



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

Gerda Ana Melnik, Sharon Peperkamp

## ► To cite this version:

Gerda Ana Melnik, Sharon Peperkamp. High-Variability Phonetic Training enhances second language lexical processing: evidence from online training of French learners of English. *Bilingualism: Language and Cognition*, 2020, pp.1-10. <10.1017/S1366728920000644>. <hal-03050640>

**HAL Id: hal-03050640**

**<https://hal.science/hal-03050640v1>**

Submitted on 9 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

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

Journal:	<i>Bilingualism: Language and Cognition</i>
Manuscript ID	BLC-19-RA--0136.R3
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Melnik, Gerda Ana; Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Etudes Cognitives, Ecole normale supérieure - PSL University; Vilnius University, Institute of Data Science and Digital Technologies Peperkamp, Sharon; Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Etudes Cognitives, Ecole normale supérieure - PSL University
Content Areas:	Speech Perception, Typical Language Development, Mental Lexicon
Linguistic Areas:	Phonology
Methods:	Behavioural Measurements
Populations:	Adult Second Language Development
Languages:	French, English
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 (Mah, Goad & Steinhauer, 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.

SCHOLARONE™  
Manuscripts

Cambridge University Press

# Running-head: Phonetic training online enhances L2 lexical processing

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

Gerda Ana Melnik<sup>1,2</sup>, Sharon Peperkamp<sup>1</sup>

<sup>1</sup> 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

<sup>2</sup> Institute of Data Science and Digital Technologies, Vilnius university, Akademijos str. 4,  
Vilnius LT-08412, Lithuania

### Supplementary materials

For supplementary materials accompanying this paper, visit

<https://www.cambridge.org/core/journals/bilingualism-language-and-cognition>

### \*Acknowledgments

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.

### Address for correspondence:

Gerda Ana Melnik  
Institute of Data Science and Digital Technologies, Vilnius University,  
Akademijos str. 4,  
Vilnius LT-08412,  
Lithuania  
email: [gerda.ana.melnik@gmail.com](mailto:gerda.ana.melnik@gmail.com)

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

## 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 (Mah, Goad & Steinhauer, 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.

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



## 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 learner's attention to the relevant phonetic cues of the sounds to be learned (Homa & Cultice, 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 of 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/-/ɛ/ (Pallier, Bosch, & Sebastián, 1997). Sebastián-Gallés & 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 /ɛ/, 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.

1  
2  
3 Importantly, within the Automatic Selective Perception model, successful perceptual  
4  
5 training also leads to the development and automatization of L2 selective perception routines  
6  
7 (Strange, 2011). This raises the hypothesis that phonetic training allows L2 learners to  
8  
9 become more efficient in processing L2 sounds even at the - cognitively more demanding -  
10  
11 lexical level. This hypothesis cannot be tested with the identification or discrimination tasks  
12  
13 that are typically used in HVPT studies, even if they employ minimal pairs of words (as is  
14  
15 often the case, see e.g. Grenon, Kubota & Sheppard, 2019; Lee & Lyster, 2016; Shinohara &  
16  
17 Iverson, 2018). Indeed, deciding whether a given stimulus corresponds, say, to the word *lock*  
18  
19 or rather to the word *rock* is not different from deciding whether a stimulus corresponds to the  
20  
21 syllable /la/ or rather to /ra/; in both cases one must just identify the first consonant (an  
22  
23 equivalent reasoning holds for discrimination tasks).<sup>1</sup> Thus, these tasks do not require lexical  
24  
25 access.  
26  
27  
28  
29

30  
31 So far, two studies on the effect of phonetic training on lexical processing have examined  
32  
33 naïve listeners' ability to learn words in a tonal language (Cooper & Wang, 2011; Ingvalson,  
34  
35 Barr & Wong, 2013). Both studies found that phonetic training improved naïve English  
36  
37 listeners' performance in a word-learning task involving difficult tone contrasts. To our  
38  
39 knowledge, however, no studies have directly assessed whether phonetic training can improve  
40  
41 word recognition in late learners. Here, we focus on this question by studying the perception  
42  
43 of the English sound /h/ by French learners of English. As /h/ does not exist in French, French  
44  
45 listeners – even those who are fluent in English – have difficulty perceiving the contrast  
46  
47 between the presence vs. absence of /h/ in English stimuli (Mah et al., 2016). At the lexical  
48  
49 level, late French learners of English tend to accept nonwords such as *usband* (cf. *husband*)  
50  
51 and, to a lesser extent, *hofficer* (cf. *officer*), as real words (Melnik & Peperkamp, 2019).  
52  
53 Similarly, when hearing such words and nonwords, low-proficiency learners fail to show an  
54  
55 N400 nonword effect, suggesting, that they process the nonwords as if they were real words  
56  
57  
58  
59  
60

