



## Issues in L2 phonological processing

Gerda Ana Melnik

### ► To cite this version:

Gerda Ana Melnik. Issues in L2 phonological processing. Linguistics. Université Paris sciences et lettres, 2019. English. NNT : 2019PSLEE007 . tel-02304656

**HAL Id: tel-02304656**

**<https://theses.hal.science/tel-02304656>**

Submitted on 3 Oct 2019

**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.



**THÈSE DE DOCTORAT**  
**DE L'UNIVERSITÉ PSL**

Préparée à l'Ecole Normale Supérieure

**Issues in L2 phonological processing**

Questions sur le traitement phonologique en langue seconde

Soutenue par

**Gerda Ana Melnik**

Le 19 juillet 2019

Ecole doctorale n° 158

**Ecole Doctorale Cerveau**  
**Cognition Comportement**

Spécialité

**Sciences Cognitives**

**Composition du jury :**

Jean-Luc, Schwartz  
Directeur de recherche,  
Université Grenoble Alpes,

*Président du jury*  
*Rapporteur*

Sophie, Dufour  
Chargé de recherche,  
Aix-Marseille Université

*Rapporteur*

Mirjam, Broersma  
Associate Professor,  
Radboud University

*Examineur*

Sharon, Peperkamp  
Directeur de recherche  
Ecole Normale Supérieure

*Directeur de thèse*

**Ecole Normale Supérieure**  
**Département d'Etudes Cognitives**

# **Issues in L2 phonological processing**

**Questions sur le traitement phonologique en langue seconde**

**Gerda Ana Melnik**

Thèse de doctorat dirigée par Sharon Peperkamp  
Laboratoire de Sciences Cognitives et Psycholinguistique  
CNRS-EHESS-ENS



# Abstract

Learning a foreign language (L2) is a difficult task, requiring considerable amounts of time and effort. One of the challenges learners must face is the processing of sounds that do not exist or are not used contrastively in their native language. The mismatch between the properties of the native language and the foreign one leads to distortions in the perception of non-native sounds and to foreign accent in their production. Moreover, these difficulties persist across levels of processing as problems in prelexical L2 sound perception and production influence the processing of words containing these sounds. Fortunately, with growing proficiency the abilities to perceive and produce L2 sounds gradually improve, although they might never attain native-like levels. This thesis focuses on L2 phonological processing and its development across modalities (perception vs. production) and across levels of processing (prelexical vs. lexical).

In the first part of the thesis, we investigate the relationship between perception and production in L2. Previous literature has provided contradictory evidence as to whether perception and production develop in parallel. We hypothesized that several methodological limitations could have brought confounds in some of these previous studies. We therefore designed an experiment that addressed these methodological issues and tested proficient English learners of French on their perception and production of the French contrast /u/-/y/ that does not exist in English. We included tasks that tap into both prelexical and lexical levels of processing in order to examine whether the link between the two modalities, if any, holds across levels of processing. Results showed that perception and production were correlated, but only when tested with tasks that tap into the same level of processing. We next explored if the developments in one modality precede developments in the other and found that good perception is indeed a prerequisite for good production.

In the second part of the thesis, we continue to investigate the phonological processing of L2 across levels by focusing on the perception of the English sound /h/ by intermediate to proficient French learners of English. We first studied if the poor perception of this sound previously reported at the prelexical level also causes problems at the lexical level. We also looked at whether asymmetries found in production (i.e. more deletions than insertions) are reflected in perception. The results revealed that French learners of English have difficulty in

perceiving /h/-initial words and non-words at the lexical level. Moreover, an asymmetry was indeed observed in their performance, which was interpreted as an indication that French learners of English have imprecise phonological representations of /h/-initial but not of vowel-initial words. Second, we carried out a training study to test if phonetic training could improve the perception of /h/ not only at the prelexical, but at the lexical level as well. We found that the High Phonetic Variability training did improve the perception of /h/ both at the prelexical and lexical levels, and that this positive effect was retained four months after training. Finally, we examined if asymmetries in the perception of /h/ at the lexical level could be explained by asymmetries at the prelexical level. The results revealed no such relationship.

Overall, this thesis demonstrates the complex and dynamic nature of the mechanisms underlying non-native speech processing and its development during learning both across modalities and across levels of processing. We discuss how future research could further explore the links between these elements of the phonological processing apparatus to get a better understanding of L2 acquisition.

# Résumé

L'apprentissage d'une langue étrangère nécessite une quantité considérable de temps et d'efforts. Les apprenants doivent faire face à de nombreux défis dans cet apprentissage, dont le traitement des sons qui n'existent pas dans leur langue maternelle. La différence entre les propriétés de la langue maternelle et de la langue étrangère entraîne des distorsions dans la perception et un accent dans la production des sons non-natifs. De plus, ces difficultés persistent à tous les niveaux de traitement, car les problèmes de perception et de production d'un son influencent le traitement des mots contenant ces sons. Heureusement, la capacité à percevoir et à produire les sons de la L2 (langue seconde) s'améliore progressivement. Cette thèse porte sur le traitement phonologique de la L2 et son développement à travers les modalités (perception vs. production) et les niveaux de traitement (niveau prélexical vs. lexical).

.Dans la première partie de la thèse, nous étudions la relation entre la perception et la production en L2. Les résultats des études précédentes ont souvent été contradictoires et nous suggérons que plusieurs limitations méthodologiques aient pu y créer des confusions. Nous avons donc pris en compte ces limitations méthodologiques et nous avons développé un paradigme expérimental afin de tester la perception et la production du contraste français /u/-/y/ par des apprenants anglophones. Nous avons utilisé des tâches qui visent le traitement prélexical et lexical afin d'examiner si le lien entre les deux modalités, s'il en existe un, est maintenu à travers les niveaux de traitement. Les résultats ont montré que la perception et la production sont corrélées, mais uniquement au niveau prélexical. De plus, nous avons trouvé que le développement de la perception précède celui de la production car il faut d'abord bien percevoir un son non-natif afin de le produire correctement.

Dans la deuxième partie, nous avons poursuivi l'étude du traitement phonologique à travers les niveaux de traitement en nous concentrant sur la perception du son anglais /h/ par des apprenants francophones. Nous avons d'abord examiné si les difficultés à percevoir ce son précédemment signalées au niveau prélexical posaient également problème au niveau lexical. De plus, nous avons examiné si l'asymétrie observée dans la production (les francophones omettent le /h/ plus souvent qu'il ne l'insèrent) était présente dans la perception. Les résultats ont révélé que les apprenants francophones ont du mal à percevoir des mots et des non-mots contenant le /h/. De plus, une performance asymétrique a été observée. Nous

avons interprété ceci comme une indication que les représentations phonologiques des mots anglais contenant le /h/ sont imprécises chez les apprenants francophones. Dans un second temps, nous avons examiné si un entraînement phonétique pouvait améliorer la perception du /h/ non seulement au niveau prélexical, mais également au niveau lexical. Nous avons démontré que l'entraînement phonétique améliorait la perception du /h/ dans les deux niveaux de traitement. De plus, cet effet positif a été maintenu quatre mois après l'entraînement. Enfin, nous avons examiné si les asymétries dans la perception du /h/ au niveau lexical pouvaient s'expliquer par des asymétries au niveau prélexical. Un tel lien n'a cependant pas été observé dans les résultats.

Dans l'ensemble, cette thèse démontre que les mécanismes sous-jacents au traitement de la parole en L2 sont complexes et dynamiques, et influencent ainsi la perception et la production tant à travers les modalités qu'à travers les niveaux de traitement. Enfin, des pistes pour les recherches futures, qui permettraient d'explorer davantage les liens entre ces éléments du traitement phonologique, sont proposées. Cela mènerait à une compréhension plus approfondie des processus impliqués dans l'acquisition de la L2.



## **Dedication**

To my Mum Danguolé, who has always been my greatest inspiration in life and research.

# Acknowledgments

I always thought that writing a dissertation is the hardest part of a PhD. And now I realize that writing acknowledgments is probably even harder – scientific writing skills do not help much when you have to thank so many great people in so few lines...

Here is a modest attempt to put my gratitude on paper.

To begin, I would like to thank Sharon, who first adopted me as a little intern coming from humanities five years ago. Not only did she teach me a lot, but she always believed in me a little more than I did, pushing me to learn things I would have never thought I was able to learn. Sharon was always there for me, but I had all the independence I needed when I could figure out things on my own. Most importantly, over these years the little intern from humanities learned to learn, to adapt, to question, to write papers and to use a variety of crazy methods, making her more confident and more and more passionate about science. I am endlessly grateful, Sharon, for providing me with all this baggage, I will carry it with me to my future destinations in research and life.

Because this PhD would have been impossible without me being at the ENS and having done the Cogmaster, I would like to thank Benjamin Spector who helped me from my very beginnings at the ENS. Being normalienne allowed me to discover a hidden world of intellectual richness and culture and to fully enjoy those years in the heart of Paris. I am also grateful to Franck Ramus and all the team of the Cogmaster teachers who inspired me tremendously and gave me the will to make cognitive science a part of my life.

