High-Variability Phonetic Training enhances second language lexical processing: evidence from online training of French learners of English*

Gerda Ana Melnik1,2 and Sharon Peperkamp1

1 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 and 2Institute of Data Science and Digital Technologies, Vilnius University, Akademijos str. 4, Vilnius LT-08412, Lithuania

Abstract

High-Variability Phonetic Training (HVPT) has been shown to be effective in improving the perception of the hardest non-native sounds. However, it remains unclear whether such training can enhance phonological processing at the lexical level. The present study tested whether HVPT also improves word recognition. Late French 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 cause difficulty both at the prelexical (Maj, Goad & Steinhauser, 2016) and the lexical level of processing (Melnik & Peperkamp, 2019). In pretest and posttest participants were administered a prelexical identification task and a lexical decision task. Results demonstrate that after training the learners’ accuracy improved in both tasks. Moreover, these improvements were retained four months after posttest. This is the first evidence that short training can enhance not only prelexical perception, but also word recognition.

1. Introduction

A common finding in second language research is that producing and perceiving non-native speech sounds is difficult (for reviews, see Piske, MacKay & Flege, 2001; Sebastián-Gallés, 2005). Nevertheless, much research has demonstrated that, with auditory training, the difficulty of perceiving even the hardest non-native sounds can be reduced. For example, numerous training studies have focused on Japanese listeners’ difficulty to perceive the English sounds /l/ and /ɹ/ (for a review, see Bradlow, 2008). These sounds are particularly difficult, as Japanese has only a single liquid consonant that is ambiguous between English /l/ and /ɹ/; consequently, Japanese listeners fail to perceive these sounds as different. Yet, auditory training on Japanese learners has proved successful (e.g., Iverson, Hazan & Bannister, 2005; McCandliss, Fiez, Protopapas, Conway & McClelland, 2002; Zhang, Kuhl, Imada, Iverson, Pruitt, Stevens, Kawakatsu, Tohkura & Nemoto, 2009), showing that, in speech perception, non-native speech sound categories can become more precise with training.

The most common training paradigm used to improve second language (L2) speech sound perception is High-Variability Phonetic Training (HVPT; Logan, Lively & Pisoni, 1991). HVPT uses multiple natural exemplars of the target sounds in a variety of phonetic environments presented in minimal pairs of either words (e.g., Hazan, Sennema, Iba & Faulkner, 2005; Logan et al., 1991; Shinohara & Iverson, 2018) or nonwords (e.g., Carlet & Cebrian, 2015; Sadakata & McQueen, 2014). 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 listener’s attention to the relevant phonetic cues of the sounds to be learned (Homa & Cullite, 1984; Logan et al., 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, 2015; Lee & Lyster, 2016), consonants (Kim & Hazan, 2010; Shinohara & Iverson, 2018), tones (Wang, Spence, Jongman & Sereno, 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, Pisoni, Yamada, Tokhura & Yamada, 1994; Okuno & Hardison, 2016). Finally, it gives rise to long-term retention of the new categories (Iverson & Evans, 2009; Lively et al., 1994), and it helps to improve L2 production (for a review, see Sakai & Moorman, 2018).

These studies and most other previous work demonstrating the effectiveness of HVPT 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 (Pallier, Colomé & Sebastián-Gallés, 2001; Weber & Cutler, 2004). For example, Spanish–Catalan bilinguals have been shown to have difficulty in perceiving the Catalan contrast /e/-/e/ (Pallier, Bosch & Sebastián-Gallés, 1997). Sebastián-Gallés and Baus (2005) demonstrated that this perceptual problem extends to the lexical level: in a lexical decision task Spanish–Catalan bilinguals had difficulty rejecting nonwords created from real words where the vowel /e/ was replaced by the vowel /e/, and vice-versa. Thus, truly successful training should also enhance performance at the lexical level. Recognizing speech sounds prelexically, however, requires different skills compared to recognizing words containing these sounds. 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 (Pisoni & Luce, 1987). Moreover, higher processing levels have higher memory demands and a larger cognitive load (Werker & Tees, 1984; Werker & Logan, 1985). Although under normal listening conditions native speakers are generally at ceiling across tasks that tap into different levels of processing, non-native listeners perform poorer on tasks that have greater lexical involvement (Díaz, Mitterer, Broersma & Sebastián-Gallés, 2012; Sebastián-Gallés & Baus, 2005).

