonetics combined. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 199–213). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225365

Lyrics as a means of raising phonetic awareness: Transcription, pronunciation, and descriptive phonetics combined

Marta Nowacka
University of Rzeszow

The present study assessed an intervention aimed at increasing phonetic awareness among 95 university English non-majors through self-selected songs. It was primarily designed to boost the students' interest and raise the attractiveness of a course of descriptive phonetics (Nowacka, 2022). The focus is twofold: to show how the students performed the tasks and to exhibit how useful they found the approach. Data collection consisted of three consecutive tasks: 1) transcription of lyrics used for reading aloud; 2) analysis of the occurrence of 51 phonetic and morphophonemic features; and 3) the students' evaluation of the attractiveness and usefulness of the activity to improve their understanding of phonetic theory.

The results show that the identification of some features was difficult. Among the issues which require further classroom intervention are: /ŋ/ followed by /g/, aspiration, <-s> forms pronounced as /-ɪz/, KIT vowel, spread vowels, centring diphthongs, and schwa spelt <o>, <u>, and <a>. Nonetheless, the respondents found this activity (extremely) attractive (71%), (extremely) useful (76%) and more attractive than other phonetic tasks (49%). It was praised for the autonomy it offered in the choice of a text/song for analysis and for the practical application of phonetics.

Keywords: lyrics, phonetic awareness, descriptive phonetics, transcription

1 Introduction

Songs have been recognised as a teaching tool in EFL for the enhancement of various skills, including pronunciation (Barrett, 2015; Hancock, 1999; Murphy, 2013; Tegge, 2018; Walker, 2006). Research on the use of songs for the improvement of pronunciation shows that singing-based intervention and focus on form positively affect pronunciation in general (Wilcox, 1995), as well as the enunciation of words (Baills et al., 2021; Saldiraner & Cinkara, 2021). The articulation of vowels has been shown to improve (Good et al., 2015) and also consonants, for example: consonant clusters by Japanese speakers (Nakataa & Shockey, 2011); final /k, g, t, d/, /l, r/ and /s, z/ by Thai speakers (Kanlayanee, 2012); dental and palato-alveolar fricatives by Indonesians (Aini et al., 2013; Stefani et al., 2015); and /θ/ and /ŋ/ by Czechs (Wodecki, 2014). Improvements in some connected speech phenomena have also been documented, such as linking by Czechs (Wodecki, 2014), as well as in relation to prosody (Degrave, 2021), fluency and naturalness of Iranians (Ashtian & Zafarghandi, 2015), aspiration and fluency of reading by Italians (Tizian, 2016), and new word segmentation by French speakers (Schön et al., 2008).

The present paper focuses on the use of transcribed lyrics in order to revise English phonetics, which, to the author's best knowledge has not yet been a target of phonetic research. A pragmatic intention behind the design of this study was to revise selected features of English phonetics by means of an out-of-class assignment at the end of the first semester, after studying segmental phonetics. The aim was to more actively engage the first-year university English students in self-study and to show them that their understanding of English and recognition of a speaker's accent could benefit from a detailed phonetic analysis of transcribed lyrics of their own choice. It was also an attempt to make the compulsory course of descriptive phonetics more entertaining, because it was perceived as being boring and useless in practical life (Nowacka, 2022).

2 Research methodology

2.1 Research questions

The study aims to answer two questions: the first regarding the students' performance of the task and the second concerning its attractiveness and usefulness.

RQ1: What phonetic features do the students identify correctly in their transcriptions of self-selected lyrics and what features are they still unable to exemplify at the end of the first semester of segmental phonetics?

RQ2: How attractive and useful do the participants find the approach?

2.2 Participants

The respondents were 95 first-year (19–20 years old; M = 29, F = 66) English students of the University of Rzeszow, Poland, at the end of a compulsory 45-hour phonetic training on segmental phonetics. The training included 15 hours of lectures and a 30-hour practical pronunciation course. They had no prior knowledge of suprasegmental phonetics, which started in the second term. All participants were coded S.1–S.95 for anonymity.

2.3 Instruments

The data was collected by means of three different tasks which were part of the obligatory, one-semester, English segmental phonetics and pronunciation course. The tasks were performed on the Microsoft Teams platform. They included:

1. Task 1 – transcription of the lyrics of a self-selected song.
2. Task 2 – a questionnaire used to analyse the occurrence of 51 phonetic and morphophonemic features in the lyrics text.
3. Task 3 – a questionnaire used to evaluate the attractiveness and usefulness of the activity using a 5-point Likert scale (1 = not attractive/useful at all; 2 = not very attractive/useful; 3 = neither attractive/useful nor unattractive/useless; 4 = attractive/useful; 5 = extremely attractive/useful). The questionnaire also included open questions:
   - What did you learn from this activity? Why was it useful?
   - Did you find this activity more attractive than other phonetic activities? Why?
   - What was attractive about the activity?
   - What did you not like about this activity?

2.4 Procedure

Task 1 was conducted first. The students were asked to prepare their transcriptions but were allowed to consult online applications, such as Photransedit[1] or others. Then, in the classroom these transcriptions were used as a warm-up exercise in reading aloud from transcription.

The core part of the study (Task 2) focused on students' ability to identify words with selected vowels, consonants, connected speech phenomena, inflected and derived forms, letter-to-sound correspondences, proper names, non-grammatical forms, as well as non-standard and accent-specific variants (see Table 1). The scope of the analysis was wide, consisting of 51 phonetic and morphophonemic features. As a revision the students were asked to analyse their transcribed texts, correct them if necessary, and to analyse them by searching for these 51 pronunciation features. This task was carried out to put the students' theoretical knowledge into practice.

Table 1

Pronunciation Features Analysed in Students' Transcriptions

Feature type	Pronunciation feature
Vocalic features	- five pure vowels: FLEECE, KIT, NURSE, TRAP, LOT[2] - schwa[3] spelt as <a >, <o> and <u>

[1] Photransedit is available at http://www.photransedit.com

[2] These are Wells' (1982, p. 120) standard lexical sets for English vowels: FLEECE: /i:/, KIT: /ɪ/, NURSE: /ɜ:/, TRAP: /æ/, and LOT: /ɒ/. They are used as keywords in the main text.

[3] The keyword commA: /ə/ (Wells, 1982, p. 120) has been replaced by the more widely known term schwa, which is used in the main text.

Feature type	Pronunciation feature
	- centring diphthongs or a corresponding monophthong /-ɪr, -er, -ʊr/ - triphthongs (e.g., *nowadays* as /ˈn**aʊə**deɪz/) or a smoothed version of (e.g., /ˈn**aə**deɪz/) - three types of vowels: spread, central, and low
Consonantal features	- final voiced obstruents - aspiration - rhoticity/non-rhoticity - dental fricatives /θ/ and /ð/ - /ŋ/ followed and not followed by /g/ - dark [ɬ]
Connected speech features	- contracted, weak, and strong forms - linking and/or intrusive /r/ - the article *the* and the preposition *to* before a vowel and a consonant
Inflected and derived forms	- plurals of nouns, possessives, 3rd person singular present tense <-s> forms pronounced as /s/, /z/, /ɪz/ - past tenses and past participles <-ed> forms pronounced as /t/, /d/, /ɪd/ - present participles <-ing> forms pronounced as /ɪŋ/ or /ɪn/
Spelling-to-pronunciation patterns	- the letter <i> as KIT vowel - <a> before a consonant as TRAP vowel - <ea> or <ee> as FLEECE vowel - <(w)or> as NURSE vowel - <oo> as GOOSE vowel - <ar> as START vowel - <or> as STRUT or NURSE vowels - a suffix <-ate> as /-ət/ in nouns or adjectives or /-eɪt/ in verbs - a suffix <-ous> as /-əs/
Other	- proper names - non-grammatical forms - non-standard and accent-specific pronunciation

In the follow-up part (Task 3) conducted at the end of the course, the students were required to evaluate how attractive and useful the task was for learning and understanding English phonetics and in comparison with other phonetic activities.

Respondents were also encouraged to provide from one to three words corresponding to a given feature, and for data analysis these were categorised into four types: right, wrong, both right and wrong, and not applicable and/or not provided. For example, the answers regarding <-ed> forms of regular verbs pronounced as /d/ were classified as one of the following options:

- right – e.g., *discovered* /dɪˈskʌvəd/, *ruled* /ruːld/, *begged* /begd/;
- wrong – e.g., *begged* /ˈbeged/, *bored* /ˈbɔːred/.

The ‘right and wrong’ category was used when students gave both in a single answer, e.g., *bored* /bɔːd/ and /ˈbɔːred/.

3 Data analysis and results

3.1 Songs chosen by students

Regarding the choice of a singer or band and their accent, 95 respondents opted for mainly British or American artists. There is a slight preference for British (*n* = 49) over American (*n* = 40) English. Among the remaining six replies respondents listed two Australians and Canadians, one singer with German English, and one with Korean English.

Altogether 86 different lyrics were chosen by 95 respondents, with Adele's (*n* = 8) and Ed Sheeran's (*n* = 5) songs the most frequently selected (see Appendix). Adele's lyrics included: Set Fire to the Rain (*n* = 3), Hello (*n* = 2), Someone Like You (*n* = 1), When We Were Young (*n* = 1), and Chasing Pavements (*n* = 1). Ed Sheeran's list included: Thinking out Loud (*n* = 2), Bad Habits (*n* = 1), Castle on the Hill (*n* = 1), and I See Fire (*n* = 1).

3.2 Results

In general, the results are optimistic. Six features, marked in grey in Table 2, constituted over 90% of correct answers, for instance, FLEECE (98%), LOT (98%), TRAP (97%), weak forms (95%), voiced dental fricatives (93%), and contractions (92%). This suggests that the vast majority of the respondents had no difficulty in exemplifying these terms. In order to emphasise those features which require further practice in the classroom, the results in Table 2 are arranged in ascending order of wrong answers.

Table 2

Percentage of Responses for Each Pronunciation Feature (All Data)[4]

Pronunciation feature	right %	wrong %	right and wrong %	not applicable / not provided %
FLEECE	98%	0%	2%	0%
NURSE	75%	0%	6%	19%
/ŋ/ not followed by /g/	80%	0%	0%	20%
<ar> = START	57%	0%	0%	43%
<-ous> as /əs/	13%	0%	0%	87%
KIT	59%	1%	40%	0%
TRAP	97%	1%	1%	1%
LOT	98%	1%	0%	1%
to before C	77%	1%	3%	19%
contractions	92%	1%	0%	7%
<-s> forms as /z/	73%	1%	3%	23%
<-ing> forms as /ɪŋ/	80%	1%	7%	12%
voiced dental fricatives	93%	2%	4%	1%
weak forms	95%	2%	0%	3%
<-ed> forms of reg. v. as /t/	32%	2%	1%	65%
<a> = TRAP	86%	2%	0%	12%
<ea>, <ee> = FLEECE	86%	2%	0%	12%
triphthongs	33%	3%	1%	63%

[4] Percentages may not total 100 due to rounding.

Pronunciation feature	right %	wrong %	right and wrong %	not applicable / not provided %
voiceless dental fricatives	87	3	3	6
the before V	33	3	0	64
<-s> forms as /s/	55	3	0	42
<i> = KIT	89	3	4	3
dark l	71	4	4	21
<or> = STRUT, NURSE	26	4	2	67
centring diphthongs	59	5	22	14
spread V	65	5	25	4
final voiced C	78	5	6	11
<-ing> forms as /ɪn/	18	5	0	77
<(w)or> = NURSE	28	5	0	66
schwa as <a>	68	6	18	7
central V	80	6	12	2
the before C	88	6	0	5
to before V	15	6	3	76
<oo> = GOOSE	42	7	0	51
<-ate> as /ət/ (n./adj.), /eɪt/ (v.)	7	7	0	85
low V	76	9	12	3
linking r	38	9	0	53
rhoticity / non-rhoticity	78	11	0	12
strong forms	66	11	2	21
intrusive r	3	11	0	86
<-ed> forms of reg v. as /d/	39	11	3	47
<-ed> forms of reg v. as /ɪd/	23	11	2	64
schwa as <o>	53	16	16	16
aspiration	41	16	19	24
<-s> forms as /ɪz/	14	21	5	60
schwa as <u>	31	22	8	39
/ŋ/ followed by /g/	7	44	0	48

The results in the bottom lines of Table 2 reveal that one feature, /ŋ/ followed by /g/, was more confusing than all the others for the respondents, as it received 44% wrong responses and only 7% right ones (e.g., *single*). The respondents mistook the consonant for the letter and listed examples of words, in which /ŋ/ was not followed by /g/, however, in which the letter <g> was included in the spelling (e.g., *thinking*).

In addition, larger numbers of wrong answers were also found with reference to such features as: schwa spelt <u> (22%), <-s> forms pronounced as /-ɪz/ (21%), aspiration (16%) and schwa spelt <o> (16%). When it comes to schwa spelt <u> (22%) (e.g., *support, success, suppose),* the most typical mistake which students made was to treat the sequence of letters <ou> as the letter <u> (e.g., the weak form of *could* /kəd/). Other less frequent incorrect examples included words with a letter <u> not pronounced as schwa, (e.g., *perf**u**me*/ˈpɜːfj**u**ːm/) or words in which schwa was found but it was represented by a letter different from <u> (e.g., with <o> in *t**o**night*). On the other hand, the most frequently given good examples comprised function words *just, but,* and *must*.

As regards the feature <-s> forms pronounced as /ɪz/, only a small minority of the respondents correctly chose words in which the suffix <-s/-es> was pronounced as /ɪz/ after sibilants (14%). The majority of those who indicated this feature (26%) gave three types of

examples: words which ended with /ɪz/ regardless of their morphological structure, words in which the suffix <-s/-es> was pronounced as /z/ after voiced sounds (e.g., *eyes*), or words in which there was no suffix and where /ɪz/ was a part of a stem (e.g., *his*).

Aspiration created problems for 16% of the informants. Its erroneous cases included: no context for aspiration (e.g., *bones*), /pl-, pr-/ and /st-, sp-/ clusters (e.g., *plans*, *start*), and medial unstressed in unaspirated position of /p/ and /t/ (e.g., *sometimes*).

The analysis of erroneous responses concerning the feature schwa spelt <o> (16%) shows that respondents misinterpret schwa when it is a part of a GOAT diphthong (e.g., *so**, *hypo**) or a triphthong (e.g., *riot**). Among the good examples are both content words (e.g., *forget*) and functional words (e.g., *of*).

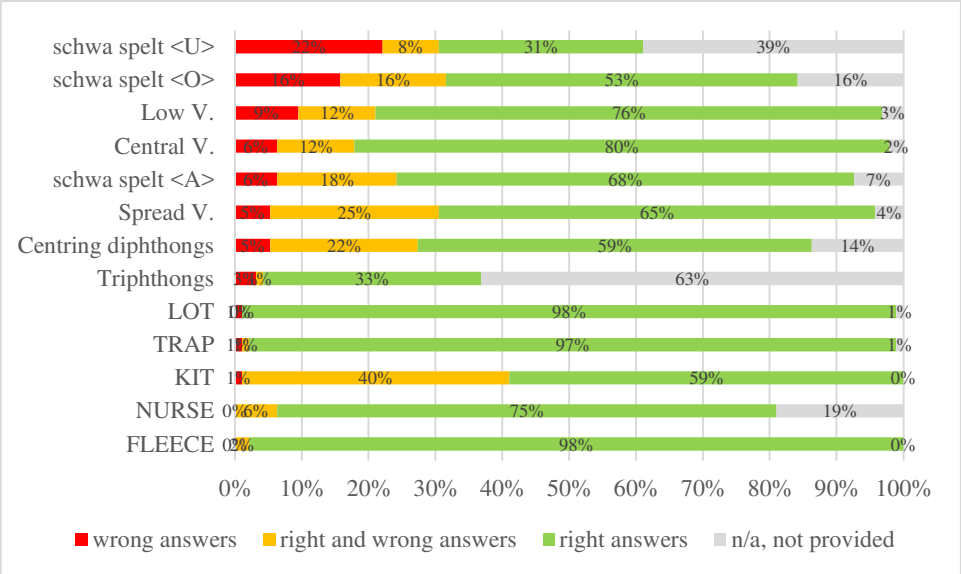
The next sub-sections (§3.2.1–3.2.5) present the results for each category of feature as listed in Table 1 (§2.4): vocalic features, consonantal features, connected speech features, inflected and derived forms, spelling-to-pronunciation patterns, and other features.

3.2.1 Responses for vocalic features

Figure 1 shows that for vowels, the percentage of 'right' answers outnumbers the combined 'wrong' and 'right and wrong' answers in all cases. Such features as FLEECE (98%), LOT (98%), TRAP (97%) and NURSE (75%), presented at the bottom of the graph, yielded a high percentage of correct examples (as shown by the green bar). Triphthongs (33%) were exemplified by one third of the group.

Figure 1

The Percentage of Four Types of Responses for Vocalic Features



Schwa spelt <u> and <o>, discussed above, comprises 30% and 32% of erroneous and partly erroneous answers, respectively. The reason for a high number of 'right and wrong' responses for the KIT vowel (40%) is the faulty inclusion of words with closing diphthongs, such as FACE (e.g., *hate*), PRICE (e.g., *die*) and less frequently CHOICE (e.g., *boy*) vowels, as well

as an occasional centring diphthong NEAR (e.g., *we're*) as an erroneous example of a word with a sole KIT vowel.

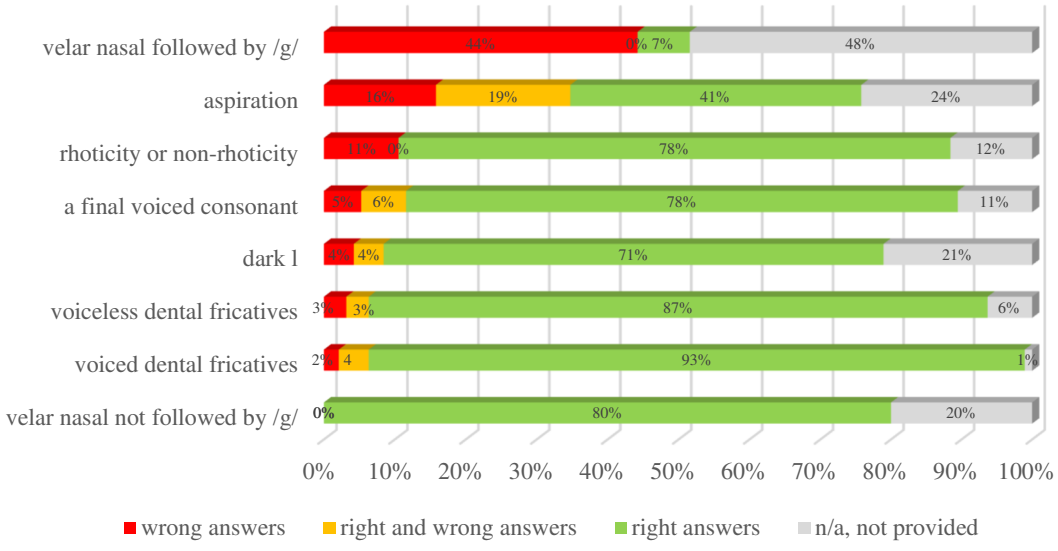
The graph also illustrates that centring diphthongs (27%) and the terms referring to the tongue position or lip shape for vowels — low (21%), central (18%) and spread (30%) — have not been understood by all the respondents. The centring diphthongs were replaced by closing diphthongs (e.g., *now*). The three vowel categories (low, central, spread) were exemplified with a random vowel; for example, low vowels were misrepresented by mid vowels such as DRESS in *never* or NURSE in *word* or diphthongs in *my*. The problem with schwa spelt <a> (24%), similarly to the above-mentioned schwa spelt <o> and <u>, was that the respondents did not pay attention to the letter <a>. Moreover, even if they did (e.g., *around, another*), they also provided other words in which schwa was represented by a letter different from <a>, or they did not recognise that schwa was a part of a diphthong NEAR (e.g., *real*) and not a sound on its own.

3.2.2 Responses for consonantal features

For consonants, summarised in Figure 2, the results show that around three-quarters of the group correctly exemplified all the features, except for: 1) /ŋ/ followed by /g/, which resulted in 44% wrong answers versus 7% correct ones; and 2) aspiration, with a similar ratio of right (41%) to wrong and partially wrong answers (35%). Among the consonantal features are: /ð/ (93%) and /θ/ (87%) correct examples, /ŋ/ not followed by /g/ (80%), (non-)rhoticity (78%), a final voiced consonant (78%) and dark [ɬ] (71%).

Figure 2

The Percentage of Four Types of Responses for Consonants



As regards the identification of whether the variety is rhotic or non-rhotic, 11% of the informants gave an erroneous answer to this question, mentioning rhotic as well as non-rhotic transcriptions (e.g., *burns* /bɜːnz/ or /bɜːrnz/), or words with the pre-vocalic /r/ (e.g., *cry*) which is common for both varieties. Interestingly, these results show that artists are using rhoticity

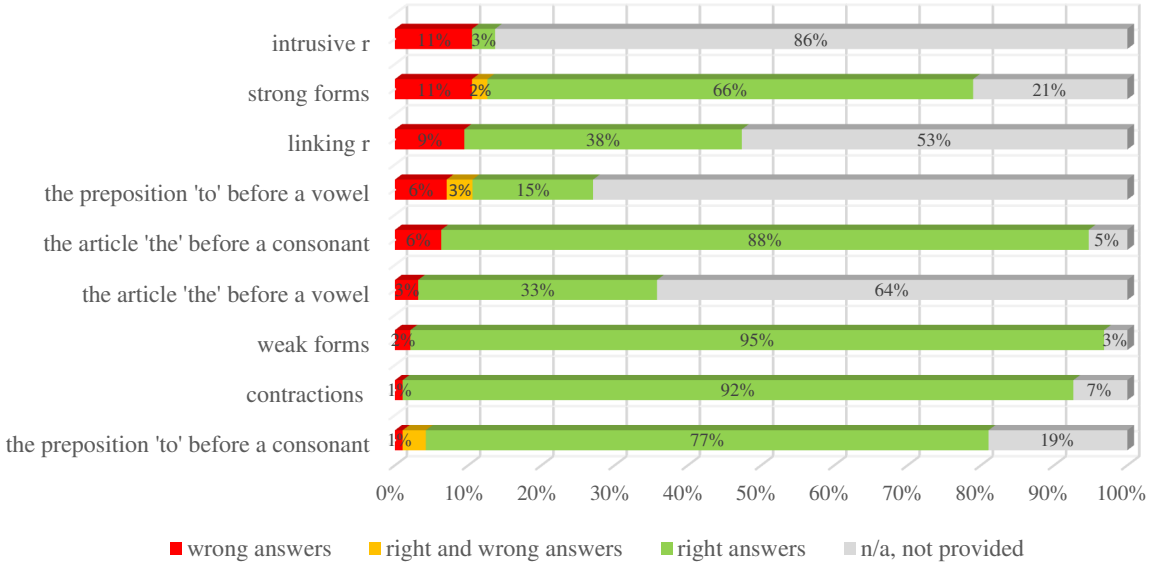
when singing, despite it not being a standard feature of their (expected) accent when speaking. Initially, when students categorised singers' accents, 55% were classified as non-rhotic while 45% as rhotic. However, 70% of the respondents listed examples of non-rhotic pronunciations and 30% of rhotic ones, which contradict the findings of Trudgill (1997), wherein British pop and rock singers reportedly tend to adopt an Americanised singing style. Thus, the results on rhoticity suggest that roughly 15% of the students might still not understand what constitutes an example of rhoticity.

3.2.3 Responses for connected speech features

The predominantly green bars of Figure 3 show reassuring results for connected speech elements; right answers outnumber wrong ones except for intrusive /r/, for which there were only 3% correct instances and 11% wrong ones. A high percentage of right answers for weak forms (95%), contractions (92%), the article *the* before a consonant (88%), the preposition *to* before a consonant (77%) and strong forms (66%) implies that the respondents understood their function in English.

Figure 3

The Percentage of Four Types of Responses for Connected Speech



One respondent whose lyrics did not include intrusive /r/, provided the following explanation "…you can actually hear the intrusive /r/ in some other Pink Floyd songs such as Astronomy Domine (e.g., *Oberon, Miranda‿**r**‿and Titania*) or Corporal Clegg (e.g., *umbrella‿**r**‿in the rain*)" (S.77). Wrong examples (11%) included: a) words with the right context for intrusive /r/, however, with no /r/ (n = 1) (e.g., *saw us* /sɔ: ʌs/); b) words with /r/ (n = 4) (e.g., *streets*); c) words with a silent letter <r> (n = 2) (e.g., *world* /wɜ:ld/); or d) an unrelated feature like linking (n = 3) (e.g., *away and*). As regards linking /r/, here 9% of the respondents gave wrong examples, which in fact exhibit rhoticity for American English (n =

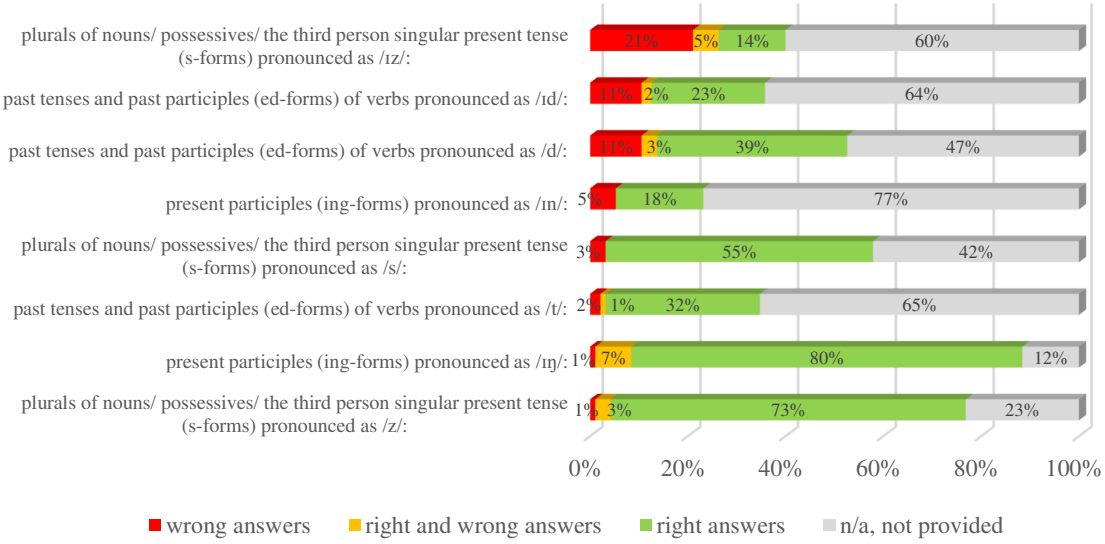
4) (e.g., *better* /ˈbetər/), non-rhoticity (*n* = 4) (e.g., *before you* /bɪˈfɔː ju/), or a wrong context for linking /r/, which is, namely, after a high vowel (*n* = 6) (e.g., *you own*, *I am*).

