Online Phonetic Training Improves L2 Word Recognition

Gerda Ana Melnik (gerda.ana.melnik@ens.fr)

Sharon Peperkamp (sharon.peperkamp@ens.fr)

Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS),
Département d'Etudes Cognitives, Ecole normale supérieure - PSL University,
29 rue d'Ulm, 75005 Paris, France

Abstract

High-Variability Phonetic Training (HVPT) has been shown to be effective in improving the perception of even the hardest second-language (L2) contrasts. However, little is known as to whether such training can improve phonological processing at the lexical level. The present study tested whether this type of training also improves word recognition. Adult proficient French late learners of English completed eight online sessions of HVPT on the perception of English word-initial /h/. This sound does not exist in French and has been shown to be difficult to process by French listeners both on the prelexical (Mah, Goad & Steinhauer, 2016) and the lexical level (Melnik & Peperkamp, 2019). In pretest and posttest participants completed an identification task as well as a lexical decision task. The results demonstrated that after training the learners' accuracy had improved in both tasks. The theoretical and applied implications are discussed.

Keywords: second language acquisition; lexical processing; word recognition; speech perception; phonetic training

Introduction

It is well known that producing and perceiving non-native speech sounds can be very challenging (for reviews, see Piske, MacKay & Flege, 2001; Sebastián-Gallés, 2005). In the realm of perception, much research has shown that with auditory training, the difficulty of perceiving even the hardest non-native sounds can be reduced. The most common training paradigm used to improve second language (L2) perception is High-Variability Phonetic Training (HVPT). HVPT uses multiple natural exemplars of the target sounds in a variety of phonetic environments. This variability enhances the process of building novel phonological categories. Importantly, perceptual training involves immediate corrective feedback that provides information to participants about their performance and promotes rapid learning by driving the learner's attention to the relevant phonetic cues of the sounds to be learned (Homa & Cultice, 1984; Logan, Lively & Pisoni, 1991). The effectiveness of this technique has been shown in many studies in a variety of languages, using several target contrasts and structures, including vowels (Carlet & Cebrian, 2014; Lee & Lyster, 2016), consonants (Kim & Hazan, 2010; Shinohara & Iverson, 2018), tones (Wang et al. 1999; Wang, Jongman, & Sereno, 2003), and syllable structure (Huensch & Tremblay, 2015). Moreover, both high- and low-proficiency speakers benefit from HVPT (Iverson, Pinet & Evans, 2012), and HVPT generalizes to new tokens and new speakers (Lively et al., 1994; Okuno & Hardison, 2016). Finally, it gives rise to long-term retention of the new categories (Lively et al., 1994), and it helps to improve L2 production (for a review, see Sakai & Moorman, 2018).

Although the effectiveness of HVPT is well studied, most previous work focused exclusively on prelexical perception, using identification or discrimination tasks. The difficulty with the perception of L2 sounds, though, is paralleled by less efficient lexical processing (e.g., Pallier, Colomé & Sebastián-Gallés, 2001; Weber & Cutler, 2004). Thus, truly successful training should also enhance performance at the lexical level. While prelexical processing only involves a phonetic analysis, lexical processing is more complex as it additionally requires mapping the incoming speech signal onto phonological representations stored in memory, and the performance gap between native and non-native listeners in L2 speech perception increases as the tasks have greater lexical involvement (Díaz et al., 2012).

So far, the only studies on the effect of prelexical auditory training on lexical processing focused on naïve listeners' ability to learn words in a tonal language (Cooper & Wang, 2011; Ingvalson, Barr & Wong, 2013). Both studies found that naïve English listeners' ability to learn words involving difficult tone contrasts improved after auditory training. To our knowledge, no studies have directly assessed the effect of auditory training on enhancing word recognition in L2 learners.

We focused on the perception of the English sound /h/ by intermediate French learners of English. As /h/ does not exist in French, French listeners – even those who are fluent in English – have difficulty perceiving the contrast between the presence vs. absence of /h/ in English stimuli (Mah et al., 2016). At the lexical level, proficient French learners of English tend to accept nonwords such as *usband* (cf. *husband*) and, to a lesser extent, *hofficer* (cf. *officer*), as real words (Melnik & Peperkamp, 2019). Thus, they have difficulty not only in perceiving the contrast between /h/ and silence, but also in distinguishing between words and nonwords that differ only in the presence vs. absence of /h/.