Differences in performance in L2 learners across tasks of different complexity are addressed in the Automatic Selective Perception model (Strange & Shafer, 2008; Strange, 2011), which posits two modes of perception – a phonological and a phonetic one. The phonological mode relies on automatic selective perception routines which do not require focused attention, and thus allow for attention to be allocated to other tasks, such as processing word meaning. It is normally used in L1 and allows the listener to rapidly extract sufficient phonologically relevant contrastive information to recognize word forms, while ignoring context-dependent phonetic details. The phonetic mode of perception, by contrast, is precisely centered on accessing context-dependent phonetic details. This mode requires attentional focus and high cognitive involvement. It is used in L1 when adjusting to a different dialect, as well as in L2 to perceive non-native sounds, provided sufficient attention and cognitive resources can be allocated to the task of phonetic decoding. Consequently, in experimental settings the performance of learners might be good on relatively simple prelexical tasks, where they can use the phonetic mode of perception and focus their attention on crucial phonetic cues. However, the same performance level might not be attained in more complex tasks, additionally requiring, for instance, attention to word meaning. In such tasks they must resort to the phonological mode of perception, which uses their L1 selective perception routines. These routines are unsuited for processing L2 categories, resulting in poor accuracy. Finally, highly experienced L2 learners might develop L2-specific selective perception routines. While these routines are typically not as optimal and fully automated as the ones for L1, they may allow for relatively good performance even in complex tasks, especially in optimal listening conditions.

Importantly, within the Automatic Selective Perception model, successful perceptual training also leads to the development and automatization of L2 selective perception routines (Strange, 2011). This raises the hypothesis that phonetic training allows L2 learners to become more efficient in processing L2 sounds even at the – cognitively more demanding – lexical level. This hypothesis cannot be tested with the identification or discrimination tasks that are typically used in HVPT studies, even if they employ minimal pairs of words (as is often the case, see e.g., Gronen, Kubota & Sheppard, 2019; Lee & Lyster, 2016; Shinohara & Iverson, 2018). Indeed, deciding whether a given stimulus corresponds, say, to the word lock or rather to the word rock is not different from deciding whether a stimulus corresponds to the syllable /la/ or rather to /ra/; in both cases one must just identify the first consonant (an equivalent reasoning holds for discrimination tasks). Thus, these tasks do not require lexical access.

So far, two studies on the effect of phonetic training on lexical processing have examined naïve listeners’ ability to learn words in a tonal language (Cooper & Wang, 2011; Ingvalson, Barr & Wong, 2013). Both studies found that phonetic training improved naïve English listeners’ performance in a word-learning task involving difficult tone contrasts. To our knowledge, however, no studies have directly assessed whether phonetic training can improve word recognition in late learners. Here, we focus on this question by studying the perception of the English sound /h/ by 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, late 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). Similarly, when hearing such words and nonwords, low-proficiency learners fail to show an N400 nonword effect, suggesting, that they process the nonwords as if they were real words (White, Titone, Genesee & Steinhauer, 2017). Thus, they have difficulty not only in perceiving the contrast between the presence and absence of /h/, but also in distinguishing between words and nonwords that differ only in the presence vs. absence of /h/.

The case of English /h/ is particularly suited to examine the effect of HVPT on lexical processing, due to the fact that 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. Moreover, they are instructed that in English – contrary to French, in which <ch> is silent – <h> is almost always pronounced, and that its pronunciation is /h/. Thus, if training results in the development of a selective perception routine for the sound /h/, French learners of English can immediately apply this routine during word recognition, because they already have metalinguistic knowledge of which words contain /h/. That is, they do not need to learn separately for which words they should update the phonological representations in their mental lexicon.