3.2.4 Responses for inflected and derived forms

The results for inflected and derived forms are presented in Figure 4. The only feature which scores low is <-s> forms pronounced as /ɪz/, discussed above under Table 2 (§3.2). In all other features, correct examples outnumber the incorrect ones, e.g., <-s> forms pronounced as: /z/ (73%) and /s/ (55%); <-ed> forms of verbs rendered as: /d/ (39%), /t/ (32%) and /ɪd/ (23%); and <-ing> forms articulated as: /ɪŋ/ (80%) or /ɪn/ (18%).

Figure 4

The Percentage of Four Types of Responses for Inflected and Derived Forms



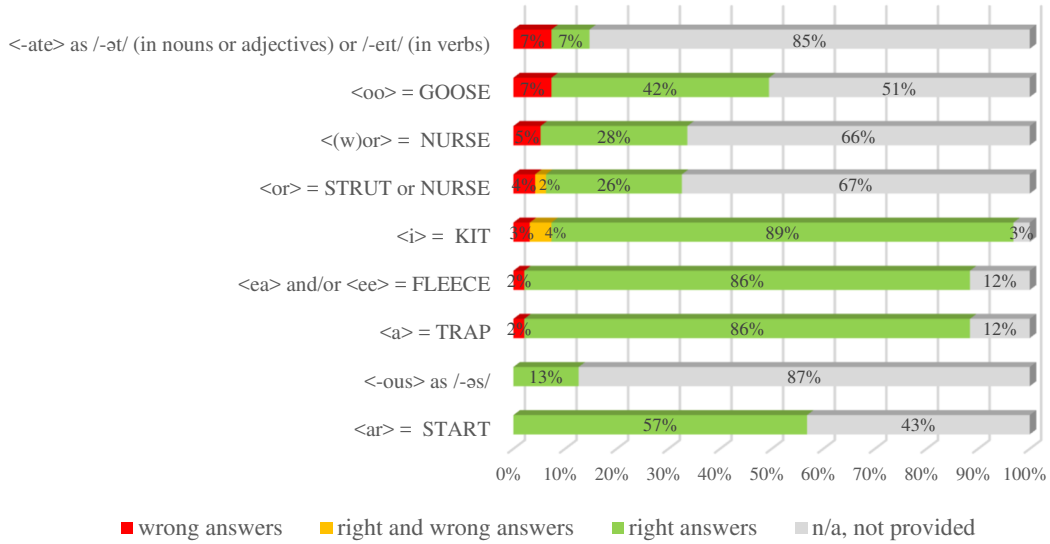
The pronunciation of <-s> forms as /z/ is a feature that is usually misarticulated by Polish respondents at the beginning of the course, owing to the lack of final voicing in the Polish language (Sobkowiak, 1996, p. 58). It is thus interesting to see the high percentage of right responses (73%). The incorrect responses for <-ed> forms pronounced as /d/ (11%) included irregular (e.g., *found*) and regular verb forms in which this suffix is pronounced as /ɪd/ after /t, d/ (e.g., *addicted*) or a modal verb (e.g., *could*). The wrong answers for <-ed> forms pronounced as /ɪd/ (11%) indicate that the informants only looked in their transcriptions for the /ɪd/ symbols without taking the morphological structure of the word into account. The erroneous examples included <-ed> forms articulated as /d/ (e.g., *cried*), a noun *side* and a contracted form *I'd*.

3.2.5 Responses for spelling-to-pronunciation patterns and other features

Letter-to-sound correspondences were less problematical for learners, with less than 7% of incorrect answers, as shown in Figure 5.

Figure 5

The Percentage of Four Types of Responses for Letter-to-Sound Correspondence



Moreover, some spelling-to-pronunciation patterns displayed high levels of correctness, for example: the letter <i> as KIT (89%), <a> as TRAP (86%), the sequences <ea> or <ee> as FLEECE (86%), <ar> as START (57%). Others, although less frequently identified, were also mostly right, for instance <oo> as GOOSE (42%), <(w)or> as NURSE (28%), <or> as STRUT or NURSE (26%) and a suffix <-ous> as /-əs/ (13%). The suffix <-ate> as /-ət/ in nouns or adjectives or /-eɪt/ in verbs obtained the same number of right (7%) and wrong (7%) answers respectively.

The questionnaire also inquired about proper names, non-grammatical forms, non-standard pronunciation and accent-specific pronunciation features. Among 24% of reported proper names there were, for example: *California*, *Elton John*, *Kool-Aid*, *Rome*, *Peter Pan*, *Bon Jovi*, *Frankie,* and *Tommy*. The listing of non-grammatical forms and non-standard pronunciation was meant as a starting point for a discussion about dialects; however, the results revealed that the respondents did not understand what it involved. Apart from four good examples of non-grammatical forms (e.g., *He don't play for respect*, *I been*, *It don't matter*, *she live*), the respondents listed such features as: a) assimilated forms (e.g., *gonna*, *wanna*, *gotta*); b) the contraction *ain't*; c) interjections and exclamations (e.g., *ugh*, *oh*, *ooh*, *whoa*, *wooh*); d) the suffix <-ing> as <-in> (e.g., *keepin'*); and e) and short forms (e.g., '*cause*, '*bout*, '*til*, and *yes* as *ya*, *yeah*). Among the non-standard pronunciation variants, the following features of casual connected speech were provided: a) suffix <-ing> as <-in> -e.g., *knowin*, *gonna*); b) *t*-elision (e.g., *li**tt**le bit*); c) a short form of *because* as *cos* and *about* as '*bout*; and d) *d*-elision (e.g., *rule**d** the*). When asked to list accent-specific pronunciation features, the informants limited their answers to either American or British English. The American features included: rhoticity

(e.g. *air*), LOT (e.g., *follow*), BATH (e.g., *can't*), *what's* as /wʌts/, the lack of a centring diphthong SQUARE in *compare* and NEAR in *year*, *worry* as /ˈwɜ:ri/ and *either* as /ˈi:ðər/. On the other hand, British features included: non-rhoticity (e.g., *California)*, LOT (e.g., *everybody*), a diphthong NEAR (e.g., *dear*), BATH (e.g., *glances*), and a diphthong SQUARE (e.g., *nightmares*).

3.3 Attractiveness and usefulness of the task

After completing the task, 76 of 95 respondents evaluated its attractiveness and usefulness. As regards the first criterion the results on a 5-point scale are as follows: attractive (63%), neither attractive nor unattractive (28%), extremely attractive (8%), not very attractive (1%), not attractive at all (0%). The usefulness of the task scored higher marks: useful (68%), neither useful nor useless (21%), extremely useful (8%), not very useful (2%), not useful at all (0%).

The detailed responses on the usefulness of the task revealed a wide selection of reasons. Each student could provide more than a single reason. The respondents reported that they had learnt the pronunciation of words (n = 17), approached phonetics by doing something enjoyable (n = 17), and explored differences between BrE and AmE and singer-unique accents (n = 12). The exercise also helped them to notice the difference in an artist's pronunciation in songs and in speech (n = 5). They also worked on transcription (n = 9), became more aware of stress, rhythm, stressed and unstressed syllables, stress-shift (n = 7), weak forms (n = 7) and sounds (n = 4), including intrusive /r/ (n = 3) and linking /r/ (n = 3). They also felt that they had improved their own pronunciation skills (n = 4) and learnt that phonetics could be useful in life (n = 3). Other single comments remarked upon the opportunity to revise theory and terms, appreciate the complexity of English phonetics, encounter different types of morphemes, and focus on accent(s).

Nearly half of the respondents (49%) found the task more attractive than other phonetics activities, 9% expressed a contrary opinion and 39% were undecided. Seventy open-ended responses show that what they liked most about the activity was the fact that they could work with the text of their own choice (71%), which gave them more control over the assignment: "the free nature of the task, no obligation regarding the type and genre of the song" (S.51). Three respondents described it as an engaging and enjoyable task, which involved more creativity and imagination. Others appreciated the combination of work with pleasure (n = 5), found it a nice change from the usual exercises (n = 6) and a more approachable way to learn. They found it interesting to learn to transcribe in this way and to look at the lyrics more thoughtfully from a different perspective. They recounted that they liked the topic itself because of the variety of issues (e.g., linking /r/) (n = 3) and admitted that this activity broadened their knowledge. They had a chance to listen to different accents and to see how an accent may change (n = 3). They were also able to compare spoken English to singing. They reported that working on lyrics with music helped them to: focus and become involved in the task, pronounce lyrics/some words correctly, remember new material, understand phonetics issues in a better way and see how they could use phonetics skills in real life. They also stressed the social aspect of the activity, as they got to know the musical tastes of their groupmates, they themselves had an opportunity to express themselves with a song of their choice, and this even helped them bond with fellow students who shared their interests.

However, 24% reported disliking the fact that the activity was time-consuming, arduous, and repetitive, involving listening and analysing (n = 12). They also disliked its complexity because the scope of the analysis was too vast, with too many features to be examined (n = 7); the self-study nature of the task (n = 2); presenting this task in front of the class (n = 2). After university lessons shifted online during the pandemic, the respondents did not appreciate

presenting their work, because of the technical challenges involved in simultaneously playing the audio and showing their transcription.

4 Discussion

This activity with a focus on form provided the students with opportunities to gain and revise phonetic knowledge at their own pace. They exercised their analytical skills by working on the lyrics of their own choice which they enjoyed.

The results show that the majority of the students were able to provide good examples of: a) the vowels LOT, FLEECE, TRAP; b) the consonants /θ/, /ð/, /ŋ/ not followed by /g/; c) the connected speech features such as weak forms, contracted forms, the article *the* pronounced as /ðə/ before a consonant; d) the letter-to-sound correspondences <i> = KIT, <a> = TRAP, and <ee> = FLEECE.

The respondents found this activity (extremely) attractive (71%), (extremely) useful (76%) and more attractive than other phonetic tasks (49%). Nevertheless, nearly one quarter of the cohort indicated the major drawback of this time-consuming task was that its scope was too large. What they appreciated most was the opportunity to exercise their phonetics expertise in a real-life activity, their autonomy in the selection of the material and the fact that they were able to learn phonetics by being actively involved in transcription and analysis, owing to their personal relationship with the song. According to the respondents, the major benefits of this task involved: learning the pronunciation of words, revising the theory of phonetics and focusing on accent(s) or the artist's non-standard pronunciation in a song. One respondent cited the example of Sting's northern rendition of BATH in British English in his 1993 song, The Shape of My Heart, where "he pronounces *the sacred geometry of chance* with *chance* as /tʃæns/ although he is British" (S.6).

There are limitations to the study and some shortcomings of the procedure. The results could have been different if the 'not provided' category had been accounted for and if it had constituted a separate category, not combined with 'non- applicable'. The design could also be improved to meet the students' needs better, most notably by reducing the vast scope of the task and the time needed for its completion. Thus, restricting the number of features would be recommended.

5 Conclusion and implications

The foremost conclusion is that the following features present obvious challenges to Polish university English first-year students: /ŋ/ followed by /g/, aspiration, <-s> forms pronounced as /ɪz/, KIT vowel, spread vowels, centring diphthongs, and schwa spelt <o>, <u>, <a>. These features should therefore be practised several times during a course.

Through this activity, the respondents reviewed features of English phonetics and showed whether they had learnt and understood the terms and concepts, which was the main aim. They gained new skills such as typing transcription and phonetic symbols, making use of the online transcription application Photransedit, and performing in front of the class or online, which involved managing audio and transcription. They reflected on different pronunciation features by paying attention to details when they compared canonical dictionary transcriptions with the actual pronunciation of the artist.

The self-study character of the activity permitted them to revise the material at their own pace. Using online tools for phonetic feature identification required autonomy and was meant to raise students' awareness of phonetics. It is hoped that they became more confident in their own phonetic expertise. In addition, they were encouraged to do some extra reading on an

artist's origin, place of residence and accent to verify whether their assumptions were right, which served as a prelude to a discussion on accent variation.

To conclude, this is a task which has its place in a phonetics course. Students appreciated being able to choose their own song to analyse, and the combination of words with music seemed to constitute a pleasant, entertaining element in their learning. As each student worked on their own text, the responsibility for providing good answers was shifted onto them. They could experience how phonetics works in the language and how they can apply it to examine speech and draw conclusions about an accent.

References

Aini, H. W. N., Basri, H., & Hastini. (2013). Improving students' pronunciation of alveopalatal sounds through English songs. *e-Journal of English Language Teaching Society* (ELTS), *1*(1), 1–13.

Ashtian, F. T., & Zafarghandi, A. M. (2015). The effect of English verbal songs on connected speech: Aspects of adult English learners' speech production. *Advances in Language and Literary Studies*, *6*(1), 212–226. https://doi.org/10.7575/aiac.alls.v.6n.1p.212

Bailles, F. Zhang, Y, Cheng, Y. Bu, Y., & Prieto, P. (2021). Listening to songs and singing benefitted initial stages of second language pronunciation but not recall of word meaning. *Language Learning 71*(2), 369–413. https://doi.org/10.1111/lang.12442

Barrett, C. (2015). *Finding our rhythm: Contextualizing second language development through music-based pedagogy*. (Publication no. 1602). [Doctoral dissertation, University of Pennsylvania]. University of Pennsylvania Repository. https://repository.upenn.edu/edissertations/1602

Degrave, P. (2021). The contribution of prosody to intelligibility, comprehensibility and accentedness in foreign language acquisition: Can music help? *Nederlandse Taalkunde 26*(2), 273–298. https://doi.org/10.5117/NEDTAA2021.2.006.DEGR

Good, A. J., Russo, F. A., & Sullivan, J. (2015). The efficacy of singing in foreign-language learning. *Psychology of Music, 43*(5), 627–640. https://doi.org/10.1177/0305735614528833

Hancock, M. (1999). *Singing grammar: Teaching grammar through songs*. Cambridge University Press.

Kanlayanee, P. (2012). *The effect of teaching English pronunciation through songs of Prathomsuksa 4 Students at Watratchaphatigaram School* [Master's thesis]. Srinakharinwirot University. http://thesis.swu.ac.th/swuthesis/Tea_Eng_For_Lan(M.A.)/Kanlayanee_P.pdf

Murphy, T. (2013). *Music and song: Resource books for teachers*. Oxford University Press.

Nakataa, H., & Shockey, L. (2011). The effect of singing on improving syllabic pronunciation: Vowel epenthesis in Japanese. In W. S. Lee & W. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII 2011): August 17-21, 2011* (pp. 1442–1445). Hong Kong: City University of Hong Kong.

Nowacka, M. (2022). English phonetics course: University students' preferences and expectations. *Research in Language*, *20*(1), 71–84. https://doi.org/10.18778/1731-7533.20.1.05

Saldiraner, G., & Cinkara, E. (2021). Using songs in teaching pronunciation to young EFL learners. *PASSA: Journal of Language Teaching and Learning in Thailand*, *62*, 119–141.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, *106*(2), 975–83. https://doi.org/10.1016/j.cognition.2007.03.005

Stefani, J., Basri, H., & Ohoiwutun, J. E. (2015). Improving the pronunciation through listening to English songs. *e-journal of English Language Teaching Society (ELTS)*, *3*(2), 1–10.

Sobkowiak, W. (1996). *English phonetics for Poles: A resource book for learners and teachers*. Bene Nati.

Tegge, F. (2018). Pop songs in the classroom: Time-filler or teaching tool? *ELT Journal*, *72*(3), 274–284. https://doi.org/10.1093/elt/ccx071

Tizian, S. (2016). *Listen and learn: An experiment on the effectiveness of songs in English L2 pronunciation* [Master's thesis]. Università degli Studi di Padova. https://thesis.unipd.it/bitstream/20.500.12608/20149/1/SILVIA_TIZIAN_2015.pdf

Trudgill, P. (1997). Acts of conflicting identity: The sociolinguistics of British pop-song pronunciation. In N. Coupland & A. Jaworski (Eds.), *Sociolinguistics: A coursebook and Reader* (pp. 251–270). Palgrave Macmillan.
Walker, R. (2006). Going for a Song. *English Teaching Professional, 43*, 19–21.
Wells, J. C. (1982). *Accents of English*. Cambridge University Press.
Wilcox, W. B. (1995). *Music cues from classroom singing for second language acquisition: Prosodic memory for pronunciation of target vocabulary by adult non-native English speakers*. [Doctoral dissertation]. University of Kansas ProQuest Dissertations Publishing. 9544866.
Wodecki, R. (2014). *Using pop songs to improve pronunciation in the EFL classroom* [Master's thesis]. Masaryk University. https://is.muni.cz/th/pcgob/?lang=en

Appendix

A List of Most Frequently Chosen Songs

Adele. (2008). Chasing Pavements [Song]. On *19*. XL Recordings; London.
Adele. (2011). Set Fire to the Rain [Song]. On *21*. XL Recordings; London.
Adele. (2011). Someone Like You [Song]. On *21*. XL Recordings; London.
Adele. (2015). Hello [Song]. On *25*. XL Recordings; London.
Adele. (2015). When We Were Young [Song]. On *25*. XL Recordings; London.
Sheeran, E. (2013). I See Fire [Song]. On *Hobbit: The Desolation of Smaug*. Water Tower Music & Decca; Burbank & London.
Sheeran, E. (2014). Thinking out Loud [Song]. Asylum & Atlantic; New York.
Sheeran, E. (2017). Castle on the Hill [Song]. On Asylum & Atlantic; New York.
Sheeran, E. (2021). Bad Habits [Song]. Asylum & Atlantic; New York.

About the author

Marta Nowacka teaches English phonetics and pronunciation at the University of Rzeszów in Poland. Her research concentrates on the Polish-accented speech of university students of English and the methodology of teaching phonetics to adults in an EFL context. She has co-authored two course-books on English pronunciation: *How much wood would a woodchuck chuck?*: *English pronunciation practice book* (Mańkowska et al., 2009) and *Sally meets Harry: Primer to English pronunciation and spelling* (Nowacka et al., 2011).

Email: mnowacka@ur.edu.pl

Pesantez, A. & Dellwo, V. (2023). A longitudinal study of individual difference in foreign language pronunciation development: The case of vowel production in Ecuadorian learners of English. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 214–224). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225313

A longitudinal study of individual difference in foreign language pronunciation development: The case of vowel production in Ecuadorian learners of English

Alejandra Pesantez
University of Zurich

Volker Dellwo
University of Zurich

This study investigated the development of the English vowel contrasts /i – ɪ/, /u – ʊ/, /ɛ – æ/, and /ɑ – ʌ/ produced by Ecuadorian Spanish speakers enrolled in the First and Second Language Teaching program at a State University in Ecuador. Recorded at six-month intervals at the end of the third, fourth, and fifth semesters, they produced 40 isolated English monosyllabic words. Each block of vowel contrasts had familiar words representing five words per vowel in a CVC and CVCC context. To examine pronunciation differences in the acoustic vowel space development over time, the Euclidean distance between the four groups of vowel contrasts was calculated based on F1 and F2 frequency values, which were Lobanov-normalised and analysed using a mixed-effects model. The results showed that Ecuadorian learners' pronunciation did not vary in the general learning trajectories over the three recordings, and did not match the native English criteria by the end of the study. However, Ecuadorian learners showed high variability *within* subjects and *between* subjects in the production of each vowel contrast over time. In summary, these results suggest that within-group, not all Ecuadorian learners followed the same developmental trajectories and varied depending on the vowel contrast.

Keywords: individual differences, L2 language, developing, vowel contrast

1 Introduction

English is a widespread international language with many second language (L2) speakers from different countries who commonly differ on how accurately they produce English vowel sounds. This level of accuracy difference has also been reported in many studies pointing out that L2 English learners' difficulty in perceiving or producing specific sounds (consonants or vowels) can vary depending on the first learners' language (Flege et al., 1997; McAllister et al., 2002). Furthermore, several studies on group observation have reported that L2 learners' pronunciation improvement may occur at the beginning of the learning process in a naturalistic environment (Flege et al., 1997; Thomson & Derwing, 2015). Although research on group data is highly informative, we believe there is much to be gained by examining individual differences in L2 pronunciation in a non-naturalistic setting. To identify specific predictors of phonological acquisition, studies on L2 learners' variability have focused on learner-internal factors such as the role of awareness in pronunciation, motivation, language use, anxiety, aptitude, among other factors (Baran-Łucarz, 2014; Elliott, 1995; Saito et al., 2019). However, the results of these studies reflect neither learner-external factors such as the quantity and quality of input and opportunity of language use (e.g., Flege & Wayland, 2019; Moyer, 1999; Muñoz, 2011), nor their impact on learner's developmental trajectories (Pesantez & Dellwo, 2022). Moreover, literature on L2 individual variation is limited and focused more on L2 perception (Kim et al., 2018; Kogan & Mora, 2022; Mayr & Escudero, 2010; Morrison, 2009; Munro et al., 2015) than L2 production (Lima, 2019; Munro et al., 2015). For example, Morrison (2009) found that L1 Spanish L2 English listeners who had exposure to the same Canadian English dialect followed an indirect developmental path in perception patterns. Furthermore, Munro et al. (2015) found high variability in consonant clusters for both onset and coda developmental performance within Mandarin and Slavic learners of English.

The current study extends previous longitudinal research on the development of English vowels produced by Brazilian learners (Lima, 2019). Both Portuguese and Spanish learners of English have problems producing vowel contrasts (Rallo Fabra & Romero, 2012). Data from several studies suggest that L1 Spanish speakers rely on vowel duration as acoustic cues to make difference between vowels, even though contrast on durational cues does not exist in the Spanish language (e.g., Escudero & Boersma, 2004). We will explore longitudinal learner performance on the English vowel contrasts /i – ɪ/, /u – ʊ/, /ɛ – æ/, and /ɑ – ʌ/ produced by a group of L1 Ecuadorian (EC) Spanish speakers. They share the same number of years of exposure in a classroom setting and have the same exposure to Spanish-accented English teachers. The choice of these vowel contrasts is intentional as they are crucial for L2 English intelligibility (e.g., Levis, 2020; Munro & Derwing, 2008; Rallo-Fabra & Romero, 2012). The study also explores individual learner development, i.e., it assesses L2 learners on their own individual pronunciation problems instead of focusing on group errors.

2 Methodology

2.1 Research questions

The study aims to answer the following research questions:

RQ1: Do Ecuadorian learners vary in their general learning trajectories (i.e., pronunciation changes over time) as they advance in their studies?

RQ2: To what extent do Ecuadorian learners vary in their developmental trajectories across the four groups of vowel contrasts?

2.2 Participants

Two groups of speakers were analysed, an L1 English group and an L2 English group. The first group was made up of six L1 American English (NE) speakers who were studying Spanish as a second language in Ecuador, including one who was an English teacher. The three females and three males had a mean age of 23.6 years (range: 19–43). The L2 group consisted of twenty-four undergraduate students enrolled in the First and Second Language Teaching program at a State University in Ecuador. They were L1 Ecuadorian (EC) Spanish monolinguals who had never been in an English-speaking country. In this paper, we present only the data for the female participants (n = 16), because vowel formant ranges vary with gender. Their ages ranged from 18 to 28 years old with a mean of 21.13. Fourteen participants reported having not studied any foreign language other than English, but two participants had studied Portuguese and Quechua. They also had experience with English learning, having had four hours per week at secondary school in Ecuador with non-native English input most of the time. All participants gave their written informed consent at each recording session and were paid for their participation.

2.3 Description of the learning context

The participants were recorded at six-month intervals at the end of the third semester (recording 1), fourth semester (recording 2), and fifth semester (recording 3). Each semester had sixteen weeks of classes (excluding weeks of mid-term exams and final exams). They studied Research Methodology, TEFL, Communicative Grammar, Reading, and Writing – among other courses – from the first to the fifth semesters, as well as Phonetics in the fourth semester and Phonology in the fifth semester. In Phonetics, the participants learnt about the sounds of the English language without a context; they focused on using phonetic symbols (vowels and consonants) to understand English pronunciation. In Phonology, they focused on the organisation of English sound patterns and those present in L2 English speech. Over the period of the three recordings, the participants had only the L2 input from L1 Spanish teachers of English in their classes. At the end of each testing session, they completed a general language background questionnaire to measure their exposure to the English language in the classroom (number of English courses taken), input (native or non-native), and the frequency of their use of English.

2.4 Production task

The experiment was conducted at the radio station of the State University in Ecuador. The participants produced 40 isolated English monosyllabic words with no carrier phrases. The instructions were written in English and appeared on the first slide. Each vowel pair tested 10 familiar words: five words per vowel in a CVC and CVCC context, and 5 disyllabic words added as distractors. The test words included:

- /i – ɪ/: *sheep–ship, cheap–chip, keys–kiss, feet–fit, seat–sit*;
- /u – ʊ/: *boot–book, fool–full, pool–pull, food–foot, suit–soot*;
- /ɛ – æ/: *bet–bat, pen–pan, men–man, send–sand, beg–bag*;
- /ɑ – ʌ/: *dog–dug, hot–hut, caught–cut, boss–bus, lock–luck.*

A Zoom H2n Handy recorder at 44.1 kHz sampling, 16-bit quantisation was used for the three recordings. Before starting the recordings, participants were familiarised with the task in a preliminary trial, where the presented words were unrelated to the words of the study. No recording was made during the familiarisation phase. Using a computer screen, the pictures

were presented in a random order (mostly tense vs. lax vowel) with their Spanish translation next to them, to avoid the effects of orthography in the production of the English segment phonemes (Bassetti, 2017; Flege et al., 1997). In cases where the participants did not say the expected word, the researcher said "No" and the participant could self-correct by saying the correct word. All the participants repeated each word two times to counter frequent hesitations during the initial production. They were asked to say the words in a natural way and at a normal volume. NE speakers followed the same procedure.