(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 <h> 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 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. /hɪlp/-/ɪlp/). Forty pairs were monosyllabic, 40 dissyllabic and 20 trisyllabic. Ten English vowels (ʌ, ɒ, ɑ, ɪ, ɛ, i:, ʌ, 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.

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.

<Insert Table 1 about here>

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/ either in the first (33.3%) or the second (66.7%) syllable. This was done such that, on the one hand, nonwords in the test condition did not stand out as being the only ones with a change in the initial syllable, and, on the other hand, there were overall as many nonwords with a change in a non-initial as in the initial syllable, ensuring that participants could not focus their attention exclusively on the initial syllable to do the task.

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 counterbalanced 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, none of the four resulting lists 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 list A were given list C, while participants who previously heard list B, were now given list D. Hence, participants heard only one of the members of each word-nonword pair throughout the whole experiment. In the retention test participants heard the same list, C or D, that they had heard in posttest.

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 time-out and the presentation of the next trial.

### 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)<sup>2</sup>, with a logit link function (R

package glmmTMB, 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.

<Insert Figure 1 about here>

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

( $\text{Mean}_{\text{male\_speaker1}} = 76.0\%$  correct,  $\text{Mean}_{\text{fem\_speaker2}} = 76.5\%$  correct,  $p > .1$ ).

**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 ( $\beta = 0.27$ ,  $\text{SE} = 0.08$ ,  $\chi^2(1) = 9.41$ ,  $p < .01$ ), 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.

<Insert Figure 2 about here>

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 ( $\beta = 0.65$ ,  $\text{SE} = 0.12$ ,  $\chi^2(1) = 25.67$ ,  $p < .001$ ) and Condition ( $\beta = 1.59$ ,  $\text{SE} = 0.13$ ,  $\chi^2(1) = 85.93$ ,  $p < .001$ ), and a Session  $\times$  Condition interaction ( $\beta = -0.95$ ,  $\text{SE} = 0.24$ ,  $\chi^2(1) =$

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).

<Insert Figure 3 about here>

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.<sup>3</sup>

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.

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/-/ɛ/ contrast, Diaz 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 ratios (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 /ɪ/-/ɪ/ 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 /ɪ/-/ɪ/ 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 /ɪ/-/ɪ/ contrast and /h/ 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/-/ɛ/ contrast for Spanish learners. As both /e/ and /ɛ/ 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 /ɛ/.

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.

## References

- Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Maechler, M., Bolker, B.M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, 9, 378–400.
- 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, 977–90.  
<https://doi.org/10.3758/BRM.41.4.977>
- Carlet, A., & Cebrian, J. (2015). Identification vs. discrimination training: Learning effects for trained and untrained sounds. In The Scottish Consortium for ICPhS 2015 (eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK: University of Glasgow.  
<http://www.icphs2015.info/pdfs/proceedings.html> (retrieved April 12, 2016).
- Cooper, A., & Wang, Y. (2011). The influence of tonal awareness and musical experience on tone word learning. In Lee, W.S. & Zee, E. (eds.), *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, 1–12.  
<https://doi.org/10.3758/s13428-014-0458-y>
- 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, 680–689.  
<https://doi.org/10.1016/j.lindif.2012.05.005>

