REVIEW PAPER

Training English Word Stress Perception and Production with Technology

Marina Ivanova

Chemnitz University of Technology, Germany

marina.ivanova@phil.tu-chemnitz.de

Abstract

Word stress is frequently afforded secondary importance in English teaching as stress placement rules are complex and because stress can be learnt along with each new word. However, training learners to pay more attention to word stress cues can support them in predicting the stress patterns of new vocabulary. Also, for speakers of fixed stress languages, perceiving and producing word stress can be more challenging, as they are accustomed to fixed stress patterns and weak acoustic stress cues. Moreover, even though L1 English speakers largely cue stress segmentally through vowel quality, speakers of languages which do not feature vowel reduction find it difficult to use this cue in stress perception and production. New methods such as High-Variability Phonetic Training (HVPT) and certain mobile learning tools have been found to benefit foreign language acquisition. These approaches have the potential to help learners master English word stress, yet most available tools focus on vowels and consonants rather than stress perception and production. This article, therefore, reviews methods for word stress teaching, both with and without technology, and presents a mock-up of a HVPT tool that focuses on both segmentals and suprasegmentals.

Keywords

Word stress teaching; High-Variability Phonetic Training; perception and production training tool

1. Introduction

Providing training in the pronunciation of segmentals like vowels and consonants, as well as suprasegmentals like word stress and intonation, requires a lot of time and facilitating individual feedback can be difficult within a packed curriculum and with large class sizes. English has variable stress, consequently, L2 English listeners often struggle with its placement. English stress rules are complex and largely determined by vowel duration (van der Hulst, 2010). Thus, learners from fixed stress language backgrounds such as Czech, languages lacking stress such as Japanese, or tonal languages such as Chinese may face difficulties with perceiving and producing word stress cues (pitch, duration, intensity, and vowel quality) as produced by L1 English speakers. Vowel duration and quality are particularly challenging to acquire by the many learner groups whose L1 employs less vowel reduction than English (e.g., Cruttenden, 2014).

Common arguments against word stress teaching are that 1) stress can be learnt with each new word, and 2) since stress rules are difficult and mis-stressing may not be problematic for intelligibility, stress does not need to receive a strong focus in language teaching where the learner is aiming to acquire English as a Lingua Franca (Jenkins, 2000, p. 135). While it is true that shifting word stress and changing the vowel's quality impairs understanding more than merely shifting stress (Cutler & Clifton, 1984), L2 English speakers often have issues with vowel quality, thus, all cues may become important for comprehension.

Technology can provide customisable solutions for perception and production training, allowing students to engage in independent learning. E-learning resources like *YouGlish*, a tool that retrieves authentic word pronunciations through excerpts from YouTube videos (Sardegna & Jarosz, 2023), can be used to expose learners to different realisations of word stress and prepare them for real-world language variation. The newest generation of pronunciation training tools such as the app *ELSA Speak* (Elsa Corp, 2023), aim at Automatic Pronunciation Assessment (APA), yet these solutions are still not reliably accurate in detecting and correcting pronunciation errors and they also have pedagogical weaknesses (Becker & Edalatishams, 2018).

This article aims to provide the first comprehensive review of the state of the art of word stress teaching methods. It addresses the research question: What are effective methods for teaching and learning English word stress with and without technology? It evaluates pronunciation training tools, e.g., *ELSA Speak* (Elsa Corp, 2023), High-Variability Phonetic Training (HVPT) tools, e.g., The *English Accent Coach* (Thomson et al., 2023) and sound articulation tools, e.g., *eNunciate!* (Gick et al., 2015). I evaluate the suitability of these tools to support the learning of word stress perception and production based on the following criteria: pedagogical approaches for dealing with suprasegmentals, the incorporation of language variation, and the provision of meaningful feedback. I then use these insights to propose a concept for a perception and production training tool that focuses on both segmentals and suprasegmentals. I aim to support curriculum developers in incorporating English stress into language curricula, which often have an insufficient focus on pronunciation (McGregor & Reed, 2018). The proposed model aims to support technology designers in developing new tools for language perception and production training.