Importantly, there is an almost perfect one-to-one mapping in English of the grapheme <h> onto the phoneme /h/. Most French L2 speakers know how to correctly write /h/-initial words. They are also instructed that <h> is rarely silent in English and that it is pronounced as /h/. If after training learners start better perceiving /h/, they might thus be able to also improve their recognition of /h/-initial English words even if they have imprecise phonological representations of such words, since they can rely on the orthography.

In the current study we trained French learners on the perception of English /h/ in a pretest–training–posttest design. In pretest and posttest, participants performed an identification task aimed at testing their phonetic perception of /h/, and a lexical decision task aimed at testing their processing of /h/ at the lexical level. In the posttest, the identification task also tested for generalization to novel items. In the identification task we used /h/- and vowel-initial nonwords as stimuli. In the lexical decision task we used words and nonwords, where the test nonwords were created from /h/-initial and vowel-initial words by removing or adding /h/, respectively.

Training was administered on-line, and consisted of eight sessions of an identification task using minimal pairs of real words (such as *air-hair*), with corrective feedback.[1] We expected the training to enhance performance in the identification task at posttest, thus replicating the findings of previous studies on the effectiveness of HVPT in improving phonetic perception of L2 sounds. Moreover, if the effect of training extends to lexical processing, performance in lexical decision should likewise improve with training.

Method

Pretest-Posttest-Generalization: Identification

Stimuli

For the pre- and posttest we selected 100 pairs of nonwords. The members of each pair differed in the presence or absence of an initial /h/ (e.g. /hasp/ – /asp/). Forty pairs were monosyllabic, 40 dissyllabic and 20 trisyllabic. Ten English vowels (ʌ, ɒ, a, ɪ, ɛ, iː, ʌɪ, əʊ, eɪ, aʊ) were used in the first (or only) syllable, thus creating a large amount of variability in phonetic context.

An additional 30 pairs of nonwords (10 monosyllabic, 10 disyllabic and 10 trisyllabic, containing the 10 vowels mentioned above) were selected to test for generalization at the end of the posttest. Half of the pairs were recorded by a male, and the other half by a female native of American English.

Procedure

Participants were tested individually in a soundproof booth. In each trial they were presented auditorily with a stimulus; their task was to press as quickly as possible the key labelled "h" with their dominant hand if they thought the nonword started with the sound /h/, and to press the key labelled "no h" with their non-dominant hand if they thought it did not start with /h/. There were 194 trials divided over two blocks. Trials were presented in a semi-random order such that no more than four trials of the same type (vowel-initial or /h/-initial) and no more than three trials recorded by the same speaker appeared in a row.

The first block started with a practice phase of six trials, during which participants received feedback. In the case of an incorrect response or no response within 2500 ms, the trial was repeated until the correct response was given. During the test phase, participants received no feedback and if they did not give a response within 2500 ms the next trial was presented. An interval of 1000 ms elapsed between the participant's response or the time-out - whichever came first - and the presentation of the next trial.

At the end of the posttest only, 60 trials with the 30 additional nonword pairs were used to test for generalization.

Pretest-Posttest: Lexical decision

Stimuli

The stimuli were the same as in Melnik & Peperkamp (2019). They consisted of 80 English test words, 40 starting with /h/ (e.g., *husband*) and 40 with a vowel (e.g. *officer*), recorded by the same male American English speaker who recorded stimuli for the identification task. They consisted of nouns, verbs and adjectives, and contained between two and four syllables. The /h/-initial and the vowel-initial words did not differ in mean frequency in the Subtlex database (Brysbaert & New, 2009) or in mean number of syllables (both $t < 1$).[2]

Each word was paired with a nonword, created by deleting or adding /h/ at the beginning (e.g. *husband* → *usband*, *officer* → *hofficer*). In addition, there were 240 English control words (nouns, verbs and adjectives), none of which starting with /h/. They were matched for mean frequency and mean number of syllables with the test words. Each control word was paired with a nonword created by replacing, deleting or inserting one phoneme other than /h/.

The test and control minimal pairs were divided into two equal groups, one for pretest and one for posttest, respecting the matching in terms of frequency and number of syllables. The pretest stimuli were further divided into two counterbalancing lists: list A and list B. Each of them contained only one member of each pretest minimal pair. For instance, if the word *husband* was in list A, its nonword counterpart *usband* was in list B. The posttest stimuli were divided into lists C and D following the same principle. Thus,

[1] Training can be done either with nonwords (e.g., Yamada, 1991) or with real words (e.g., Logan et al., 1991). Here, we chose to use real words because repeated exposure to a large number of nonwords during training might have induced a bias to excessively accepting nonwords in the lexical decision task in pre- and posttest.