Of course, I thank all the members of the LSCP for their support, scientific stimulation and friendship. Thank you Anne Christophe, for being such a kind director and creating such a family-like environment in the lab. I am very grateful to Alex Cristia, my former teacher and lab tutor, for all the scientific support and smart advice you gave me.

Merci Isabelle Brunet, pour ton aide avec les participants et pour les conversations qui ont rendu les heures de test au sous-sol bien plus agréables. Merci à Vireack UI et Michel Dutat de votre aide et de votre bonne humeur. Merci à Radhia Achheb de m'avoir dépannée avec tous ces monstres de l'administration française. Et merci à Anne-Caroline Fiévet, d'avoir toujours été aussi gentille avec moi.

Alice Latimier, merci d’avoir été une vraie amie. I am also very happy and grateful to have shared one office and many amazing conversations with Matthieu Koroma, Benjamin Rebouillat, Jean-Maurice Leonetti and Nicolas Barascud. Thank you to all the remaining members of the PhD-family: Camilia Scaff, Ava Guez, Milica Denic, Monica Barbir, Georgia Loukatou, Charlotte Vandendriessche, Rahma Chaabouni, Si Berrebi, Hualin Xiao, Amir Anvari and others. Thank you Mireille and Naomi Havron for being the “big sisters” whom I approached many times to ask for advice. Rory Turnbull, thank you for helping me to overcome my fear of statistics. I would also like to thank former members of the lab – Alex Martin, Alex de Carvalho, Julia Carbajal, Adriana Guevara Rukoz, Yue Sun who will remain for me the respectable “senior PhDs” forever.

Thank you Maria Giavazzi for this first amazing and enriching teaching experience I had with you.

I am also grateful to many more people in the lab, in the DEC and the ENS – I will probably have to write another thesis to put all the remaining acknowledgements!

Finally, I would like to thank my family and family friends. Un énorme merci à la famille Laumoniers qui m’ont encouragé et soutenu tout au long de mon aventure parisienne. Vous aviez été et vous serez toujours pour moi ma deuxième famille.

Merci à ma maman, Danguolė Melnikienė, d’être la meilleure maman et la meilleure chercheuse qu’on puisse imaginer. Merci à mon papa, Jaroslav Melnik, de m’avoir appris à creuser plus loin et à croire en moi. Thank you to my little brother, Alexandras Melnik, for having discussed with me all my scientific projects whenever I was stuck - I truly believe you will make an outstanding researcher. Thank you to my grandparents, Ona Juozapavičienė, kad visada mane išklausei ir palaikėi, and Йосип Мельник, що навчив берегти добрий настрій і співати, коли важко. Et finalement, un gigantesque merci à Maxime Leroy de m’avoir soutenu psychologiquement et gastronomiquement pendant toute ma thèse et surtout pendant la rédaction – je ne l’aurais pas finie sans tes encouragements.

# Publications

## Journal Articles

- Melnik, G. A., Peperkamp, Sh. (2019). Perceptual deletion and asymmetric lexical access in second language learners. *Journal of the Acoustic Society of America*, 145(1), EL13 – EL18. doi: 10.1121/1.5085648
- Di Liberto, G. M., Wong, D.E., Melnik, G. A., de Cheveigné, A. (2019)<sup>1</sup>. Cortical responses to natural speech reflect probabilistic phonotactics, *NeuroImage*, 196, 237 – 247. doi: 10.1016/j.neuroimage.2019.04.037

## Conference Proceedings

- Melnik, G. A., Peperkamp, Sh. (to appear, 2019). Online Phonetic Training Improves L2 Word Recognition, *Proceedings of the Annual Conference of the Cognitive Science Society*.

## Manuscripts

- Melnik, G. A., Turnbull, R., Peperkamp, Sh. (submitted). On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u/-/y/ contrast. Submitted to: *Second Language Research*.
- Melnik, G. A., Peperkamp, Sh. (submitted). High Variability Phonetic training enhances L2 lexical processing: evidence from training French learners of English online. Submitted to: *Bilingualism: Language and Cognition*

---

<sup>1</sup> This paper is not included in the thesis, nor in the appendices, as its topic is not related to the topic of this dissertation.

## **Table of contents**

<b>Abstract.....</b>	<b>I</b>
----------------------	----------

<b>Résumé.....</b>	<b>III</b>
--------------------	------------

<b>Acknowledgments .....</b>	<b>VI</b>
------------------------------	-----------

<b>Publications .....</b>	<b>VIII</b>
---------------------------	-------------

<b>1 Chapter 1: Introduction.....</b>	<b>1</b>
---------------------------------------	----------

1.1 L2 phonological processing and its relationship to L1 .....	2
---	---

1.1.1 Models of L2 perception and production .....	2
--	---

1.1.2 Perceptual distortions in prelexical and lexical levels.....	5
--	---

1.1.3 Distortions in L2 speech production .....	9
---	---

1.2 Perceptual asymmetries .....	13
----------------------------------	----

1.2.1 Directional asymmetries in vowels .....	13
---	----

1.2.2 Directional asymmetries in consonants.....	15
--	----

1.2.3 The link between prelexical and lexical asymmetries.....	17
--	----

1.3 Training .....	24
--------------------	----

1.3.1 Classical HVPT procedure .....	24
--------------------------------------	----

1.3.2 Testing the robustness for “real-life processing” .....	25
---	----

1.3.3 New methods – ecologically realistic environments.....	27
--	----

1.4 Outline of the following chapters .....	29
---	----

<b>2 Chapter 2: The relationship between perception and production: the processing of French /u/-/y/ by English natives ..</b>	<b>30</b>
--	-----------

2.1 Introduction .....	30
------------------------	----

2.2 On the relationship between perception and production of L2 sounds: Evidence from Anglophones’ processing of the French /u/-/y/ contrast .....	30
---	----

2.2.1 Introduction .....	31
--------------------------	----

2.2.2	Methods .....	38
2.2.3	Results and discussion.....	40
2.2.4	General discussion.....	47
2.2.5	Notes.....	52
2.2.6	Appendix .....	54
2.2.8	References .....	58
2.4	Conclusion .....	65
<b>3</b>	<b>Chapter 3: Non-native sound perception across levels of processing: the perception of English /h/ by French natives.....</b>	<b>66</b>
3.1	Introduction .....	66
3.2	Perceptual deletion and asymmetric lexical access in second language learners – Melnik & Peperkamp (2019).....	67
3.3	The effect of phonetic training on L2 word recognition .....	74
3.3.1	Introduction .....	74
3.3.2	Methods.....	78
3.3.3	Results and discussion.....	84
3.3.4	General discussion.....	88
3.3.5	References .....	91
3.4	The relationship between prelexical and lexical asymmetries in L2 perception .....	96
3.4.1	Introduction .....	96
3.4.2	The phonetic properties of /h/ and predictions of existing models .....	99
3.4.3	Results from the training study .....	101
3.4.4	Discussion .....	104
3.4.5	Comparing results from lexical decision in the Melnik & Peperkamp (2019) study and the training study .....	107
3.5	Conclusion .....	110
<b>4</b>	<b>Chapter 4: General discussion .....</b>	<b>111</b>

4.1	The relationship between perception and production.....	112
4.1.1	Summary of empirical work.....	112
4.1.2	Further questions .....	112
4.2	Non-native sound perception across levels of processing .....	115
4.2.1	Summary of empirical work.....	115
4.2.2	Further questions .....	116
4.3	Insights from perception on the development of production.....	122
<b>Appendix A: The role of domain-general cognitive capacities in</b>		
<b>L2 phonological learning.....</b>		<b>124</b>
A.1	Introduction .....	124
A.2	Methods .....	125
A.3	Results .....	127
A.4	Discussion .....	129
<b>Appendix B: Online Phonetic Training Improves L2 Word</b>		
<b>Recognition – Melnik &amp; Peperkamp (to appear, 2019).....</b>		<b>132</b>
<b>References.....</b>		<b>139</b>

# Chapter 1: Introduction

Many of us have experienced the difficulty of learning a foreign language and the great amounts of effort this learning requires. Such difficulties occur not only at the levels of word-learning or sentence-comprehension, but also in dealing with sounds of the second language (L2). Namely, adult foreign speech is often accented. Moreover, a form of “accent” also exists in perception as we fail to accurately perceive some sounds or sound sequences of the L2.

A better understanding of why these problems occur, as well as whether and how they could be solved is of great importance both theoretically, to understand the cognitive mechanisms behind L2 phonological acquisition and processing, and practically, to improve foreign language teaching methodologies. This thesis addresses these questions in two parts. First, we consider the relationship between perception and production in L2 phonological processing and the nature of this relationship across levels of processing. Specifically, we test native English learners of French on the difficult French contrast /u/-/y/ at both the prelexical and the lexical level (Chapter 2). Second, we further investigate the perception of non-native sounds across levels of processing by focusing on the perception of the English sound /h/ by French learners of English. We examine the perceptual asymmetries that occur in the lexical processing of this sound and we test the effects of auditory training on inducing changes in the perception of /h/ across levels of processing. In addition, we study the relationship between prelexical and lexical asymmetries in /h/ perception (Chapter 3).