2. Background

2.1. Teaching word stress

Determining suitable teaching methods depends on the associated teaching and learning goals. Although nativeness may be a goal for many English learners due to their favorable attitudes towards British and American English (e.g., Brabcová & Skarnitzl, 2018), there is considerable research supporting the intelligibility principle, as nativeness is very difficult to obtain for adults. Regarding word stress, Jenkins (2000) notes that intelligibility problems occur only when a stress shift is combined with another phonological error. Acknowledging the complexity of stress assignment rules in English, Jenkins (2000, p. 151) suggests providing general guidelines such as "two-syllable noun – first syllable stress, verb – second syllable stress" and stress-determining suffixes. Szpyra-Kozłowska (2008, p. 168) notes that these proposals may be difficult to apply in practice, especially for learners coming from a fixed-stress language, e.g., Polish. Interestingly, some suprasegmental features such as vowel reduction can even hinder intelligibility (Jenkins, 2000, p. 135) due to the strongly stress-timed rhythm in English. Thus, learners aiming at intelligibility may need a weaker focus on suprasegmentals and a stronger focus on sounds.

Common word stress teaching methods include the learning of simplified rules based on word class and phonological structure (Checklin, 2012), analogies with other words with similar stress patterns, or a focus on predictable patterns such as stress-attracting or determining suffixes (<-ee>, <-ity>) (Setter & Sebina, 2017, p. 148). Field (2005, pp. 419–421) recommends perception exercises, weak quality exercises, presentation by rule and by vocabulary item, analogy exercises, and applying lexical stress to segmentation. Sadat-Tehrani (2017) has successfully taught stress rules related to word class, verb-noun pairs, suffixes, and compound nouns. Kenworthy (1987) suggests activities such as shifting the stress in students' names as a start (e.g., *MAry* vs. *maRIE*). A comprehensive set of stress rules is presented by Dickerson (2015) and applied in studies such as that undertaken by Sardegna and Dickerson (2023). A useful resource for ranking stress placement errors is Ghosh and Levis' hierarchy (2021).

Regarding the focus on rules, there is evidence that explicit learning of stress rules is more beneficial than implicit learning (Sardegna & Dickerson, 2023). However, the declarative knowledge of stress assignment rules may not necessarily lead to accuracy (O'Brien, 2019). Rather, O'Brien (2019) advocates the development of self-assessment skills and a sensitivity to stress rules, even if the students cannot formally describe them. With higher sensitivity to stress regularities, learners will need to actively recall assignment rules less, and instead naturally predict the stress of a new word and apply it correctly.

Moreover, the needs of English learners from different language backgrounds should be considered. The "stress deafness" theory argues that learners have different sensitivity to stress, based on stress exceptions in their L1 (Peperkamp et al., 2010). Thus, some English learners may need more targeted word stress training. For instance, in the context of Polish learners of English, Rojczyk and Porzuczek (2019, p. 19) recommend that teachers explain the use of stress cues, the relative strength of stress, and stress placement. In the Japanese context, Saito and Saito (2017) have shown that English learners can be trained to recognise suprasegmental patterns through instruction, pointing out cross-linguistic differences.

In order to train their response to phonetic and phonological stress cues, English learners can benefit from controlled exposure to language variation. For French learners of English, phonetic training involving the identification of stressed syllables accentuated with different duration increases has been shown to significantly reduce stress insensitivity (Carpenter, 2015). This approach is reminiscent of HVPT.

2.1.3. HVPT and word stress

HVPT is a useful method for providing more exposure to different cues and developing phonological categories that are missing or different in the L1. Learners are exposed to variants of the same sound in different contexts, e.g., produced by various speakers or the same speaker. Research on HVPT mostly focuses on phoneme perception (Cheng et al., 2019) but there are also a few studies that incorporate word stress (Couper, 2021; Uchihara et al., 2022).

There is predominantly positive evidence for HVPT (Cheng et al., 2019; Couper, 2021; Thomson, 2018; Uchihara et al., 2022), however, some studies also show mixed results (cf. Georgiou, 2021; Sadakata & McQueen, 2014). While Thomson and Derwing (2015) find improvements in both the perception and production of sounds, Ingvalson et al. (2013) and Sadakata and McQueen (2014) highlight the influence of individual aptitude on the effectiveness of perceptual training. Nevertheless, an analysis of 27 studies on foreign language perceptual training by Rato and Oliveira (2022) revealed successful transfer and partial retention of perceptual skills. However, most of the examined studies focused on segmentals, therefore, these results should be applied to word stress with caution.