[2] The familiarity of these words was evaluated by a separate group of 45 adult French learners of English in an online rating questionnaire. The /h/- and vowel-initial words that were chosen for the experiment did not differ in mean familiarity ($t = 1.0$, $p > 0.1$).

no list contained both members of a given word–nonword pair. Each of the four lists contained 10 /h/-initial and 10 vowel-initial words, 10 /h/-initial and 10 vowel-initial nonwords, as well as 60 control words and 60 control nonwords. Finally, for a practice phase there were two additional words and two additional nonwords, none involving /h/.

Procedure

In pretest half of the participants were randomly assigned to one of the two pretest lists (list A or list B). In posttest, participants who previously heard the list A were given the list C, while participants who previously heard the list B, were now given the list D. Hence, participants heard only one of the members of each word-nonword pair throughout the whole experiment.

The procedure was identical to that in Melnik & Peperkamp (2019): Participants performed a speeded auditory lexical decision task. In each trial they heard a word or a nonword and had to answer if the item was an English word. They were instructed to use their dominant hand for "yes"- and their non-dominant hand for "no"-responses on a button box. There were 160 trials divided over two blocks, each containing the same number of test and control stimuli. Trials were presented in a semi-random order such that between one to three control trials appeared between two experimental ones, and that no more than four trials of the same type (word or nonword) appeared in a row.

The first block started with a practice phase of four trials with control items, during which participants received feedback ('correct' or 'wrong' written on the screen). In the case of an incorrect response or no response within 2500 ms, the trial was repeated until the correct response was given. During the test phase, participants received no feedback and if they did not give a response within 2500 ms the next trial was presented. An interval of 1000 ms elapsed between the participant's response or the time-out and the presentation of the next trial.

Training: Identification

Stimuli

We selected 59 minimal pairs of real words differing in the presence or absence of an initial /h/. Given the limited number of such minimal pairs, we used both frequent words (e.g. *hair-air*) and infrequent ones (e.g. *hosier-osier*) words. However, word frequency was not considered to have an impact, as the task used in training was prelexical.

Four different speakers, two men and two women, recorded the items. One of the male speakers and one of the female speakers were those who recorded the stimuli for the nonword identification task used in pretest and posttest, with the male speaker having also recorded the stimuli for the lexical decision task.

Procedure

The training consisted of eight high-variability phonetic training sessions. In the first four sessions participants heard one speaker per session. In the following four sessions they heard a pair of speakers in each session, such that all four male-female combinations were used.

All training sessions were run at the participants' homes through internet. The online training sessions were designed using the JsPsych library (de Leeuw, 2015) in JavaScript. Before each training session participants received by email a link to the corresponding training session webpage. Stimuli were presented at a comfortable listening level, set individually. The details of each training session (e.g., participant details, day and time of completion, RTs and responses) were automatically sent to the MySql database after the completion of each session. Participants could only do one session per day and there could be no more than one day in between two sessions. Thus, the whole course of training was completed in eight to fifteen days.

In each trial participants first saw the two response alternatives written on the screen (e.g. "hair – air"). The word starting with /h/ was always displayed on the left, and the word without /h/ always on the right. The auditory stimulus was played 800 ms later. The task was to press as quickly as possible the left arrow key if the word started with /h/ and the right arrow key otherwise. When the participant pressed the key, the corresponding word was highlighted in bold. If the response was correct, the word "Correct" written in green appeared in the middle of the screen, in between the two alternatives. If it was incorrect, the word "Wrong" written in red appeared on the screen, followed after 1000 ms by auditory feedback of the form: "*The word was not: XXX. It was: YYY*", spoken by the same speaker as the stimulus itself. For instance, if the stimulus played was the word "hair" but the participant chose instead the word "air", the word "Wrong" was displayed on the screen and the phrase *"The word was not: air. It was: hair"* was played.

If no response was given within 2500 ms, the words "Too slow" appeared on the screen. An interval of 1000 ms elapsed between the participant's response or the time-out - whichever came first - and the presentation of the next trial. There were 118 trials in each session, and trials were presented in a random order. Each session lasted from 15 to 20 min, depending on the accuracy of the participant.