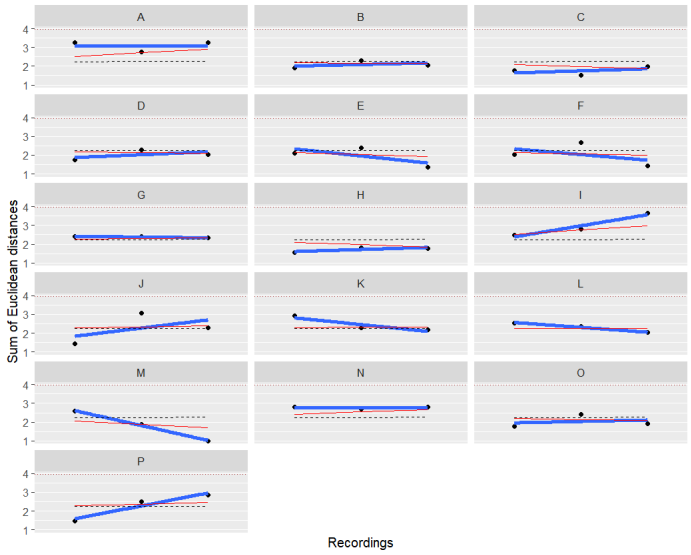
2.5 Acoustic measurements

Our approach to evaluating L2 speakers' performance relies on Euclidean distance (ED) values of the vowel contrasts. We compared the ED values with values from English native speakers which served as baseline. However, the native speakers' data was not used in the statistical models. The NE values were only used for comparison, as the aim of the study was not to test participants' native-likeness.

To compare the individual changes in the spectral characteristics over time, the onset and offset of each vowel segment were manually annotated by following standard phonetic criteria, using waveform, spectrogram, and auditory discrimination cues in Praat (Boersma & Weenink, 2022). The F1, F2, and F3 formants means were extracted using an automatic formant measurement method in Praat. The same parameters used by Kartushina and Martin (2019) were also applied in this study. The F1 and F2 mean values obtained by each vowel setting were examined to see if their values were within the speaker's vowel formant gender ranges. F1 and F2 formants were Lobanov-normalised by semester and grouped by speakers using the visible vowel website (Heeringa & Van de Velde, 2018). The ED between the four groups of vowel contrasts was calculated based on F1 and F2 frequency values. Then the ED values for each vowel contrast were plotted to trace individual vowel development over time (see Figure 1). The distance values were fitted into a linear-mixed effects model to the data.

Figure 1

Mixed-Effects Model of the Sum of ED of the Four Pairs of English Vowels by Recording for Each Speaker



Notes. Time is ordered, from left to right. The speaker code-letter is in the strip above each panel.

3 Results

R core Team (2022) was used for statistical analysis and figure plotting. For the purposes of RQ1 (i.e., to look at the variability in the general vowel development of the group), the sum of the ED of the four target pairs of vowels was used to fit a linear mixed-effects model (LMM) to the data. The fixed effect was coded as Recordings, and a random effect for subject and a random effect for the slope with respect to recordings (see Bates, 2010, for a review). In Figure 1, the three black dots are the sums of ED for each recording, and the blue line (which is the tendency line for each speaker from a simple within-subject linear model) indicates little evidence of vowel development. The general tendency of the group (the black dashed line) which is repeated in every individual plot, does not reflect an increasing slope over the three recordings. The red line is the result of the LMM fitted for each speaker's data. It indicates no subject-to-subject variation over the three recordings. Finally, the dotted line, with no slope and repeated in all individual plots at 3.96, marks the sum of the ED of the six NE speakers showing how each participant differs from NE data. The results of the LMM showed no significant individual difference in the general learning trajectories over the three recordings $F(1, 28.79) = 0.0041$; $p > 1$.

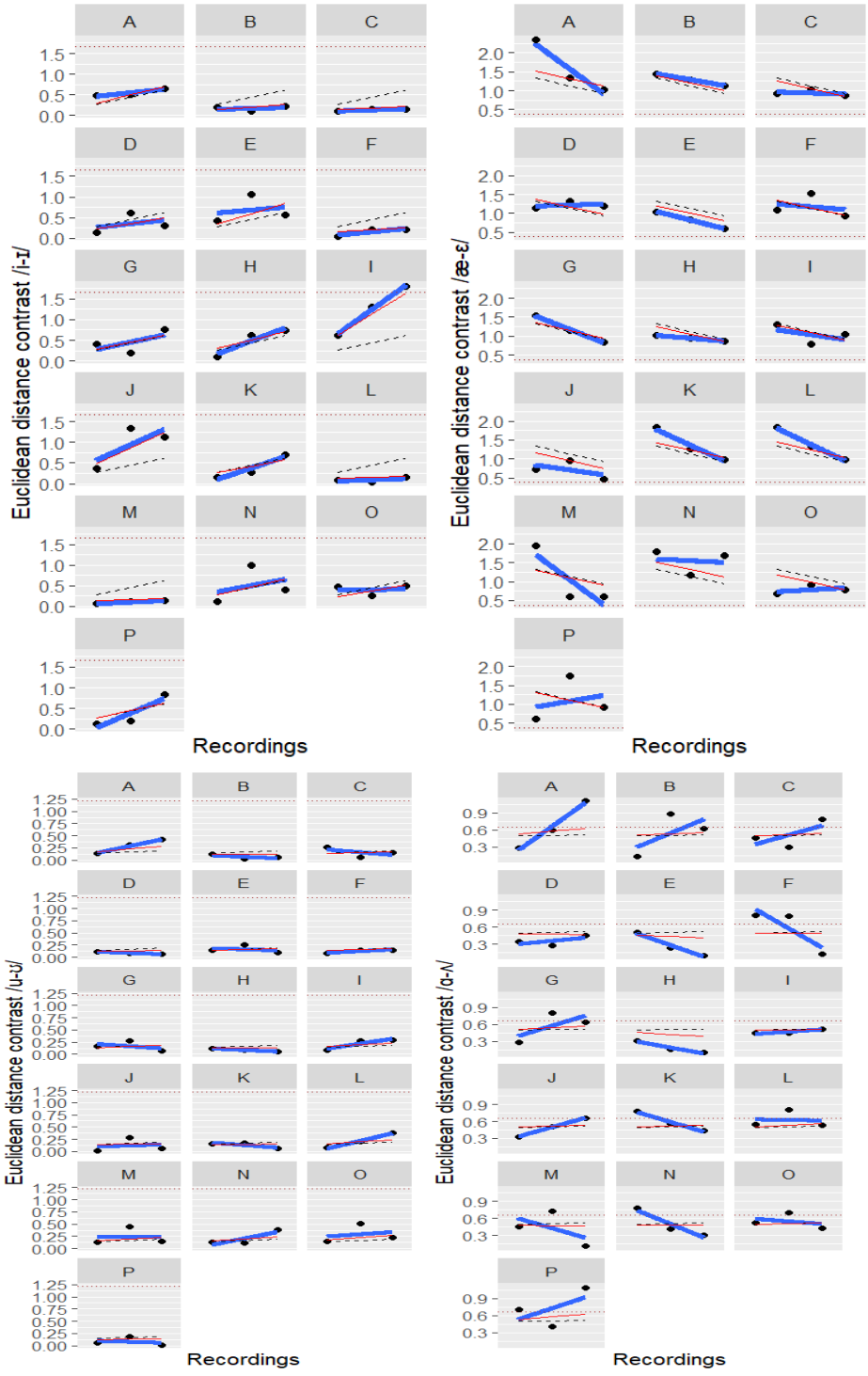
We can also observe different learning trajectories within-subject (the blue line). For example, of the sixteen EC participants, six started with ED values comparable to the mean of the group and four of them showed no improvement in recording 2 and 3 (B, D, G, O). Interestingly, only three participants improved the ED values over the three recordings (I, J, P). Nonetheless, there were some participants who decreased their ED values (E, F, K, L, M) over the three recordings. Furthermore, four participants got ED values above or below the

population (A, C, H, N). In general, these data suggest that there are all types of L2 developmental routes within subjects, even though in the group no general different change rate was found.

For the purposes of RQ2, i.e., to explore how EC learners vary in their developmental trajectories across the four groups of vowel contrasts, we used an LMM per each vowel contrast. Figure 2 shows changes in the ED values (three black dots) per each participant, and with the tendency to increase (blue line) for some participants (simple within subject linear model) for the contrast /i – ɪ/ and /ɑ – ʌ/, a clear trend of decreasing for the /æ – ɛ/ and a steady trend for /u – ʊ/. What stands out in these figures is the difference in the subject-to-subject variation over the three recordings (red line, representing the result of the LMM fitted for each speaker's data) and the variability within-subject with respect to the general tendency of the group. Finally, the dotted line with no slope for the /i – ɪ/, /u – ʊ/ (left panels), for /ɑ – ʌ/ and /æ – ɛ/ (right panels) represent the NE's ED at 1.66, 1.22, 0.66, and 0.38, respectively. The results of the LMM were significant for the /i – ɪ/ F (1, 38.015) = 10.381; $p < .01$, and for the /æ – ɛ/ F (1, 31) = 11.239; $p < .01$, but not for the /u – ʊ/ F (1, 44.372) = 1.271; $p > 1$, and /ɑ – ʌ/ F(1, 42.651) = 0.0637; $p > 1$.

Figure 2

Mixed-Effects Model of ED for the /i/–/ɪ/, /æ/–/ɛ/, /u/–/ʊ/ *and* /ɑ/–/ʌ/ *of the English Vowels by Recording for Each Speaker*



three dots = ED blue line = within-subject red line = mixed model

black dashed line = group dotted line = native speaker

Note. Time is ordered, from left to right. The speaker code-letter is in the strip above each panel.

Of the four groups of vowel contrasts, the /i – ɪ/, /æ – ɛ/ and /ɑ – ʌ/ showed higher within-subject variability in their L2 production. As observed in Figure 2, for the /i – ɪ/ contrast, six participants (B, D, G, K, N, O) started with ED corresponding to the mean of the group and only four participants followed the developmental performance of the group. Interestingly, only one participant (I) reached NE standards in recording 3. For the contrasts /æ – ɛ/, three participants (B, F, I) started with ED comparable to the mean group and three followed the developmental performance of the group. For the /ɑ – ʌ/ pair, there were three participants (C, E, I) who started with ED similar to the group mean, but only one (I) followed the development of the group. For the /u - ʊ/ pair, most of the participants started and kept the development of the group, but only four participants slightly increased the ED over the three recordings (A, I, L, N). These findings mean that these participants show higher variability in L2 speech development in front vowel contrasts than the back vowels over the three recordings.

4 Discussion

Our study provides evidence of EC learner's pronunciation not changing over time. This finding is consistent with Herrero and Delicado (2022) whose L2 learners also showed minimal change in the quality of the vowels. The relatively stable L2 productions in the sum of the ED found here may reflect that EC learners need more years of English instruction to improve their L2 production of the vowel contrast (Muñoz, 2011). However, it might be also related to a degree of early fossilisation that occurred during the first or second semester, as has been reported in other studies (Derwing & Munro, 2014; Munro & Derwing, 2008). To determine which vowel contrasts show high variability in the developmental performance between EC learners, we analysed each vowel contrast separately. The results for the front vowel contrasts showed statistically significant differences in learners' developmental path, but not for the back vowel contrasts. For the contrasts /æ – ɛ/, we expected that EC learners would start with larger spectral difference than NE speakers due to the influence of Spanish /e/ in the English /ɛ/ production. This finding is consistent with that of Flege et al. (1997), where Spanish participants produced larger spectral differences than did the group of native speakers. The results in our study differ from Lima (2019) who found that most participants created separate categories for the vowel pairs over the four recordings and maintained them in most cases. In our study, participants did not produce vowel contrasts with similar ED values to NE speakers over the three recordings, and those who could increase the ED values lost them in the second or third recording. Moreover, in Lima's (2019) study, many participants started in the first recording by producing a vowel contrast for the /i – ɪ/, whereas none of our participants started with higher ED values in recording 1 for this vowel pair. However, both studies show the high degree of variability in L2 acquisition and in the type of developmental patterns L2 learners follow. We did not analyse intra-subject variability statistically per each vowel, but as Figure 2 shows, most EC learners vary in their own developmental trajectory (blue line) for the vowel pair /ɑ – ʌ/. For example, speaker P started with an ED (three dots) comparable to NE speakers in recording 1, changed the order in recording 2, and surpassed NE in recording 3.

In general, this study showed that over the three recordings fewer than four participants followed the mean trend of the group for each vowel pair. Given this finding, a more individualised approach would be a more effective pedagogical strategy for these speakers, especially for /i – ɪ/ and /æ – ɛ/. In the context of foreign language acquisition with major L2 foreign-accented input, it would be better to start with phonetics and phonology classes from the beginning of instruction to avoid fossilised errors in pronunciation. In addition, integrating applications of automatic speech recognition into pronunciation instruction can help teachers

to pinpoint problems in L2 sounds, and at the same time make L2 learners aware of their own pronunciation challenges.

5 Conclusion

This study explored EC speakers' variability in the development of eight English vowels in an instructed foreign language context, with most input supplied by non-native English teachers over five semesters. In most such classrooms, L2 learners have major exposure to foreign-accented input. The nature of that L2 input is crucial in the process of speech learning, with adults' performance depending on the quantity and quality of input (Flege & Bohn, 2021). Studies tracking L2 learners and their non-native English teachers might give a better understanding of how foreign-accented input activates L2 phonetic acquisition.

Valuable insight is gained about learners of English in the instructed foreign language context of Ecuador, a country underrepresented in the field of L2 English. However, the small sample size means that conclusions cannot be generalised to all L1 Spanish learners. Nonetheless, this paper contributes to an important line of L2 pronunciation research related to individual differences and their changes over time, showing that not all vowels are acquired at the same rate and that very few learners follow the mean of the group. More research with larger sample sizes and more words representing each vowel needs to be conducted in both perception and production, to further examine long-term L2 development.

Acknowledgements
This project was funded partially by the University of Zurich-GRC (Graduate Campus) Travel Grants, 2019-2020.

References

Baran-Łucarz, M. (2014). The link between pronunciation anxiety and willingness to communicate in the foreign-language classroom: The Polish EFL context. *Canadian Modern Language Review*, *70*, 445–473. http://doi.org/10.3138/cmlr.2666

Bassetti, B. (2017). Orthography affects second language speech: Double letters and geminate production in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1835–1842. https://doi.org/10.1037/xlm0000417

Bates, D. M. (2010). *lme4: Mixed-Effects Modelling with R.* Available online at: http://lme4.r-forge.r-project.org/lMMwR/lrgprt.pdf

Boersma, P., & Weenink, D. (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.21. http://www.praat.org/

Derwing, T., & Munro, M. J. (2014). Once you have been speaking a second language for years, it's too late to change your pronunciation. In L. Grant & D. Brinton (Eds.), *Pronunciation myths: Applying second language research to classroom teaching* (pp. 34–55). University of Michigan Press.

Elliott, A. R. (1995). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *The Modern Language Journal*, *79*(4), 530–542. https://doi.org/10.1111/j.1540-4781.1995.tb05456.x

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 551–585. https://doi.org/10.1017/S0272263104040021

Flege, J. E., Bohn, O-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470. https://doi.org/10.1006/jpho.1997.0052

Flege, J. E., & Wayland, R. (2019). The role of input in native Spanish late learners' production and perception of English phonetic segments. *Journal of Second Language Studies*, *2*(1), 1–44. https://doi.org/10.1075/jsls.00004.fle

Flege, J. E., & Bohn, O-S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901

Heeringa, W. & Van de Velde, H. (2018). Visible Vowels: A tool for the visualization of vowel variation. In *Proceedings CLARIN Annual Conference 2018* (8-10 October, Pisa, Italy). CLARIN ERIC.

Herrero de Haro, A., & Delicado Cantero, M. (2022). Longitudinal study of Spanish vowel acquisition by Australian students. *Language Teaching Research*. Advance online publication. https://doi.org/10.1177/13621688221140106

Kartushina, N., & Martin, C. D. (2019). Third-language learning affects bilinguals' production in both their native languages: A longitudinal study of dynamic changes in L1, L2 and L3 vowel production. *Journal of Phonetics*, *77*, 1–21. https://doi.org/10.1016/j.wocn.2019.100920

Kim, D., Clayards, M., & Goad, H. (2018). A longitudinal study of individual differences in the acquisition of new vowel contrasts. *Journal of Phonetics*, *67*, 1–20. https://doi.org/10.1016/j.wocn.2017.11.003

Kogan, V. V., & Mora, J. C. (2022). The effects of individual differences in native perception on discrimination of a novel non-native contrast. *Laboratory Phonology*, *24*(1), 1–28. https://doi.org/10.16995/labphon.6431

Lima, R. M., Jr. (2019). A longitudinal study on the acquisition of six English vowels by Brazilian learners. In S. Calhoun, P. Escudero, M Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 3180–3184). Australasian Speech Science and Technology Association Inc.

Levis, J. (2020). Revisiting the intelligibility and nativeness principles. *Journal of Second Language Pronunciation*, *6*(3), 310–328. https://doi.org/10.1075/jslp.20050.lev

Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition*, *13*, 279–297. https://doi.org/10.1017/S1366728909990022

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*(2), 229–258. https://doi.org/10.1006/jpho.2002.0174

Morrison, G. S. (2009). L1-Spanish speakers' acquisition of the English /i/−/ɪ/ contrast II: Perception of vowel inherent spectral change. *Language and Speech*, *52*(4), 437–462. https://doi.org/10.1177/0023830909336583

Moyer, A. (1999). Ultimate attainment in L2 Phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition, 21*(1), 81–108. https://doi.org/10.1017/S0272263199001035

Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, *58*(3), 479–502. https://doi.org/10.1111/j.1467–9922.2008.00448.x

Munro, M. J., Derwing, T. M., & Thomson, R. I. (2015). Setting segmental priorities for English learners: Evidence from a longitudinal study. *International Review of Applied Linguistics in Language Teaching*, *53*(1), 39–60. https://doi.org/10.1515/iral-2015-0002

Muñoz, C. (2011). Input and long-term effects of starting age in foreign language learning. *International Review of Applied Linguistics in Language Teaching*, *49*, 113–133. https://doi.org/10.1515/iral.2011.006

Pesantez, A., & Dellwo, V. (2022). L2 phonetic accuracy development in a foreign language setting: a longitudinal study. In B. Blecua, J. Cicres M. Espejel & M. J. Machuca (Eds.), *Propuestas en fonética experimental: enfoques metodológicos y nuevas tecnologías* (pp. 225–228). https://dugi-doc.udg.edu/handle/10256/20770

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rallo Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics*, *40*, 491–508. https://doi.org/10.1016/j.wocn.2012.01.001

Saito, K., Suzukida, Y., & Sun, H. (2019). Aptitude, experience and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition, 41*, 201–225. https://doi.org/10.1017/S0272263117000432

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326–344. https://doi.org/10.1093/applin/amu076

About the authors

Alexandra Pesantez is a PhD student at the Computational Linguistics Lab-University of Zurich. Under the supervision of Professor Dr. Volker Dellwo, she is carrying out a longitudinal study to explore Ecuadorians' L2 language speech development. Through acoustic analysis, she has examined which English vowels are easily acquired and which ones are acquired at a later stage. These data are part of an extensive project on foreign language acquisition.

Email: alejandra.pesantezpesantez@uzh.ch

Volker Dellwo is chair of Phonetics at the University of Zurich. He is interested in a wide variety of phenomena related to human and machine communication with speech. Most of his research is about speaker specific (indexical) information in the speech signal and how humans and machines can use such information to recognise individuals and/or process speech.

Email: volker.dellwo@uzh.ch

Rupp, L., Das, A., Kamps, A. & Acosta, E. (2023). English Pronunciation in a Global World: A MOOC course for pronunciation teaching. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 225–235). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225221

English Pronunciation in a Global World: A MOOC course for pronunciation teaching

Laura Rupp
Vrije Universiteit Amsterdam

Amrita Das
Vrije Universiteit Amsterdam

Alisah Kamps
Vrije Universiteit Amsterdam

Ericka Acosta
Vrije Universiteit Amsterdam

The paper presents the Massive Open and Online Course (MOOC) English Pronunciation in a Global World (Rupp, 2019). We outline the context in which the MOOC was created, one in which English is a global language yet the majority of the world's population do not speak English (Crystal, 2017), and one in which linguistic misunderstandings occur and accent discrimination persists (Lev-Ari, 2021). We outline the aims of the MOOC and describe how we sought to achieve these aims using linguistic and psychological theories (Derwing & Munro, 2015; Ryan & Deci, 2000), as well as methods from the Mixed Classroom Model in Practice (Ramdas et al., 2019). We report on the findings from running the MOOC thus far, some of which we anticipated (e.g., the new type of pronunciation data it generates) and some of which were not immediately foreseen (e.g., participants spontaneously mentoring other participants). In the discussion we reflect on the advantages and challenges associated with using a MOOC for pronunciation teaching, and address the opportunities that it provides.

Keywords: Massive Open and Online Course (MOOC), English pronunciation, inclusive English education, English accents, English pronunciation research

1 Introduction

English is the most widely spoken language in the world and using English as a Lingua Franca has provided numerous opportunities for people to connect to each other. However, many 'disconnections' have arisen at the same time.[1] First, estimates of the number of English speakers in the world (Crystal, 2017) suggest that the majority of the world's population cannot speak (much) English because they have not had the opportunity to learn it. Second, speakers of English speak with various accents, from which linguistic misunderstandings may arise. One specific example that led to a near-conflict happened in 2008, when Bernard Kouchner, the then French minister of Foreign Affairs, warned that one country would 'eat' another country. This sounded very aggressive and caused a great upset. Kouchner had wanted to say 'hit' but he had dropped 'h' and pronounced a different vowel, as many French speakers would do. He felt compelled to apologise for the phonetic confusion that he had caused.[2] Third, accents are associated with particular personal characteristics and this may lead to discrimination based on accent (Lev-Ari, 2021). For example, testimonies of witnesses have been discredited because of the way they speak English.[3]

2 The MOOC and its aims

In this context, we created a Massive Open and Online Course (MOOC) called English Pronunciation in a Global World to see whether it could be an instrument in addressing these three matters related to English pronunciation. The MOOC therefore has three aims, both linguistic and social:

1. Aim 1 – To provide an academic, research-informed course on English pronunciation that is freely accessible and can be attended by anyone in the world on any device.
2. Aim 2 – To enhance linguistic understanding of both learners' own accents and variation in English accents to help prevent misunderstandings. Here we adhere to the Intelligibility Principle: "The notion that the goal of pronunciation instruction should be to help learners to become more understandable by focusing on those aspects of an accent that interfere with listener comprehension." (Derwing & Munro, 2015, p. 178)[4]
3. Aim 3 – To enhance sociolinguistic appreciation of variation in English accents to help combat social issues associated with English pronunciation e.g., accent discrimination. To quote one MOOC participant[5]: "The most remarkable thing that I learnt … is the fact that it's ok for me to have my own identical accent ...I'm glad that maybe I won't feel so embarrassed now every time I speak aloud."

At a more general level, our aims relate to three larger goals:

1. The United Nations Sustainable Development Goals (UN SDG) #4 Quality Education (Ensure inclusive and equitable quality education and promote lifelong learning

[1] In this paper, we will not venture to address the reasons for this, one of which has to do with the colonial roots of English. See Philipson (1992) for discussion and the notion of linguistic imperialism.
[2] http://news.bbc.co.uk/1/hi/world/middle_east/7658728.stm
[3] https://www.linguisticsociety.org/sites/default/files/Rickford_92_4.pdf
[4] Levis (2005) was the first to introduce the notion of Intelligibility Principle.
[5] All examples from MOOC participants are quoted in their original version. Grammatical errors and non-idiosyncratic word choices have not been corrected in the text.

opportunities for all) and #10 Reduced Inequalities (Reduce inequality within and among countries).[6]
2. Sociolinguistic principles like Labov's (1982) Principle of Debt Incurred, that "Linguists have an obligation to expose misunderstandings and misinterpretations about language to the attention of the widest possible audience" (p. 173).
3. Community engagement seeking to "generate activities that build capacity and affirm human dignity through sustainable and ***reciprocal*** [emphasis added] collaborations with communities who experience disadvantage or marginalisation."[7]

3 Previous research on the topic

One important tenet of the MOOC is that it should be informed by academic research. Therefore, to achieve our aims in the creation of content for the MOOC, we took account of linguistic research on the intelligibility of, and attitudes towards (Beinhoff, 2013), pronunciation features in speakers of English. Regarding intelligibility research, we were aware of the difficulty of establishing the intelligibility of pronunciation features in the tradition of Jenkins's (2000) Lingua Franca Core (see Henderson, 2008, p. 6); nonetheless we felt the best approach was to take into account the evidence that has been reported in the literature, expressing appropriate caveats in the MOOC. We adhere to the distinction proposed by Derwing and Munro (2009) whereby "accent is about difference, comprehensibility is about the listener's effort, and intelligibility is the end result: how much the listener actually understands" (p. 480).

Furthermore, to ensure successful learning, we considered psychological and educational approaches to learning. First, to foster motivation and development in the learners, in the absence of a physically present teacher, we deployed Ryan and Deci's (2000) Self-Determination Theory (SDT). This psychological theory suggests that people have three needs, i.e., a sense of competence, relatedness, and autonomy; when these three needs are fulfilled, they can make choices and stay motivated. Second, to ensure that the diverse group of learners in a MOOC would form a community, engaging with one another and developing together, we deployed the VU Mixed Classroom Educational Model (MCEM). This educational approach "builds upon differences to enrich the learning experience for all students present" (Ramdas et al., 2019, p. 5). It works in a three-phase process (Ramdas et al., 2019, p. 9):

1. *Sensitising* – students need to be sensitised to their own frame of reference and the existing diversity in the classroom, and a safe learning environment to do so should be created;
2. *Engaging* – students should interact constructively with different perspectives present in the classroom; and
3. *Optimising* – every student's learning process should be optimised by having them capitalise on different perspectives and approaches.

Thus, the MCEM seeks to "to open up to differences, to co-create an inclusive environment, and capitalise on different perspectives in order to create value" (Ramdas et al., 2019, p. 5).