The current Chapter is divided into three sections. We will first review findings on the influence of the native language on the phonological processing of L2 and the implications this has on perception and production, separately. Second, we will go over previous research on perceptual asymmetries found at the prelexical level in vowel and consonant perception. We will then discuss how these prelexical asymmetries could relate to asymmetries found at the lexical level of processing. Third, we will give a brief overview of the most common training procedure, the High Variability Phonetic Training. We consider studies which prove its robustness for “real life processing” and those that apply it in more ecologically realistic environments.



## **1.1 L2 phonological processing and its relationship to L1**

### **1.1.1 Models of L2 perception and production**

A great number of studies have shown that adult speakers face major difficulty when producing and perceiving speech sounds that are not used in their native language (for reviews, see Piske et al. 2001; Sebastián-Gallés 2005; Dufour & Nguyen, 2008). This difficulty stems from the particularities of the language acquisition and processing mechanisms. In one of the early studies on L2 acquisition, Trubetzkoy (1939) hypothesized that our native language (L1) phonology acts as a “phonological sieve” and filters out those properties of the L2 speech signal which are not relevant to the phonological system of the native language. This has been confirmed by evidence from L1 acquisition studies, showing that although newborns are sensitive to phonological contrasts of any language (Eimas et al., 1987), speech perception becomes attuned to the contrastive sounds of the native language very early in life. For instance, Kuhl et al. (2006) tested 6 to 12 months old infants and found that performance increased for the native-language contrast during the first year, but it declined for non-native contrasts over the same period of time. This attunement to one's maternal language might result in distorted perception of foreign sounds that differ from the L1 sounds in some phonological characteristics. In L2 production, the problems caused by the attunement to one's native language are even more evident than in perception. Namely, foreign accentedness is one of the most salient features that accompanies foreign language learning in adulthood.

Although some researchers argued that the capacity to tune one's perceptual system to foreign sounds is limited to a critical period that ends at puberty (Lenneberg, 1967), much evidence later showed that even adults can – to a certain extent – modify their speech perception and production patterns (e.g., Flege et al., 1997). It has been suggested that instead of completely impeding L2 learning, L1 strongly influences this process (Birdsong, 2006). For example, Iverson & Evans (2007) tested if participants whose native languages (Spanish, French, German, and Norwegian) differ in vowel system size also differ when learning the English vowel system. Results showed that participants who have larger L1 vowel systems performed better on L2 English vowels in different perceptual tasks, such as vowel identification in quiet and in noise. In line with this finding, the best known models of non-native speech production and perception – the Perceptual Assimilation Model (PAM; Best, 1995) and the Speech Learning Model (SLM; Flege, 1995) assume that the success of

learning to produce and perceive non-native sounds depends on the phonetic similarities and dissimilarities between L1 and L2 segments. Phonetic similarity or dissimilarity is defined in terms of the articulatory and acoustic characteristics of the linguistically relevant speech sounds (Rojczyk, 2010). According to the SLM, identical sounds will be learned easily, new sounds will cause more difficulty, while similar sounds will create most perceptual problems and will be very hard to acquire. For instance, in perception, Flege et al. (1995) examined the identification of English liquids /ɹ/-/l/ by inexperienced and experienced Japanese learners of English. This contrast has been shown to be very difficult for Japanese learners, as Japanese has only one liquid /r/, which is acoustically somewhere in between the English /ɹ/-/l/ contrast. Perceptually the Japanese /r/ has been shown to be closer to the English /l/ than to the English /ɹ/ (Takagi, 1993; Hattori & Iverson, 2009). Thus, according to the SLM, the /l/ is a similar sound to Japanese /r/, and thus should be harder to learn than the “new” sound /ɹ/. Guion & Flege (2000) confirmed this hypothesis, as in their study proficient Japanese learners of English were able to discriminate English /ɹ/ from the Japanese /r/ at significantly above-chance rates, while they were unable to discriminate English /l/ from the Japanese /r/.

In production, a similar pattern was reported in Flege (1987). The study tested three groups of American English learners of French differing in their level of proficiency on their production of French phonemes /u/ and /y/. In accordance to the predictions of SLM, French /y/ is a “new” phone for native American English speakers as this sound does not exist in English and thus, it should be easy to learn. This is indeed what was found in the study: English learners of French had no major difficulty in producing /y/ accurately even at the beginning of the learning and only slightly differed from native French speakers. In contrast, the SLM predicts that French /u/ should be considered a “similar” sound as it is similar to the English /u/ and thus, it should be difficult to learn. The results of Flege (1987) confirmed this prediction, as even highly experienced learners of French did not produce French /u/ in a native-like manner, although the production accuracy increased across groups with language proficiency.

Even though the SLM predicts different levels of difficulty for different sounds of an L2, with sufficient exposure, new phonological categories and the corresponding mental representations can still be created as the learning mechanisms remain intact over the life span and thus can be used by L2 learners. This does not mean, though, that the accuracy on non-native sounds will be nativelike (Flege, 1995).

The Perceptual Assimilation Model (PAM), on the other hand, postulates that articulatory gestures are the primitives on which speech perception is based. L2 listeners

perceptually classify L2 sounds into categories depending on the degree of articulatory similarity perceived between the native and non-native sounds. This in turn determines how the L2 sounds are assimilated to native categories. According to the model, three scenarios can occur:

1. L2 sounds are perceived as existing L1 categories. Within this scenario, three situations might occur:
  - a. Two-Category Assimilation (TC): Both L2 sounds are perceived as exemplars of two distinct L1 categories, leading to excellent discrimination.
  - b. Category-Goodness Assimilation (CG): Both L2 sounds are perceived as tokens of a single L1 phoneme, but one of the L2 sounds fits better. This leads to intermediate accuracy in discrimination.
  - c. Single-Category Assimilation (SC): Both L2 sounds are perceived as equally good or bad tokens of a single L1 category, leading to poor discrimination.
2. L2 sounds are perceived as speech sounds not corresponding to any concrete L1 category. Within this scenario two situations are possible:
  - a. Uncategorized-Categorized assimilation (UC): One L2 sound is perceived as a good exemplar of an L1 category, while the other L2 sound is not categorized in terms of the L1 inventory. The discrimination is predicted to be very good as it involves distinguishing between a clear category and a sound that does not belong to it.
  - b. Uncategorized-Uncategorized (UU): Both L2 sounds remain uncategorized in terms of the L1. The discrimination of these sounds might vary in accuracy.
3. Non-Assimilable: L2 sounds will be perceived as a non-speech sound, which will result in good to excellent discrimination.

PAM proposes that improvement in L2 perception (and production) is possible, as listeners continue to refine their perception of speech gestures over time. Thus, L1 phonetic categories can evolve to incorporate the additional properties of the L2 categories. This can lead to slightly shifting phonetic details of the L1 and L2 versions of the phoneme compared to those of monolinguals of either language. Thus, despite the potential difficulties in L2 sound perception, experience with the L2 can trigger perceptual readjustments and improvements in perception.

### 1.1.2 Perceptual distortions in prelexical and lexical levels

The mismatch between the properties of the native language and the foreign one occurs at two levels: segmental/suprasegmental and phonotactic. Thus, either certain segments or suprasegments of the L2 do not exist in the L1, or certain sequences of L2 sounds are not allowed in the L1. In order to match the phonological constraints of the L1 such mismatches are “repaired” by the perceptual system. Three types of “repair” strategies have been attested: changing the illegal sound or sequence, deleting the disallowed sound or part of the sequence or inserting a sound to correct the disallowed sequence (the later strategy has been attested only in phonotactic repair) (Sebastián-Gallés, 2005; Davidson & Shaw, 2012). This results in perceptual distortions or “illusions”. We will first give a short overview of such “repair” processes at the prelexical level of processing.