- Grenon, I., Kubota, M., & Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. *Journal of Phonetics*, 72, 17–34.  
<https://doi.org/10.1016/j.wocn.2018.10.005>
- Hazan, V., Sennema, A., Iba, M., and Faulkner, A. (2005). Effect of audio-visual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47, 360–378.  
<https://doi.org/10.1016/j.specom.2005.04.007>
- 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.  
<https://doi.org/10.1017/S0142716403000250>
- 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.  
<https://doi.org/10.1016/j.wocn.2015.06.007>
- 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, 1045.  
[https://doi.org/10.1044/1092-4388\(2012/12-0024\)](https://doi.org/10.1044/1092-4388(2012/12-0024))
- Iverson, P., & Evans, B. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *Journal of the Acoustical Society of America*, 126, 866–877.  
<https://doi.org/10.1121/1.3148196>
- Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: a comparison of methods for teaching English /r/-/l/ to Japanese adults. *The*

- 1  
2  
3 *Journal of the Acoustical Society of America*, 118, 3267–3278.  
4  
5 <https://doi.org/10.1121/1.2062307>  
6  
7  
8 Iverson, P., Pinet, M., & Evans. B. G. (2012). Auditory training for experienced and  
9  
10 inexperienced second-language learners: Native French speakers learning English vowels.  
11  
12 *Applied Psycholinguistics*, 33, 145–160.  
13  
14 <https://doi.org/10.1017/S0142716411000300>  
15  
16  
17 Kim, Y. H., & Hazan, V. (2010). Individual variability in the perceptual learning of L2 speech  
18  
19 sounds and its cognitive correlates. In Dziubalska-Kołaczyk, K., Wrembel M. & Kul, M.  
20  
21 (eds), *Proceedings of the Sixth International Symposium on the Acquisition of Second*  
22  
23 *Language Speech*, pp. 251–256. Poznań, Poland.  
24  
25  
26 Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in  
27  
28 adverse conditions: a review. *Speech Communication*, 52, 864–886.  
29  
30 <https://doi.org/10.1016/j.specom.2010.08.014>  
31  
32  
33 Lee, A. H., & Lyster, R. (2016). Effects of different types of corrective feedback on receptive  
34  
35 skills in a second language: A speech perception training study. *Language Learning*, 66,  
36  
37 809–833.  
38  
39 <https://doi.org/10.1111/lang.12167>  
40  
41  
42 Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency  
43  
44 discrimination acuity on the phonetic training of English vowels for native speakers of  
45  
46 Greek. *Journal of the Acoustical Society of America*, 128, 3757–3768.  
47  
48 <https://doi.org/10.1121/1.3506351>  
49  
50  
51 Leong, C. X. R., Price, J. M., Pitchford, N. J., & van Heuven, W. J. B. (2018). High  
52  
53 variability phonetic training in adaptive adverse conditions is rapid, effective, and  
54  
55 sustained. *PLOS ONE*, 13, e0204888.  
56  
57 <https://doi.org/10.1371/journal.pone.0204888>  
58  
59  
60



- 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, 2076–2087.  
<https://doi.org/10.1121/1.410149>
- 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.  
<https://doi.org/10.1121/1.1894649>
- 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.  
<https://doi.org/10.3389/fpsyg.2016.01469>
- McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience*, 2, 89–108. <https://doi.org/10.3758/CABN.2.2.89>
- Melnik, G. A., Peperkamp, S. (2019). Perceptual deletion and asymmetric lexical access in second language learners. *Journal of the Acoustic Society of America*, 145, EL13–EL18  
<https://doi.org/10.1121/1.5085648>
- Okuno, T., & Hardison, D. M. (2016). Perception-production link in L2 Japanese vowel duration: Training with technology. *Language Learning & Technology*, 20, 61–80.
- Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64, B9–B17.  
[https://doi.org/10.1016/S0010-0277\(97\)00030-9](https://doi.org/10.1016/S0010-0277(97)00030-9)