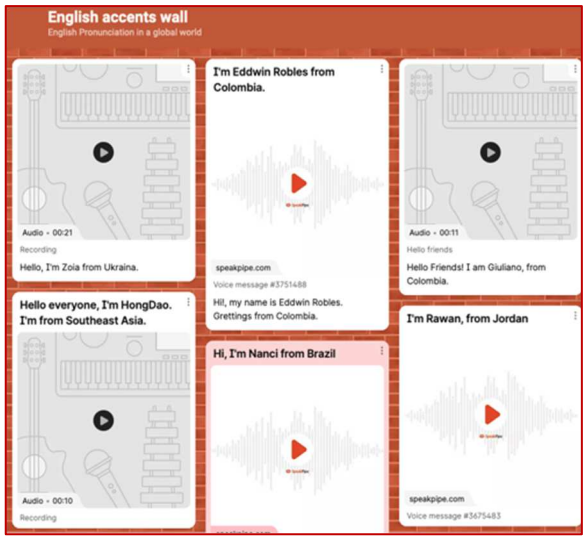
[6] https://sdgs.un.org/goals
[7] https://www.acu.edu.au/about-acu/community-engagement

4 Research methodology

To achieve the aim of providing inclusive English education, we ran the MOOC on the platform FutureLearn.[8] In FutureLearn, learners have free access for the duration of the course (four weeks) and the following six weeks.[9] More importantly, FutureLearn offers technical support, ensuring that the course can be attended on any device anywhere in the world. Adhering to the MCEM's sensitising phase and the SDT component need of relatedness, we created an inclusive learning environment through introductory activities in the MOOC in which the learners get to know one another. For example, learners share their English accents by posting audio clips on a virtual Padlet wall (see Figure 1).

Figure 1

English Accents Wall in the MOOC: Learners Upload Recordings to Introduce Themselves



Hello everyone … my name … I am from Indonesia … I am working as a teacher … and I am very happy to meet you in this opportunities to learned from person in other countries.

On another Padlet wall, learners can voice prior experiences speaking English. They find commonalities with other participants such as employment prospects, educational ambitions, or a lack of confidence in speaking English, as in the examples (1a) and (1b) from two MOOC participants:

(1a) Hi, I'm from Malaysia. … I'm from the city that mainly speak Chinese thus my pronunciation is not that proper and afraid to talk in English.

[8] https://www.futurelearn.com/. One advantage of the FutureLearn platform is that one can make use of their experience regarding online learning, expressed as four pillars of social learning: 1) telling stories; 2) designing for conversation; 3) developing skills; and 4) celebrating progress.

[9] For longer access and other advantages, learners need to purchase an upgrade. However, we seek as much as possible to distribute upgrades to disadvantaged learners (amongst others, Sakhi for Girls Education; https://sakhiforgirlseducation.org), as well as via our mentor programme (see §7). Note further that material in the MOOC that we have created ourselves is Creative Commons (BY-NC-CA). This material can be freely downloaded and adapted, as long as the source is acknowledged. We hope that English teachers find the materials useful for their own English classes.

(1b) Greetings! I'm from India and as you all know India had been a British colony and their influence is clearly seen in India, even after 75 years of independence. At present it has become kind of necessary to learn English here in India otherwise you end up being the inferior one.

Learners can join our social media channels and thus remain part of the MOOC's community of learners also after completing the course.[10]

The introductory activities are addressed in week 1 and are part of the first main topic, i.e., what is important in English pronunciation. The other three topics in weeks 2, 3 and 4 are: English vowels, English consonants, and suprasegmental features. Each week contains a number of steps. Each step contains a type of activity that helps achieve the MOOC's aims of enhancing participants' understanding of English pronunciation features and their appreciation of different English accents. At every step, learners can meet and engage with other learners, share their ideas, join in by leaving comments, adding their perspectives. For instance, in the activity 'Discussions', learners learn about linguistic and social aspects of English pronunciation, in particular the notions of intelligibility, credibility, and identity in English. There is also a range of practice activities in the MOOC that contain, amongst other tools: explanatory videos (which explain the IPA and why English sounds different from the way it is written by outlining phonological changes that have occurred in the history of English), listen-and-repeat exercises, analytical assignments of real or near-real life data, quizzes, etc. By introducing learners to a variety of practice tools, we hope to boost their autonomy (in line with Ryan and Deci's (2000) Self-Determination Theory - SDT) and to empower them by showing where to look for such tools themselves, so they can continue to develop their pronunciation after the course.[11] Also, there is a peer-review exercise in which learners: 1) set pronunciation goals for themselves according to the kind of accent they wish to develop; 2) make a recording of a word list and a reading passage at the beginning of the course; 3) make a second recording at the end of the course, focusing on their goals and attempting to apply what they have learnt; and 4) peer-review one another's recording. In this way we aim to enhance learners' feelings of competence (consistent with SDT) and to optimise their learning from one another (consistent with MCEM) so that they: 1) they take agency over their own accents; 2) find out on their own about differences between their accents and others'; 3) help others by giving feedback regarding what is difficult to understand and what they like about the other person's accent; and 4) support and encourage each other. Examples (2a) and (2b) illustrate these endeavours:

(2a) I noticed you didn't pronounce the final -es in words like 'increases' and 'focuses'.

(2b) Once again, congrats!

Finally, the MOOC team act as mentors, offering support to learners by responding to a selection of posted comments. In two runs every year, we facilitate the course, which means that we offer participants the opportunity to directly communicate with us, thus providing a bridge between online learning and the classroom. For example, we hold an online live session in which learners can join us and we answer questions, reflect, and award the winner of the most insightful comment.

[10] www.facebook.com/globalenglishlearn; www.instagram.com/moocenglishpro.
[11] The MOOC draws as much as possible from available material from the Internet that is of good quality, and has complemented this material with self-created material.

5 Results and discussion

In three and a half years, the MOOC has reached over 117,000 learners from over 191 countries (see UN SDGs #4 and #10). We have seen a great deal of interaction amongst the learners, with some of the comments liked around 200 times and giving rise to threads of nearly 60 comments. Interestingly, learners are not only looking to us for advice or help, but also acting as a community and giving each other advice and helping each other out, an unexpected level of collegiality. For example, in (3a), one MOOC participant explains the pronunciation feature of voicing to another learner, and in (3b), one participant explains how to upload their recording.

(3a) A: I got confused between /t/ and /d/ sound and also /b/ and /p/.
B: If you put your hand on your throat and say first /p/ then /b/ you'll feel the throat vibrate on /b/… the same with /t/ and /d/, the first is voiceless and the second voiced.

(3b) A: How can I share my audio. I recorded through speak pipe. But I don't know how to send.
B: … did you … If so, you can share that link here in the comments.

Regarding our Aim 2, we found that the MOOC participants developed insights about intelligibility, and optimised their understanding of pronunciation features by capitalising on differences between their accents and those of fellow learners (see the MCEM's optimising phase). The examples (4a) and (4b) illustrate this point:

(4a) The course brought me a greater awareness of different accents. The accent itself is not a mistake, as long as it meets certain pronunciation features necessary for good communication. It made me lose some fears, and risk more to speak without fear. I believe that this can lead me to improve the pronunciation of the English language.

(4b) I have realised that the position of stress is different from mine.

Regarding our Aim 3, we found that the MOOC participants constructively interacted with each other to derive a nuanced understanding of the notion of credibility and the prejudice that goes with it (see the MCEM's engaging phase). The following examples illustrate and exchange of ideas beginning with an insight into this topic (5) and then a discussion of it between different learners in a thread (6a, 6b, 6c, 6d):

(5) I often have the feeling that people question my knowledge comparing to people with better pronunciation and fluency. Until now I never read there was actually some research on credibility and accent relation. At least I understand it better and I feel my perception of those situation was correct.

(6a) Yeah, I was also impressed about accent affecting credibility. When I watched the video, I considered what happens in my country and I agree about accent causing discrimination. … Actually, these pronunciation features are wrongly related to … laziness, illiteracy …

(6b) I guess it is something seen and experienced in many places (if not all). Where people are judged (criticised) on the basis of their accents.

(6c) I never thought about it … I did not want to change my accent because I am Spanish, what can I do? I have a Spanish accent …However, recently I am finding it challenging to progress in my career, and I think language might be something that could help me. Everyone says "what a lovely accent you have" but I am not sure is in a patronising or genuinely way.

(6d) Around the world happens the same discrimination about the bad accent … we could be better people because, in this way we study to understand the need[s] of other people from different places and help them[,] like in this group. More educated day after day.

To summarise, we have created an academic, research-informed MOOC that is freely available (Aim 1). We examined if a MOOC could help make English education accessible and enhance understanding of English pronunciation features and appreciation of different English accents. We believe the results discussed in this section support these ideas. In this context, the three key findings are:

1. The MOOC reaches out to tens of thousands of learners around the world (contributing to the UN SDGs #4 and #10), who interact with the material in the MOOC and with one another (Principle of Community Engagement; see also §6);
2. Learners express understanding of the notion of intelligibility, and of the pronunciation features that affect the intelligibility of their own English accents and that of other speakers of English (see Aim 2);
3. Learners express understanding of the notion of credibility, and an appreciation of diversity in other English accents (see Aim 3).

In this way, a MOOC can play an important role in Labov's (1982) Principle of Debt Incurred.

6 Conclusion and implications

Having concluded that a MOOC is a suitable tool for making English education accessible, and for enhancing understanding of English pronunciation features and the appreciation of different English accents, we can now apply the MOOC in various ways, both in pronunciation research and in pronunciation teaching.

In research, the MOOC provides a new type of data. Comments that participants make in the MOOC are self-reported and not affected by the presence of a teacher or a native speaker. Therefore, they provide a window on various aspects of pronunciation features directly from the perspective of the learner, for example revealing the salience of pronunciation features. A broad definition of salience is that it concerns linguistic features that are 'noticeable' or 'stand out'. Taking into account work by Trudgill (1986), Auer et al., (1998, p. 167) distinguished between two types of salience: objective and subjective. Objective salience involves language-internal properties, such as the phonetic distance or phonological contrast between two pronunciation features. Subjective salience involves extra-linguistic properties, such as a feature being a high-prestige variant or an overtly stigmatised or stereotyped one. We conducted a pilot study amongst Spanish speakers of English, examining their responses to: 1) a step in the MOOC in which they set pronunciation goals, and 2) one of the questions in a questionnaire concluding the course (What feature(s) of your English accent would you still like to improve? Why?). In 235 responses from Spanish speakers of English, we found that 25% put rhythm and stress first (Kamps, 2021) — suprasegmental features which apparently stood out for the learners but may not fit either category of salience straightforwardly.

Therefore, we could explore MOOC data to gain more insight into the nature of a notion such as salience. Additionally, it has been argued that salience plays a role in a number of linguistic processes, such as language change, dialect accommodation and second language learning (see the overview in Jansen, 2014). Accordingly, we could take account of features that learners conceive of as salient themselves when we prioritise the pronunciation features we teach. Furthermore, we may use the MOOC data to probe other pronunciation issues, such as the perceived intelligibility of particular pronunciation features or the comprehensibility of English accents. For instance, in one step in the MOOC, learners learn about rhoticity and comment whether they consider rhotic or non-rhotic English accents more comprehensible, as shown in examples (7a), (7b), and (7c):

(7a) Rhotic speakers are easier to understand (MOOC participant from the Philippines)

(7b) The non-rhotic is very British. I like it! But it's a little bit difficult to understand. (MOOC participant from Spain)

(7c) I lived in New Zealand for 4 years, and they have a non-rhotic accent. It was difficult to understand … (MOOC participant from Chile)

Another application of the MOOC in pronunciation research is how we can learn about the pronunciation of English by speakers of languages that are less familiar. For example, one of the MOOC participants commented that "My language Ibibio, spoken across the Niger Delta region of south Nigeria does not have the /ɜ:/ sound". Such a comment illustrates the way in which the MOOC is meant to constitute a true form of Community Engagement; we share the pronunciation material that we have with learners, and conversely, we learn from our learners and can improve the content of the MOOC through the input they provide.

Third, research into e-learning could be conducted, by setting up studies in which the MOOC is used in pronunciation teaching in an experimental group and a comparison group receives in-class pronunciation teaching. Some steps in the MOOC are suitable for evaluating learners' progress in the learning of pronunciation features and the understanding of other speakers' accents. Examples of this are steps in which learners make a recording of their English pronunciation and peer-review the recording of another learner, respectively, and steps where they make a second recording and conduct a second peer-review, applying what they have learnt in the course.

In teaching, we have capitalised on the finding that MOOC participants provide feedback to other participants of their own accord. Namely, we have established a network of external mentors, contacting MOOC alumni in the first instance, and now drawing from a pool of MOOC alumni who volunteer as external mentors in a Google form in the MOOC. We aim to have external mentors speaking a range of different native languages because they will be able to provide tailor-made feedback to learners who share the same native language (as well as providing expert input to the MOOC that we do not have ourselves). External mentors give feedback in the MOOC for four weeks and write a reflection of their experiences as a MOOC mentor; they are then awarded an upgrade to the MOOC and a certificate. More recently, we have engaged current university students in mentoring, too, offering them a teaching opportunity and hoping for them to gain valuable experiences. According to the 'Lernen durch Lehren' principle (Martin & Oebel, 2007), students may learn about pronunciation differences by trying to explain these to others. Furthermore, students may struggle to make their theoretical knowledge feel tangible and may not feel educated enough on a topic to legitimately join the conversation. Mentoring in the MOOC may help to bridge that gap for them, because

they get first-hand experience of pronunciation research pursuits and the effects of this research on both a real-life curriculum and, importantly, in real people's lives.

The importance of native speakers also participating in the MOOC should not be forgotten, as one observed: "Knowing what I know now from this course I am so grateful to be a native English speaker. I had no idea of the intricacies of pronunciation and the challenges posed to learners of English in this regard." Real-world communication involves speakers from diverse backgrounds and proficiency levels, so encompassing such diversity with the MOOC is key to truly achieving inclusion and promoting understanding.

Acknowledgments

We would like to thank the students from Concordia University, Canada and Université-Grenoble Alpes, France, who have been working as external MOOC mentors. They joined us following the EPIP 7 conference[12], demonstrating the important function which conferences can have. We also greatly benefited from exchanges with colleagues, in particular Pavel Trofimovich of Concordia University. We are grateful to two anonymous reviewers for their insights; these have significantly improved the paper. We would also like to thank all learners in the MOOC for their valuable input from which we have learnt so much. The authors acknowledge support from VUvereniging [grant AB/rk/2020/088]. We are also indebted to the editors of this volume for their critical eye and constructive input. Any remaining inadequacies in the paper are our own.

References

Auer, P., Barden, B., & Grosskopf, B. (1998). Subjective and objective parameters determining 'salience' in long-term dialect accommodation. *Journal of Sociolinguistics 2*(2*)*, 163–187. https://doi.org/10.1111/1467-9481.00039

Beinhoff, B. (2013). *Perceiving identity through accent: Attitudes towards non-native speakers and their accents in English*. Peter Lang.

Crystal, D. (2017). *Making sense: The glamorous story of English grammar*. Oxford University Press.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*(4), 476–490. https://doi.org/10.1017/S026144480800551X

Derwing, T. M. & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspective for L2-teaching and research*. John Benjamins.

Jenkins, J. (2000). *The phonology of English as an international language: New models, new norms, new goals*. Oxford University Press.

Henderson, A. (2008). Towards intelligibility: Designing short pronunciation courses for advanced field experts. *ASp, 53–54*, 1–22. https://doi.org/10.4000/asp.369

Jansen, S. (2014). Salience effects in the North-West of England. *Linguistik Online, 66*(4), 91–110. http://dx.doi.org/10.13092/lo.66.1574.

Kamps, A. (2021). *Salience of pronunciation features in Spanish L2 learners of English*. [Unpublished Honours project]. Vrije Universiteit Amsterdam.

Labov, W. (1982). Objectivity and subjectivity in linguistic science. *Language in Society, 11*(2), 165–201.

Lev-Ari, S. (2021). Here's why people might discriminate against foreign accent: New research. *The Conversation*. https://theconversation.com/heres-why-people-might-discriminate-against-foreign-accents-new-research-172539

[12] https://epip7.sciencesconf.org/

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*(3), 369–377.

Martin, Von J.-P. & Oebel, G. (2007). Lernen durch lehren: Paradigmenwechsel in der Didaktik. *Deutschunterricht in Japan, 12*, 4–12.

Ramdas, S. K. K., Slootman, M. W., & Van Oudenhoven-van der Zee, K. (2019). *The VU mixed classroom educational model*. Mixed Classroom Booklet. https://assets.vu.nl/d8b6f1f5-816c-005b-1dc1-e363dd7ce9a5/d7847606-cfa2-482b-8cde-c6e7b1bb7e49/Mixed_Classroom_booklet_tcm270-935874.pdf

Ryan, R. M., & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037//0003-066X.55.1.68

Rupp, L. (2019). *English pronunciation in a global world* [MOOC]. FutureLearn. https://www.futurelearn.com/courses/english-pronunciation.

Philipson, R. (1992). *Linguistic imperialism.* Oxford University Press.

Trudgill, P. (1986). *Dialects in contact*. Blackwell.

About the authors

Laura Rupp is a Senior Lecturer in English Language and Linguistics and director of the Centre for Global English at Vrije Universiteit Amsterdam (VU), The Netherlands. At VU, she teaches, amongst other subjects, Global English and English Pronunciation. She created the online course MOOC English Pronunciation in a Global World in order to address pressing issues that arise from the use of English as a Lingua Franca: both linguistic (the need for mutual understanding in English) and social (accent discrimination and inequality/non-inclusion from lack of access to (good quality) English education). Her research has also focused on grammatical and discourse-pragmatic variation in traditional dialects of English; she has published various articles and co-authored a monograph in this field.

Email: l.m.rupp@vu.nl

Amrita Das is the Programme Manager of Didactics & Communication at VU's Centre for Teaching & Learning, overseeing the spectrum of teacher professionalisation, from early career development to senior management and leadership. She worked on the MOOC as a mentor and coordinator, facilitating several runs and helping learners. She is the co-coordinator of Vrije Universiteit's University Teaching Qualification programme and specialises in the VU's Mixed Classroom Educational Model. She has used the Model to understand better the aspects of inclusion and belonging in the MOOC, and wishes to extend its uses across curriculum reflections and other educational practices. Her extensive academic background includes a doctorate in cultural mobility and transnational poetry, specialising in the American ghazal.

Alisah Kamps is a Masters student of sociolinguistics at the Universiteit van Amsterdam, The Netherlands. She completed her undergraduate degree in Communication & Information Studies at Vrije Universiteit Amsterdam (VU) with a specialisation in English Language and Linguistics. Alisah has been part of the team of the MOOC English Pronunciation in a Global World since 2020 after she conducted her honours programme project on the saliency of pronunciation features using MOOC data. Currently, she is an internal mentor whose role is to guide learners in the MOOC. After co-presenting the MOOC at EPIP7, Alisah has taken up a new research project that probes the deployment of VU educational Mixed Classroom methods for informing the linguistic attitudes of learners in the MOOC.

Ericka Acosta completed a BA degree in applied linguistics from the Communication & Information Studies programme at Vrije Universiteit Amsterdam (VU), The Netherlands. Her specialisation is in language education, and she has a combined minor in neuro studies and digital humanities. In the MOOC English Pronunciation in a Global World, Ericka serves as an internal mentor and a research assistant. As an internal mentor, Ericka supports learners and addresses their questions and concerns. In her role as a research assistant, she organises and enriches MOOC data and she assist in managing social media accounts. Ericka was part of the team that presented the MOOC at EPIP7 in Grenoble. Ericka is to pursue an MA programme in Text Mining at VU.

Sanvicente, F., Ferragne, E., Navarro, S., & Guyot Talbot, A. (2023). French learners of English struggle to identify nasalised American vowels in CVN words. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 236–248). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225140

French learners of English struggle to identify nasalised American vowels in CVN words

Francesca Sanvicente
Paris-Cité University and ENS Paris-Saclay

Emmanuel Ferragne
Paris-Cité University

Sylvain Navarro
Paris-Cité University

Anne Guyot Talbot
Paris-Cité University

English and French have different phonological systems (Roach, 2009; Walker, 2001) and French learners have difficulties perceiving some contrasts, such as the approximants /w/ – /r/ (Hallé et al., 1999). They also tend to assimilate multiple English vowels to a single L1 category (Iverson & Evans, 2007). General American exhibits a raising of /æ/ in nasal environments (Carignan et al., 2016), thus creating an acoustic proximity between /æ/ and /ɛ/. The purpose of the present study was to assess the influence of CVN contexts on the perception of vowels by French university students.

We addressed this issue as part of a pedagogical project called SEPALE, which consists of identification and discrimination exercises involving English phonemic contrasts and within which we explored the issue of nasalisation. Thirteen Californian speakers were recorded reading CVC and CVN words, to create stimuli used by undergraduate students in discrimination and perception tasks. Their answers were recorded in logfiles, used in our analysis.

We found a higher proportion of errors when /æ/ – /ɛ/ and /ɑ/ – /ʌ/ were in a pre-nasal position and the first contrast yielded better identification and discrimination rates than /ɑ/ – /ʌ/. This finding could inform teachers' choices of perception or production exercises.

Keywords: French learners, nasalisation, identification tasks, perception, discrimination

1 Introduction

General American (hereafter GA) vowels can be affected by nasalisation when followed by a nasal consonant. Thus, we decided to examine the possible perception impairment that they could provoke in French learners. Before /m/, /n/, and /ŋ/, the tongue follows a rising trajectory whose peak differs depending on the consonant; the velar differs from the other two, but for all three a lowering of the first formant of low vowels is generally observed. The alteration of the perceived height of low vowels when followed by a nasal (Mielke et al., 2017) begs the larger question of how a nasal consonant influences the perception of vowels by French learners of English.

SEPALE, a pedagogical project, was used as a platform to look into the possible impact of a CVN (consonant – vowel – nasal) environment on the perception of GA vowels by French learners. SEPALE[1] is a project consisting of exercises to train the perception of those vocalic contrasts which are supposed to be difficult for French learners of English. The project is part of the participant's curriculum, so the exercises were completed weekly over a period of 12 weeks, although some students were less assiduous. The results analysed in the present study are the uploaded answers to the perception tasks as done by the participants within the wider SEPALE project.

2 Previous research on the topic

There is a large body of academic work pertaining to vowel nasalisation in English. Several authors have notably shown that nasalisation may affect perceived vowel height (e.g., Beddor, 2007; Krakow et al., 1988; Wright, 1975) because of the resulting acoustic shift in the F1 region. Generally speaking, nasalisation entails a lowering of high and mid-vowels and raising of low vowels (Wright, 1975, 1986). Studies combining acoustics and ultrasound tongue imaging have demonstrated that depending on the vowel, tongue height adjustments may be used to compensate or enhance the acoustic effect of nasality. For instance, Carignan et al. (2011) showed that /i/ is produced with a compensatory higher tongue position in pre-nasal context in order to counteract the resulting raising of F1 and to avoid neutralisation with neighbouring /ɪ/. They observed no such adjustment in pre-nasal /ɑ/, where the perceived variation in vowel height induced by nasalisation is acceptable without any risk of misidentifying the vowel. Conversely, Mielke et al. (2017) found that the tongue raising movement of /æ/ in pre-nasal environments may constitute a strategy to enhance the acoustic effect of nasalisation. Carignan et al. (2016) explain that while tongue raising before velar consonants /g/ and /ŋ/ is described as the result of coarticulation (as evidenced by a gradual raising of the tongue in anticipation of the velar constriction), such raising before other nasals peaks around the midpoint of the vowel, indicating "a distinct phonetic target" (p. 316).

Most accents of North American English, including GA, exhibit a raising of /æ/ in pre-nasal environments (Carignan et al., 2016; Labov et al., 2006), the realisation of which is generally described as involving a raising-falling trajectory before /m/ and /n/ (e.g., *pan* [peən]) and a raising trajectory before /ŋ/ (e.g., *bang* [beɪŋ]). As shown by Krakow et al. (1988), the discrimination of nasalised /ɛ/ and /æ/ in a contextual nasal environment (i.e., when the nasalised nucleus is followed by a nasal consonant) is unproblematic for native speakers. However, the acoustic proximity of the nasalised nuclei in e.g., *bend* and *band* might make it difficult for French learners to distinguish /ɛ/ vs. /æ/ in those environments. This difficulty

[1] SEPALE is a project that was developed at Université Paris Cité thanks to funding devoted to pedagogical innovation received from 2021 to 2022.

usually bears no equivalence with British varieties where /æ/-raising is largely absent (Mielke et al., 2017).

Given the absence of compensatory tongue adjustment to the nasality induced raising of /ɑ/ (Carignan et al., 2011) and lowering of mid-vowels in pre-nasal environments (Wright, 1975), we also observe difficulties in French learners to correctly discriminate pre-nasal /ɑ/ and /ʌ/ (e.g., *pond* vs. *punned*).

We hypothesised that French learners of L2 English would have more difficulties in perceiving the /æ/ – /ɛ/ and /ɑ/ – /ʌ/ contrasts in pre-nasal environments (e.g., *sand – send* and *pond – punned*) than in other pre-consonantal environments (e.g., *bat – bet* and *cot – cut*).

3 Methodology

3.1 Participants

In this study, 118 second-year students from Université Paris Cité participated in the experiment during the second semester of the academic year. The students were between 19 and 21 years old and their proficiency was at or near a B2 level of English, which is the required level for admission to the *Licence LLCER Anglais* degree programme in French universities.

3.2 Stimuli creation

As the purpose of the discrimination study was to determine whether the nasal after the nucleus had an influence on the students' perception of the vowel, a list of both CVC (consonant – vowel – consonant) and CVN one-syllable words was established. The words included the /æ, ɛ, ɑ, ʌ, ɪ, i, ɔ/ vowels as a nucleus. Some items with /ɪ/ and /i/ from the final list of stimuli were adapted from Krzonowski (2020), and other CVC and CVN words were added with /æ, ɛ, ɑ, ʌ, ɔ/. We will focus on the /æ/ – /ɛ/ and /ɑ/ – /ʌ/ contrasts, but the /ɪ/ – /i/ contrast will be included as a baseline because no such nasalisation effect is known to impact its realisation.

We avoided the postvocalic liquid /l/ as mergers have been observed in this context, for example, *pull – pool – pole* may all sound alike due to /l/ (Arnold, 2015), and similarly in *feel – fill* (Labov et al., 2006), where the distinction between /i/ and /ɪ/ is often neutralised for the same reason. We also decided against items with a pre-velar context, such as *pag – peg*. Although there is usually no /æ/-raising before a velar stop in GA, the existence of a raising pattern with a rising tongue trajectory before /g/ has already been observed. As a consequence, in some regions of North America, the word *bag* is pronounced [bejg] (Mielke et al., 2017). The list of stimuli is provided in the Appendix.

Thirteen native speakers of GA (seven women) were recruited in Southern California, all but three of whom were from California. Of those three, one stated that they were from all over the US, another from New Orleans, and someone else from New Jersey. Their ages ranged from 20 to 73 years of age, the average age being 40. The calculated standard deviation was 16, and the median value was 39.