At the segmental level, phonological change is probably the most widely attested type of illusions, that results in L2 sound assimilation (Flege et al., 1999; Hayes-Harb & Masuda, 2008; Escudero et al., 2012, etc.). It occurs when an L2 learner does not perceive the difference between two L2 sounds as both of those sounds are mapped to a single L1 category. One of the most studied examples is the case of Japanese learners who fail to perceive the difference between English liquids /ɹ/ and /l/ as Japanese has only one liquid /r/ (Goto, 1971; Hattori & Iverson, 2009). Concerning the phonotactic mismatches between L1 and L2, phonological change arises when one phoneme is perceptually changed into another to correct a phonotactically illegal sequence of sounds (Hancin-Bhatt, 1994; Brannen, 2002; Cutler et al. 2004, etc.). For instance, French native speakers have been shown to perceptually transform the Hebrew onset clusters /tʃ/ and /dʃ/, into onset clusters /kʃ/ and /gʃ/ (Hallé & Best, 2007). The second type of phonological illusions, epenthesis, occurs when an L2 learner perceptually inserts an illusory (epenthetic) phoneme in order to correct a sequence of segments that violates the phonotactic constraints of the L1 (Davidson & Shaw, 2012; Durvasula & Kahng, 2016; Guevara-Rukoz et al., 2017, etc.). For example, Japanese speakers perceive an illusory vowel /u/ in disallowed consonant clusters (Dupoux et al., 1999). Finally, the last type of repair strategies, deletion, has been scarcely studied. Deletion could arise as an alternative to change in order to “repair” an illegal segment. Similarly, it could be an alternative to epenthesis in cases where the L2 syllabic structure should be modified to match the L1 phonotactic constraints. These two uses of deletion are indeed attested in loanword adaption where deletion typically concerns less salient sounds or sounds in less salient positions (see Kang, 2011). As there seems to be a close parallel between the perceptual

“repair” processes happening in loanword adaptation and those found in speech perception (Peperkamp et al., 2008), it seems likely that deletion in loanword adaptation should as well have its counterpart in perception. One such example has been attested by Cho et al. (2008) who tested Korean learners of English on the perception of the English diphthong [ow], using a phonetic transcription task. The results demonstrated that participants misperceived [ow] as [o], suggesting that they perceptually deleted the second element of the diphthong. Similarly, Mah et al. (2016) carried out an EEG experiment on the perceptual deletion of /h/ by French learners of English and native English speakers. When in an oddball paradigm participants listened to series of “um” and “hum” syllables, MMN was detected only for native French speakers, but not for French learners of English. Thus, due to perceptual deletion the learners did not perceive the difference between the presence vs. absence of /h/. Concerning phonotactics, an example of deletion is attested in Thai, where vowel-adjacent liquids delete, as such sound sequences are disallowed in this language (Yun, 2014).

Importantly, these distortions occurring in non-native sound perception can also influence the perception of L2 words containing these sounds. Previous studies used a variety of tasks (lexical decision (Darcy et al., 2013); lexical decision with long-term repetition priming (Pallier et al., 2001); cross-modal priming (Broersma & Cutler, 2011); word identification (Diaz et al., 2012); eye-tracking (Cutler et al., 2006); semantic relatedness judgment (Ota et al., 2009)) to show that perceptual problems at the prelexical level can severely impair the processing of the non-native language at the lexical level. Hence, the three perceptual repair strategies observed at the prelexical level of processing have also been found to surface at the lexical level. Most of the studies that looked at how perceptual distortions affect lexical processing focused on perceptual assimilation. For example, Ota et al. (2009) investigated how the perceptual assimilation of English /ɪ/-/I/ that has been reported in Japanese learners at the prelexical level affects their lexical processing of English minimal pairs that only differ in /ɪ/-/I/. They used /ɪ/-/I/ minimal pairs (e.g., ROCK – LOCK) and /p/-/b/ control minimal pairs (e.g., PEACH–BEACH) in a visual semantic relatedness judgment task. In this task participants saw in each trial two written words (e.g., KEY – LOCK) and had to judge whether they are semantically related or not. Japanese participants made significantly more errors when the trial contained a word involving the sound /ɪ/ or /I/; i.e. when they saw a pair such as KEY – ROCK, they wrongly answered that these words are related (this suggests that the word LOCK was activated when seeing ROCK). This is an indication that Japanese learners have inaccurate phonological representations of English words containing /ɪ/ and /I/ which get activated despite the information provided by orthography.

A few studies tested the negative impact that the two other types of perceptual repair strategies can have on lexical access and word recognition. Regarding perceptual epenthesis, its effect on lexical processing was tested by Dupoux & Pallier (2001). Note, that instead of using stimuli from a foreign language they tested Japanese native speakers on Japanese words and nonwords. The participants performed, among other things, a lexical decision task which showed that they tended to accept nonwords such as *\*sokdo* but not *\*mikdo* as real words (cf. *sokudo* ‘speed’ and *mikado* ‘emperor’). This suggests that Japanese speakers perceptually inserted an illusory vowel /u/ in the nonwords, which for *\*sokdo* but not *\*mikdo* resulted in a real word. Thus, because of perceptual illusions, Japanese speakers do not perceive the difference between words and nonwords that differ only in the presence or absence of /u/ in a consonant cluster. A similar pattern of results was obtained by White et al. (2017) who studied the difficulty in lexical processing caused by the perceptual deletion of /h/ in French listeners. In an EEG study, French learners of English performed a semantic classification task on words and nonwords, where the nonwords were created from /h/- and vowel-initial words by removing or adding /h/, respectively. Crucially, the participants were not informed that the items contained nonwords as well as real words. Results revealed that low-proficiency learners did not show an N400 nonword effect; thus, they processed the nonwords as if they were real words, suggesting that the misperception of /h/ resulted in impaired lexical processing of h-initial words and nonwords.

Taken together, the studies reviewed in this section show that the perceptual repair strategies, which occur in order to “repair” L2 sounds that do not match the constraints of the L1, can have a great negative impact on perception both at the prelexical and lexical levels of processing.

### **1.1.2.1 Increasing perceptual difficulty with increasing levels of processing**

The strength of these perceptual problems might be modulated by experience and proficiency in L2 (Flege et al., 1997). However, within the same level of proficiency, performance on L2 perceptual tasks can also vary depending on the level of processing being tested. That is, in order to decode the incoming acoustic signal into meaningful words the listener has to succeed in accurately performing throughout several stages, starting from auditory processing, phonetic and phonological analysis, to word recognition and lexical access (Pisoni & Luce, 1987). Although under normal listening conditions the accuracy of native speakers of a given language is generally at ceiling across tasks that tap into different levels of processing, there is evidence that early bilingual speakers who succeeded in one task

might not succeed in the others. This was first tested by Sebastián-Gallés & Baus (2005) who focused on highly proficient early Spanish-Catalan bilinguals. They performed three perceptual tasks involving the Catalan contrast /e/-/ɛ/ which does not exist in Spanish. The tasks used in this experiment were chosen such as to test the robustness of a range of phonological representations: a categorization task with isolated and synthesized stimuli, a gating task, and a lexical decision task. The results showed that for Spanish-dominant participants the performance on the tasks was increasingly difficult. While many bilinguals (68%) reached native-like accuracy in phonological categorization, only a few of them (18%) reached this level in the lexical decision task. In a further study, Díaz et al. (2012) tested if the same effects can be observed in late learners of an L2. Dutch L2 learners of English were tested on their processing of the English /æ/-/ɛ/ contrast with tasks that tap into different levels of processing: categorization, lexical decision and word identification tasks. As in Sebastián-Gallés & Baus (2005), they found that many more participants succeeded to perform at a nativelike level in the phonetic task (categorization) than in tasks that involve lexical processing (lexical decision and word identification). This is likely due to the fact that different tasks that involve different levels of processing levels, require different skills.

Another set of studies by Werker and her colleagues (Werker & Tees 1984b; Werker & Logan 1985) demonstrated similar results by using one task but variable ISIs (Interstimulus Intervals) to tap into different levels of processing. Werker & Logan (1985) implemented an AX discrimination task to test native English speakers on a difficult retroflex/dental contrast, which does not exist in English. They used three ISI conditions (250 ms, 500 ms and 1500 ms) and hypothesized that variable memory demands and cognitive load in the task will trigger different processing levels. Thus, an ISI of 250 ms would tap an auditory-acoustic level, one of 500 ms a phonetic level, and one of 1500 ms a phonological level. The results confirmed their predictions, as participants performed differently depending on the ISI condition. Importantly, in the shortest 250 ms ISI condition, native English speakers could discriminate the difficult contrast, as well as within-category phonetic differences. With longer ISIs, however, their performance decreased, suggesting that a task becomes more difficult when it taps into a higher level of processing.

In accordance with these results, several other studies also showed that even the hardest non-native contrasts can be perceived at the low acoustic level. For instance, Dupoux et al. (1997) demonstrated that naïve French listeners can discriminate between Spanish stimuli that vary only in the position of stress in a low-level task, i.e., AX discrimination. However, as soon as the tasks tap into higher order of processing, such as ABX discrimination



(Dupoux et al., 1997), sequence recall, or lexical decision (Dupoux et al., 2008), French participants exhibit “stress deafness”, or inability to perceive stress contrasts.

For certain contrasts the difficulty in performing accurately in more complex tasks involving lexical access can persist even after many years of practice. Pelzl et al. (2018) tested highly proficient learners of Mandarin Chinese on several tasks involving Mandarin tones. While L2 speakers were very accurate at tone identification in isolated syllables (experiment 1), they performed extremely poorly on lexical decision (experiment 2) compared to native speakers of Mandarin. Moreover, Pelzl et al. (2018) carried out a third experiment using EGG, where participants performed a sentence judgment task on sentences containing disyllabic real words or tonal nonwords, depending on the condition. Results showed that L2 learners did not show N400 pseudoword effect when hearing sentences with nonwords, suggesting that they did not perceive the difference between words and nonwords.