Since HVPT usually introduces input from different speakers, it needs to be administered with computer technology. The following section looks at the design principles and existing solutions for technologies offering training in stress production and perception.

2.2. Technology in word stress teaching

2.2.1. Designing perception and production training technologies

Computer and Mobile-Assisted Language Learning (CALL and MALL) are well-researched frameworks which can inform the technology-assisted learning of word stress, for instance, through Computer-Assisted Pronunciation Training (CAPT). CAPT has some constraints, mostly due to the limitations of computers in relation to adaptability, perceptual accuracy, and feedback (Lee et al., 2015). However, technologies have developed considerably, and some methods such as HVPT are dependent on technology.

The development of pronunciation tools starts from setting learning goals, the language model, and the "golden speaker". This traditionally used to be the teacher or L1 speaker recordings. Now stress production can be demonstrated through diverse authentic language materials like *YouGlish* (Sardegna & Jarosz, 2023). Students can also improve their motivation and performance in producing prominence contrasts by repeating acoustically modified recordings of themselves (Henderson & Skarnitzl, 2022).

After choosing the "golden standard", the tolerance threshold for deviations from this standard needs to be determined. CAPT systems should be able to prioritise errors based on their frequency, salience, and intelligibility (Rogerson-Revell, 2021, p. 197). It is indeed difficult to distinguish errors from natural sociolinguistic variation, yet both teachers and technologies need to make these decisions rapidly and reliably. For automatic stress detection, progress has been made in the classification of stress into primary, secondary, and unstressed syllables (Chen & Wang, 2010; Ferrer et al., 2015; Li et al., 2018), and these technologies can be applied in CAPT in the future.

Overall, educators and language technology developers are confronted with many decisions on best practice in pedagogy and technology. A review of the available solutions is necessary to determine useful established features.

2.2.2. Available solutions

Few e-learning tools address stress production and perception. I tested a word stress lesson on *ELSA Speak* (Elsa Corp, 2023) and found some useful feedback such as "Emphasize the first syllable by making it louder and longer. The other syllable should be softer and quicker" when I mispronounced "happy" as /hæˈpiː/. Here, the instructions successfully emphasise the duration and intensity features of the syllables. However, a

disadvantage of the speaking practice lessons is the focus on only one target feature. For instance, in a lesson on final sounds, I mis-stressed "biggest" as /bɪgˈɛst/. Then, in a word stress lesson, I mispronounced the diphthong in "going" as /ˈgoɪŋg/. These mistakes were not corrected and I received the feedback "Excellent! Your pronunciation is spot on!". While it may be useful to ignore mistakes that do not not specifically pertain to the lesson in order to maintain learner motivation, it may ultimately slow down the learning progress (see Coulange, 2023). Nevertheless, apps such as *ELSA Speak* are a step in the right direction for enabling autonomous language practice.

The *English Accent Coach* (Thomson et al., 2023) is the most popular free online tool for HVPT. Variability is introduced by including different speakers pronouncing single vowels and consonants in isolated environments. However, the tool does not address word stress.

Sound articulation software is useful for practicing production but it can also assist with perception. The *Sounds of Speech* website and app (University of Iowa, 2014) visualise speech sounds with animations and speaker video recordings of their articulation. A similar tool is *eNunciate!* (Gick et al., 2015), which provides ultrasound overlay videos and sound articulation animations. These tools can be combined with input on suprasegmentals and integrated in language e-learning tools.

Overall, combined teaching of stress rules and HVPT powered by new technological advances is likely to support stress acquisition. The following section proposes ways to integrate stress in a HVPT tool alongside traditional segmental perceptual training.
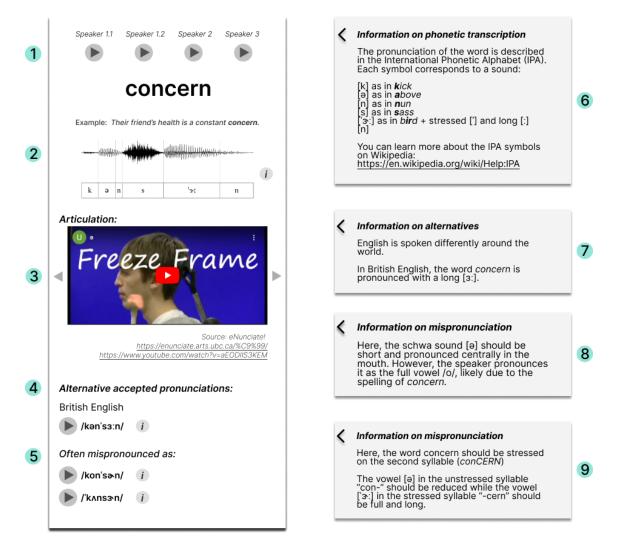
3. HVPT tool concept

The proposed concept is a perception and production training tool which incorporates both segmental and suprasegmental aspects of spoken language. Pronunciation or word stress rules should be learnt additionally. Figure 1 presents a mock-up of the tool.