- 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, 445–449.  
<https://doi.org/10.1111/1467-9280.00383>
- Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, 97, 593–608.  
<https://doi.org/10.1121/1.412282>
- Piske, T., Mackay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191–215.  
<https://doi.org/10.1006/jpho.2001.0134>
- Pisoni, D., & Luce, P. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, 25, 21–52.  
[https://doi.org/10.1016/0010-0277\(87\)90003-5](https://doi.org/10.1016/0010-0277(87)90003-5)
- Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, 5, 1–15.  
<https://doi.org/10.3389/fpsyg.2014.01318>
- 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, 187–224.  
<https://doi.org/10.1017/S0142716417000418>
- Sebastián-Gallés, N. (2005). Cross-language speech perception. In Pisoni, D. B. & Remez, R. E. (eds.), *The Handbook of Speech Perception*, pp. 546–566. 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.  
<https://doi.org/10.1016/j.wocn.2017.11.002>
- Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39, 456–466.  
<http://dx.doi.org/10.1016/j.wocn.2010.09.001>
- Strange, W., Shafer, V.L. (2008). Speech perception in late second language learners: the re-education of selective perception. In Zampini, M. & Hansen, J. (eds.). *Phonology and Second Language Acquisition*, pp. 153 – 191. Cambridge University Press.
- Tamminen, H., Peltola, M. S., Kujala, T., & Näätänen, R. (2015). Phonetic training and non-native speech perception — New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures. *International Journal of Psychophysiology*, 97, 23–29.  
<https://doi.org/10.1016/j.ijpsycho.2015.04.020>
- 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.  
<http://asa.scitation.org/doi/10.1121/1.1531176>
- 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.  
<https://doi.org/10.1121/1.428217>
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50, 1–25.  
[https://doi.org/10.1016/S0749-596X\(03\)00105-0](https://doi.org/10.1016/S0749-596X(03)00105-0)
- Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech

- 1  
2  
3 perception. *Perception & Psychophysics*, 37, 35–44. <https://doi.org/10.3758/BF03207136>  
4  
5  
6 Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language  
7  
8 speech perception. *The Journal of the Acoustical Society of America*, 75, 1866–1878.  
9  
10 <https://doi.org/10.1121/1.39098>  
11  
12 White, E. J., Titone, D., Genesee, F., & Steinhauer, K. (2017). Phonological processing in late  
13  
14 second language learners: The effects of proficiency and task. *Bilingualism: Language and*  
15  
16 *Cognition*, 20, 162–183.  
17  
18 <https://doi.org/10.1017/S1366728915000620>  
19  
20  
21 Yamada, J. (1991). The discrimination learning of the liquids /r/ and /l/ by Japanese speakers.  
22  
23 *Journal of Psycholinguistic Research*, 20, 31–46.  
24  
25 <https://doi.org/10.1007/BF01076918>  
26  
27  
28 Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., Kawakatsu, M.,  
29  
30 Tohkura, Y., & Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A  
31  
32 magnetoencephalography study. *NeuroImage*, 46, 226–240.  
33  
34 <http://doi.org/10.1016/j.neuroimage.2009.01.028>  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Notes**

<sup>1</sup> Lively 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.

<sup>2</sup> 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.

<sup>3</sup> The results remain the same without the six participants who were taking some type of English classes during the period of testing and training.

## Figure captions

Figure 1. Boxplots of A' scores in the identification task in pretest, posttest, and generalization

(a), and retention test (b). The black crossmarks indicate mean A' scores.

Figure 2. Boxplots of A' scores in the lexical decision task in pretest, posttest, and

generalization (a), and retention test (b). The black crossmarks indicate mean A' scores.

Figure 3. Scatter plot showing individual participants' gains in the identification and the

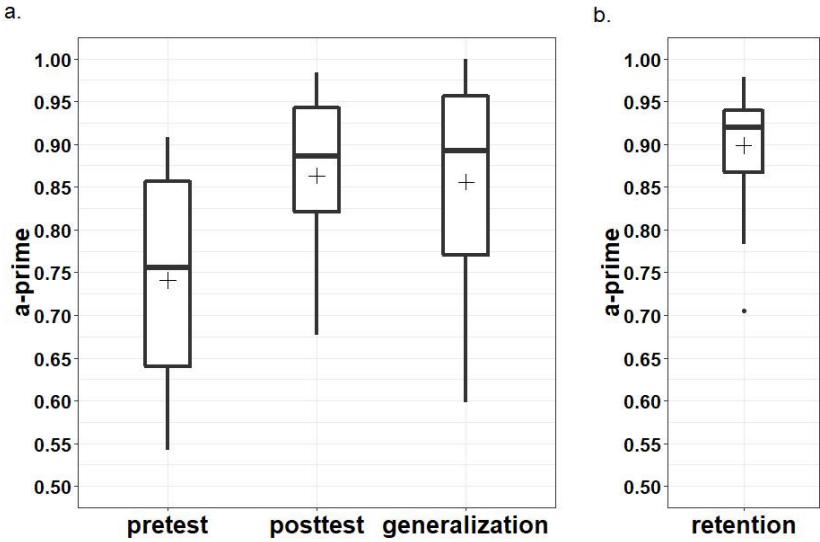
lexical decision tasks. The grey shading represents confidence intervals.