Participants

Participants were French intermediate learners of English, recruited from among university students (about half of which in an English department). In order to avoid ceiling performance or insufficient knowledge of English vocabulary, only participants whose accuracy in pretest was below 80% in the identification task and above 70% on control items in the lexical decision task went through the training and posttest. Of the 51 participants who did the pretest, 25 satisfied these criteria, out of whom a total of 24 completed the study and were included in the data analysis. Among these participants, there were 12 women and 12 men, aged between 19 and 32 (mean: 22.3), who had started learning English at school. They filled in a questionnaire to self-evaluate their speaking, listening, reading, vocabulary and grammar skills in English and French, on a scale from 1 to 10. The overall mean score was 6.4 (SD = 1.6) for English and 9.4 (SD = 0.9) for French.

None of the participants reported a history of speech or language problems. They received a small payment after the pretest, and those who underwent training received a second, larger, payment when they came back to the laboratory for the posttest.

Results

Pretest-Posttest-Generalization: Identification

Prior to analysis, we discarded responses with a reaction time of 0 ms. Figure 1 displays the identification accuracy of participants in pretest, posttest and generalization. As the identification task is a signal detection task, we used the A' statistic, which provides a non-parametric, unbiased, index of sensitivity (here: to the difference between words and nonwords), with 0.5 indicating chance performance and 1.0 perfect performance. A repeated measures ANOVA by participant with the factor Session (Pretest vs. Posttest vs. Generalization), revealed a main effect of Session ($F(2,46) = 26.75$, $p < .001$), with the accuracy improving from an average A′ score of 0.74 in pretest to 0.86 in posttest and 0.86 in generalization. Bonferroni-adjusted pairwise t-tests revealed that there was a significant difference between pretest and posttest ($p < .01$), as well as between pretest and generalization ($p < .01$). There was no difference between the performance in the posttest and in the generalization ($p = .82$).

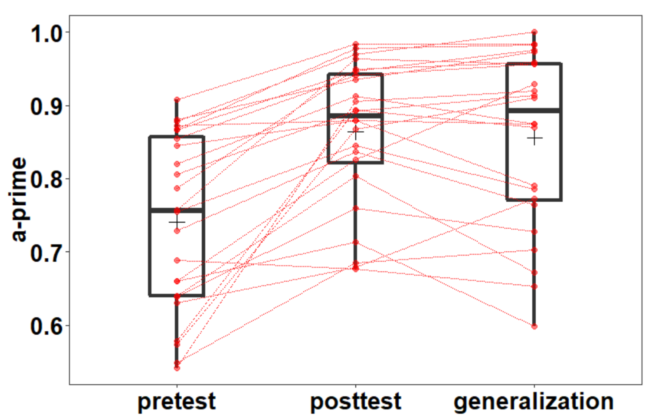

Figure 1. Boxplots of A′ scores in the identification task in pretest, posttest, and generalization. The red dots represent individual participants; the lines link each participant's performance in the three sessions. The black cross marks indicate mean A′ scores in each session.

Pretest-Posttest: Lexical Decision

Prior to analysis, we discarded responses with 0 ms reaction time. Figure 2 displays the accuracy of participants on the test items in pretest and posttest. As the participants had a strong bias for 'yes'-responses (shown by their low accuracy scores on test nonwords), we used the A' statistic as in the analysis of performance in the identification task.

We carried out a repeated measures ANOVA by participant with the factors Session (pretest vs. posttest), Condition (test vs. control) and Lists (AC vs. BD), as well as an interaction between Session and Condition. We found main effects of Session ($F(1, 23) = 39.36$, $p < .001$) and Condition ($F(1, 23) = 73.93$, $p < .001$), and a Session X Condition interaction ($F(1, 23) = 30.87$, $p < .001$). Pairwise t-tests revealed that the interaction was due to the fact that in control items, the effect of Session was not significant, while in test items, there was a significant difference between pretest and posttest ($p < .001$), with the accuracy improving from an average A′ score of 0.62 in pretest to 0.82 in posttest. There was no effect of the counterbalancing factor Lists.