Recordings were made using ROCme! (Ferragne et al., 2012), a program fed with an HTML file to prompt words on a computer's screen. The speakers' productions were recorded with an AT2020 USB microphone. After each item, the speakers were invited to press a gamepad button to switch to the next word. The stimuli were presented in a randomised order to avoid redundancy and fatigue. In total, they each read 161 words, twice. We used both recordings when usable; some were trimmed, others were removed when the background noise was too distracting, or when the beginning of the recording was too close to the onset.

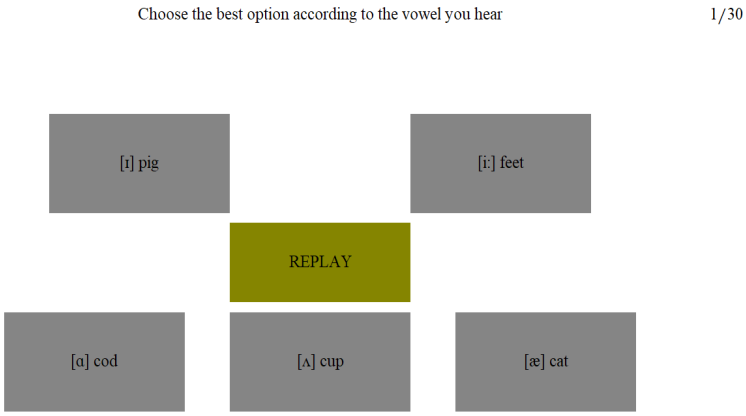
3.3 Procedure

Four types of tasks were included in the experiment. Each exercise was divided into sub-types which corresponded to the vowels involved, or the pre-nasal position of the vowel. In this way, students could choose the sub-type they wanted to focus on, and each attempt was composed of 30 prompts selected randomly out of the 161 words recorded by the 13 speakers.

The alternative forced choice exercises 2AFC and 5AFC played one word and provided the participants with two or five boxes, respectively, from which to pick a match for the vowel heard in the word played. For example, in the 5AFC exercise, participants heard a one-syllable word, and had to determine which vowel corresponded to the nucleus. Five options were provided in the form of boxes, as shown in Figure 1. The purpose of these tasks was to test the participant's ability to identify the vowels involved.

Figure 1

5AFC Window



In the AX exercises, participants heard two stimuli and had to determine if they were the same words or if they were different. In this task, two boxes labelled "Same" and "Different" were shown on the screen, and participants had to choose the option corresponding to their answer. This exercise evaluated their ability to differentiate vowel sounds in two minimal pairs. Finally, in the Oddity exercises, three stimuli were heard and the odd one out had to be found. Thus, there were three boxes carrying a number and corresponding to each stimulus. In this last task, their capacity to discriminate sounds was tested.

Participants could hit the "REPLAY" button as many times as they wished. At the end of each attempt, they received feedback in the form of a pop-up window, indicating: the number of wrong and right answers, the total time spent doing the exercise, and the number of times the "REPLAY" button was pressed. Answers were recorded in the logfiles created at the end of each attempt, as well as the number of errors and the response times.

4 Data analysis and results

Logfiles were analysed with SQLite[2] (Gaffney et al., 2022) and Python[3] functions. Table 1 presents the error percentage per type of exercise throughout the study. The exercise type Oddity has the lowest percentage of errors, with AX at the other side of the spectrum.

Table 1

Percentage of Errors per Type of Exercise

Type of exercise	Error percentage (%)
Oddity	12.38
2AFC	19.83
5AFC	23.19
AX	25.49

The same observation emerges from Table 2, which shows the percentage of errors in both a non pre-nasal and in a pre-nasal context for the different types of exercises. The lowest percentage of errors in both contexts was found in the Oddity type of exercise, and the highest for CVN was recorded in AX exercises. The impact of nasalisation is shown in the last column, where the subtraction of the CVN error percentage from the CVC's is given. For each of the 118 participants, the mean percentage of errors across all exercises was computed for CVC and CVN contexts separately. In order to determine whether performance was, as expected, poorer in the CVN context, we ran a paired Wilcoxon signed rank test with continuity correction. The test corroborated our hypothesis, as it showed a significant value ($V = 232$, $p < 0.0001$).

Table 2

Percentage of Errors per Type of Exercise in a CVC and CVN Context

Type	Error percentage when CVC (%)	Error percentage when CVN (%)	Difference (%CVN - %CVC)
Oddity	10.59	16.46	5.87
5AFC	19.09	26.78	7.69
2AFC	14.93	24.99	10.06
AX	18.79	31.10	12.31

2 SQLite (Version 3.31.1) [Computer software]. SQLite Development Team. Retrieved November 11, 2022. Available at https://www.sqlite.org/download.html

3 Python Software Foundation. Python Language Reference, version 2.7. Available at https://www.python.org/downloads/release/python-3810/

Table 3 displays the error percentage for the vowels involved in the study in CVC and CVN contexts. The percentage of errors increases for every item, except for /i/ which was inversely affected by its pre-nasal position, since students actually performed better when the vowel was nasalised. On the other hand, /ʌ/ appears to be the most affected vowel, with an increase in misperceptions of 17.82% when followed by a nasal; /æ/ was also quite affected (16.72%), as was /ɪ/ (15.21%).

Table 3

Percentage of Errors per Vowel in CVC and CVN Contexts and Impact of Nasalisation (Last Column) in 2AFC and 5AFC Exercises

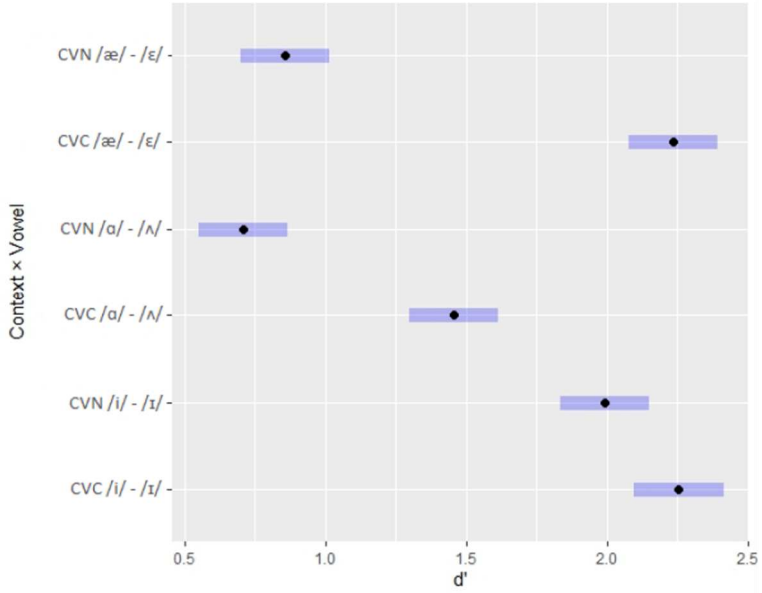
Vowel	Error percentage when CVC (%)	Error percentage when CVN (%)	Difference (%CVN - %CVC)
/i/	25.33	9.92	-15.41
/ɑ/	24.80	28.19	3.39
/ɛ/	15.47	29.49	14.02
/ɪ/	15.10	30.31	15.21
/æ/	6.26	22.98	16.72
/ʌ/	12.63	30.45	17.82

We then focused on the discrimination (AX) exercises, as they naturally lend themselves to the computation of the relatively unbiased d′ sensitivity index from Signal Detection Theory (Green & Swets, 1966). A mixed-effects model was fitted to the data in R software with the lmer function from the lmerTest package, with d′ as a dependent variable. We added Context and Vowel as fixed effects, as well as the interaction between Context and Vowel. We included Participant as a random effect. Both factors and their interaction were significant ($p < 0.0001$). Post-hoc comparisons were performed with the emmeans[4] package and Figure 2 shows these findings. Three distinct groups appear: 1) the /æ/ – /ɛ/ contrast, for which students performed better when the vowel was not in a pre-nasal position; 2) /ɑ/ – /ʌ/, which showed moderate results in a CVC environment and poor results when the vowels were followed by a nasal; and 3) /ɪ/ – /i/, which obtained good results in both CVC and CVN contexts.

4 https://cran.r-project.org/web/packages/emmeans/index.html

Figure 2

Sensitivity Indices (d') for Each Combination of the Factors Vowel and Context in the AX Experiments.



*Note***.** Higher d′ values show better discrimination. Blue bars show 95% confidence intervals: non-overlapping bars reflect significant differences.

Figures 3 and 4 display the average rate of errors for the /ɑ/ – /ʌ/ and /æ/ – /ɛ/ contrasts. The dotted line corresponds to the linear regression and shows a subtle decreasing trend. On average, students seem to have made more mistakes when the vowel was in a pre-nasal position. The percentage of errors for these pairs followed by a nasal approaches the 0.5 line, while in a non-nasal environment the students seem to have performed better, with a lower average rate of errors for each contrast. Moreover, the line corresponding to the CVN environment never overlaps with the other, the two trends remain almost parallel throughout the period.

Figure 3

Average Rate of Errors During the Second Semester of 2022 for the /ɑ/ – /ʌ/ Contrast

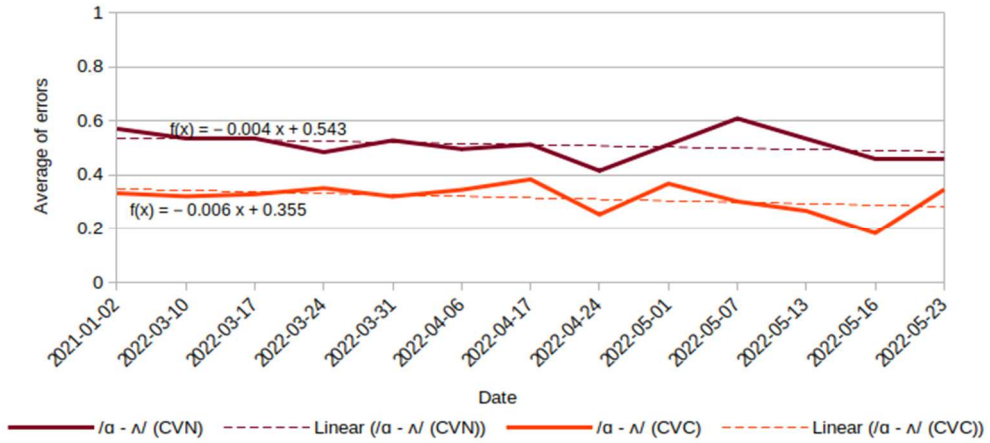
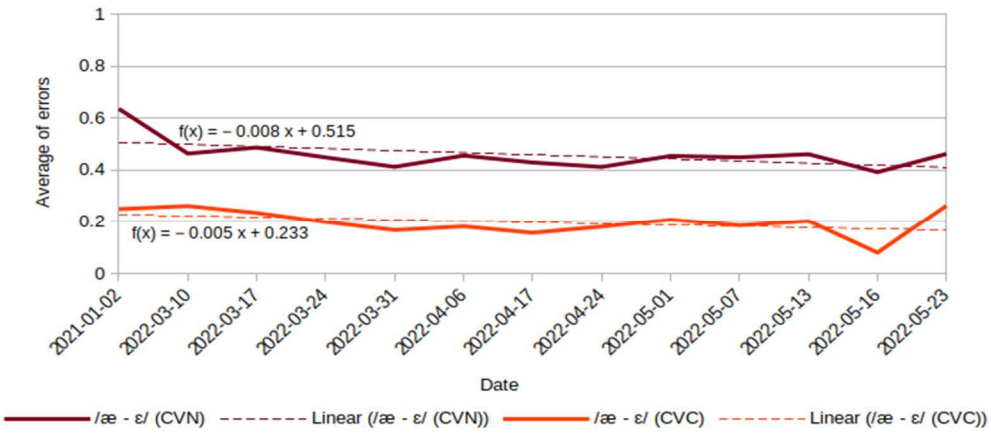


Figure 4

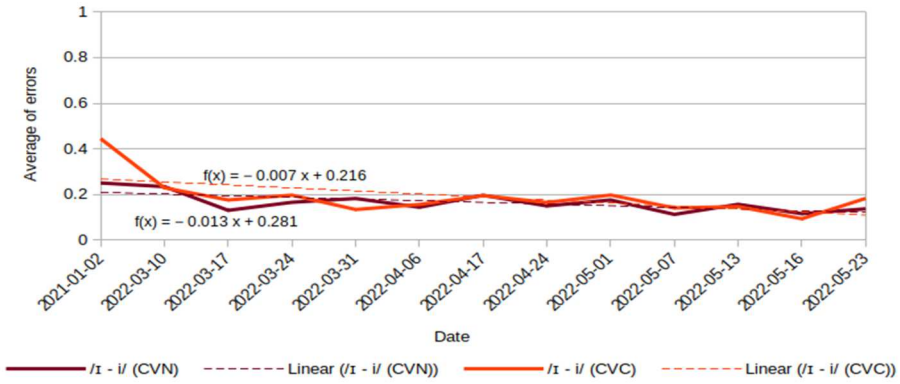
Average Rate of Errors During the Second Semester of 2022 for the /æ/ – /ɛ/ Contrast



By way of comparison, Figure 5 shows the average rate of errors through the course of the semester for the /ɪ/ – /i/ vowels, for which the trend of the linear regression is also negative, therefore indicating a slight decrease. The lines corresponding to the CVC and CVN environments overlap, and the average rate of errors, independent of the context, never goes above 0.3.

Figure 5

Average Rate of Errors During the Second Semester of 2022 for the /ɪ/ – /i/ Contrast



5 Discussion and implications

In the course of this study, we sought to determine whether the pre-nasal position of /æ/ – /ɛ/ and /ɑ/ – /ʌ/ vowels could have an influence on the perception of said vowel by French learners of English. Our results show that the presence of a nasal after these vowels has an impact on the perception of the nucleus by French learners. The data showed an increase in the percentage of errors for every item, except for /i/, for which the percentage was surprisingly low in a CVN environment. The tongue height adjustment strategy at play in nasalised /i/ does seem to favour the identification of /i/, as evidenced by the radical drop in misidentified /i/ tokens in 2AFC and 5AFC exercises. The other vowels were impacted by their pre-nasal position, ranging from a 14.02% to a 17.82% increase in the error percentage in a CVN context for /ɛ, ɪ, æ, ʌ/. In contrast, /ɑ/ shows relative stability but yields poor performance even in a CVC context.

The influence of nasalisation is especially visible for the /æ/ – /ɛ/ and /ɑ/ – /ʌ/ vowels. The results demonstrated that these pairs were better recognised in a non-nasal environment than when followed by a nasal. Similarly, the /æ/ – /ɛ/ contrast seems to have been less difficult to recognise than the other pair, since it accounted for fewer errors than the /ɑ/ – /ʌ/ pair whether in a CVC or a CVN environment.

When we compare these results to the average rate of errors for the /ɪ/ – /i/ contrast, we find a clear difference between this contrast and the other two. The superimposition of the two lines corresponding to the CVC and CVN environments indicates that in the case of this pair, vowels do not seem to be affected by their pre-nasal position. This comparison clearly suggests that nasalisation has a major impact on the /æ/ – /ɛ/ and /ɑ/ – /ʌ/ contrasts, supporting the findings of previous studies. In fact, the raising-falling trajectory of /æ/ in pre-nasal environments mentioned by Carignan et al. (2016) and Labov et al. (2006) brings the vowel closer to the realisation of /ɛ/, making the distinction between both sounds more difficult for French learners. Similarly, our findings concur with the observation of a lowering of mid-vowels (such as /ʌ/) in pre-nasal positions (Wright, 1975), and with the absence of compensatory tongue adjustment to the raising of nasalised /ɑ/ (Carignan et al., 2011), since participants performed less well when the vowels were followed by a nasal.

In terms of pedagogical implications, these findings could be used to assess the extent to which example word lists or minimal pairs in textbooks are appropriate for French learners of English. Teachers could then sequence lexical items in relation to their CVC or CVN context, or at least be aware that the CVN context may prove more challenging for learners to perceive.

For example, in the first section of Hancock (2012), the vowel sound /æ/ is explored through different words. A note draws together words such as *ask* and *dance,* and explains that the vowel is generally produced with /æ/. In addition, it is stated that the sound may be /ɑː/ in South-East England. In this case, a supplementary note could have indicated that in GA, the vowel sound in a pre-nasal position is closer to what Carignan et al. (2016) transcribe as [eə]. This would raise awareness about the dialectal differences in the pronunciation of /æ/ vowels in a CVN context.

6 Conclusion

The purpose of this paper was to analyse the influence of nasalisation on the perception of GA vowels by French undergraduates studying English. SQL queries, tables and plots were used to isolate tendencies in the effects of a nasal after a vowel, revealing a greater impact on the /æ/ – /ɛ/ and /ɑ/ – /ʌ/ vowels. Our paper signals the importance of drawing the attention of teachers to the influence of nasalisation on GA vowels and to the way they are perceived by French learners, since we have shown that nasalised /æ/ – /ɛ/ and /ɑ/ – /ʌ/ vowels are more difficult to discriminate.

A potential improvement to our study would be to consider the effect of the speaker's gender on vowel perception. Krzonowski et al. (2018) highlighted, in their acoustic study, a difference in the production of vowels by female and male English native speakers. As an example, a slight overlap was observed in the production of the /ɪ/ – /i/ vowels when produced by English female speakers, and this was exacerbated when pronounced by male speakers. It may therefore be the case that the difference between /ɪ/ – /i/ when produced by male speakers in our SEPALE exercises is harder for our students. We have included the speaker gender in the scripts which create the database, and further SEPALE studies could request that participants indicate their gender prior to doing the exercises. This way, the difference in terms of vocal tract sizes and vowel frequency between women and men could be compared to the success rates in vowel perception.

Acknowledgements

This study was funded by "Innovations pédagogiques" ANR-18-IDEX-0001 - IdEx Université Paris Cité.

References

Arnold, L. (2015). Multiple mergers: Production and perception of three pre-/l/ mergers in Youngstown, Ohio. *University of Pennsylvania Working Papers in Linguistics*, *21*(2), 2.

Beddor, P. S. (2007). Nasals and nasalization: The relation between segmental and coarticulatory timing. In J. Trouvain & W.J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 249–254). Saarbrücken: Saarland University.

Carignan, C., Shosted, R., Shih, C., & Rong, P. (2011). Compensatory articulation in American English nasalized vowels. *Journal of Phonetics*, *39*(4), 668–682. https://doi.org/10.1016/j.wocn.2011.07.005

Carignan, C., Mielke, J., & Dodsworth, R. (2016). Tongue trajectories in North American English /æ/ tensing. In M. -H. Cote, R. Knooihuizen, & J. Nerbonne (Eds.), *Future of dialects: Selected papers from methods in dialectology XV* (pp. 313-319). https://doi.org/10.17169/langsci.b81.157

Ferragne, E., Flavier, S., & Fressard, C. (2012). ROCme!: logiciel pour l'enregistrement et la gestion de corpus oraux [ROCme!: software for the recording and management of oral corpora]. In L.

Besacier, H. Blanchon & G. Sérasset (Eds.), *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012, vol. 5: Software Demonstrations* (pp. 19–20). ATALA/AFCP.

Gaffney, K. P., Prammer, M., Brasfield, L., Hipp, D. R., Kennedy, D., & Patel, J. M. (2022). SQLite: past, present, and future. *Proceedings of the VLDB Endowment*, *15*(12), 3535–3547. https://doi.org/10.14778/3554821.3554842

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley.

Hallé, P. A., Best, C. T., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, *27*(3), 281–306. https://doi.org/10.1006/jpho.1999.0097

Hancock, M. (2012). *English pronunciation in use (Intermediate, with answers)*. Cambridge University Press.

Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *Journal of the Acoustical Society of America*, *122*(5), 2842–2854. https://doi.org/10.1121/1.2783198

Krakow, R. A., Beddor, P. S., Goldstein, L. M., & Fowler, C. A. (1988). Coarticulatory influences on the perceived height of nasal vowels. *Journal of the Acoustical Society of America, 83*(3), 1146–1158. https://doi.org/10.1121/1.396059

Krzonowski, J., Pellegrino, F., & Ferragne, E. (2018). Étude acoustique de la production de voyelles de l'anglais par des apprenants francophones. In *Journées d'Études sur la Parole* (pp. 523–531). http://doi.org/10.21437/JEP.2018-60

Krzonowski, J. (2020). *Perception et production de monophtongues de l'anglais par des apprenants francophones: Effets d'entraînements en perception et en production* [Perception and production of English monophthongs by French speaking learners: Effects of perception and production trainings] [Unpublished doctoral dissertation]. University of Lyon.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology, and sound change: A multimedia reference tool*. Walter de Gruyter.

Mielke, J., Carignan, C., & Thomas, E. R. (2017). The articulatory dynamics of pre-velar and pre-nasal /æ/-raising in English: An ultrasound study. *The Journal of the Acoustical Society of America*, *142*(1), 332–349. https://doi.org/10.1121/1.4991348

Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.

Walker, D. C. (2001). *French sound structure*. University of Calgary Press.

Wright, J. T. (1975). Effects of vowel nasalization on the perception of vowel height. In C. A. Ferguson, L. M. Hyman, & J. J. Ohala (Eds.), *Nasalfest: Papers from a symposium on nasals and nasalization* (pp. 373–388). Stanford University.

Wright, J. T. (1986). The behavior of nasalized vowels in the perceptual vowel space. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology*, (pp. 45–67). Orlando, FL: Academic Press.

Appendix

List of Stimuli

Context	/æ/	/ɛ/	/ɑ/	/ʌ/	/ɔ/	/ɪ/	/i/
/_p/	cap	-	cop	cup	-	kip	keep
/_t/	cat mat pat tat	met pet	bot cot not pot rot tot	but cut mutt nut putt tut	bought caught naught taught wrought	-	-
/_k/	back knack lack pack	beck neck peck	bock lock pock stock tock	buck luck puck	stalk talk baulk chalk	bik lick pick	beak leak peak
/_d/	cad dad had mad rad sad	dead head med red said	bod cod mod nod sod	bud cud	baud gnawed sawed	did	deed
/_m/	cam ham stamp tam	hem	bomb pomp stomp tom	bum chump come hum pump	-	dim	deem
/_n/	ban bland can dan dance land man pan ranch sand tan	ben dense ken lend men pen send ten wrench	blonde con don fond non pond	bun done fund nun pun punned tun	-	bin kin min	bean keen mean
/_ŋ/	-	-	bong honk long	bung hunk lung	-	-	-
/_f/	-	-	cough	cuff	-	-	-
/_s/	-	-	-	crust	crossed	-	-
/_r/	-	-	bar bard card farm par parch spar star	-	bore bored form porch pore spore store tort	-	-

About the authors

Francesca Sanvicente is a Master's degree student at Université Paris Cité and ENS Paris-Saclay

Email: francesca.sanvicente@ens-paris-saclay.fr

Emmanuel Ferragne is a Professor of English phonetics and has been at Université Paris Cité since 2009. His current research is concerned with sociophonetics, the acquisition of English phonology by French learners, and forensic voice comparison.

Email: emmanuel.ferragne@u-paris.fr

Sylvain Navarro is an Associate Professor of English phonology and has been at Université Paris Cité since 2015. He coordinates the PAC (*Phonologie de l'Anglais Contemporain*) research program. His research interests revolve around phonological and phonetic variation in varieties of English, with a particular focus on phenomena relating to rhoticity.

Email: sylvain.navarro@u-paris.fr

Anne Guyot Talbot is an Associate Professor of English phonetics and phonology and has been at Université Paris Cité since 2007. Her current research focuses on the link between morphology and lexical stress, lexical stress location, and English intonation.

Email: anne.talbot@u-paris.fr

Thir, V. (2023). Understood or not? Issues in using orthographic transcription for assessing intelligibility to international listeners of English. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 249–260). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225289

Understood or not? Issues in using orthographic transcription for assessing intelligibility to international listeners of English

Veronika Thir
University of Passau

The most popular method for measuring intelligibility of L2 pronunciation is orthographic transcription (Munro & Derwing, 2015). This paper discusses the construct of intelligibility underlying this method, as well as its strengths and weaknesses compared to other methods of assessing intelligibility, such as shadowing. It highlights issues arising when using orthographic transcription with large, linguistically diverse samples of listeners at different proficiency levels, as is often the case in English as a lingua franca (ELF) research. This is not to say that orthographic transcription should not be used for ELF intelligibility studies, in particular quantitative ones. Instead, a case is made for carefully considering its potential biases in particular research contexts, so that appropriate measures to counteract them can be taken to ensure a satisfactory level of validity and reliability. This argument is illustrated with examples from a large-scale study ($N = 508$) involving both native and non-native listeners from over 80 different L1 backgrounds at different proficiency levels.

Keywords: intelligibility, orthographic transcription, research methods, English as an international language (EIL), English as a lingua franca (ELF)

1 Introduction

Since intelligibility is frequently named as the primary goal of L2 pronunciation teaching, its assessment has particular relevance to researchers and practitioners in L2 pronunciation pedagogy. However, since it is a perceptual phenomenon, its measurement constitutes a challenge. Various methods for assessing intelligibility have been proposed, such as comprehension questions, responses to true/false statements, listener summaries, scalar ratings, shadowing or even focused interviews (see Kang et al., 2018; Munro & Derwing, 2015). The most popular one in L2 pronunciation research is orthographic transcription (Munro & Derwing, 2015), whereby subjects listen to an audio stimulus and write down what they hear. The number of correctly transcribed words (which may be minimal pairs, key words or entire sentences, depending on the research questions) is taken to reflect the amount of understanding that has taken place.

This paper takes a closer look at orthographic transcription by examining the construct of intelligibility underlying it, as well as its validity and reliability in comparison to other widespread methods of measuring intelligibility, in particular shadowing. In doing so, it scrutinises two approaches to coding orthographic transcriptions: the 'exact word match' technique and the allowance of spelling errors. The latter seems preferable with non-native listeners, who may be penalised when using a strict exact word match. However, if spelling errors are accepted, it may sometimes be difficult to distinguish them from transcriptions reflecting unintelligibility. Such 'ambiguous' transcriptions become more frequent when working with large, linguistically heterogenous listener samples, as is the case in quantitative studies on international intelligibility, i.e., intelligibility among English as a lingua franca (ELF) users. The present paper highlights issues and challenges in this respect by reporting on the process of data coding in an ELF intelligibility study which involved over 500 native and non-native listeners from a wide variety of L1 backgrounds at different proficiency levels. It is argued that neither coding technique is inherently superior, but that careful consideration of the sample and research context is necessary before deciding on a technique, which also needs to be considered in the interpretation of results.