In the familiarisation section, the learner first listens to pronuciations of the target word "concern" by the same speaker and by two different speakers (1). This method has been similarly applied in HVPT tools such as the *English Accent Coach* (Thomson et al., 2023) where different speakers pronounce the target sounds. However, in the proposed tool, speakers of different varieties of English such as American English, Nigerian English, and Polish English, pronounce the whole word. Presenting different varieties aims to expose the learners to real-life language variation. Then, a sample sentence together with the audio waveform and the IPA transcription of the target word are presented (2). The waveform aims to show the intensity distribution of the sounds, where the learners can observe that /ə/ has lower intensity than /ˈɝː/. In pop-up (6), the learner can consult the meaning of each IPA symbol used in the transcription. Next, videos by *eNunciate!* (Gick et al., 2015) present the articulation of the word's sounds through which the reader can navigate (3). In section (4), alternative accepted pronunciations are presented based on the Cambridge and Oxford English Dictionaries. Pop-up (7) explains the alternative British English pronunciation where the /ɝː/ is not rhotacised. Section (5) presents frequent mispronunciations such as the full vowel /o/ instead of /ə/ in /konˈsɝːn/ and the shifted stress in /ˈkʌnsɝn/, which are accordingly explained in (8) and (9).

Figure 1

Mock-up of a HVPT app incorporating (1) pronunciations of the same and different speakers, (2) a waveform with IPA transcription, (3) an articulation video from eNunciate! (Gick et al., 2015), (4) alternative pronunciations, and (5) frequent mispronunciations. Additional information is provided through the pop-ups (6–9).
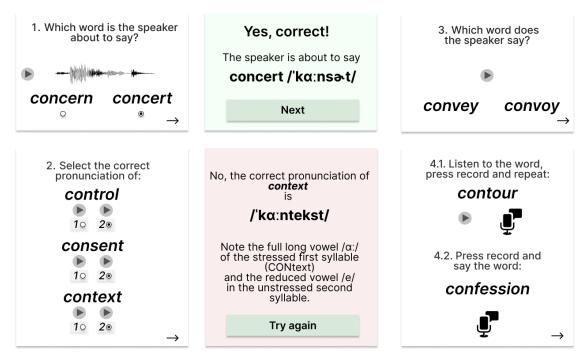

Next, in the practice section[1], the learner practices perception and production through tasks associated with the targeted words (Figure 2).

---

[1] Waveform created in Praat (Boersma & Weenink, 2023). Microphone image by M7MD k7hald, https://commons.wikimedia.org/w/index.php?curid=100268520

Figure 2

Examples of tasks and feedback within the HVPT training programme.

In Task 1, the learner listens to an audio where the last sound has been cut and the segment "concer-″ either as in "concern″ /kənˈsɜː/ or as in "concert" /ˈkɑːnsə/ is presented. This task is common in psycholinguistic experiments such as those undertaken by Cooper et al. (2002). It is suitable for language learning scenarios since the learner has to use stress cues to identify which word has been said. In Task 2, the learners listen to pronunciations of the target words "control″, "consent″, and "context″ with correct and shifted stress (/kənˈtroʊl/–/ˈkontroʊl/*, /kənˈsent/–/ˈkonsent/*, /ˈkɑːntekst/–/konˈtekst/*) and select the correct pronunciation. The green and red windows in the middle of Figure 2 present automatic feedback in the case of a correct answer to the first task and an incorrect answer to the second task. These are common colours associated with correct and incorrect responses in language learning tools such as *ELSA Speak.* A feedback improvement proposed by the HVPT tool is the embedding of the particular sounds that were perceived wrongly (here /ɑː/ and /e/). In Task 3, the learner has to distinguish the vowel and stress differences between "convey″ /kənˈveɪ/ and "convoy" /ˈkɑːnvɔɪ/. In Task 4, the learner can record their own pronunciation of the word "contour" after listening to it (4.1) and the word "confession" without assistive input (4.2). Here, the two words again, purposefully, have different stress but the same initial syllable.