We trained late French learners of English on the perception of English /h/ in a pretest–training–posttest design, using the classical version of HVPT (Logan et al., 1991). In pretest, 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 identification task we used /h/- and vowel-initial nonwords as stimuli. On each trial participants had to decide whether the nonword they heard started with the sound /h/ or not. In the lexical decision task we used words and nonwords, where the test nonwords

1Lively et al., (1993) used both word and nonword minimal pairs in an identification task and indeed observed neither a main effect of lexical status nor an interaction of this factor with the factor pre- vs. posttest.
were created from /h/-initial and vowel-initial words by removing or adding /h/, respectively, and the control nonwords by either changing, deleting or inserting one phoneme. For each item participants had to reply whether it is a word or not in English. Given the difficulty of the /h/ sound for French speakers, they are expected to have particular difficulty with these critical items, making more "no"-replies to the real words (misses) and “yes”-replies to the nonwords (false alarms) compared to the control items, (as previously shown in Melnik & Peperkamp, 2019). In posttest, both tasks were repeated. For the identification task we used the same stimuli as in pretest, supplemented by trials with novel items, such as to test for generalization. For the posttest of the lexical decision task we used new stimuli. Four months after the posttest, participants returned for a long-term retention test, which was identical to the posttest.

We did not include a control group of non-trained participants who are only tested in pre- and posttest. Studies comparing trained to control participants using identification tasks have provided ample evidence that the HVPT paradigm is effective in improving non-native sound perception (Iverson et al., 2005; Lee & Lyster, 2016; Okuno & Hardison, 2016), and several recent studies using the paradigm no longer included such a control group either (Grenon et al., 2019; Leong, Price, Pitchford & van Heuven, 2018; Sadakata & McQueen, 2014; Shinohara & Iverson, 2018; Tamminen, Peltola, Kujala & Näätänen, 2015). As to the lexical decision task, we included a control condition with nonwords that involve other changes than the deletion or addition of /h/. Note also that in this task we used different items in pre- and posttest, thus ensuring that potential improvements in this task can only be due to the training.

While the pre-, post- and retention test were run in our lab in Paris, training was administered online at participants’ homes. It consisted of eight sessions of an identification task using minimal pairs of real words (such as air-hair), with corrective feedback. 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 due to the automatization of L2 selective perceptual routines (Strange, 2011), performance in lexical decision should likewise improve with training. We also expected the effects of training to be robust at the prelexical level, and hence to be observable in the identification task four months after the posttest, as previously found in several training studies which used only prelexical tasks (Lee & Lyster, 2016; Lively et al., 1994). Importantly, additional retention of the positive effects of training at the lexical level would be an indication that phonetic training can have long-term benefits on processing at the lexical level.

2. Methods

2.1. Pretest-Posttest-Generalization: Identification

Stimuli

For the pre- and posttest we selected 100 pairs of items; the great majority of these items were nonwords but a few were low-frequency real words (see Appendix S1, Supplementary Materials). The members of each pair differed in the presence or absence of an initial /h/ (e.g., /h/ʌp/-/ʌp/). Forty pairs were monosyllabic, 40 disyllabic and 20 trisyllabic. Ten English vowels (ʌ, ɒ, ə, ɪ, ı, e, ɪ, ι, ɔ, et, ai) 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.

One member of each pair was recorded by a male, and the other by a female native speaker of American English, with the proviso that each speaker recorded equal numbers of /h/-initial and vowel-initial nonwords. Table 1 shows average duration (ms) and intensity (dB) of the sound /h/, as well as the ratio between /h/ and the initial /hV/-portion, in the /h/-initial nonwords used in the identification task.

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 arrow key labelled “h” with their dominant hand if they thought the nonword started with the sound /h/, and to press the arrow key labelled “no h” with their non-dominant hand if they thought it did not start with /h/. Participants were explicitly told that the items they would hear are nonwords. There were 200 trials divided over three 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 six trials of the training block, three /h/- and three vowel-initial, served as practice phase, 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 (two blocks of 97 trials each), participants received no feedback and there was a time-out of 2500 ms: if participants did not give a response within the allotted time 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, participants performed the same task in 60 trials with the 30 additional nonword pairs.