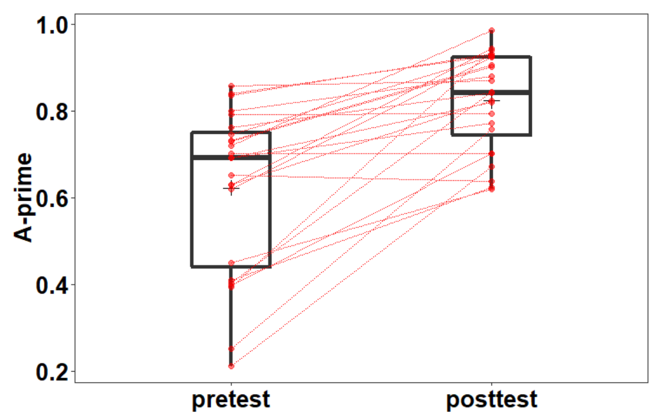

Figure 2. Boxplots of A′ scores in the lexical decision task in pretest and posttest. The red dots represent individual participants; the lines link each participant's performance in both sessions. The black cross marks indicate mean A′ scores in each session.

Discussion

The present study examined if phonetic training can enhance the recognition of words that contain a difficult non-native sound. We tested French learners with intermediate proficiency in English on both their prelexical perception and their lexical processing of stimuli containing /h/. This sound does not exist in French, and French listeners tend to confuse it with silence (Mah et al., 2016). The participants underwent eight sessions of High-Variability Phonetic training, and were tested in pretest and posttest by means of an identification and a lexical decision task.

We found that participants improved in both tasks in posttest compared to pretest. For the identification task, we also observed generalization to new items. The results for this task are in accordance with results from previous studies that used HVPT. Concerning the lexical decision task, this is the first piece of evidence that HVPT can improve not only prelexical but also lexical processing. As mentioned in the introduction, successful word recognition depends on the correct decoding of the speech signal and the matching of this percept to the phonological representation stored in long-term memory (Pisoni & Luce, 1987). If listeners have difficulty with at least one of those aspects, then word recognition might be less effective. Evidence that this is the case is shown by the fact that in the lexical decision task during pretest, the test items involving the difficult sound /h/ yielded higher error rates than the control items. Note that performance on control items was very good in both pre- and posttest (mean A′ score 0.94). As the test and control items were matched in frequency, this indicates that the difficulty participants encountered with the test items was caused by the presence of /h/ and not by a lack of English vocabulary. Importantly, this difficulty was clearly reduced after training, as in posttest participants made less errors on the test items with /h/ than in pretest, while their performance did not change on control items.

Our findings have both theoretical and practical implications. From a theoretical point of view, they shed light on the relationship between prelexical and lexical processing in L2 learning. It is generally agreed upon that speech processing involves several stages, ranging from auditory processing, phonetic and phonological analysis, to word recognition and lexical access (Pisoni & Luce, 1987). In a study on Dutch L2 learners' processing of the English /æ/-/ɛ/ contrast, Díaz et al. (2012), found that the performance gap between native and non-native listeners increases as the tasks have greater lexical involvement. This is likely due to the fact that different perceptual tasks tap into different processing levels, thus requiring different skills and involving different amounts of cognitive load. Our finding that improvement in prelexical perception is paralleled by an improvement in lexical processing suggests a bottom-up sequential order in learning. Although at a specific time point in learning the proficiency in prelexical perception might be ahead of that in lexical processing, a rapid improvement in the former might give rise to change in the latter. This is in accordance with the Automatic Selective Perception model (Strange, 2011), which proposes that L2 phonological processing is less automatic and therefore requires more attentional resources than phonological processing in L1. Consequently, while the performance of learners might be good on relatively simple prelexical tasks, where they can exclusively focus their attention on crucial phonetic cues, the same performance level might not be obtained in tasks requiring the processing of more complex stimuli and attention to other information, such as word meaning. According to this model, the processing of simple tasks becomes more automatic and nativelike as proficiency grows. Thus, in our study, training possibly rendered the prelexical processing more efficient, thus allowing participants to allocate more cognitive resources to the lexical level of processing.

A similar finding on the benefit of phonetic training for higher processing levels was reported in a study on the perception of L2 speech in noise (Lengeris & Hazan, 2010). Adverse listening conditions such as a high signal-to-noise ratios (SNRs) have been shown to involve increased cognitive load and to have greater negative effects for speech perception in non-native than in native listeners (for a review, see Lecumberri et al., 2010). In this study, it was shown that HVPT in quiet improves the perception of a difficult L2 sound in noise.