2 What type of intelligibility are we measuring with orthographic transcription?

Intelligibility is a complex concept that is defined in various ways within and across different fields (e.g., Derwing & Munro, 1997; Schiavetti, 1992) and different methods of assessing intelligibility might more validly reflect certain intelligibility constructs than others. When listeners transcribe the words they hear, they are essentially engaging in the process of word identification. This most closely resembles Smith's (1992) definition of *intelligibility* as "word/utterance recognition" (p. 76) in his framework; the other two dimensions are *comprehensibility* (i.e., understanding the literal meaning of a word or utterance) and *interpretability* (i.e., understanding the pragmatic meaning behind a word or utterance). One weakness of this framework, however, is its implicit bottom-up view of the comprehension process: the three components are regarded as "degrees of understanding on a continuum, with intelligibility being lowest and interpretability being highest" (Smith, 1992, p. 76). This leads to issues in empirical operationalisation, since

> the comprehension of speech is not a linear process: when a word cannot be recognized on the basis of bottom-up cognitive processes, the listener may nonetheless be able to 'fill it in' by exploiting top-down strategies. (Munro & Derwing, 2015, p. 378)

Thus, orthographic transcription "reflects more than just low-level speech processing" (Munro & Derwing, 2015, p. 378), i.e., it measures more than intelligibility as defined by Smith. Zielinski (2012) also points to "the highly interdependent nature of the different components of the Smith framework" (p. 405) and, likewise, to difficulties in empirical application:

> […] researchers might find it difficult to measure intelligibility as separate from comprehensibility because listeners will use what is a natural process of listening and refer to the meaning and context of an utterance to identify difficult words within it. (Zielinski, 2012, p. 405)

There is abundant empirical evidence supporting the use of context in the process of spoken word recognition (see e.g., Bent et al., 2019; Kennedy & Trofimovich, 2008; Thir 2020a, 2020b). Thus, if we interpret the components in Smith's framework as interrelated rather than hierarchical, this more accurately reflects the nature of speech comprehension. Intelligibility defined as spoken word recognition is therefore best understood as a cognitively interactive process that involves going back and forth between bottom-up and top-down processing (see also Mirman, 2017). Following Magnuson (2017), intelligibility is defined here as the process of "[…] map[ping] phonological forms to intended words in memory" (p. 76). With these adjustments, orthographic transcription seems well suited to measuring intelligibility as spoken word recognition (henceforth SWR). It also seems to meet quality criteria in academic research, such as validity and reliability.

3 Validity and reliability of orthographic transcription in measuring intelligibility as SWR

Validity, here mainly understood as construct validity (e.g., Gass, 2015), refers to the quality of accurately capturing the phenomenon under investigation: "a test is valid if it measures what it is supposed to measure" (Dörnyei, 2007, p. 51). One major advantage of orthographic transcription is that it allows SWR to be quantified more accurately than, for instance, listener ratings of intelligibility. Listener ratings lack validity since they only reflect how much listeners *think* they understood (Munro & Derwing, 2015) and risk confounding intelligibility with similar but separate phenomena such as "ease or effort of understanding" (termed comprehensibility in Munro & Derwing, 2015, p. 380). Compared to methods such as comprehension questions, summaries, and sentence verification, orthographic transcription allows for a more precise focus on the phenomenon of SWR. Moreover, it makes it possible to measure the intelligibility of decontextualised words (e.g., by presenting individual words or nonsense sentences), which may be necessary in certain research contexts (see §4).

One might argue that the same is true for shadowing, another common technique to measure SWR. This involves listeners instantly repeating an audio stimulus, with correct repetition being treated as signifying understanding. However, it is debatable whether correct repetition can be regarded as a sign of correct word identification (i.e., of having assigned the auditory input to the intended entry in the mental lexicon), or whether it measures phoneme (or even phone) recognition. Clearly, it is possible to correctly repeat a sound sequence without understanding what it means. At least in the case of English, it is, however, more difficult to correctly transcribe a word without knowing what the sound sequence refers to, so the danger of measuring sound rather than word recognition seems lower in the case of orthographic transcription.

For both methods, a relevant question is how precisely they measure SWR: do they only measure listeners' ability to identify spoken words, or do other skills factor into them?

Orthographic transcription obviously presupposes some competence in standard orthography, and the extent to which it also measures orthographic skills depends on how potential spelling errors are treated in data analysis. When examining previous research, two approaches emerge in this respect. The exact word match (e.g., Derwing & Munro, 1997; Munro et al., 2006) only accepts perfect matches to standard orthography. The allowance of spelling errors[1] (e.g., Bradlow & Pisoni, 1999; Field, 2005; Kennedy & Trofimovich, 2008), however, arguably offers greater validity since it does not conflate orthographic skills and SWR. This seems particularly important when working with non-native listeners and other groups prone to orthographic weaknesses, who might be penalised when a strict exact word match is used. However, the allowance of spelling errors does not necessarily meet the criterion of reliability, though this may be counteracted to some extent by using multiple coders when analysing listeners' transcriptions.

Hence, some may be quick to recommend shadowing in lieu of orthographic transcription when working with non-native listeners, to overcome the issue of additionally measuring orthographic skills. However, shadowing presupposes something else which many L2 listeners lack: an ability to articulate target words to a recognisable extent. As L2 pronunciation researchers and practitioners can well attest, numerous non-native listeners may very well correctly identify a word or phrase but may not necessarily be able to pronounce it in a recognisable way themselves. Thus, shadowing may lack validity in that it measures not only SWR, but also pronunciation skills. This is particularly problematic when studying the intelligibility of minimal pair words, which L2 listeners might be able to distinguish perceptively, but at the same time might not be able to distinguish productively. A similar issue may arise when working with native listeners whose L1 dialect contains different phonemic distinctions than the variety under study. Since orthography is far more standardised than pronunciation, and L2 learners are typically more competent in spelling than pronunciation, especially when it comes to minimal pair words, orthographic transcription seems to offer greater validity in such cases.

The second criterion, reliability, refers to consistency in measurement (see, for example, Dörnyei, 2007). Regarding raters or coders, it may be defined as "degree of consistency with which instances are assigned to the same category by different observers or by the same observer on different occasions" (Hammersley, 1992, p. 67), which refers to inter- and intra-rater reliability respectively. These types of reliability may be improved by using predetermined, precisely defined criteria when coding or rating participant responses. However, depending on the type of responses, a certain amount of personal judgment by the rater may still be necessary, potentially compromising inter- and intra-rater reliability. One major advantage of orthographic transcription is that it typically necessitates less such judgment than other measures of intelligibility that involve a greater amount of open-ended input by respondents, such as summaries or comprehension questions. Shadowing does not fare better either, since, provided no acoustic analysis is performed, it involves the auditory impressions of raters. Orthographic transcription is superior in this respect especially when using the exact word match technique; comparing participant entries to standard orthography is a straightforward and unambiguous approach ensuring high inter-rater and intra-rater reliability.

With the allowance of spelling errors, however, some variation in raters' judgments is to be expected. The extent to which inter-rater reliability may thus be compromised depends on

[1] The distinction between these approaches is not entirely clear cut. For example, Bent and Bradlow (2003) measured intelligibility by "counting the number of keywords transcribed perfectly" (p. 1605), with missing or superfluous morphemes resulting in words being coded as incorrect. However, "obvious spelling errors were not counted as incorrect" (p. 1605). Similarly, Derwing et al. (2002) describe their coding scheme as "an exact word match technique", but with the proviso that they "ignore[d] spelling errors" (p. 252).

factors such as the amount of orthographic variation in the data and the proficiency level of listeners. With moderately sized samples of educated native speaker listeners or highly proficient L2 listeners, such variation is typically comparatively minor, and inter-rater reliability is unlikely to be unacceptably low. A different picture emerges when investigating large, linguistically heterogenous samples of L2 listeners at different proficiency levels, who may all bring their own L1-specific orthographic weaknesses with them. This is particularly relevant to the ever-growing field of ELF intelligibility research, which, by definition, aims at studying intelligibility for internationally diverse populations consisting mostly of L2 listeners. In such cases, researchers may face several challenges when coding orthographic transcriptions.

4 Problems with the allowance of spelling errors: A large-scale ELF intelligibility study

Before discussing the issues that arose in the process of coding orthographic transcriptions in my own study (see §4.2), we need to consider the research context, i.e., the study's aim and underlying hypothesis, the exact method adopted, and the nature of the sample of listeners (see §4.1). This is necessary to comprehend the significance and potential implications of the issues arising in the coding process for the study's findings.

4.1 Research context

The study investigated the effect of co-text and context on the international intelligibility of two features of an Austrian accent in English (for a rationale, see Thir, 2020a, 2020b). These were the realisation of TRAP /æ/ as [e] (thus conflating the TRAP – DRESS distinction) and the realisation of NURSE /ɜː/ as [øə].[2] A subsidiary aim of the study was to investigate the effect on intelligibility of these two sound substitutions (Thir, 2020a), as well as differences in intelligibility between mono- and disyllabic words, since longer words have been found to be more intelligible than shorter words (see e.g., Howes, 1957).

A listening experiment consisting of four different conditions was developed. Three of these involved the presentation of target words in sentence co-text in the form of a cloze test (for details see Thir 2020a, 2020b), while the remaining one was a control condition where words were presented in isolation. In the conditions involving a cloze test, participants saw a single sentence on their screens with a gap in the place of the target word. In the SYN CONDITION, carrier sentences were semantically neutral and merely indicated the part of speech (POS) of the target word (e.g., *It's quite* _____.), i.e., they contained a syntactic cue. In the SYN+SEM CONDITION, in addition to indicating the POS of the target word, carrier sentences also included a semantic cue (e.g., *feather* for the target word *bird* in *They found the feather of a* _____.). In the SYN+SCH CONDITION, carrier sentences were semantically neutral but an additional schematic cue in the form of a short description under which the statement was made was presented before each sentence (e.g., *At the airport* before the carrier sentence *I need to pick up my* _____., for the target word *bag*).

Intelligibility was measured as orthographic transcription, i.e., after hearing the entire sentence, participants had to type the target word into the gap on their screens before moving on to the next sentence. Each condition contained six different types of target words which were intermixed with nine distractor words, as shown in Table 1:

[2] These are Wells' (1982, p. 120) standard lexical sets for English vowels: TRAP: /æ/, DRESS: /e/, and NURSE: /ɜː/, which are used as keywords in the main text.

Table 1

The 24 Target Words in the Experiment

CONDITION	NURSE monosyll.	NURSE disyll.	TRAP monosyll. non-MP	TRAP disyll.	TRAP MP 'different POS'	TRAP MP 'same POS'
C	*birth*	*worker*	*rat*	*palace*	*sand* (n) – *send* (v)	*land* (v) – *lend* (v)
SYN	*nurse*	*curtain*	*flat*	*massive*	*bad* (adj.) – *bed* (n)	*gas* (n) – *guess* (n)
SYN+SEM	*bird*	*purple*	*van*	*chapter*	*dad* (n) – *dead* (adj./adv.)	*pan* (n) – *pen* (n)
SYN+SCH	*firm*	*servant*	*cab*	*servant*	*bag* (n) – *beg* (v)	*pants* (n pl.) – *pence* (n pl.)

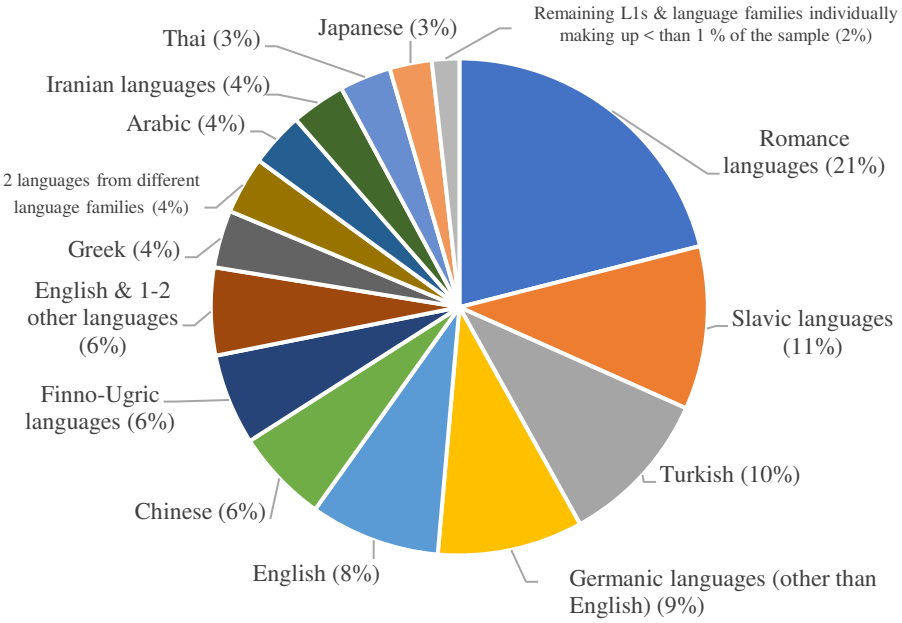
Note. For minimal pair (MP) words, the corresponding DRESS word and the part of speech (POS) is provided where n = noun, v = verb, adj. = adjective, adv. = adverb.

To obtain data from participants who were not necessarily familiar with the Austrian accent, the experiment was conducted via the internet with the help of the survey tool *SoSciSurvey*[3]. Listeners were recruited via e-mail, social media and the author's international contacts. The sample consisted of 508 (M = 175, F = 330, Other = 3) native listeners (NL: n = 66; 13%) and non-native listeners of English (NNL: n = 442; 87%), aged 18–74 years (M = 29.4). The listeners came from 81 different L1 backgrounds[4], as summarised in Figure 1. Most had a Romance language as their L1 (n = 107; 21%), followed by Slavic languages (n = 54; 11%), Turkish (n = 52; 10%) and a Germanic language other than English (n = 48; 9%).

[3] SoSciSurvey https://www.soscisurvey.de/
[4] Different combinations of two or three first languages were counted as different L1 backgrounds; for a full list, see Thir (2020b).

Figure 1

Overview of L1 Backgrounds Grouped into Selected Language Families in the Sample



Participants were asked to assess their listening proficiency using a slightly adapted version of the self-assessment scale for listening in a foreign language of the *Common European Framework of Reference* (Council of Europe, 2018, p. 167). Table 2 shows that most listeners considered themselves quite advanced (67% combined at C1 or C2 level), some reported being at an intermediate level (29% combined at B1 or B2) and the remaining participants assessed themselves at a beginner level (4% combined at A1 or A2).

Table 2

Participants' self-assessed listening proficiency

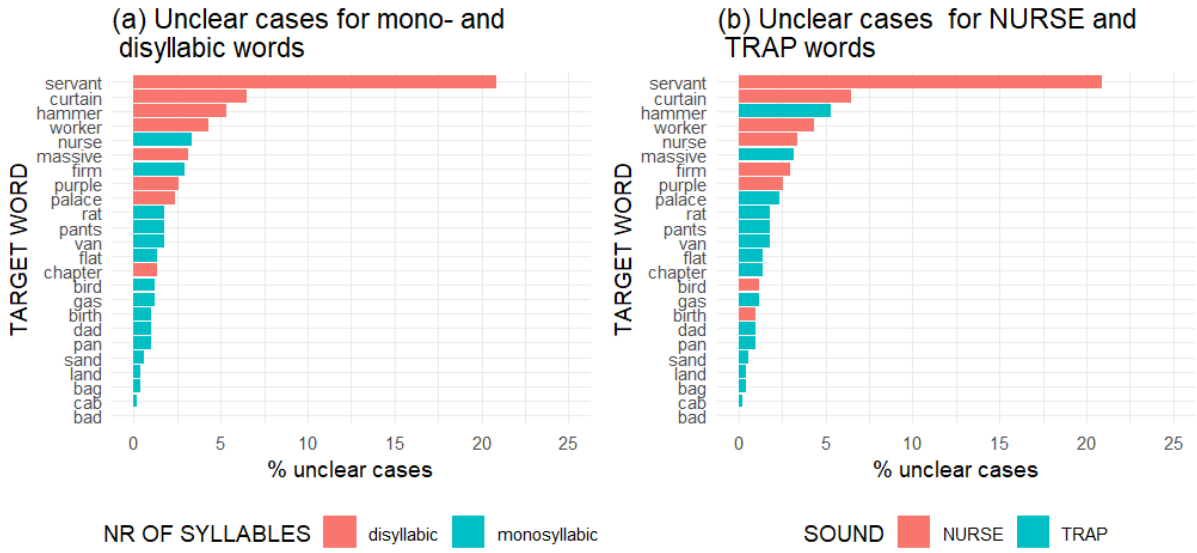
CEFR level	*n*	%
A1	8	2
A2	12	2
B1	52	10
B2	95	19
C1	129	25
C2	212	42
Total	*508*	

4.2 Issues in coding orthographic transcriptions

To increase the validity of the chosen measurement of intelligibility, the option of allowing spelling errors and missing or added morphemes was initially considered in the coding process. As a first step, 217 ambiguous transcriptions were identified, i.e., although they did not perfectly match the dictionary entry of a word, they could potentially be classified as indicative of intelligibility, which amounts to 1.8% of all 12.192 entries. Notably, certain target words exhibited a much higher percentage of such unclear cases than others. Figures 2a and 2b list all target words in descending order according to their proportion of unclear cases. The order is the same in Figure 2a (which compares mono- and disyllabic words) and Figure 2b (which compares NURSE and TRAP words). By far the highest proportion of unclear cases was identified for *servant* (20.9%), followed by *curtain* (6.5%) and *hammer* (5.3%).

Figures 2a and 2b

Percentage of Unclear Cases for Each Target Word: a) Mono- vs. Disyllabic Words; and b) NURSE vs. TRAP words



As shown in Figures 2a and 2b, there was a tendency for disyllabic words and NURSE words to exhibit a higher number of ambiguous transcriptions. Since differences in intelligibility between mono- and disyllabic words and between NURSE and TRAP words were part of the research questions investigated in Thir (2020a, 2020b), ensuring reliability in coding such cases was crucial to avoid biasing the results. Therefore, all ambiguous transcriptions were presented to eight researchers at the University of Vienna at an ELF research meeting. It soon became clear that inter-rater reliability was compromised, since there was considerable disagreement. For certain words, there was a continuum of ambiguous spellings (e.g., *purple* in Table 3), and it was impossible to decide where to draw the line between acceptable and unacceptable ones. For others, there were fewer, but equally tricky cases. For example, *van* (pronounced as [ven]), was transcribed as <vane> by six listeners. This could simply constitute a misspelling, but could also point to listeners having incorrectly identified the word as /veɪn/, transcribing it in

analogy to *pane*. Considering these circumstances, a strict exact word match approach[5] was chosen to ensure coding reliability, though this inevitably penalised (some of) the non-native listeners in the sample. Naturally, this needed to be taken into consideration when interpreting the results, especially in relation to differences in performance between native and non-native listeners, or listeners at different proficiency levels.

Table 3

Ambiguous Spellings for the Word Purple

Spelling	Frequency
purpel	1
purpal	1
purpul	1
purpule	1
purpole	1
puprle	1
perpleo	1
perpul	1
puple	1
pupple	1
purper	2
perpur	1

Note. The last two options were considered ambiguous since they resembled the German word *Purpur* (signifying *purple*).

There were, however, a few exceptions to this strict approach, which seemed clear cases of intelligibility. These included capitalisations (e.g., *Worker* or *wORKER* for *worker*), insertion of punctuation marks or of a numeral key (e.g., *flat4* for *flat*) and if parts of a carrier sentence had additionally been transcribed (e.g., *a van* for *van)*. In one case, a listener had noticed the phonological ambiguity in the stimulus and provided two options (*Land* and *Lent* for *land* in the control condition), which was also accepted.

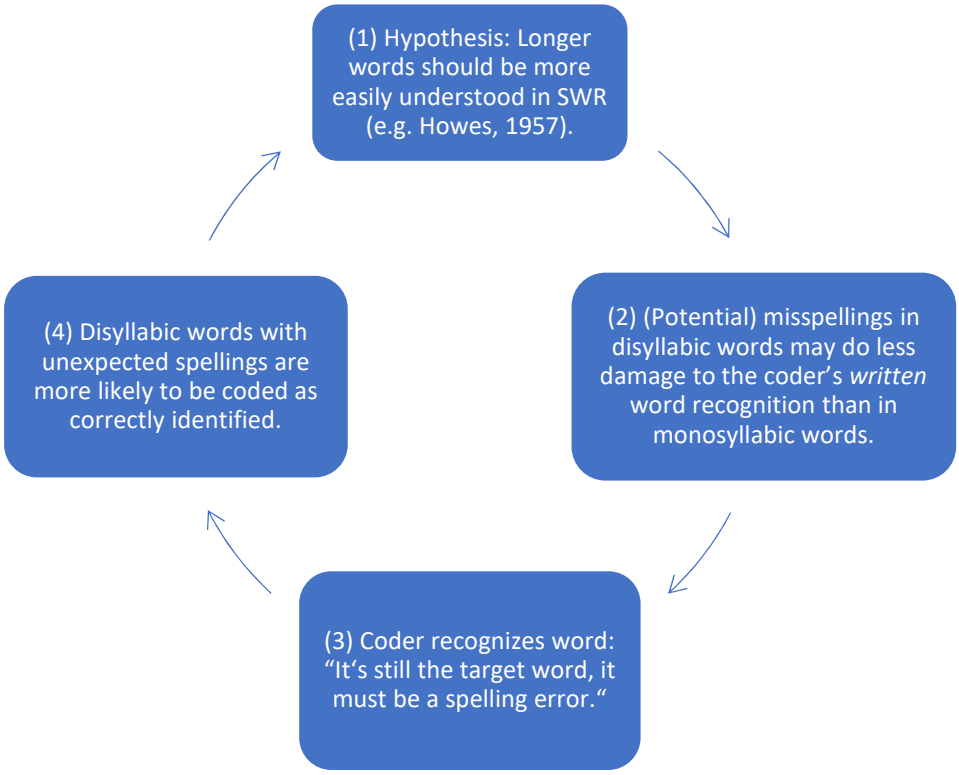
Another reason why the exact word match seemed most appropriate for this study relates to the issue of objectivity. During the coding process, it seemed more tempting to accept spelling errors in longer (i.e., disyllabic) words, since they were still more easily recognisable as the intended words. For example, it seemed obvious that **messive* was a misspelling of *massive*, while **ret* for *rat* or **lend* for *land* clearly seemed to indicate incorrect SWR, although all cases involved the substitution <e> for <a>. Especially in the case of TRAP minimal pair words, there seemed to be little reason to classify such substitutions as anything other than incorrect word identification. Since differences in intelligibility between mono- and disyllabic words and between minimal pair and non-minimal pair TRAP words were part of the research questions examined, treating the substitution <e> for <a> differently depending on the word type would have biased the results. Thus, there would have been a danger of circularity, or of falling prey to a self-fulfilling prophecy by using a coder's *written* word recognition of the target word as

[5] The choice was made in consultation with Prof. Barbara Seidlhofer, a leading scholar in the field of ELF.

a criterion in coding transcriptions. Figure 3 illustrates this issue using the example of the hypothesis that spoken longer words are more intelligible than shorter ones (caption 1). Partly influenced by this hypothesis, and partly influenced by their ability to recognise the intended word more easily in the case of disyllabic (i.e., longer) words (caption 2), the coder will likely conclude that an ambiguous transcription of a disyllabic word is a manifestation of orthographic weakness rather than failed SWR (caption 3). Consequently, disyllabic words are more likely to be classified as correctly identified than monosyllabic words (caption 4), which then results in the confirmation of the initial hypothesis (caption 1).

Figure 3

Example of Circularity in the Coding Process When Coder's Written Word Recognition Is Used as Criterion for Classifying Transcriptions



5 Discussion and conclusion

This paper highlighted several advantages of orthographic transcription to measure intelligibility if the construct is defined in terms of SWR, notably over the method of shadowing, which seems problematic when working with non-native listeners. At the same time, it stressed that the validity and reliability of orthographic transcription depends on which approach to coding is adopted: the exact word match or the allowance of spelling errors, with the former offering greater reliability but lower validity than the latter. The present paper highlighted how reliability and objectivity might be compromised when allowing spelling errors while working with large, linguistically heterogeneous samples of listeners at various proficiency levels, as is often the case in quantitative ELF intelligibility studies. Crucially, the

conclusion that should be drawn from this analysis is not that the exact word match technique is superior per se. Rather, it is necessary to carefully weigh the considerations discussed here in each particular research context, that is, in relation to the research questions and hypotheses examined, the nature of the stimuli material (e.g., minimal pair words) and the size and nature of the sample. Clearly, there will always be a trade-off between validity, reliability and also feasibility, and the exact nature of this trade-off can only be determined in each specific research context and needs to be taken into consideration when interpreting the results.

This paper also demonstrated that even in quantitative studies, some issues in data analysis might only emerge when taking a closer look at one's raw data, such as what participants typed in as transcriptions. This highlights the importance of poring intently over one's raw data, especially when working with under-researched populations for whom certain well-established research methods might turn out to be (somewhat) problematic. International listeners, especially at intermediate and lower proficiency levels, constitute such a population in quantitative intelligibility research. However, it is precisely these listeners who need to receive greater attention due to their ever-increasing use of English for cross-cultural communication worldwide and the ensuing issues of intelligibility in international contexts. This is not to say that well-established methods should not be used with such populations, but that certain adaptations might be necessary to ensure an appropriate balance of validity and reliability.

References

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, *114*(3), 1600–1610. https://doi.org/10.1121/1.1603234

Bent, T., Holt, R. F., Miller, K., & Libersky, E. (2019). Sentence context facilitation for children's and adults' recognition of native- and nonnative-accented speech. *Journal of Speech, Language, and Hearing Research, 62*(2), 423–433. https://doi.org/10.1044/2018_JSLHR-H-18-0273

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America, 106*(4), 2074–2085. https://doi.org/10.1121/1.427952

Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching assessment*. Council of Europe. https://rm.coe.int/cefrcompanion-volume-with-new-descriptors-2018/1680787989

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*(1), 1–16. https://doi.org/10.1017/S0272263197001010

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*(3), 399–423. https://doi.org/10.2307/3588487

Gass, S. (2015). Experimental research. In B. Paltridge & A. Phakiti (Eds.), *Research methods in applied linguistics: A practical resource* (2nd ed., pp. 101–117). Bloomsbury Academic.

Hammersley, M. (1992). *What's wrong with ethnography? Methodological explorations*. Routledge.