2.2. Pretest-Posttest: Lexical decision

Stimuli

The stimuli were the same as in Melnik and Peperkamp (2019) (see Appendix S1, Supplementary Materials). 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. A group of 45 adult French learners of English whose L2 proficiency is comparable to that of the participants in the current study rated the words as highly familiar (mean familiarity score: 4.95 on a scale from 1 to 5 (SD = 0.1)). The /h/-initial and the vowel-initial words did not differ in mean number of syllables, in mean frequency in the Subtlex database (Brysbaert & New, 2009), (both t < 1), or in mean familiarity in the rating questionnaire (t = 1.0, p > 0.1).

Table 1 shows average duration (ms) and intensity (dB) of the sound /h/, as well as the ratio between /h/ and the initial /hV/-portion, in the /h/-initial words and nonwords used in the lexical decision task.

Each word was paired with a nonword, created by deleting or adding /h/ at the beginning (e.g., husband – *husband, officer –
Results

The procedure was identical to that in Melnik and Peperkamp (2019): participants performed a speeded auditory lexical decision task, using 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 timeout and the presentation of the next trial.

Table 1 shows average duration (ms) and intensity (dB) of the sound /h/, as well as the ratio between /h/ and the initial /hV/-portion, in the /h/-initial stimuli used in the test and training tasks (numbers in parentheses are standard errors).

<table>
<thead>
<tr>
<th>Task</th>
<th>Speaker</th>
<th>Condition</th>
<th>Duration (ms)</th>
<th>Intensity (dB)</th>
<th>Ratio: h/hV</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEST: Identification in nonwords</td>
<td>Male1</td>
<td>/h/</td>
<td>59.6 (2.6)</td>
<td>0.40 (0.02)</td>
<td>0.45 (0.01)</td>
</tr>
<tr>
<td></td>
<td>Fem2</td>
<td>/h/</td>
<td>54.2 (1.6)</td>
<td>0.25 (0.01)</td>
<td>0.44 (0.01)</td>
</tr>
<tr>
<td></td>
<td>t-test</td>
<td>NS</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST: Lexical decision</td>
<td>Male1</td>
<td>Words</td>
<td>93.3 (3.4)</td>
<td>0.52 (0.01)</td>
<td>0.44 (0.002)</td>
</tr>
<tr>
<td></td>
<td>Nonwords</td>
<td>111.1 (2.9)</td>
<td>0.56 (0.01)</td>
<td>0.44 (0.002)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t-test</td>
<td>***</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRAINING: Identification in real words</td>
<td>Male1</td>
<td>Words</td>
<td>147.9 (3.4)</td>
<td>0.56 (0.02)</td>
<td>0.42 (0.001)</td>
</tr>
<tr>
<td></td>
<td>Male2</td>
<td>97.0 (4.0)</td>
<td>0.45 (0.02)</td>
<td>0.45 (0.01)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fem1</td>
<td>107.1 (3.1)</td>
<td>0.36 (0.02)</td>
<td>0.42 (0.003)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fem2</td>
<td>93.9 (2.8)</td>
<td>0.34 (0.02)</td>
<td>0.40 (0.004)</td>
<td></td>
</tr>
</tbody>
</table>

* * *: * p < .001

### 2.3. Training

**Stimuli**

We selected 59 minimal pairs of real words differing in the presence or absence of an initial /h/ (see Appendix S1, Supplementary Materials). Given the limited number of such minimal pairs, we used both frequent words (e.g., hair-air) and infrequent ones (e.g., hosier-osier). However, word frequency was not considered to have an impact, as the task used in training did not require lexical access.

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. Table 1 shows average duration (ms) and intensity (dB) of the sound /h/, as well as the ratio between /h/ and the initial /hV/-portion, in the /h/-initial words used in the training task.
Procedure
The training started one to three days after the pretest, and 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 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 session (e.g., participant details, day and time of completion, RTs and responses) were automatically sent to the MySql database after its completion. 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 corresponded to the word “hair” but the participant chose instead the word “air”, the word “Wrong” was displayed on the screen and the auditory feedback “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 participant’s accuracy. The last session was separated from the posttest by one or two days.