4. Discussion

Informed by previous research on phonetic training, the HVPT tool includes both single speaker and multi-speaker variation by presenting the target word 1) in different intonational contexts by the same speaker and 2) by speakers of different varieties of English.

Since multimodal input is beneficial for pronunciation learning (Rogerson-Revell, 2021), the tool incorporates multiple visualisations: a waveform as a visual intensity cue, sound boundaries representing duration, IPA symbols, and capitals corresponding to stress in the information window (9). For the segmentals, the tool embeds the ultrasound overlay videos from *eNunciate!* (Gick et al., 2015) and thus allows learners to practice articulating the individual sounds of the word.

The tasks are informed by the word stress teaching suggestions described by Field (2005), such as training perception and weak forms. They are also based on common HVPT exercises like same–different discrimination (Qian et al., 2018) and knowledge of correct stress placement. Tasks 4.1 and 4.2 involve recording the learner's own production which is reviewed by the teacher. Automatic pronunciation assessment (APA) would have been more efficient, however, the current state of the art does not guarantee reliable results and APA is still met with a lack of trust by students and educators despite its technological advances (Thi-Nhu Ngo et al., 2023).

When evaluating pronunciation, both human teachers and technological solutions eventually encounter issues of language variation, as it is difficult to determine the border between natural variation and pronunciation error. It is still a long-term goal to train tools to tolerate the minor, persistent features of the user's language variety that do not hinder intelligibility. Korzekwa et al. (2021) have taken a step in this direction and developed a model for pronunciation error detection that evaluates the severity of the errors. Still, the wider application of such solutions is yet to come.

One benefit of technology is the customisation of the tasks to the individual needs of the learner based on their language background. A needs analysis can help to organise the HVPT as a whole and insights from second language acquisition can inform the individual components. A contrastive approach, as in Saito and Saito (2017), could be helpful to show the differences and similarities between the L1 and the target language.

Moreover, language training requires time and motivation. Longer treatments produce greater effects (Lee et al., 2015), so the learners need to be engaged in multiple sessions. In terms of generating motivation, the proposed mock-up is minimalistic compared to apps like *ELSA Speak* or *Duolingo* which employ gamification. Rewards like points for correct answers can be incorporated into the HVPT tool as well. Gamification has been shown to be an efficient technique for improving user engagement and motivation but there is mixed evidence regarding its impact on students' affective states and learning outcomes (Boudadi & Gutiérrez-Colón, 2020). While the learning goal, i.e., pronunciation improvement, should not be compromised in the service of gamification, an appealing modern design with a reward system is expected to benefit learner motivation.

Finally, the tool is directed at intermediate and advanced learners and is not suitable for beginners or learners with special educational needs. Since the tool relies on detecting fine acoustic differences through listening, it has limited accessibility for people with hearing impediments and future efforts for improving its accessibility are necessary.

5. Conclusion

Word stress teaching and learning can be supported by technologies that facilitate training in perception and production. Simplified stress placement rules, together with exposure to authentic input, can help the acquisition of word stress placement regularities. HVPT has been shown to have largely beneficial effects on vowel and consonant acquisition but its effect on stress still needs to be explored. I have proposed a concept for a HVPT-based tool for training the perception and production of English as a foreign language incorporating both segmentals and suprasegmentals. The concept relies on exposing learners to single speaker and multi-speaker variation and providing them with multimodal information on pronunciation and articulation. The concept also proposes sample tasks that test the distinction and pronunciation of English words where learners need to rely on both segmental and suprasegmental cues.

The tool can thus be further developed and integrated into existing curricula following frameworks such as that proposed by McGregor & Reed (2018). Moreover, the tool can be adapted to the needs of language learners from different backgrounds: e.g., considering that Slavic and German learners of English perceive stress position shifts differently (Ivanova et al., 2023), Slavic learners of English can receive HVPT with more emphasis on word stress, vowel reduction, and stress-timed rhythm, whereas German learners of English can focus more on HVPT for vowel quality or consonant pronunciation. Practice with the HVPT tool complements other learner-centred approaches such as active

learning. The tool thereby fits into different educational contexts like language learning at secondary and tertiary level.