Table 1. Average duration (ms) and intensity (dB) of the sound /h/, as well 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).

Task	Speaker	Condition	Duration (ms)		Intensity (dB)	
			/h/	Ratio: h/hV	/h/	Ratio: h/hV
TEST:	Male1		59.6	0.40	56.7	0.45
Identification			(2.6)	(0.02)	(1.1)	(0.01)
in nonwords	Fem2		54.2	0.25	55.9	0.44
			(1.6)	(0.01)	(1.3)	(0.01)
<i>t-test</i>			<i>NS</i>	<i>***</i>	<i>NS</i>	<i>NS</i>
TEST:	Male1	Words	93.3	0.52	55.8	0.44
Lexical			(3.4)	(0.01)	(0.3)	(0.002)
decision		Nonwords	111.1	0.56	55.8	0.44
			(2.9)	(0.01)	(0.3)	(0.002)
<i>t-test</i>			<i>***</i>	<i>*</i>	<i>NS</i>	<i>NS</i>
TRAINING:	Male1		147.9	0.56	51.5	0.42
Identification			(3.4)	(0.02)	(0.4)	(0.001)
in real words	Male2		97.0	0.45	48.1	0.45
			(4.0)	(0.02)	(0.9)	(0.01)
	Fem1		107.1	0.36	53.09	0.42
			(3.1)	(0.02)	(0.4)	(0.003)
	Fem2		93.9	0.34	46.02	0.40
			(2.8)	(0.02)	(0.5)	(0.004)

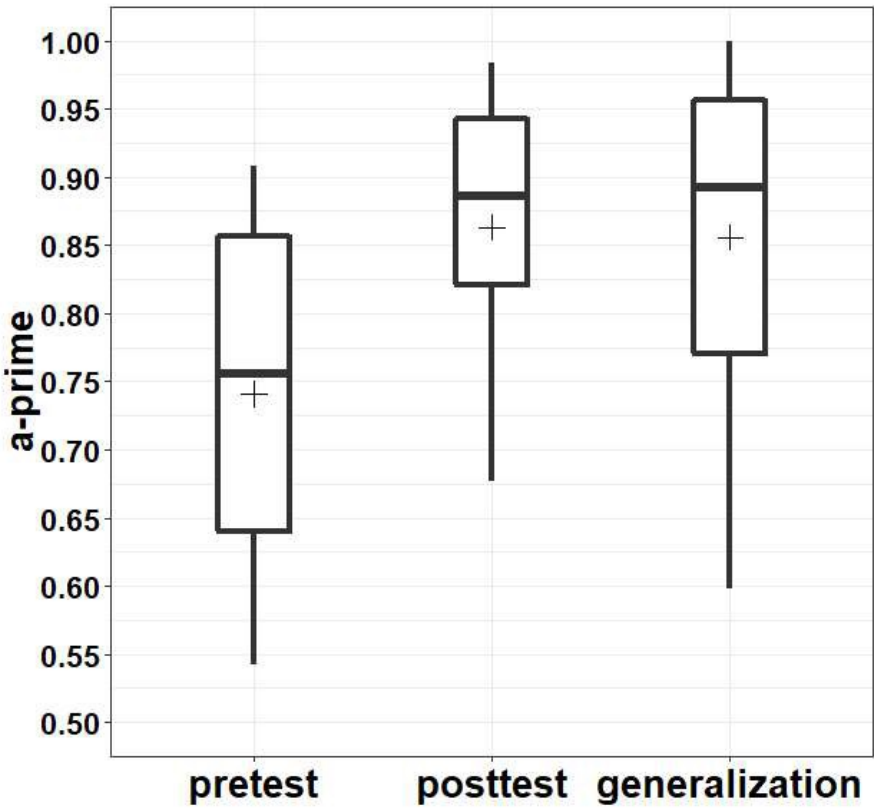
\*:  $p < .05$

\*\*\*:  $p < .0001$





a.



b.