2.4. Participants
Participants were French intermediate learners of English, born in France and recruited from among university students in Paris. Three of them were students in an English language 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 training and posttest and were hence included in the data analysis. These participants, 12 women and 12 men aged between 19 and 32 (mean: 22.3), had started learning English at school. Before coming to the pretest they filled in an online questionnaire to self-evaluate their speaking, listening, reading, vocabulary and grammar skills in English, French, and any other language they knew, 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 indicated being more proficient in another foreign language than English. Twenty-one of them returned to the laboratory for a retention test four months after the posttest (mean number of days = 115.3, SD = 5.4).

None of the 51 participants who did the pretest reported a history of speech or language problems. They received a small payment after the pretest. The 24 who underwent training received a second, larger, payment when they came back to the laboratory for the posttest, and the 21 who came for the retention test received a bonus payment at the end of the retention test.

3. Results and discussion
As both identification and lexical decision are signal detection tasks, we used individual A’ scores as dependent measure for all our analyses. A’ provides a non-parametric, unbiased, index of sensitivity, with 0.5 indicating chance performance and 1.0 perfect performance. We analyzed the datasets using generalized mixed effects regression modeling for beta distribution (since A’ is continuous between zero and one), with a logit link function (R package gllvmTMB, Brooks et al., 2017). For the interested reader, the mean percentages of correct responses for each task can be found in Appendix S2 (Supplementary Materials).

3.1. Identification
Pretest, Posttest, Generalization
Prior to analysis, we discarded trials with no response or time-out (2.5% of the data in pretest, 2.2% in posttest, 2.1% in generalization). Figure 1a displays the participants’ A’ scores in pretest, posttest, and generalization.

We constructed a model with Session (Pretest vs. Posttest vs. Generalization) as a contrast-coded fixed effect, and intercepts for Participants as random factor. P-values were obtained by likelihood ratio tests of the full model against the model without the effect or interaction in question. The analysis revealed that the fixed effect of Session was significant ($\beta = -1.11, SE = 0.16, \chi^2(1) = 32.02, p < .0001$), 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 ($p = .82$).

We also examined if performance was influenced by the acoustic properties of the stimuli produced by each speaker. Recall from Table 1 that of the four acoustic measures (i.e., the average duration of /h/, the average duration ratio between /h/ and the initial /hV/, the average intensity of /h/, and the average intensity ratio between /h/ and the initial /hV/), only the average duration ratio between /h/ and the initial /hV/-portion was significantly different for the two speakers. Specifically, it was smaller in the stimuli produced by the female speaker than in those produced by the male speaker (0.25 versus 0.40; t(79.65) = 7.83, p < .001). Performance on stimuli produced by each of the two speakers, however, did not differ (Mean_{male_speaker} = 76.0% correct, Mean_{female_speaker} = 76.5% correct, p > .1).

Note that in all sessions, none of the participants had an A’ score of exactly 0 or 1 in a given task or condition; the requirement for beta distributions was thus met for all datasets.
Retention

Prior to analysis, we discarded trials with no response or time-out (2.3% of the data). Figure 1b displays the identification accuracy in the retention test of the 21 participants who returned to the laboratory four months after posttest. We constructed a model with Session (Posttest vs. 4-months delayed posttest) as fixed factor and random intercepts for Participants. There was a significant effect of Session ($\beta = 0.27$, SE = 0.08, $\chi^2(1) = 9.41$, $p < .01$), with the accuracy improving from an average A’ score of 0.87 in posttest to 0.90 in 4-months delayed posttest. Thus, the performance of participants in identification did not decline after a period of four months.

3.2. Lexical Decision

Pretest-Posttest

Prior to analysis, we discarded responses with no response or time-out (1.5% of the data in pretest, 1.2% in posttest). Note that the high error rates on test nonwords made it impossible to analyze the reaction times, as such analyses are done on the correct responses only. Figure 2a displays participants’ A’ scores on the test and control items in pretest and posttest.

We constructed a model with fixed factors Session (pretest vs. posttest), Condition (test vs. control) and Lists (AC vs. BD), as well as an interaction between Session and Condition. The model included random intercepts for Participants. We found significant effects of Session ($\beta = 0.65$, SE = 0.12, $\chi^2(1) = 25.67$, $p < .001$) and Condition ($\beta = 1.59$, SE = 0.13, $\chi^2(1) = 85.93$, $p < .001$), and a Session × Condition interaction ($\beta = -0.95$, SE = 0.24, $\chi^2(1) = 14.59$, $p < .001$). Pairwise comparisons 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, which was therefore omitted from further analyses.