While I have discussed many proposals for word stress teaching and learning based on previous research, developing the HVPT tool and testing its efficiency was not feasible within the scope of this study. Nevertheless, future studies can use the proposed concept and mock-up to create the tool and test the experience of language learners and teachers working with it.

Overall, this article has outlined common practices and considerations in English word stress teaching with a focus on the use of technology. It has shown how determining goals and models in language learning influences nearly every decision in the creation of pronunciation learning materials. Moreover, intelligibility is seen as a valid model and variation as a natural part of every sociolinguistic environment, therefore language variation should have a place in teaching and learning English pronunciation around the world. Finally, exposure to this variation has promising potential for the training of perceptual sensitivity and production intelligibility.

Ethical statement

The author declares no competing interests.

References

Becker, K., & Edalatishams, I. (2018). 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* (pp. 434–438). Iowa State University.

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer*. http://www.praat.org/

Boudadi, N. A., & Gutiérrez-Colón, M. (2020). Effect of Gamification on students' motivation and learning achievement in Second Language Acquisition within higher education: A literature review 2011-2019. *EUROCALL Review*, *28*(1). https://doi.org/10.4995/eurocall.2020.12974

Brabcová, K., & Skarnitzl, R. (2018). Foreign or native-like? The attitudes of Czech EFL learners towards accents of English and their use as pronunciation models. *Studie z Aplikované Lingvistiky*, *1*, 38–50.

Carpenter, A. D. (2015). Phonetic training significantly mitigates the stress 'deafness' of French speakers. *International Journal of Linguistics*, *7*(3), 94–108.

Checklin, M. (2012). What in the world do we know about word stress? A review of what it is and how to teach it. *TESOL in Context*, *S3*, 1–13.

Chen, J.-Y., & Wang, L. (2010). Automatic lexical stress detection for Chinese learners' of English. In *7th International Symposium on Chinese Spoken Language Processing* (pp. 407–411). IEEE. https://doi.org/10.1109/ISCSLP.2010.5684859

Cheng, B., Zhang, X., Fan, S., & Zhang, Y. (2019). The role of temporal acoustic exaggeration in High Variability Phonetic Training: A behavioral and ERP study. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01178

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, *45*(3), 207–228.

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). https://doi.org/10.5281/ZENODO.8137754

Couper, G. (2021). Teaching and testing perception of word stress: Many shades of perception. In J. Levis & A. Guskaroska (Eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*. Iowa State University.

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

Cutler, A., & Clifton, C. E. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 183–196). Lawrence Erlbaum.

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

Elsa Corp. (2023). *ELSA Speak*. https://apps.apple.com/us/app/elsa-english-language-app/id1083804886

Ferrer, L., Bratt, H., Richey, C., Franco, H., Abrash, V., & Precoda, K. (2015). Classification of lexical stress using spectral and prosodic features for computer-assisted language learning systems. *Speech Communication*, *69*, 31–45. https://doi.org/10.1016/j.specom.2015.02.002

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*(3), 399–423.

Georgiou, G. P. (2021). Effects of phonetic training on the discrimination of second language sounds by learners with naturalistic access to the second language. *Journal of Psycholinguistic Research*, *50*(3), 707–721. https://doi.org/10.1007/s10936-021-09774-3

Ghosh, M., & Levis, J. M. (2021). Vowel quality and direction of stress shift in a predictive model explaining the varying impact of misplaced word stress: Evidence from English. *Frontiers in Communication*, 6. https://doi.org/10.3389/fcomm.2021.628780

Gick, B., Hall, K., Rullmann, H., & Wiltschko, M. (2015). *eNunciate! A visual language learning tool*. https://enunciate.arts.ubc.ca/

Henderson, A. J., & Skarnitzl, R. (2022). "A better me": Using acoustically modified learner voices as models. *Language Learning and Technology*, *26*(1), 1–21.

Ingvalson, E. M., Barr, A. M., & Wong, P. C. M. (2013). Poorer phonetic perceivers show greater benefit in phonetic-phonological speech learning. *Journal of Speech, Language, and Hearing Research*, *56*(3), 1045–1050. https://doi.org/10.1044/1092-4388(2012/12-0024)

Ivanova, M., Neubert, C. R., Schmied, J., & Bendixen, A. (2023). ERP evidence for Slavic and German word stress cue sensitivity in English. *Frontiers in Psychology 14*. https://doi.org/10.3389/fpsyg.2023.1193822

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

Kenworthy, J. (1987). *Teaching English pronunciation*. Longman.