However, there is evidence that for some contrasts accuracy across levels of processing can improve with proficiency, although the speed and size of the improvement will not necessarily be the same at different levels of processing. Darcy et al. (2013) carried out a set of experiments testing American English learners of Japanese on the Japanese contrast between singleton and geminate consonants, and American English learners of German on the German contrast between front and back rounded vowels. They tested intermediate and advanced learners as well as native speakers of each language in an ABX discrimination task and a lexical decision task. In both language settings, all participants performed with high accuracy in the ABX task, so that the performance of learners and native speakers did not differ. According to the authors, this is an indication that the phonological categorization of a hard contrast can be learned to a nativelike level. The pattern of results was, however, very different in the lexical decision task. Groups differed significantly, and only native speakers performed in consistence with their performance in the ABX task. All groups of learners made much more mistakes in this task, compared to the ABX. Nevertheless, proficient learners had significantly better results than intermediates, pointing to the possibility to improve one’s perception even at the lexical level of processing.

### **1.1.3 Distortions in L2 speech production**

Differences in L1 and L2 phonological inventories also impact L2 speech production and result in perceived foreign accent. Foreign accent refers to accent at the segmental level and to global accent at the sentence or utterance level (Riney & Flege, 1988). The difficulty in



L2 production can arise in two, not mutually exclusive, ways: first, some sounds can be perceived inaccurately, which can lead to wrong production. Second, the pronunciation of certain sounds requires to use some articulators that are not used in the production of L1 sounds, thus resulting in motor difficulty. Concerning the first of these reasons, the perceptual problems described in section 1.1.2. are mirrored in inaccurate production. For instance, the “repair” strategy consisting in perceptually changing one sound into another results in confusions of English /ɪ/ and /I/ in the productions of Japanese speakers. Flege et. al. (1995) tested two groups of Japanese learners of English on English words containing the sounds /ɪ/ and /I/ in a reading and a spontaneous speech task. The authors found that /ɪ/ and /I/ tokens produced by inexperienced Japanese learners were often misidentified by native English judges, whereas productions of advanced learners were much more accurate and did not differ significantly from that of native speakers’. This suggests that the perceptual problems encountered with this difficult contrast were transferred to production. With extensive experience, however, these problems can be overcome. Similarly, the illusion of epenthetic /u/ in consonant clusters, observed in the perception of Japanese learners of English, has its counterpart in production. Masuda & Arai (2008) tested monolingual Japanese speakers and proficient Japanese speakers of English on the production of nonwords containing consonant clusters. They found that both less and more proficient groups of speakers inserted a vowel /u/ in English consonant clusters, but the rate of insertions was much higher in monolingual Japanese speakers (in 80 % of items) than in highly proficient speakers of Japanese (in 12% of items), indicating the possibility of improvement with raising proficiency. Finally, deletions in production have been reported by Janda & Auger (1992) who showed that in English conversation French speakers of English delete /h/ from 5 up to 55% of the time, depending on the speaker. This mirrors patterns in /h/ perception, where French learners do not perceive the difference between the presence and absence of /h/ in English stimuli. Moreover, there is evidence that French learners sometimes use hypercorrection strategies and insert an /h/ in the wrong place (Janda & Auger, 1992; John & Cardoso 2008). This points to the fact that the difficulty with this sound in production stems from imprecise perception, not from the articulatory complexity of /h/, as French learners of English are capable of producing this sound accurately.

Turning to the articulatory difficulty with L2 sounds, pronunciation is considered to be the only “physical” aspect of language that involves complex neuromuscular demands (Scovel, 1988). Therefore, speech production is most affected by physiological limitations, compared to speech perception, morphology or syntax (Simmonds et al., 2011b). Moreover,

learning to pronounce an L2 sound that does not exist in the native language requires to retune the neural circuits involved in the motor control of articulation, which is necessary to perform rapid unfamiliar sequences of movements (Simmonds et al., 2011a). One example of such difficulty in pronunciation are click sounds, used in Bantu languages, such as Xhosa. Lewis et al. (1994) tested adult English learners of Xhosa on their production of clicks, and found that learners encountered major difficulties in articulating clicks sounds and differed significantly in intelligibility judgments from the native speakers of Xhosa. Note, that these problems in production are likely to be due to articulatory difficulty and not perceptual problems.

Although there are no studies on how English learners perceive clicks in Xhosa, Best et al. (1988) conducted a well-known study on the perception of clicks by English natives in another Bantu language, i.e., Zulu. They showed that English speakers can accurately discriminate between pairs of Zulu clicks, although these sounds do not exist in English or other Indo-European languages. This suggests that despite accurate perception, some L2 sounds cannot be produced accurately because of physiological constraints in articulation.

The level of accentedness can also depend on the amounts of exposure and experience with the foreign language. One of experience-related factors is the length of residence (LOR). For instance, Flege et al. (1997) tested speakers with different native languages (German, Mandarin, Spanish, and Korean) on the production of L2 English vowels. L2 learners were divided into experienced and inexperienced groups, depending on their length of residence in the USA. Results showed that experienced participants were more accurate in their productions of L2 vowels compared to the less experienced ones. The results suggest that L2 pronunciation can improve through practice. Another important factor, often considered in L2 studies, is the age of learning (AOL). It refers to the age at which the learner was first exposed to the L2. The general assumption is that the earlier the AOL, the better the outcomes of learning are. For example, Flege (1993) tested Chinese participants on the production of vowel length before word-final consonants /t/ and /d/ in English words. In such contexts English natives produce longer vowels before /d/ than before /t/, and thus vowel length becomes the cue to differentiate between /t/ and /d/ which sound the same due to final devoicing. Results revealed that the Chinese participants who arrived to the USA in adulthood significantly differed in their productions from native English speakers and Chinese participants who arrived to the USA before the age of 10. Thus, starting to learn a foreign language early can indeed lead to better accuracy in the production of the sounds of this language.

Finally, the degree of perceived foreign accent depends on a variety of other factors, such as gender, formal instruction, motivation, language learning aptitude, amount of native language (L1) use and communicative pressure (Piske et al. 2001).

To sum up, differences between the phonologies of the L1 and L2 might lead to distortions when perceiving L2 sounds. These distortions can affect perception of non-native sounds across levels of processing, these perceptual problems being more difficult to overcome at higher levels of processing than at lower ones. Similarly, mismatches between L1 and L2 phonological inventories result in foreign accent when speaking the L2. Depending on the cases, these difficulties in L2 production might stem from inaccurate perception and/or from articulatory constraints. Although much research has been conducted on L2 phonological processing, both in perception and production, in order to understand the underlying mechanisms behind the acquisition of L2 phonology, many questions remain unanswered. One of them is the relationship between perception and production within and across levels of processing. We will address this question in Chapter 2 of this thesis.

## 1.2 Perceptual asymmetries

### 1.2.1 Directional asymmetries in vowels

The perception of some non-native contrasts can cause more difficulties than the perception of others. Moreover, within one L2 contrast both sounds might be perceived with differing degrees of difficulty, such as in the perception of the English contrast /ɪ/-/I/, where Japanese learners perceive /ɪ/ better than /I/ (Hattori & Iverson, 2009). Interestingly, some such perceptual asymmetries have also been reported in L1 perception. A series of studies on L1 and cross-language vowel perception in infants showed that for many between-category vowel contrasts, the order of stimuli presentation may influence the discrimination accuracy, i.e., when a pair of vowels is presented in one direction, they are easier to discriminate than when they are presented in the opposite direction (for a review, see Polka & Bohn, 2003; Polka & Bohn, 2011). For example, Polka & Bohn (1996) showed that both German-learning infants and English-learning infants discriminated better between the German contrast /u/-/y/ when the contrast changed from /y/ to /u/, than in the reverse change from /u/ to /y/. Evidence from this and other studies on other languages revealed that these directional asymmetries are consistent cross-linguistically in that infants have better performance when the vowel changes from a more central to a more peripheral one in the F1-F2-F3 vowel space. Based on this observation Polka & Bohn (2011) framed a model of early phonetic development, the Natural Referent Vowel (NRV) model, which emphasizes the importance of peripheral vowels that serve as reference or perceptual anchors (Polka & Bohn, 2003) in the development of the native vowel inventory. The peripheral vowels act as referents because they are perceptually more salient due to formant frequency convergence, or focalization (Schwartz et al., 2005). As formants get closer to each other, the acoustic energy raises and increases the amplitude of each formant. In this way, a universal bias to attend to more salient peripheral vowels helps infants to perceive the differences between vowels and to consequently create stable categories.

The NRV hypothesizes that directional asymmetries in vowel perception reduce with experience with the phonological system of the L1. That is, the perceptual biases favoring peripheral vowels can be over-ridden in order to optimize native language perception. A study on Danish-learning infants conducted by Polka & Bohn (2011) supports this hypothesis. Six- to twelve-months-old infants divided into younger and older participants were tested on the British contrast /ɒ/-/ʌ/, as well as the Danish contrast /e/-/ɛ/. Results revealed that for the

British contrast, both groups of infants performed more accurately when the vowel changed from the more central /ʌ/ to the more peripheral /ɒ/. For the Danish contrast, all infants also showed a directional asymmetry as they discriminated better the change from /ɛ/ to /e/ than the reverse. However, the asymmetry in the older age group for the Danish contrast was much smaller than in the younger one. This suggests that as the perceptual system gets attuned to the L1, the bias to favor more peripheral categories fades because of L1-specific constraints. However, this bias is maintained for the non-native categories.