On the practical side, the current findings could have implications for language teaching. The above-mentioned aspects of speech processing – lexical perception and perception of speech in noise – are inherent elements of "real life" language processing. The fact that they can be improved by relatively short HVPT is encouraging. Moreover, our training was administered online and not in a well-controlled laboratory setting; it can thus easily complement traditional language teaching methodologies. Finally, we note that participants of our study reported that being trained on real words was very motivating, as they had the occasion not only to enhance their perception but to learn new words as well.

To conclude, we showed that even short online HVPT can improve both prelexical and lexical processing of a difficult L2 sound. Future research should test if these improvements are retained in the long term. Furthermore, although we observed significant improvements, only some participants were at ceiling in posttest. Thus, further studies should look at the effect of training length on learning outcomes. This would help us understand if there is an upper limit of improvement in lexical processing that training can induce.

Acknowledgments

This research was supported by a Ph.D. fellowship from Ecole normale supérieure to G. A. Melnik and by grants from the Agence Nationale de la Recherche (Grant Nos. ANR-17-EURE-0017 and ANR-17-CE28-0007-01).

References

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977-90.

Carlet, A. & Cebrian, J. (2014). Training Catalan speakers to identify L2 consonants and vowels: A short-term high variability training study. *Proceedings of the 7th International Symposium on the Acquisition of Second Language Speech* (pp. 85-98). Concordia University Working Papers in Applied Linguistics.

Cooper, A., & Wang, Y. (2011). The influence of tonal awareness and musical experience on tone word learning. *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 512-515). Hong Kong, SAR, China.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1-12.

Díaz, B., Mitterer, H., Broersma, M., Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, *22*(6), 680-689.

Homa, D., & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology*, *10*, 83-94.

Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics*, *52*, 105–120.

Ingvalson, E. M., Barr, A. M., & Wong, P. C. (2013). Poorer Phonetic Perceivers Show Greater Benefit in Phonetic-Phonological Speech Learning. *Journal of Speech Language and Hearing Research, 56*(3), 1045.

Iverson, P., Pinet, M., Evans. B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(1), 145-160.

Kim, Y. H. & Hazan, V. (2010). Individual variability in the perceptual learning of L2 speech sounds and its cognitive correlates. *Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech* (pp. 251-256). Poznań, Poland.

Lecumberri, M. L. G., Cooke, M., Cutler, A. (2010). Non-native speech perception in adverse conditions: a review. *Speech Communication*, *52*(11-12), 864-886.

Lee, A. H. & Lyster, R. (2016). Effects of Different Types of Corrective Feedback on Receptive Skills in a Second Language: A Speech Perception Training Study. *Language Learning*, *66*(4), 809-833.

Lengeris, A. & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *Journal of the Acoustical Society of America*, *128*(6), 3757-3768.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tokhura, Y., Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96*(4), 2076-2087.

Logan, J.S., Lively, S.E., Pisoni, D.B., (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *Journal of the Acoustical Society of America*, *89*, 874-886.

Mah, J., Goad, H., Steinhauer, K. (2016). Using event-related brain potentials to assess perceptibility: The case of French speakers and English [h]. *Frontiers in Psychology*, *7*, 1-14.

Melnik, G. A., Peperkamp, S. (2019). Perceptual deletion and asymmetric lexical access in second language learners. *Journal of the Acoustical Society of America, 145*(1), EL13-EL18.

Okuno, T. & Hardison, D. M. (2016). Perception-production link in L2 Japanese vowel duration: Training with technology. *Language Learning & Technology*, *20*(2), 61-80.

Pallier, C., Colomé, A., Sebastián-Gallés, N. (2001). The Influence of Native-Language Phonology on Lexical Access: Exemplar-Based Versus Abstract Lexical Entries. *Psychological Science*, *12*(6), 445-449.

Piske, T., Mackay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*(2), 191-215.

Pisoni, D. & Luce, P. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, *25*(1-2), 21-52.

Sakai, M. & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187-224.

Sebastián-Gallés, N. (2005). Cross-language speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception*. Oxford: Blackwell.

Shinohara, Y., Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. *Journal of Phonetics*, *66*, 242-251.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*, 456-466.

Wang, Y., Jongman, A., Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, *113*, 1033-1043.

Wang, Y., Spence, M., Jongman, A., Sereno, J. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106*, 3649-3658.

Weber, A., Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1-25.

Yamada, J. (1991). The discrimination learning of the liquids /r/ and /l/ by Japanese speakers. *Journal of Psycholinguistic Research, 20*(1), 31-46.