Korzekwa, D., Lorenzo-Trueba, J., Drugman, T., Calamaro, S., & Kostek, B. (2021). Weakly-supervised word-level pronunciation error detection in non-native English

speech. In *Interspeech 2021* (pp. 4408–4412). ISCA. https://doi.org/10.21437/Interspeech.2021-38

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

Li, K., Mao, S., Li, X., Wu, Z., & Meng, H. (2018). Automatic lexical stress and pitch accent detection for L2 English speech using multi-distribution deep neural networks. *Speech Communication*, *96*, 28–36. https://doi.org/10.1016/j.specom.2017.11.003

McGregor, A., & Reed, M. (2018). Integrating pronunciation into the English language curriculum: A framework for teachers. *The CATESOL Journal*, *30*(1), 69–94.

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

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, *38*(3), 422–430. https://doi.org/10.1016/j.wocn.2010.04.001

Qian, M., Chukarev-Hudilainen, E., & Levis, J. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *Language Learning and Technology*, *22*(1), 69–96.

Rato, A., & Oliveira, D. (2022). Assessing the robustness of L2 perceptual training: A closer look at generalization and retention of learning. In U. Kickhöfel Alves & J. I. Alcantara de Albuquerque (Eds.), *Second language pronunciation* (pp. 369–396). de Gruyter. https://doi.org/10.1515/9783110736120-014

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

Rojczyk, A., & Porzuczek, A. (2019). EFL pronunciation teaching to Polish English studies majors. In D. Gabryś-Barker (Ed.), *Challenges of foreign language instruction in the university context* (pp. 9–25). Wydawnictwo Uniwersytetu Śląskiego.

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01318

Sadat-Tehrani, N. (2017). Teaching English stress: A case study. *TESOL Journal*, *8*(4), 943–968. https://doi.org/10.1002/tesj.332

Saito, Y., & Saito, K. (2017). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, *21*(5), 589–608. https://doi.org/10.1177/1362168816643111

Sardegna, V. G., & Dickerson, W. (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. 91–107). Multilingual Matters.

Sardegna, V. G., & Jarosz, A. (2023). Learning English word stress with technology. In R. I. Thomson, T. M. Derwing, J. M. Levis, & K. Hiebert (Eds.), *Proceedings of the 13th Pronunciation in Second Language Learning and Teaching Conference.* Iowa State University.

Setter, J., & Sebina, B. (2017). English lexical stress, prominence and rhythm. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds*.), The Routledge handbook of contemporary English pronunciation* (pp. 137–153). Routledge. https://doi.org/10.4324/9781315145006-9

Szpyra-Kozłowska, J. (2008). Lingua franca core, phonetic universals and the Polish context. In K. Dziubalska-Kołaczyk & J. Przedlacka (Eds.), *English pronunciation models* (pp. 151–176). Peter Lang.

Thi-Nhu Ngo, T., Hao-Jan Chen, H., & Kuo-Wei Lai, K. (2023). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 1–18. https://doi.org/10.1017/S0958344023000113

Thomson, R. I. (2018). High Variability [Pronunciation] Training (HVPT). *Journal of Second Language Pronunciation*, *4*(2), 208–231. https://doi.org/10.1075/jslp.17038.tho

Thomson, R. I., & Derwing, T. M. (2015). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le, E. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference* (pp. 88–97). Iowa State University.

Thomson, R. I., Derwing, T. M., & Munro, M. J. (2023). *English Accent Coach*. https://www.englishaccentcoach.com/home

Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2022). The effects of talker variability and frequency of exposure on the acquisition of spoken word knowledge*. Studies in Second Language Acquisition*, *44*(2), 357–380. https://doi.org/10.1017/S0272263121000218

University of Iowa. (2014). *Sounds of Speech*. https://soundsofspeech.uiowa.edu/german

van der Hulst, H. (2010). Word accent systems in the languages of Europe. In H. van der Hulst, R. Goedemans, & E. van Zanten (Eds.), *A survey of word accentual patterns in the languages of the world* (pp. 429–507). De Gruyter.