This has been further tested in experiments focusing on adult vowel perception. For instance, adult native speakers of English and German were tested on the perception of the German contrast /u/-/y/ (Polka & Bohn, 2011). Results showed that while German native speakers showed no directional asymmetry in the perception of the contrast, English native speakers perceived the contrast better when it changed from /y/ to /u/, compared to the reverse change from /u/ to /y/. Thus, adult German speakers did not show any asymmetry for their native contrast, suggesting that the universal bias for the referent vowel has been overridden (see above, Polka & Bohn, 1996). However, adult speakers of English maintained this bias as the contrast in question is not used in their L1.

Kriengwatanaa & Escudero (2017) further tested the assumptions of the NRV by investigating the role of experience on directional asymmetries in adult L2 speakers. Namely, they investigated if gaining experience in L2 can override the universal biases and help in reducing asymmetries in L2 sound perception, just like experience with the native language inventory in infancy helps to override the perceptual biases in L1 and results in diminishing directional asymmetries. Spanish learners of Dutch were tested and trained on the Dutch contrast /ɑ/-/a:/ which does not exist in English. The study showed that Spanish listeners had indeed an asymmetrical perception of the Dutch vowel contrast, as they performed better when the vowel changed from /a:/ to /ɑ/ (a central to a peripheral vowel), than in the other direction. Nevertheless, this asymmetry remained stable after training, although the categorization accuracy of participants improved significantly. The authors suggest that this results could be due to two reasons. First, it could be the case that the training and exposure to the vowel contrast was not sufficient to induce changes in the L2 learners they tested. The NRV does not predict explicitly what amount of exposure is required to overcome perceptual biases. Conversely, the universal perceptual biases could be modifiable only during infancy and remain intact in adult speech acquisition, despite increasing levels of proficiency of learners.

Even though asymmetries might not be modulated by experience in adult L2 speakers, it is possible that certain L2 contrasts have a higher probability to be perceived asymmetrically than others. Tyler et al. (2014) investigated this hypothesis by applying the PAM (Perceptual Assimilation Model) framework. According to them, the presence of the asymmetry in perception will depend on the perceptual assimilation of the L2 contrast in question. If the two L2 sounds are assimilated to two distinct L2 categories (Two-Category Assimilation) or one of the L2 sounds is assimilated to an L2 category, while the other sound of the contrast remains uncategorized (Uncategorized-Categorized), no asymmetry should be observed, according to PAM. If, however, the two L2 sounds are assimilated into a single L1 category, such as in Category-Goodness Assimilation or in Single-Category Assimilation, an asymmetry in perception might occur. To test these hypotheses Tyler et al. used an ABX discrimination task and tested American English native speakers on Norwegian, French and Thai contrasts. As predicted, results revealed no perceptual asymmetry for the TC and UC contrasts. An asymmetry was found for the SG contrast. However, contrary to predictions, there was no asymmetry for the CG contrast. The authors proposed that the asymmetry was not observed because the CG contrast was represented by too few data points in the experimental design, thus decreasing the power of the statistical analyses. Taken together, these findings suggest that not all L2 contrasts might be subjected to perceptual biases and this seems to depend on how the L2 contrasts are assimilated.

## **1.2.2 Directional asymmetries in consonants**

### **1.2.2.1 The Native Language Magnet Theory (NLM)**

Although one of the main focuses of the NRV are perceptual asymmetries, this model exclusively concentrates on vowel perception. Thus, in order to address the question of perceptual asymmetries in consonants we have to look into a different theoretical framework, namely, the NLM (Native Language Magnet model) (Kuhl, 1991; Kuhl et al., 2008). This model mainly focuses on the development of speech perception but it also addressed the question of asymmetries both in vowels and consonants. According to it, language abilities are innate and young infants have an ability to discriminate between sounds of any language. At this stage some directional asymmetries caused by universal perceptual bias can be observed. However, experience with multiple tokens of the L1 categories results in the formation of prototypes, or areas in the perceptual space that serve as “category centers”. These prototypes act as “perceptual magnets” by attracting new exemplars of the category

towards them and becoming a reference point for their generalization. This mechanism “warps” perception, as it diminishes the perceived distance of tokens that are close to the prototype, and increases the perceived distance at the edges of the categories. This enhances perception in the native language but can hinder perception of non-native sounds. More precisely, L2 sounds that fall close to the L1 prototype in the perceptual space will be attracted to it, making those L2 sounds less distinct. Thus, the more similar an L2 sound is to the L1 prototypes, the greater the native language magnet effect and therefore the poorer discrimination or the greater the difficulty to distinguish these sounds from L1 sounds.

Asymmetries occur when discriminating sounds which fall within an L1 category and which differ in level of prototypicality. First, the L2 sound that is closer to the L1 prototype will be harder to discriminate and to learn than an L2 sound that is less prototypical. Second, if in a discrimination task the more prototypical stimulus is played first in the pair, it activates the native category which acts as a perceptual magnet and attracts the second less prototypical sound which will be assimilated. If these sounds are played in the reverse order, the less prototypical sound does not activate the native category, resulting in less assimilation and better discrimination.

#### **1.2.2.2 Evidence on asymmetries in consonant perception**

Although a considerable number of studies looked into directional asymmetries in vowel perception, only a few of them investigated such asymmetries in consonants. For example, Tsushima et al. (2003, 2005) tested adult native speakers of Japanese on the English contrast /ba/-/va/. They reported that participants performed more accurately when discriminating the change from /v/ to /b/ than the change in the other direction. As Japanese has no sound /v/ and a sound /b/ that is phonetically different from the English /b/, the authors hypothesized that, similarly to referent vowels, consonants that are perceptually closer to a native phoneme might act as perceptual anchors. Thus, the English /b/ played the role of a referent sound in this experiment, causing the above described directional asymmetry.

In a recent study Nam & Polka (2016) tested 5-6 months-old English- and French-learning infants on the same /v/-/b/ contrast, which exists in both languages. The results revealed an asymmetry in the same direction as in Tsushima et al. (2003, 2005), namely, infants from both language groups noticed when /v/ changed to /b/ but not when the change occurred in the other direction. Nam & Polka suggested that the asymmetry occurred because stops are acoustically more salient than fricatives and thus, they serve as natural referents in the perception of consonant manner-of-articulation. However, Dar et al. (2018) reported a



different pattern of results after testing infants and adults on non-native consonant perception. In the first part of this study two groups of English-learning infants (7-months-old and 11-months-old) were tested on the Urdu affricate contrast /tʃ<sup>h</sup>/-/tʃ/ which does not exist in English. Results showed no asymmetry in the discrimination of younger infants, and a clear asymmetry in the performance of older infants as they discriminated the contrast only when the /tʃ<sup>h</sup>/ was presented first. The finding that only older children performed asymmetrically is not consistent with the predictions of the NRV, as both groups of children should show an asymmetry when perceiving non-native sounds. Furthermore, the second part of the study tested adult speakers of Urdu and English on the same contrast. Although the English natives performed worse on the contrast than the Urdu natives, asymmetries were not found in any of the groups. Again, the NRV would correctly predict no asymmetry in Urdu natives but wrongly predict an asymmetry in English natives. Thus, it is not clear whether these findings contradict the NRV, or whether in order to apply this model to consonant perception, it would require incorporating some modifications. Concerning the direction of the asymmetry, Dar and colleagues suggested that in accordance with the Native Language Magnet Model, /tʃ/ is a more prototypical category for English-learning infants as English has words starting with /tʃ/, and thus hearing a non-prototypical sound after the prototypical one blocks its discrimination. It is not clear what direction of the asymmetry should be predicted by the NRV, as it has not yet been defined how referents should be chosen in consonants. If they are chosen in terms of acoustic saliency, specific measurements should be picked to evaluate this salience. Nam & Polka (2016) suggested that Amplitude rise time (ART) could be used as a reliable cue to distinguish stops and fricatives. However, it is not clear how to compare two affricates. Thus, more research is need to better understand the nature and direction of asymmetries in consonant perception.

### **1.2.3 The link between prelexical and lexical asymmetries**

All of the above mentioned studies looked at directional asymmetries in vowel and consonant perception at the prelexical level of processing. Polka & Bohn (2011) investigated at which stage of prelexical processing (acoustic or phonetic-phonological) perceptual bias occurs. They found that asymmetries in the perception of the German contrast /u/-/y/ are observed in English native speakers only when tested in a long ISI condition. If, however, the ISI is reduced to 500 ms, the performance becomes symmetrical. As short ISI condition has been shown to tap into the acoustic level of processing, the authors conclude that the lack of



asymmetries in this condition suggest that asymmetries are not caused by bias in auditory processing. Instead, they stem from bias in phonetic-phonological processing.