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *The Journal of the Acoustical Society of America, 29*(2), 296–305. https://doi.org/10.1121/1.1908862

Kang, O., Thomson, R., & Moran, M. (2018). Empirical approaches to measuring the intelligibility of different varieties of English in predicting listener comprehension. *Language Learning, 68*(1), 115–146. https://doi.org/10.1111/lang.12270

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review, 64*(3), 459–489. https://doi.org/10.3138/cmlr.64.3.459

Magnuson, J. S. (2017). Mapping spoken words to meaning. In G. Gaskell & J. Mirkovic (Eds.), *Speech perception and spoken word recognition* (pp. 76–96). Routledge.

Mirman, D. (2017). Zones of proximal development for models of spoken word recognition. In G. Gaskell & J. Mirkovic (Eds.), *Speech perception and spoken word recognition* (pp. 97–115). Routledge.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition, 28*(1), 111–131. https://doi.org/10.1017/S0272263106060049

Munro, M. J., & Derwing, T. M. (2015). Intelligibility in research and practice: Teaching priorities. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 377–396). Wiley-Blackwell. https://doi.org/10.1002/9781118346952.ch21

Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R. D. Kent (Ed.), *Intelligibility in speech disorders* (pp. 11–34). John Benjamins.

Smith, L. E. (1992). Spread of English and issues of intelligibility. In B. B. Kachru (Ed.), *The other tongue: English across cultures* (2nd ed., pp. 75–90). University of Illinois Press.

Thir, V. (2020a). International intelligibility revisited: L2 realizations of NURSE and TRAP and functional load. *Journal of Second Language Pronunciation, 6*(3), 458–482. https://doi.org/10.1075/jslp.20012.thi

Thir, V. (2020b). *Phonological intelligibility in ELF communication: The role of co-text and context* [Unpublished doctoral dissertation]. University of Vienna.

Wells, J. C. (1982). *Accents of English*. Cambridge University Press.

Zielinski, B. (2012). Intelligibility in World Englishes: Theory and Application by C. L. Nelson. *World Englishes, 31*(3), 404–406. https://doi.org/10.1111/j.1467-971X.2012.01765.x

About the author

Veronika Thir holds a PhD in English Applied Linguistics from the University of Vienna. Her research interests focus on issues of intelligibility in ELF communication, in particular as they relate to the use of linguistic co-text and extralinguistic context by international listeners. Her work has been published in the *Journal of Second Language Pronunciation* and as a chapter in *Language Change: The Impact of English as a Lingua Franca* (2021, CUP). She is currently working as a post-doctoral researcher at the University of Passau, in a research project on cultural differences regarding personal data disclosure.

Email: Veronika.Thir@uni-passau.de

Xi, X., Li, P., & Prieto, P. (2023). Reducing acoustic overlap of L2 English vowels through gestures encoding lip aperture. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 261–270). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225191

Reducing acoustic overlap of L2 English vowels through gestures encoding lip aperture

Xiaotong Xi
Pompeu Fabra University, Barcelona

Peng Li
University of Oslo

Pilar Prieto
Pompeu Fabra University, Barcelona
Catalan Institution for Research and Advanced Studies, Barcelona

This study aims to investigate whether hand gestures mimicking the lip aperture of non-native vowels can improve learners' production accuracy after audiovisual perceptual phonetic training. Sixty-six Catalan/Spanish bilingual learners of English were randomly assigned to either the No Gesture or Gesture group for training on the challenging English vowels /æ/ and /ʌ/. In the Gesture group, participants saw the instructor perform gestures mimicking the lip aperture of the low vowel /æ/ and mid vowel /ʌ/. Participants in the No Gesture group only saw the instructor produce the speech. Pronunciation performance was evaluated before and after training using paragraph reading, picture naming, and word imitation tasks. Pillai scores were used to measure the acoustic overlap between /æ/ and /ʌ/. The results showed that although both training groups showed less overlap between the two vowels after training, gestural training had a greater effect than no gesture training in the picture naming and paragraph reading tasks. These findings suggest that hand gestures mimicking visible articulatory features, such as lip aperture, can improve the pronunciation accuracy of non-native sounds in L2 learners.

Keywords: hand gestures, English vowels, L2 production, acoustic overlap, phonetic training

1 Introduction

1.1 Pronunciation training with multiple modalities

One of the difficulties adults face as they learn an L2 is the pronunciation of non-native sounds. While these difficulties can be attributed to factors such as the transfer from the L1 phonetic realisation to the L2 production (Kartushina & Frauenfelder, 2014), the limited quantity of the L2 input (Muñoz, 2008), and the quality of the L2 pronunciation instruction (Derwing & Munro, 2015), phonetic training can help overcome these difficulties.

According to the Dual Coding Theory (Paivio, 2014), learners may retain and retrieve information coded through verbal and non-verbal channels more easily than through one modality. Focusing on L2 phonological acquisition, audiovisual phonetic training can boost L2 speech sound learning by providing visual information from facial gestures (e.g., Hazan et al., 2005; Inceoglu, 2016). Furthermore, multimodal phonetic training with embodied cues may provide richer visual sources for L2 speech learning. This is supported by the Embodied Cognition theory, which holds that physical actions can shape cognitive processes (Wilson, 2002). Shapiro and Stolz (2019) as well as Sullivan (2018) argue that embodiment offers benefits in educational settings, where the role of hand gestures in learning has received significant attention. Specifically, while teachers' gestures can help lighten learners' cognitive load by shifting the burden from verbal to visuospatial processing, learners' self-performed gestures indicate whether they have fully comprehended the concept being taught (Shapiro & Stolz, 2019). The benefits of hand gestures have been extensively investigated in different learning domains, such as mathematics, science, and languages, among many others (for a review, see Novack & Goldin-Meadow, 2015). Regarding L2 pronunciation, hand gestures mimicking phonetic features have revealed positive effects on L2 speech production. In the following section, we will review the results of studies investigating the effectiveness of hand gestures in training L2 sound pronunciation.

1.2 Phonetic training with hand gestures

While hand gestures are frequently used by teachers in L2 classrooms to teach pronunciation (Hudson, 2011; Smotrova, 2017), their effectiveness in the learning of L2 sounds has only been tested in a few empirical studies, with mixed results. Some studies have provided positive results of using hand gestures. For example, horizontal sweep gestures mimicking durational features improved the pronunciation accuracy of L2 Japanese long vowels (Li et al., 2020). Similarly, hand gestures mimicking the air burst of stop consonants led to a more accurate pronunciation outcome right after training (Amand & Touhami, 2016; Xi et al., 2020) and at delayed posttest (Li et al., 2021). In contrast, Hoetjes and van Maastricht (2020) used hand gestures to mimic lip rounding and tongue protrusion in order to train Dutch speakers to produce Spanish /u/ and /θ/. However, only the gestures cueing lip rounding of /u/ showed positive effects on pronunciation. Therefore, more evidence is needed to assess the role of hand gestures on the pronunciation of non-native sounds when they mimic the articulatory features of the sounds.

1.3 Current study

The present study examines the effectiveness of multimodal phonetic training with hand gestures encoding articulatory features (specifically, lip aperture) in producing English /æ/ and /ʌ/. Since

Catalan/Spanish speakers lack the /æ/ and /ʌ/ vowels in their native language, they tend to perceive them as a single /a/ category and produce them with smaller spectral differences compared to English speakers (Aliaga-Garcia & Mora, 2008). Formant analysis showed that Catalan/Spanish learners' productions of the /æ/ – /ʌ/ pair were closer to each other compared to native speakers. However, high-variability auditory input training had limited effects on improving their production (Aliaga-Garcia & Mora, 2008). Since better articulatory control is crucial for accurate sound production, including relevant audiovisual and gestural information in phonetic training paradigms could enhance the benefits.

To accurately pronounce the pair of English vowels /æ/ and /ʌ/, two main articulatory features are crucial, namely, tongue height and tongue backness. The first vowel is an open-mid-to-open front vowel, whereas /ʌ/ is an open-mid, central vowel (Carley & Mees, 2020). This indicates that /æ/ is pronounced with a larger lip aperture and a more fronted tongue position than /ʌ/. Following Hudson's (2011) classroom observations, we designed a hand gesture to mirror the lip aperture of the two vowels. The thumb and fingers represent the lower and upper lips, respectively. The vertical distance between them indicates the size of the lip aperture, with a larger distance for /æ/ and a shorter distance for /ʌ/ (see Figure 1).

Figure 1

Hand Gestures for the English Vowels /æ/ (Left Panel) and /ʌ/ (Right Panel)

Based on previous research (e.g., Hoetjes & van Maastricht, 2020; Xi et al., 2020), we hypothesise that training with hand gestures mimicking lip aperture would boost Catalan/Spanish speakers' pronunciation of English /æ/ and /ʌ/ more than training without such gestures.

2 Research methodology

2.1 Participants

In a between-subjects study with a pretest–posttest–delayed posttest design, we recruited 66 Catalan/Spanish bilinguals (54 females, M_{Age} = 19.7 years, *SD* = 1.8) from a public university in Catalonia to learn the English vowel pair /æ/ and /ʌ/. They reported having an intermediate English proficiency level and started learning English at an average age of 5.2 years (*SD* = 2.0). None of them reported any hearing or speech impairments. All students volunteered to participate in the experiment and signed a consent form that allowed us to process their data.

Participants were randomly assigned to either the No Gesture (NG) group (n = 33, female = 26) or the Gesture (G) group (n = 33, female = 28). Both groups completed the same tests and differed only in the type of training they received. To ensure homogeneity, a series of ANOVA tests confirmed that no significant differences were found in age, age of L2 acquisition, extracurricular hours of English courses, study abroad weeks, and visuospatial working memory between the two groups (all $p > .05$).

2.2 Audiovisual materials for the phonetic training

The audiovisual materials for the phonetic training were recorded in a soundproof room. A General American English male speaker acted as the instructor and model speaker. For each training group, we prepared a familiarisation and training video. In the familiarisation video, the instructor provided an explicit explanation in English regarding the differences in lip aperture and tongue position between the two vowels. Especially for the G group, the instructor explained that the hand gesture represented lip aperture, in order to help participants to understand the articulatory feature and map hand gestures to articulatory information. For the training videos, we selected 6 pairs of English CVC minimal word pairs contrasting in /æ/ and /ʌ/ (e.g., *cat–cut*) and created 12 short sentences embedding each of the words (e.g., *A **cat** walks by*). The instructor produced the 12 words and 12 sentences while being video recorded. Again, he performed hand gestures in the G condition when producing the target vowels.

All the video clips were uploaded to the Tobii Pro Lab software[1] to generate the training project for each training group with the familiarisation video preceding the training video. The software allows researchers to easily create multimodal materials (images, videos, webpages, etc.) for carrying out eye-tracking research. The clips of words and sentences were repeated three times. The whole session lasted around 15 minutes for each group.

2.3 Procedure

The phonetic training was performed individually in a quiet room. The NG group watched the training video without gestural input during the training, while the G group watched the video with gestural input. Participants did not imitate the speech or hand gestures. They all took the test at three points in time: before, immediately after, and one week after the training with the same tasks. To assess the learning outcome from different angles, we included three tasks: a) a less-controlled paragraph reading task — where participants read an English paragraph[2] containing 14 instances of /æ/ and /ʌ/; b) a spontaneous picture naming task — where participants named 10 simple black-white line drawings[3] designed to elicit 10 words containing /æ/ and /ʌ/; and c) a well-controlled word imitation task — where participants imitated 6 minimal pairs of English CVC words contrasting only in /æ/ and /ʌ/ following the instructor's model speech. While the words in the paragraph reading and picture naming tasks never appeared in the training session, half of the words used in the word imitation task were trained. Participants were audio-recorded during the three testing sessions. After the production tasks, participants did two controlled tasks: a) a language background questionnaire; and b) a symmetry span task to measure the visuospatial

[1] https://www.tobii.com/

[2] The paragraph was adapted from the textbook *Phonetic Words and Stories*, Book 5 by Kathryn J. Davis. https://www.soundcityreading.net

[3] The drawings are from https://arasaac.org/

working memory capacity (Blacker et al., 2017).These tasks were included as previous research shows that language experience is associated with L2 pronunciation acquisition (Derwing, 2008), and visuospatial working memory correlates with the learning outcomes through instruction with hand gestures (Aldugom et al., 2020).

3 Data analysis and results

The first author manually annotated 6,794 target vowels using Praat (Boersma & Weenink, 2022). Among these, 129 items containing abnormal formant frequencies and 58 items with mispronunciation (i.e., the participants produced a non-target word) were excluded from the analysis. In addition, eight participants did not do the delayed posttest, and 46 items were not recorded due to technical problems. Thus, 6607 vowels were analysed.

We measured the acoustic distance between the production of the two vowels using Pillai scores. Pillai scores strongly correlate with native listeners' perceptual judgments and are useful in capturing phonetic training gains (Mora, 2021). The Pillai score ranges from 0 to 1, which reflects a speaker's overlap between two vowel realisations (Nycz & Hall-Lew, 2014). A Pillai score closer to 0 means greater acoustic overlap, whereas a score closer to 1 means greater frequency distance. We calculated the Pillai scores of the three tasks for each group at three different testing time points, using the midpoint of the vowels' first and second formants (F1 and F2). The formant values were exported using a script with the maximum number of formants set to 5 and the formant ceiling set to 5500Hz for female speakers and 5000Hz for males. Next, to account for the gender differences in vocal tract length, we normalised the formant values following the Nearey Intrinsic method with phonTools, version 0.2-2.1 (Barreda, 2015), and then averaged the values by the number of testing items. We conducted eighteen Multivariate Analyses of Variance (MANOVA) in RStudio with the normalised mean F1 and F2 mid-point values as the dependent variable and the vowel (2 levels: /æ/ and /ʌ/) as the independent variable. The summary output of each MANOVA gave the Pillai score (see Figure 2).

Table 1 shows the mean Pillai scores across the three tasks at the pretest, posttest, and delayed posttest. For the paragraph reading task, the NG group demonstrated increased overlap immediately and one week after training compared to the pretest. However, the G group showed decreased overlap after training, even though this improvement was not sustained. In the picture naming task, both groups showed decreased vowel overlap after training. The NG group demonstrated an increase in Pillai score at the posttest, followed by a decrease to the pretest level at the delayed posttest. The G group had a larger improvement from the pretest to the posttest, and although their Pillai score decreased after one week, it remained higher than the pretest score. Finally, in the word imitation task, both training groups demonstrated similar improvements in Pillai scores from the pretest to the posttest. However, at the delayed posttest, their Pillai scores decreased to a lower level than the pretest. Overall, the results suggest that hand gestures mimicking articulatory movements have limited effects on improving vowel production in the word imitation task. However, they are effective for improving vowel production in paragraph reading and picture naming tasks.

Figure 2

Vowel F1-F2 Plots with Normalised Mean Formant Values for /æ/ and /ʌ/ across Conditions and Tests from Paragraph Reading (Upper Panel), Picture Naming (Middle Panel), and Word Imitation Tasks (Lower Panel) across Groups and Tests

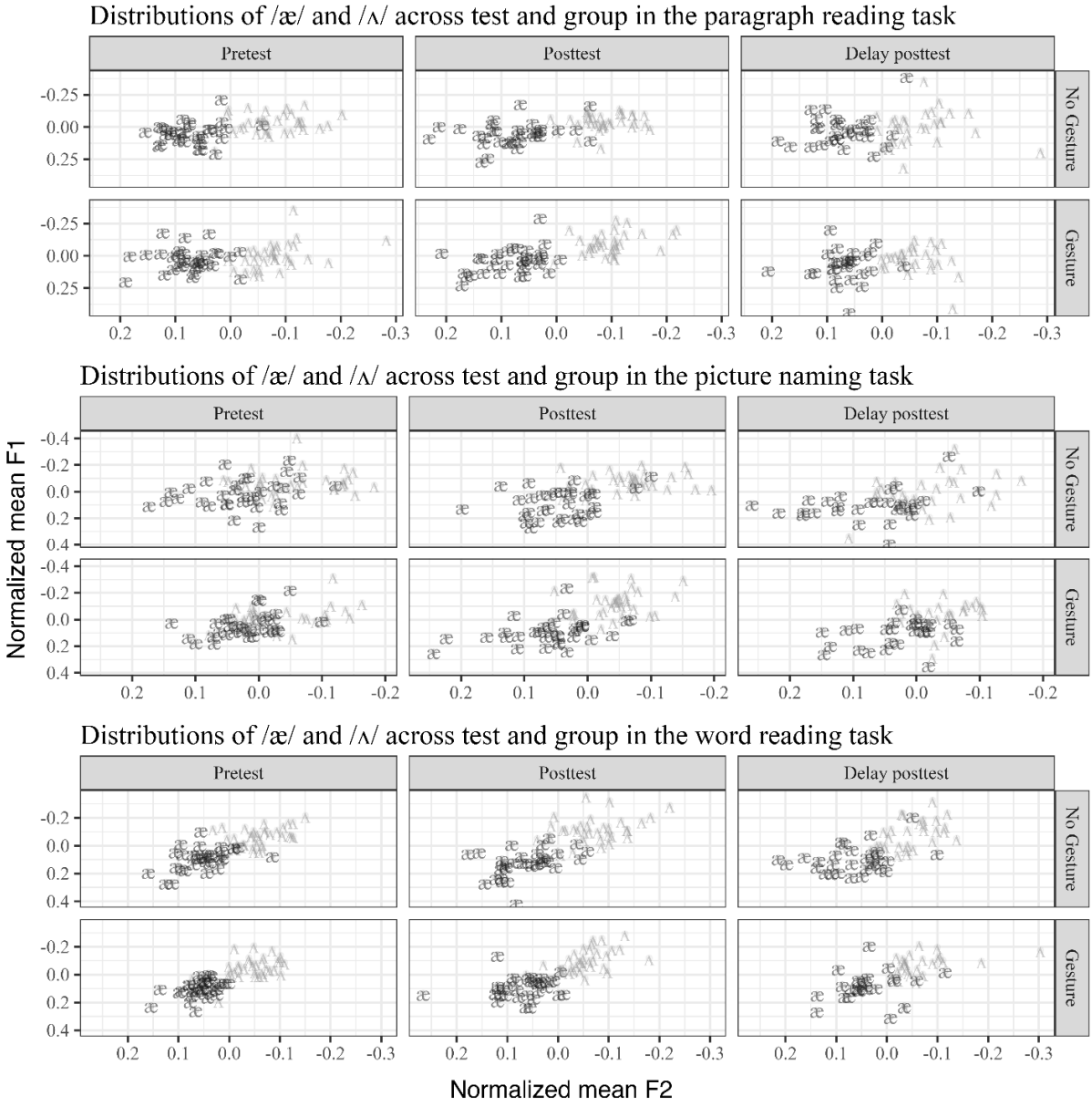


Table 1

Pillai Scores across Training Conditions and Testing Time in the Three Production Tasks

	No Gesture			Gesture		
	Pretest	**Posttest**	**Delayed posttest**	**Pretest**	**Posttest**	**Delayed posttest**
Paragraph reading task	0.70	0.69	0.60	0.71	0.74	0.71
Picture naming task	0.32	0.59	0.31	0.17	0.61	0.42
Word imitation task	0.66	0.70	0.53	0.68	0.71	0.52

4 Discussion and conclusion

The present study aimed to investigate the effectiveness of hand gestures that mimic lip aperture on improving the pronunciation of non-native English vowels /æ/ and /ʌ/. A total of 66 Catalan/Spanish learners of English participated in the study, and their pronunciation performance was evaluated through three tasks: a paragraph reading task, a picture naming task, and a word imitation task. The learners' pronunciation improvement was measured using Pillai scores, which are considered a reliable indicator of vowel overlaps in L2 speech production.

The results showed that the effects of hand gestures on vowel production varied depending on the task. In the paragraph reading and picture naming tasks, training with hand gestures reduced the vowel overlap from pretest to posttest more than training without hand gestures. Although after one week the vowel overlap of /æ/ and /ʌ/ became larger (shown by the reduced Pillai score), the overlap of the NG group became larger than the pretest level. In contrast, that of the G group was either back to pretest level in the paragraph reading task or was smaller than the pretest in the picture naming task. These findings suggest that hand gestures can help reduce L2 learners' acoustic overlap of English /æ/ and /ʌ/ in the less-controlled reading task and the spontaneous production task.

However, in the more controlled word imitation task, both training groups showed similar performance, with vowel overlaps of /æ/ and /ʌ/ decreasing immediately after training but increasing beyond the pretest level after one week. This suggests that neither training method helped L2 learners sustain the improvement in the pronunciation of the target vowels. These limited effects of hand gestures on imitating non-native sounds contradict previous studies (e.g., Li et al., 2020; Xi et al., 2020), and the discrepancy could be due to differences in learners' L2 proficiency and task difficulty. While the previous studies trained naïve learners with no experience in the target L2, our participants had intermediate L2 proficiency. Therefore, the controlled imitation task might have been too easy to detect improvements.

Our results are also consistent with a previous phonetic intervention study which identified more benefits from pronunciation intervention in the discourse-reading task than in imitation tasks by English learners with elementary to intermediate levels (Ozakin et al., 2022). In addition, speech production by L2 learners in controlled tasks, such as imitation, may not necessarily reflect their productive knowledge of non-native sounds (Llompart & Reinisch, 2019). Thus, the imitation task may not be an appropriate tool to test the outcome of phonetic training gains in L2 learners with good proficiency levels.

Taken together, the results of the present study suggest that multimodal phonetic training with hand gestures mimicking lip aperture may promote more native-like pronunciation of non-native vowels. Specifically, it helps reduce the acoustic overlap of difficult vowels in spontaneously produced target words and read sentences. Crucially, although participants trained with hand gestures still differed in their production from native English speakers (Pillai score: 0.857, Perry & Tucker, 2019), they showed improvement after a short training session (15 min). Future research could incorporate longer multimodal training sessions to potentially obtain further pronunciation gains. It should be noted that our observation is not based on the significance test, but on the change in Pillai score. As is shown in Table 1, the gestural effects were smaller in the paragraph reading task compared to the picture naming task. This suggests that multimodal training may yield larger effects in spontaneous speech rather than controlled speech production.

To conclude, encoding relevant articulatory information through hand gestures may help learners retrieve information from memory given that the mental representations of the speech sounds are constructed from more than one channel. This interpretation would support the Dual Coding theory (Paivio, 2014). The findings of this study also support the Embodied Cognition Theory (Wilson, 2002) and highlight the importance of involving embodiment in the L2 pronunciation classroom (Sullivan, 2018). Hence, this study supports multimodal pedagogical strategies that encourage teachers to include various sensory modalities in pronunciation instruction (Wrembel, 2011). Hand gestures are effective pedagogical tools that can be easily incorporated into the L2 pronunciation classroom.

Acknowledgments

This study was supported by the Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación and Fondo Europeo de Desarrollo Regional [PGC2018-097007-B-l00], and Agència de Gestió d'Ajuts Universitaris i de Recerca [2017 SGR_971]. The first author is supported by the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya and the European Social Fund under the Grant for the recruitment of early-stage research staff [2021FI_B 00137]. The second author is supported by the Research Council of Norway through its Centres of Excellence funding scheme [223265]. The authors sincerely thank Patrick L. Rohrer (Universitat Pompeu Fabra/ Nantes Université) for his help in creating the training materials. Many thanks to the audience at the EPIP7 conference for their useful comments on this study.

References

Aldugom, M., Fenn, K., & Cook, S. W. (2020). Gesture during math instruction specifically benefits learners with high visuospatial working memory capacity. *Cognitive Research: Principles and Implications*, *5*(27), 1–12. https://doi.org/10.1186/s41235-020-00215-8

Aliaga-Garcia, C., & Mora, J. C. (2008). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), *Recent research in second language phonetics/phonology: Perception and production* (pp. 2–31). Cambridge Scholars Publishing.

Amand, M., & Touhami, Z. (2016). Teaching the pronunciation of sentence final and word boundary stops to French learners of English: Distracted imitation versus audio-visual explanations. *Research in Language*, *14*(4), 377–388. https://doi.org/10.1515/rela-2016-0020

Barreda, S. (2015). *phonTools: Functions for phonetics in R* (R package version 0.2-2.1).

Blacker, K. J., Weisberg, S. M., Newcombe, N. S., & Courtney, S. M. (2017). Keeping track of where we are: Spatial working memory in navigation. *Visual Cognition*, *25*(7–8), 691–702. https://doi.org/10.1080/13506285.2017.1322652

Boersma, P., & Weenink, D. (2022). *Praat: doing phonetics by computer* (Version 6.1.51).

Carley, P., & Mees, I. M. (2020). *American English phonetics and pronunciation practice*. Routledge.

Derwing, T. M. (2008). Curriculum issues in teaching pronunciation to second language learners. In M. L. Zampini & J. G. Hansen Edwards (Eds.), *Phonology and second language acquisition* (pp. 347–369). John Benjamins Publishing Company.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: evidence-based perspectives for L2 teaching and research*. John Benjamins.

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, *47*(3), 360–378. https://doi.org/10.1016/j.specom.2005.04.007

Hoetjes, M., & van Maastricht, L. (2020). Using gesture to facilitate L2 phoneme acquisition: the importance of gesture and phoneme complexity. *Frontiers in Psychology*, *11*, 575032. https://doi.org/10.3389/fpsyg.2020.575032

Hudson, N. (2011). *Teacher gesture in a post-secondary English as a second language classroom: a sociocultural approach* [Doctoral dissertation, University of Nevada Las Vegas]. UNLV Theses, Dissertations, Professional Papers, and Capstones. https://doi.org/http://dx.doi.org/10.34917/2432927

Inceoglu, S. (2016). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics*, *37*(5), 1175–1199. https://doi.org/10.1017/S0142716415000533

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, *5*, 1246. https://doi.org/10.3389/fpsyg.2014.01246

Li, P., Baills, F., & Prieto, P. (2020). Observing and producing durational hand gestures facilitates the pronunciation of novel vowel length contrasts. *Studies in Second Language Acquisition*, *42*(5), 1015–1039. https://doi.org/10.1017/S0272263120000054

Li, P., Xi, X., Baills, F., & Prieto, P. (2021). Training non-native aspirated plosives with hand gestures: Learners' gesture performance matters. *Language, Cognition and Neuroscience*, *36*(10), 1313–1328. https://doi.org/10.1080/23273798.2021.1937663

Llompart, M., & Reinisch, E. (2019). Imitation in a second language relies on phonological categories but does not reflect the productive usage of difficult sound contrasts. *Language and Speech*, *62*(3), 594–622. https://doi.org/10.1177/0023830918803978

Mora, J. C. (2021). Assessing L2 vowel production gains after high-variability phonetic training: Acoustic measurements vs. perceptual judgements. In J. Romero (Eds.), *Proceedings of the 3rd International Symposium on Applied Phonetics – ISAPh 2021* (pp. 9–18). ISCA. https://doi.org/10.21437/ISAPh.2021-2

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, *29*(4), 578–596. https://doi.org/10.1093/applin/amm056

Novack, M., & Goldin-Meadow, S. (2015). Learning from gesture: How our hands change our minds. *Educational Psychology Review*, *27*(3), 405–412. https://doi.org/10.1007/s10648-015-9325-3

Nycz, J., & Hall-Lew, L. (2014). Best practices in measuring vowel merger. *Proceedings of Meetings on Acoustics*, *20*(1), 06008. https://doi.org/10.1121/1.4894063

Ozakin, A. S., Xi, X., Li, P., & Prieto, P. (2022). Thanks or Tanks: Training with tactile cues improves learners' accuracy of English interdental consonants in an oral reading task. *Language Learning and Development,* 1−16. Advance online publication. https://doi.org/10.1080/15475441.2022.2107522

Paivio, A. (2014). *Mind and its evolution: A Dual Coding Theoretical Approach*. Psychology Press.