In order to test if there was a relationship between the amount of improvement in prelexical and lexical levels, we carried out a Pearson correlation test between gains in identification task and gains in the lexical decision task (gain was calculated by subtracting the pretest score from the posttest score for each participant and each task). Results revealed that there was a moderate but significant correlation between the two ($r = .41$, $p = .04$) (see Figure 3).

We also examined if performance was influenced by the acoustic properties of word and nonword stimuli. Recall from Table 1 that while words and nonwords did not differ with regard to the intensity of /h/, both the average duration of /h/ and the average duration ratio between /h/ and the initial /hV/-portion were larger in the nonword than in the real word stimuli (Mean duration: words = 93.3 ms, nonwords = 111.1 ms, $t(75.38) = 4.00$, $p < .0001$; Mean ratio: words = .52, nonwords = .56, $t(76.32) = 2.33$, $p = .02$). Yet, performance was worse on nonwords than on words. This difference in performance can thus not be accounted for by the acoustic properties of the stimuli.

Retention

Prior to analysis, we discarded trials with no response or time-out (1.3% of the data). Figure 2b displays the accuracy on the test and control items in the retention test for lexical decision of the 21 participants who returned to the laboratory four months after posttest. To examine retention of the training improvements after four months in the 21 participants who returned for the retention test, we constructed a model with fixed factors Session (posttest vs. 4-months delayed posttest), Condition (test vs. control), as well as an interaction between Session and Condition. The model included random intercepts for Participants. We found a significant effect of Condition ($\beta = 1.29$, SE = 0.14, $\chi^2(1) = 52.11$, $p < .001$), but no effect of Session and no interaction. Thus, overall performance in the lexical decision task did not significantly reduce 4 months after the immediate posttest.
4. General discussion

The current study investigated if phonetic training can lead to better recognition of words that contain a difficult non-native sound. We tested late French learners of English on both their prelexical perception and their lexical processing of stimuli containing /h/. As this sound does not exist in French, French listeners tend to confuse the presence of /h/ with its absence. Eight sessions of High-Variability Phonetic training (HVPT) were administered to the participants online at their homes. They were tested in pretest, posttest and a retention test by means of an identification and a lexical decision task.

The results of the pretest show that French learners of English have difficulty in perceiving the difference between the presence and absence of /h/ at both the prelexical and lexical level of processing; these results are thus in accordance with Mah et al. (2016) and Melnik and Peperkamp (2019), respectively. Crucially, we found that participants’ performance in both tasks was improved 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 those from many previous studies that used HVPT (e.g., Hazan et al., 2005; Lee & Lyster, 2016; Lively et al., 1994; Shinohara & Iverson, 2018). Importantly, they show that training does not need to be administered in a well-controlled laboratory setting to be effective. Concerning the lexical decision task, our results provide 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 fewer errors on the test items with /h/ than in pretest, while on control items their performance did not change. The gain from training in the identification task and the lexical decision task were correlated, suggesting that the more effective training is on prelexical perception, the greater the transfer effects onto lexical perception. Note that it is highly unlikely that the improvement from pretest to posttest was due to learning outside the lab and not to training. Indeed, the posttest was separated from the pretest by only 10 to 20 days, and all participants were tested in Paris, where they mostly use French in their daily life. Moreover, only three of them were enrolled in studies in a department of English language, and only three others were taking English classes during the period of testing. All remaining participants (N = 18) reported that they were not attending English classes during the academic year in question.

Finally, results from the retention test showed that the positive effects of training did not decrease four months after the posttest. This suggests that learning induced by phonetic training is robust not only at the prelexical processing level, as reported earlier in other case studies (Iverson & Evans, 2009), but also at the lexical processing level. In identification (but not lexical decision) there was even a small but significant improvement in the retention test compared to the posttest. One possible explanation is that participants heard the stimuli for the third time and that this slightly facilitated their task (note, though, that they were not at ceiling). Alternatively, the training might have triggered a learning trajectory, such that they further improved their perception of /h/ somewhat over the course of four months with subsequent exposure to English in their daily lives.