If perceptual asymmetries occur while perceiving speech sounds, this might have an influence on how we recognize words that contain these sounds. As prelexical perception is closely linked to perception at the lexical level, one could expect to observe asymmetries at the lexical level as well. To our knowledge, no studies working in the NRV or NLM frameworks looked at the possible influence of asymmetries in prelexical perception on lexical processing. Though it is true that most of the initial studies testing asymmetries focused on preverbal infants, much of the recent research looked at adults without trying to investigate the effects of perceptual bias on lexical processing.

Besides the NRV and NLM models and research investigating early phonetic-phonological development, there is an ongoing discussion on asymmetries in the field of L2 phonological acquisition. Nevertheless, this literature focused almost exclusively on asymmetries in lexical processing. Namely, several studies reported that in lexical tasks L2 learners perform more accurately on one member of an L2 contrast than on the other (Weber and Cutler, 2004; Escudero et al., 2008; Cutler et al., 2006; Broersma and Cutler, 2011; Diaz et al., 2012; Zhang et al., 2012; Darcy et al., 2013). The theoretical questions raised by this literature aim at identifying the reasons explaining the occurrence of these lexical asymmetries: is the perception inaccurate and this leads to the activation of the wrong word? Or is the perception good but the representations of words in the lexicon are inaccurate, which leads to the activation of wrong representations? However, these studies did not look at prelexical asymmetries, and thus did not address the possibility that lexical asymmetries might occur because of or at least might be influenced by asymmetries in prelexical perception.

There thus seems to be a gap in research on this question. As no studies investigated asymmetries at different levels of processing on the same participants, we will briefly review several studies that found asymmetries at the prelexical level of processing and will compare them with data from other studies on the lexical level of processing of the same contrasts whenever such data is available. Although such a comparison cannot lead to robust conclusions as the prelexical and lexical perception were tested on different participants, who possibly differed in levels of proficiency, it can shed some light on the link or absence of it between perceptual bias in prelexical and lexical levels of processing. We will focus on two cases of difficulties with an L2 contrast: the perceptual problems of Japanese speakers with

the English /ɪ/-/I/ contrast and the difficulties of Dutch speakers with the English /æ/-/ɛ/ contrast.

### **1.2.3.1 Japanese on /ɪ/-/I/**

A number of studies looked at prelexical perception of the English /ɪ/-/I/ contrast by Japanese learners. In an identification task Hattori & Iverson (2009) demonstrated that Japanese learners perceive /ɪ/ (82% correct) more accurately than /I/ (58%) and they confuse /I/ with /ɪ/ a little more than /ɪ/ with /I/ (16% vs. 22%). The authors also show that the confusion of English /I/ and Japanese /r/ occurs more often (19% of the times) than between English /ɪ/ and Japanese /r/ (2% of the times). Finally, there was no significant difference between /ɪ/-/I/ and /r/-/I/ confusions, while the difference was significant between /ɪ/-/r/ and /r/-/I/. This suggests that /I/ was more strongly assimilated than /ɪ/ to the Japanese /r/.

Hattori & Iverson (2010) found that in identification Japanese speakers correctly identified English /ɪ/ and /I/ with similar accuracy (71 % vs 72 % of the times). However, in the discrimination of the F3 dimension they reported that Japanese speakers were significantly less accurate on /I/ than on /ɪ/.

Guion & Flege (2000) tested three groups of Japanese learners of English differing in proficiency level and a group of monolingual English speakers on the discrimination of English and Japanese contrasts. Although they did not test for asymmetry for the English /ɪ/-/I/ contrast, they looked at asymmetric assimilation of those English sounds. They showed that all participants performed poorly when discriminating between the English /I/ and the Japanese /r/, but the English natives and the more proficient groups of Japanese learners performed significantly above chance and significantly better than the low proficiency Japanese group when discriminating between the English /ɪ/ and the Japanese /r/.

Some studies also investigated the possibility to improve the perception of the /ɪ/-/I/ contrast by using training. Bradlow et al. (1997) tested Japanese learners of English on their identification of this contrast in a pretest and posttest after training. They found that participants performed better on /ɪ/ than on /I/, but the improvement from pretest to posttest was larger for /I/. Thus in posttest accuracy for /I/ approached accuracy for /ɪ/. However, a training study by Hazan et al. (2005) revealed no significant asymmetry neither in pretest (/I/ was correctly identified 59% and /r/ 61% of the times), nor in posttest (/I/ was correctly identified 79.0% and /r/ 76.2% of the times).

At the lexical level, only one study looked at the asymmetrical processing of English /ɪ/ and /l/ in Japanese learners. Cutler et al. (2006), tested Japanese learners of English in an eye-tracking paradigm. The test stimuli were English pairs of words (target and distractor). One member of each pair started with /l/, the other with /ɪ/, and the remaining part of the first syllable were similar in both words of a pair (e.g. *locker-rocket*). The results revealed an asymmetry: when the target played was *rocket*, Japanese participants looked longer at a distractor picture of a *locket* than at unrelated distractors, while this pattern did not occur in the reverse condition when the target word was *locket* and the distractor was a picture of a *rocket*. Thus, when the /ɪ/-initial words were played, the /l/-initial words got activated as competitors to the target words.

Overall, the prelexical perception of the English /ɪ/-/l/ contrast by Japanese speakers seems to be asymmetrical. They tend to perceive less accurately the English /l/, as it is more assimilated to the native Japanese category /r/. We do not know what predictions the NRV would make for this contrast, as we should measure the perceptual salience of both liquids. However, these findings are in line with the NLM, which predicts that /l/ is a closer sound to the native prototype /r/, and thus it will be harder to perceive than the English /ɪ/.

Concerning the lexical level of processing, whether the word played was /ɪ/- or /l/- initial, the words first activated in the mental lexicon were /l/-initial words. Thus, words starting with a more prototypical sound /l/, which is better assimilated to the native category /r/, are accessed more easily. If we compare asymmetries at the prelexical vs. lexical levels, the sound /l/ as a closer sound to the native Japanese category is harder to perceive than /ɪ/, however, it seems that in a lexical task the reverse asymmetry is observed: word recognition is more accurate when listening to /l/-initial words than when listening to /ɪ/-initial words.

### 1.2.3.2 Dutch learners on /ɛ/-/æ/

At the prelexical level of processing, Cutler et al. (2004) tested Dutch and English native speakers on their identification of a variety of English vowels and consonants. The authors did not provide statistical analyses for pairwise comparisons, however we can consider the performance means for each group for /ɛ/ and /æ/. Dutch participants confused /ɛ/ for /æ/ 25% of the times and /æ/ for /ɛ/ 34% of the times in VC position. In CV position, Dutch participants confused /ɛ/ for /æ/ 22% of the times and /æ/ for /ɛ/ 39% of the times. Even though we do not know whether these differences are statistically significant, a trend for an asymmetry can be observed in this data. Namely, /æ/ seems to cause more confusions than /ɛ/. Concerning native English speakers, their confusion rates were rather low as they

identified native phonemes. In VC positions they confused /ε/ for /æ/ 3% of the times and /æ/ for /ε/ 4% of the times. In CV position, English participants confused /ε/ for /æ/ 9% of the times and /æ/ for /ε/ 12% of the times. Thus, there is a slight asymmetry in natives' perception that goes in the same direction as for Dutch learners.

Escudero et al. (2012) tested the categorization of the English /ε/-/æ/ contrast by native speakers of two varieties of Dutch (North Holland and Flemish). They found that the performance of both groups of participants was asymmetrical, in that they performed more accurately on the vowel /ε/ than on the vowel /æ/ (/æ/ was confused with /ε/ in 32% (North Holland speakers) and 26% (Flemish speakers) of the tokens, while /ε/ was confused with /æ/ in only 9% and 8 % of the tokens).

At the lexical level of processing, the results are much more puzzling. Weber and Cutler (2004) used eye-tracking with a visual world paradigm to test English word recognition by highly proficient Dutch learners of English and a group of native English controls. The test stimuli were English nouns pairs (target and distractor) which had similar onsets that differed only in one vowel, i.e. /ε/ or /æ/ (for example, *pencil* – *panda*). The results revealed an asymmetry: when the target played was *panda* (containing the vowel /æ/, which does not exist in Dutch), Dutch participants looked longer at a distractor picture of a *pencil* (containing the vowel /ε/) than at unrelated distractors; conversely, this pattern did not occur when the target word was *pencil* and the distractor was a picture of a *panda*. Thus, only words containing English /ε/ were initially activated in Dutch listeners, independently of whether they heard /ε/ or /æ/. Weber and Cutler (2004) argue that this asymmetry stems from poor perception. According to them, lexical representations of words containing these difficult vowels are distinct. Nevertheless, as the English /ε/ has a close Dutch counterpart /ε/, while the English /æ/ has none, stimuli with both /ε/ or /æ/ are perceived as containing /ε/, leading to an initial activation of the /ε/-containing representation. There was no such asymmetry in native English participants.