Perry, S. J., & Tucker, B. V. (2019). L2 production of American English vowels in function words by Spanish L1 speakers. *Canadian Acoustics*, *47*(3), 94–95. https://jcaa.caa-aca.ca/index.php/jcaa/article/view/3328

Shapiro, L., & Stolz, S. A. (2019). Embodied cognition and its significance for education. *Theory and*

Research in Education, *17*(1), 19–39. https://doi.org/10.1177/1477878518822149
Smotrova, T. (2017). Making pronunciation visible: Gesture in teaching pronunciation. *TESOL Quarterly*, *51*(1), 59–89. https://doi.org/10.1002/tesq.276
Sullivan, J. V. (2018). Learning and embodied cognition: A review and proposal. *Psychology Learning and Teaching*, *17*(2), 128–143. https://doi.org/10.1177/1475725717752550
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. http://view.ncbi.nlm.nih.gov/pubmed/12613670
Wrembel, M. (2011). Cross-modal reinforcements in phonetics teaching and learning: An overview of innovative trends in pronunciation pedagogy. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 104–107). City University of Hong Kong. http://hdl.handle.net/10593/12093
Xi, X., Li, P., Baills, F., & Prieto, P. (2020). Hand gestures facilitate the acquisition of novel phonemic contrasts when they appropriately mimic target phonetic features. *Journal of Speech, Language, and Hearing Research*, *63*(11), 3571–3585. https://doi.org/10.1044/2020_JSLHR-20-00084

About the authors

Xiaotong Xi is a PhD student at Universitat Pompeu Fabra in Spain. Her main research interests include second language pronunciation acquisition, phonetic training, and hand gestures.

Emails: xiaotong.xi@upf.edu

Peng Li is a postdoc fellow at University of Oslo in Norway. His research focuses on the second language speech learning, multimodal pronunciation training, and prosodic studies.

Emails: peng.li@iln.uio.no

Pilar Prieto is an ICREA research professor at Universitat Pompeu Fabra in Spain. Her main research goal is to understand the role of prosody and co-speech gestures in human communications from a crosslinguistic, developmental, and cognitive perspective.

Emails: pilar.prieto@upf.edu

Young, R. and Messum, P. (2023). Pointing on charts is a third way of working in the pronunciation classroom. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 271–280). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8225744

Pointing on charts is a third way of working in the pronunciation classroom

Roslyn Young
Pronunciation Science Ltd.

Piers Messum
Pronunciation Science Ltd.

In the teaching of L2 pronunciation, speech and writing are the two principal ways to present content and correct errors. In our workshop, we described how their characteristics differ and the pedagogical significance of the differences. Speech is always ephemeral but sometimes too fast for learners to follow, whereas writing persists but is slow to produce. Both have their place, but sometimes in class neither is well suited to the particular demands of the moment. Pointing on charts is a third way of working which is fast to execute, visible like writing but ephemeral like speech. In our experience, pointing facilitates the creation of joint attention and learner involvement, and well-designed charts make it easy to work with precision. These and the other characteristics of pointing on charts show the value of this third way of working on L2 pronunciation.

Keywords: pointing, joint attention, phonemic chart, wall charts, Silent Way

1 Introduction

This article is about teachers using pointing on wall charts as a teaching tool in the foreign language classroom, a way of working developed by Gattegno (1963) in his Silent Way approach. We start by noting three ways in which pointing is central to the story of human development and how pointing relates to teaching through its capacity to create joint attention. We discuss how charts are best designed and how in the pronunciation classroom pointing can facilitate learner participation and engagement. Finally, before concluding, we compare pointing to speaking and writing as an effective, complementary way of presenting content and of correcting learners' mistakes.

2 Pointing as a way of creating joint attention

Morrison (2020a) identifies pointing as an essentially human activity which emerged several million years ago, and argues that it has been fundamental to human development. He proposes that pointing and a proto-language co-evolved at some time after the divergence of the chimpanzee and hominid lineages, about six million years ago. Chimpanzees did not develop pointing (although chimpanzees in captivity do learn this skill), whereas the hominids learned to use pointing to create joint attention. However, additional signals were sometimes useful to establish what the pointing gesture was meant to indicate, and the disambiguation of pointing provided a specific, concrete starting point for protolanguage (see also Morrison, 2020b).

Other researchers have developed the Cooperative Eye Hypothesis (Kano, 2022; Kobayashi & Kohshima, 2001) describing how in hominids a physical attribute evolved that enabled another form of pointing. As shown in Figure 1, humans have eyes with visible sclerae (the whites of their eyes), while the sclerae of animal eyes are not usually visible. As a result, the position of the human iris shows the direction of gaze and human eyes can effectively point and thus create joint attention. Animal eyes are made for seeing; human eyes are made both for seeing and for being seen.

Figure 1

Human Eyes and Animal Eyes

Human babies learn to both point and follow pointing by the end of their first year (Butterworth, 2003). Thus, pointing as a way of creating joint attention predates any use of speech or writing for this purpose ontogenetically as well as phylogenetically. Tomasello et al. (2007) describe the cognitive sophistication of infant pointing and claim that pointing is integral to what it means to be human.

3 Pointing in language teaching: The development of charts

By the time Homo erectus emerged about 2.3 million years ago, toolmaking had become so sophisticated that teaching was necessary to pass the skills required from generation to generation, according to Morrison (2020a). He defines human teaching as "a joint attentional activity in which a relative expert … goes out of her or his way to help a relative novice acquire some new and generally useful component of knowledge or skill" (p. 17). He observes that "directing attention is a core and indispensable tactic in any human teacher's repertoire" (p. 17). The link between pointing and teaching is thus clear:

> Physical pointing (and later, symbolic pointing, as I am doing here, by making the phrase 'physical pointing' my subject) is fundamental to teaching because it establishes, and refers to, an object of joint attention, without which teaching cannot take place" (Morrison, 2020a, p. 213).

The best-known use of pointing on charts in language teaching is found in Gattegno's Silent Way work. His first wall charts published in 1962 evolved out of the Words in Colour charts he had developed for teaching reading and writing. He laid out the function words of L2 (originally for English, French, and Spanish), colour coded for their pronunciation, on a set of word charts. He also showed all the sound-spelling correspondences of each language on a chart that he called a Fidel, borrowing the Amharic word for a syllabary. Teachers and learners can interact with these charts, typically using telescopic pointers, to create sequences of sounds/spellings that generate words, and sequences of words that generate sentences (Figure 2[1]). With the Silent Way charts, therefore, anything that can be said in the L2 can be pointed to on the charts and the pronunciation (and its relationship to the spelling of the words) can be worked on.[2]

[1] Screen capture from https://www.youtube.com/watch?v=0YAqt-gtWcI

[2] Examples of pointing on charts by teachers and learners can also be seen in the video "Pointing on charts: A technique from the Silent Way" at https://www.youtube.com/watch?v=_3BQJcXx7S0

Figure 2

Learners Interacting with Charts and with Each Other, Independently of a Teacher

Note. The set of PronSci English charts shown contains Word charts (left and centre-right), a Rectangle (phonemic) chart (centre-left) and Spelling chart (Fidel) (right).

Creating joint attention within a whole class was one significant benefit of working this way. Other benefits for classroom dynamics[3] include:

- Heads are up and everyone is 'speech-ready'. This creates a likelihood of participation and a potential for spontaneous interaction between the learners once the pointing has finished.
- The teacher can see the learners' faces and what they reveal about each person's learning.

In 1978, Gattegno abstracted the sounds of L2 and their colours from his American English Fidel to create a phonemic chart. Using this, words could be generated without involving spelling. Since then, many people have developed phonemic charts, usually in 'pigeonhole' designs with each cell labelled with an IPA symbol. These can be used as pointing materials in the same way as Gattegno's chart. Gattegno's Word charts, intended for beginners, have been supplemented by charts designed for more advanced learners. Other types of charts have been developed to address, for example, verb conjugation and grammar in French.

In the case of pronunciation, pointing might involve a phonemic chart, but the teacher could also use a midsagittal section of the head for the placement of the tongue or a physical model designed to help learners develop the articulatory setting of L2. Each of these materials provide information that goes beyond the simple lists of sounds found in many language textbooks. Such lists indicate what sounds are used in a language, but they do not provide information about how to articulate the sounds or the relationships between them. Similarly, a list of city

[3] For more detailed explanations, see Messum (2018).

metro stations is not as useful as a map of the metro system; the former indicates that a destination exists, but the latter helps a traveller to see how to get there.

Phonemic charts have the potential to do the same for the learning of sounds, although in our view most designs fall short of what is possible (see §4). For example, charts such as the one proposed by the British Council (see Figure 3) are somewhere between an inventory and a map, in that they do make a start at showing relationships between sounds but only in a basic way. To explore chart design and rationale in more detail, the next section will compare the British Council (BC) chart with the PronSci chart (Figure 4).

Figure 3

British Council Phonemic Chart for British English[4]*, with Sounds Organised in Pigeonholes*

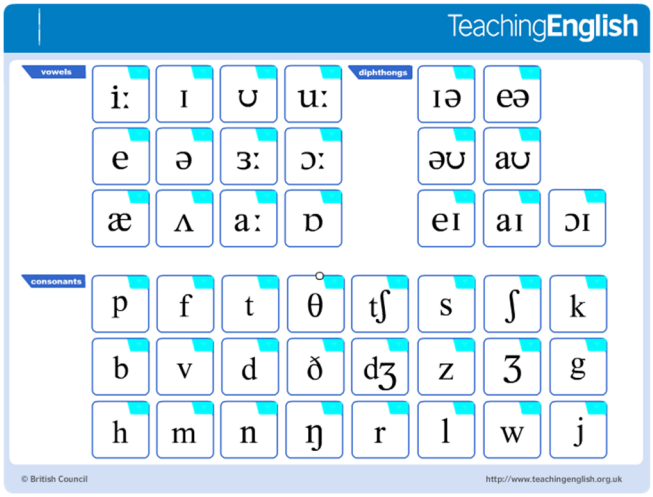
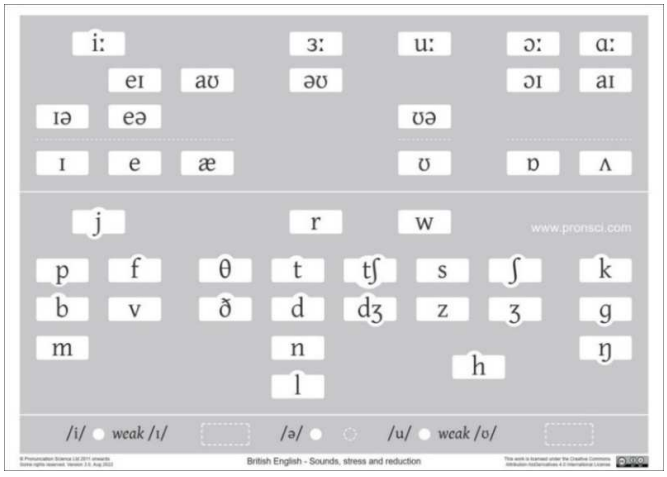


Figure 4

PronSci Phonemic Chart for British English[5] *with Sounds Arranged to Create a Stylised Map*



[4] The British Council chart is available as a free download at https://www.teachingenglish.org.uk/teaching-resources/teaching-secondary/teaching-tools/phonemic-chart

[5] The PronSci phonemic chart is available as a free download at https://www.pronunciationscience.com/guides/

4 Comparison of phonemic charts

A phonemic chart has a role to play in the learning of sounds for both production and perception, in the presentation of new material (the pronunciation of new words and phrases) and in the diagnosis and correction of mistakes. So when we refer to learning new L2 speech sounds, we are referring to the learner producing new sounds in isolation but also within sequences of sounds. For the presentation of whole words and of sequences of sounds in sentences, more is involved, as English has both free lexical stress and free sentence stress. The teacher has to integrate prosody into the presentation of the bare sequence of sounds.

The British Council (BC) chart is a typical example of a phonemic chart intended for use in a classroom. For reasons of brevity, we will only compare the Pronunciation Science (PronSci) chart to this one, and only do so in three respects.

First, the layout of the BC chart provides an inventory of the sounds of southern British English (SBE) in a set of pigeonholes and has a high informational density. The sounds are presented in two dimensions rather than as a list, which makes it easier to apprehend as a whole. The second dimension allows for some relationships and characteristics of the sound system to be indicated. Such information is carried by adjacency: someone who already knows about the phonology of SBE can quickly see the logic behind the arrangement, with evident relations between some horizontal and vertical neighbours. But a pigeonhole design also ends up forcing sounds that have no relation into adjacency: e.g., while /ŋ/ has a relationship of nasality with /n/, it has nothing in common with /ð/ or /r/; /h/ has nothing in common with /b/ or /m/. For a novice learner, therefore, adjacency is not a reliable guide to relationships on this chart. In contrast, the PronSci chart shows /ŋ/ in a relation with the other nasals, and with /k/ and /g/. The relative isolation of /h/, a breathy onset to a vowel, indicates that it has no relation to any other consonant in English. On the PronSci chart, adjacency is a reliable indicator of significance for learners, because the chart uses space to separate sound cells meaningfully[6].

Second, the treatment of schwa in both charts is significant, given its role in spoken English. In the BC chart, schwa is treated as just another vowel, whereas the PronSci chart gives schwa a distinct place within the inventory, highlighting its importance to the prosodic system of English. It is shown as a reduced vowel at the bottom of the chart (separate and different, therefore, from full vowels), and joined there by weak [i] and [u] sounds, as found in advanced learner's dictionaries (e.g., Deuter et al., 2015). Furthermore, schwa is located in a part of the chart that only shows unstressed vowels. Splitting vowels into those that can form the nucleus of a stressed syllable (pointed in the upper part of the chart), and those that can form the nucleus of an unstressed syllable (pointed in the lower part), introduces the stress pattern of any word that is pointed. In a language with free word stress, its prosody is as characteristic of a word as its segmental composition. When a new word is pointed on the PronSci chart, both its pronunciation and its stress pattern are simultaneously shown. Thus, in pointing to polysyllabic words, the action of pointing schwa in the bottom line of the chart is visually striking and highlights its distinctive quality as a reduced, low energy sound (see also Young & Messum, 2022).[7]

Our third criticism of the BC-type design relates to the organisation of the vowels. The BC chart arranges the pure vowels of SBE in a way that approximates to their positions on the IPA quadrilateral. These positions, in turn, approximate to the highest point of the tongue in the mouth, which correlates with the acoustically significant point of maximum constriction in the mouth for most, but not all, vowels. However, L2 learners are not phoneticians; they have no sense of formants as components of a vowel sound (first formant for vowel height, second

[6] For a full description of the relationships portrayed, see Messum and Young (2014).
[7] The PronSci chart also resolves the issue of pointing full but unstressed vowels (see Messum & Young, 2017).

formant for degree of vowel backness), and relative tongue heights are not proprioceptively available to them. It is not easy to become sensitive to either. If there were no alternative ways of organising vowels, then a pedagogical chart might as well follow a tool designed for phonetic analysis, but a more useful approach is for an arrangement to represent articulatory gestures to which learners can become sensitive.

The PronSci chart therefore reflects Esling's (2005) organisation of vowels into front, central, raised, and retracted regions that reflect natural actions of the tongue given its muscle physiology. Thus, vowels are shown in four groups, left to right. The vertical dimension of a non-pigeonhole design is then exploited to separate tense and lax vowels, a distinction which is of pedagogical significance for all the characteristics which this distinguishes in English: phonotactics (vowels in free and checked syllables), length (in stressed syllables), and degree of constriction (tense vowels requiring lingual gestures that move the tongue closer to the wall of the vocal tract than for lax vowels).

As we have found in our teaching and teacher training experience, the PronSci design also has two practical advantages for classrooms where pointing is going to be used extensively, First, the less regular layout makes it easier for teachers and learners to become familiar with the place of each sound. And second, the gaps between the cells can be helpful; when learners have to follow a pointer, it needs to be clear exactly which sound is being pointed at; pointing near a boundary in a pigeonhole design can be confusing for those at the back of the class.

In general, even a chart with a pigeonhole design has four important benefits if it is on the classroom wall in constant view and, most importantly, in use. First, the learners can see the task ahead of them: to master the production of all these sounds. They cannot fail to notice the existence of each of the sounds in the language, including sounds they cannot yet distinguish. Second, they can start sorting the multitude of sounds they hear in speech into a limited number of phonemes. Third, they know on a continuing basis which sounds they have mastered and which they still have to work on. Finally, when learners are pointing to the sounds in a word, they see all the possibilities in front of them and have to decide which one they can or must choose. This sharpens their awareness of sounds.

Nonetheless, a pigeonhole design limits the number of relationships that can be portrayed. We have found that when sounds are presented on a chart which shows as many of the significant relationships between them as possible, learners are more likely to be intelligent in the approach they take to learning sounds. It generates interest in the task and commitment to it. The Pronunciation Science (PronSci) phonemic charts[8] for British English, American English, French, Spanish, and other languages abandon the pigeonhole convention and make use of space to delimit groups of sounds with similar features and other relationships.

5 Pointing's potential impact on learner participation and engagement

Learners can be invited to point on a phonemic chart very early in a course, sending a positive signal about participation and taking responsibility for one's own learning. When a learner is pointing, the atmosphere in the class often becomes more collegial and some anxious learners (Horwitz et al., 1986) feel free to participate more. For the teacher, learners' pointing and speaking is revealing of their present competence. Learners can be invited to point after just six or seven sounds have been introduced, when they might begin to feel a pressure to memorise the relationship between the colours or symbols on the chart and the sounds they represent. Work that promotes familiarity rather than memorisation is helpful at this point. The aim, after all, is not to learn the chart, but to learn to pronounce the sounds.

[8] The Pronunciation Science charts have been developed by the authors (Messum & Young, 2014).

Once learners understand how a chart presents the spoken language, they can use it to ask questions by initiating pointing activities themselves:

- They check their pronunciation with the teacher.
- They explore alternatives to a given sequence which are often generated by different registers or rates of speech, e.g., 'February' pronounced in 4, 3, or 2 syllables.
- They ask questions about things they have heard in natural speech which puzzle them (e.g., contractions, the use of weak forms, liaisons, etc.). Pointing on a chart allows the teacher or the learners to show visually what is actually said in connected speech, for instance, how contractions change sounds, how stress influences reduction, how rate of speech modifies strings of sounds, and other phenomena.

To work on these questions efficiently without a chart, learners would have to know how to write phonemic symbols, whereas just pointing to these symbols on a chart is easier. This may lower the barrier to asking about, for example, a particular sound or word, or connected speech phenomena.

Pointing reflects the ephemeral quality of spoken language. It requires the learners to remain attentive: they have to actively notice what is taking place because as soon as the pointer leaves a cell, no trace of its passage remains. Rather than reading a word written in full on the board, they have to mentally note which sounds are being touched and in what order by sub-vocalising as the pointing takes place. This has the effect of enhancing retention. If they can't remember what was pointed, they cannot say it. The teacher knows this immediately, and can give some small hints, or choose to do the work again if this seems to be necessary.

Teachers need certain skills to point effectively: choosing a suitable pointer, where to stand, how to point as the teacher, introducing sounds and building sound sequences, getting learners to point, etc. (see Young, 2018). When learners are pointing they have to engage deeply with pronunciation. As an activity where performance is required, pointing forces learners to develop criteria for correctness. Standing at the chart, pointer in hand, they have to consider the choices available to them (selection of sounds, pattern of stress and reduction, etc.). The other learners in the class work on the same problem at the same time, and will make it obvious if they think the learner who is pointing has made a mistake or if they do not understand the learner's choices.

6 Pointing vs. speaking and writing

To work on a pronunciation problem, learners must examine what has been said. Language classes generally work on problems in two ways: spoken discussion and/or writing on the board. Both have their place, but neither is entirely satisfactory. Speech allows for a quick intervention, but is often too ephemeral to allow students to examine the language. The permanence of writing allows errors to be reflected upon, but writing is cumbersome and it reduces a speaking task to a reading one. For both of these reasons, the work loses intensity and learners tend to become distracted. Pointing to a sequence of sounds or words on a chart, whether done by the teacher or the learners, provides a third way of working on errors or presenting content, midway between speech and writing. First, pointing takes place more slowly than speech for those times when speech is too fleeting for its details to be followed, but it is faster than writing, saving time because sounds need only be 'touched' or tapped. Second, pointing is less ephemeral than speech because the learners do 'see' the sounds. However, pointing is not as permanent as writing because each sound is left behind when the pointer moves on.

Speaking in order to correct has two further disadvantages that pointing overcomes:

- An oral correction is usually addressed to the learner who has the problem and other learners in the class may not feel involved. Pointing on a chart turns a private problem into a class-wide lesson by creating joint attention on the issue.
- It is well known that listeners hear a foreign language through the filters of their L1. The teacher cannot be sure that the learners have heard the spoken correction as it was intended. In contrast, pointing is unambiguous.

Writing also has two key disadvantages that pointing avoids:

- To know the pronunciation of a word in L2, a learner has to know exactly what speech sounds it is made up of. Presenting a transcription of the word in a textbook or written on the board does not make sufficient demand on many learners. It relieves them of the need to create an image of the word in their minds. They only need to read the transcription to the point of basic recognition. However, in order to make progress, they need to be mentally active, moving their articulators while they read, vocalising the phonetic symbols or sub-vocalising them. When reading a word on the board, many learners do not make the effort this requires. In contrast, the ephemerality of pointing requires learners to do this work.
- Writing breaks the flow of the lesson: the teacher is (usually) not looking out at the class, and they can only hope that the class remains attentive, but they are unable to control this. If the teacher is pointing, they are facing their learners and can see whether or not the learners are involved.

Thus, pointing on a chart can be better adapted than either of the other two ways of working. On the one hand, it gives speech enough permanence to be examined in detail; on the other hand, it gives the learners a visible version of a spoken sentence without the permanence of writing.

7 Conclusion

When pronunciation is being dealt with, pointing on charts is an effective way of working in the classroom. Using even the simplest phonemic chart has its benefits, but a well-designed chart coupled with spelling and word charts is a comprehensive toolkit for addressing a range of pronunciation issues.

In our own teaching, we have found that learners like the various challenges of working with charts: they enjoy following the teacher pointing, they enjoy pointing themselves, and they vicariously enjoy pointing when fellow learners are working with a chart. Many classroom interventions relating to pronunciation, whether initiated by the teacher or the learners, may be improved by pointing.

References

Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 9–33). Psychology Press. https://doi.org/10.4324/9781410607744

Deuter, M., Turnbull, J., Bradbery, J., Hey, L., & Holloway, S. (Eds.). (2015). *Oxford advanced learner's dictionary of current English* (9th ed.). Oxford University Press.

Esling, J. H. (2005). There are no back vowels: The laryngeal articulator model. *The Canadian Journal of Linguistics / La Revue Canadienne de Linguistique*, *50*(1), 13–44. https://doi.org/10.1353/cjl.2007.0007

Gattegno, C. (1963). *Teaching foreign languages in schools: The Silent Way* (1st ed.). Educational Explorers.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*, 125–132. https://doi.org/10.2307/327317

Kano, F. (2022). Evolution of the uniformly white sclera in humans: Critical updates. *Trends in Cognitive Sciences*, S1364661322002327. https://doi.org/10.1016/j.tics.2022.09.011

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: Comparative studies on external morphology of the primate eye. *Journal of Human Evolution*, *40*(5), 419–435. https://doi.org/10.1006/jhev.2001.0468

Messum, P. R. (2018). Why we should use a chart and a pointer for teaching pronunciation. *Speak Out!*, *58*, 53–60.

Messum, P. R., & Young, R. (2014). *The PronSci rectangle charts: Phonetic guide*. Pronunciation Science Ltd. https://www.pronunciationscience.com/guides/

Messum, P. R., & Young, R. (2017). *The PronSci English Rectangle charts. Part 2, Using the advanced features*. https://www.pronunciationscience.com/guides/

Morrison, D. M. (2020a). *The coevolution of language, teaching, and civil discourse among humans*. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-48543-6

Morrison, D. M. (2020b). Disambiguated indexical pointing as a tipping point for the explosive emergence of language among human ancestors. *Biological Theory*, *15*(4), 196–211. https://doi.org/10.1007/s13752-020-00355-6

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*(3), 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x

Young, R. (2018). How to use a chart and a pointer for teaching pronunciation. *Speak Out!*, *59*, 20–26.

Young, R., & Messum, P. R. (2022). Teaching schwa: Using intentional 'stuttering' to improve English pronunciation. In J. Levis & A. Guskarova (Eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharines, ON. https://doi.org/10.31274/psllt.13367

About the authors

Roslyn Young worked at the University of Franche-Comté in the Applied Linguistics Centre from 1968 until she retired, teaching spoken English and sometimes spoken French in intensive courses. She ran her first 'pronunciation clinic' in 1978 and has worked continuously since then on ways of improving the teaching of pronunciation. She wrote her doctoral thesis on Caleb Gattegno's model of learning and its relevance to his work in language teaching. She has published numerous articles on the teaching of pronunciation and on how people learn.

Email: roslynyoung@gmail.com

Piers Messum has taught English in Japan, France and the UK. His PhD in Phonetics from University College London investigated how children learn L1 pronunciation. He co-authored *How We Learn and How We Should be Taught* (2011) with Roslyn Young. He now teaches in London, and does teacher training and materials development for Pronunciation Science Ltd.

Email: p.messum@gmail.com

Index

V

W