3The results remain the same without the six participants who were taking some type of English classes during the period of testing and training.
Our findings 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/-e 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 in this study is paralleled by an improvement in lexical processing suggests a bottom-up sequential order in learning. That is, while at a specific learning stage 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.

How exactly might HVPT improve lexical processing in L2 learners? We interpret this result in light of the Automatic Selective Perception model (Strange & Shafer, 2008; Strange, 2011). Recall that this model posits both a phonetic and a phonological mode of speech perception. The phonetic mode is cognitively demanding, focused on attention to phonetic detail. The phonological mode, by contrast, uses automatized language-specific speech perception routines, focused on detecting sufficient information for the rapid and robust identification of (sequences of) speech sounds; this mode thus allows the listener to allocate cognitive resources to the lexical processing level. L2 learners typically rely on the phonetic mode for the purposes of perceiving non-native sounds in tasks such as identification and discrimination. Their performance in those tasks is better than that in lexical tasks such as lexical decision, for which they lack the relevant speech perception routines. Our results, then, suggest that HVPT aids the development and/or automatization of such routines in L2, allowing learners to gradually rely more on the phonological and less on the phonetic mode, and hence to crucially improve not only their prelexical but also their lexical perception.

A similar finding on the benefit of phonetic training on the automatization of L2 speech processing was reported in a study on the perception of speech in noise (Lengeris & Hazan, 2010). Adverse listening conditions such as a high signal-to-noise ratio (SNRs) have been shown to involve increased cognitive load (Pichora-Fuller, Schneider & Daneman, 1995). This is one of the reasons why, just as in higher-processing level tasks, environmental signal distortion has greater negative effects for speech perception on non-native than on native listeners (for a review, see Lecumberri, Cooke & Cutler, 2010). Lengeris and Hazan (2010) showed that HVPT in quiet improves the perception of difficult L2 sounds in noise (multi-talker babble). Another finding worth noting is that of Shinohara and Iverson (2018) on the improved production of the English /θ/-/ð/ contrast by Japanese learners as a result of phonetic training. They suggest that training might induce automaticity of phonetic processes which in turn allows L2 learners to pronounce the correct acoustic contrasts.

The case of the English /θ/-/ð/ contrast for Japanese learners and the one we investigated here share an important aspect in light of the transfer of improved prelexical perception to improved lexical processing. That is, both the /θ/-/ð/ contrast and /θ/ are transparently encoded in English orthography. Therefore, Japanese and French learners have metalinguistic knowledge of which words contain the difficult sound(s), and can therefore use their newly developed perception routine to immediately update the phonological representations of these words. Compare this to, for instance, the Catalan /e/-/ε/ contrast for Spanish learners. As both /e/ and /ε/ are spelled <e>, Spanish learners who have developed a speech perception routine for this contrast during HVPT should still learn word-by-word whether the grapheme <e> corresponds to /e/ or to /ε/.

We end with a note on practical implications of our results in the realm of language teaching. Word recognition is an inherent element of “real life” language processing. The fact that it can be improved by relatively short HVPT is encouraging. Moreover, our training was administered online and can thus easily complement traditional language teaching methodologies. Our low drop-out rate (only one participant did not finish the set of training sessions, and only four did not come back for the retention test in the lab) holds promise of successfully introducing our method not only in the classroom but also in a self-study setting.

To conclude, we showed that even short online HVPT can improve both prelexical and lexical processing of a difficult L2 sound. Moreover, we demonstrated that these improvements are retained in the long term. However, 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.

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728920000644.

Acknowledgements. This research was supported by grants from the Agence Nationale de la Recherche (ANR-17-EURE-0017, ANR-17-CE28-0007-01) and a Postdoctoral fellowship from Vilnius University to Gerda Ana Melnik. We would like to thank Ewan Dunbar, Mollie Hamilton, Monica Hegde, and Jeremy Kuhn for recording the stimuli.

References

Bilingualism: Language and Cognition


Downloaded from https://www.cambridge.org/core. Ecole normale superieure on 09 Apr 2021 at 12:15:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1366728920000644