However, when Dutch learners of English were tested in a lexical decision task no such asymmetry was found (Broersma & Cutler, 2011). The test words in the lexical decision task were English words containing the English vowel /ε/ or /æ/. Nonwords were created by replacing one vowel with the other (e.g., *lamp* --> *\*lemp*, *chest* --> *\*chast*). Although Dutch participants accepted nonwords as real words more often than English natives, there was no difference between the false alarm rate in /ε/- vs. /æ/-nonwords in either of the groups.

Furthermore, a subsequent study by Diaz et al. (2012) used the same stimuli for the lexical decision task as Broersma & Cutler (2011) and found an asymmetry in both native

Dutch and native English participants. Although they did not show analyses for words and nonwords separately, both groups responded more accurately to æ-type (*lamp*, \**lemp*) than to ε-type (*chest*, \**chast*). This finding goes in the opposite direction of the above described Weber & Cutler (2004) study, as the more prototypical and closer to native English category /ε/ caused more difficulties in lexical perception than the more distant /æ/.

Overall, Dutch learners seem to perceive the English contrast /ε/-/æ/ asymmetrically, as they have better performance on the more Dutch-like vowel /ε/ than on the vowel /æ/. From the perspective of the NRV, this is unexpected, as /æ/ is more peripheral than /ε/ in the vowel space. This is also unexpected from the perspective of the NLM, which would predict the /ε/ to be harder to discriminate than the /æ/, as it is a perceptually closer sound to the native Dutch category /ε/.

At the lexical level, the results are contradictory. Three studies showed different findings: from no asymmetry, to asymmetry to one direction and asymmetry to the opposite direction. Thus, if there is no asymmetry in lexical processing, it remains unclear why the prelexical bias disappears at the lexical level. On the other hand, if following findings of Weber & Cutler (2004) we consider that there is an asymmetry in lexical processing, it seems to go in the same direction as the prelexical bias. Namely, prelexically English /ε/ as a closer counterpart of a Dutch category is perceived better than English /æ/. Similarly, at the lexical level, word recognition was more accurate when listening to /ε/-words compared to /æ/-words. Nevertheless, this pattern is not consistent with the asymmetries in the above-described perception of English /ɪ/-/I/ by Japanese learners, where performance at the prelexical level was better on the less prototypical sound. Finally, the direction of the lexical asymmetry found by Diaz et al. (2012) (/æ/ > /ε/, better performance on the less prototypical sound) does not match the prelexical perceptual asymmetry in Dutch (/ε/ > /æ/, better performance on the more prototypical sounds), nor the lexical asymmetry in Japanese (/I/ > /ɪ/, better performance on the more prototypical sound). Of course, such comparisons across languages and across studies with different participants are not ideal. Nevertheless, they give an indication that the question of perceptual asymmetries is very complex and more studies are needed to shed more light on it.

To summarize, the reviewed literature shows that prelexical and lexical asymmetries are well documented perceptual phenomena found across languages. However, the nature of these asymmetries, as well as the relationship between asymmetries across levels of processing are not fully understood. In section 3.2 we will focus on the reasons for the occurrence of lexical asymmetries in L2 perception. We will reconsider the questions

previously raised by L2 literature on lexical access: are representations wrongly activated because of bad perception or are the representations wrong per se, and thus even good perception cannot guarantee accurate lexical access? In sections 3.3 and 3.4 we will investigate the link between asymmetries at the prelexical vs. lexical levels.

## **1.3 Training**

### **1.3.1 Classical HVPT procedure**

The perceptual problems encountered by L2 learners can be reduced with increasing levels of proficiency or by means of auditory training. In the early 90s a series of studies developed a training procedure called the High Variability Phonetic Training that has been shown to improve the perception of even the hardest non-native sounds (Logan et al., 1991; Lively et al., 1993; Lively et al., 1994; Bradlow et al., 1997). The effectiveness of this paradigm relies on several factors, such as the type of stimuli, the feedback and the task used. First, the HVPT uses different talkers and different phonetic environments in stimuli in order to create variability during training. It has been shown that variability in stimuli helps perceptual learning by enhancing robust category formation. For instance, Logan et al. (1991) trained Japanese native speakers on the English /ɪ/-/I/ contrast with stimuli produced by multiple speakers and involving several phonetic contexts. Results showed that after training participants improved in posttest compared to the pretest. Moreover, Lively et al. (1993) replicated this study and included a generalization task in posttest. They found that after training participants not only improved on already heard tokens, but they were also able to perceive novel words produced by an old and a new speaker. Conversely, in a second experiment they trained a group of participants on stimuli produced by only one talker and found that after this training participants improved on already heard stimuli but there was no generalization to new words produced by a new talker. Thus, listening to multiple talkers helps participants to get used to the variability present in the input and to create abstract representations of the L2 sounds.

Second, learning of the new categories is enhanced by the use of corrective feedback during training. Trial-by-trial feedback drives learners' attention to relevant phonetic cues in the stimuli. McCandliss et al. (2002) trained participants on either a paradigm with feedback or a paradigm without feedback. They showed that the group trained with feedback learned much better and faster. Moreover, only in this group did the improvements generalize to new stimuli. While most L2 perception training studies used simple feedback, that convey a written right-or-wrong message, all types of feedback are not equally effective. Namely, Lee & Lyster (2016) proposed that different types of feedback involve different cognitive processes (Lyster & Izquierdo, 2009) and thus they could have differential impact on learning. They compared four types of corrective feedback in an auditory training experiment and

demonstrated that auditory feedback that combined target (right) and nontarget (wrong) forms was the most effective. Such feedback allows learners to hear the right and the wrong words one after the other and notice the subtle but important differences between the two stimuli.

Third, the task used in training can also impact the learning outcomes. Most of the studies used either identification or discrimination tasks in training sessions. Following the studies by Strange & Dittmann (1984) and Logan et al. (1991), it was first believed that identification training is more effective than discrimination training. More precisely, the former study used discrimination training and found no generalization to untrained stimuli, whereas the later used identification and succeeded in triggering generalization to new stimuli. Nevertheless, these two studies also differed in many ways, such as the absence (Strange & Dittmann) vs. presence (Logan et al.) of variability in the stimuli. Subsequent studies who directly compared these two procedures in a controlled manner found no significant difference in the improvement of groups who underwent identification vs. discrimination training. For instance, Flege (1995) trained two groups of Mandarin native speakers on the identification or discrimination of English /t/ and /d/. He found that in both groups the improvement did not differ significantly and the effect of training generalized to untrained words. Furthermore, similar results were reported by Wayland & Li (2008) who trained native Chinese and native English speakers on Thai contrasts, using identification or discrimination tasks. They showed that both training procedures were similarly effective. Finally, Shinohara & Iverson (2018) trained Japanese speakers on the English /ɪ/-/I/ contrast on both identification and discrimination tasks. They found that following such mixed training the improvement in identifying the difficult L2 sounds was not larger than in studies which used only identification training. This suggests that the addition of both tasks does not increase the effectiveness of HVP training.

### **1.3.2 Testing the robustness for “real-life processing”**

Laboratory-based HVP training has been shown to be effective on vowels (Carlet & Cebrian, 2014; Lee & Lyster, 2016), consonants (Kim & Hazan, 2010; Shinohara & Iverson, 2018), tones (Wang et al. 1999; Wang, Jongman, & Sereno, 2003), and syllable structure (Huensch & Tremblay, 2015). Moreover, it gives rise to long-term retention of the new categories (Lively et al., 1994). These findings show that in speech perception, non-native speech sound categories can become more precise and native-like with training. Besides the above-described positive effects, it is important to test what is the relevance of laboratory



training for more naturalistic L2 learning and processing. For instance, Iverson et al. (2012) trained a group of French learners of English staying in England and a group of French learners of English staying in France in order to test if HVPT has additional advantages to natural exposure. They hypothesized that learners who stayed in England had already received much exposure to natural variability in their environment and thus they might benefit less from short training. However, results showed that both groups of participants improved to similar degrees. According to the authors, this suggests that learning during training is different from natural learning situations as in addition to providing variability, it drives the learners' attention to specific phonetic cues. Thus, auditory training can supplement natural learning and bring improvement even at higher proficiency levels.

Furthermore, if HPVT is helpful in enhancing the perception of difficult L2 sounds, can it also directly impact their production? Sakai & Moorman (2017) published a meta-analysis on 18 studies (retained after applying exclusion criteria) which looked at the effects of perception training on production. They reported medium-size improvement in perception ( $d = 0.92$ ,  $SD = 0.96$ ) and small improvement in production ( $d = 0.54$ ,  $SD = 0.4$ ) following perception training. This confirms that perception training can induce improvements in production without explicit production training. Moreover, a recent study by Lengeris (2018) tested if auditory training can improve pronunciation in more ecological conditions, i.e. in spontaneous speech. As in spontaneous speech no written input is provided, learners have to retrieve the correct words from their mental lexicon and construct a syntactically valid sentence, in addition to accurately pronouncing the non-native sounds. Thus, pronouncing sounds in spontaneous speech is a more difficult task compared to production tasks usually used in laboratory settings. Two groups of Greek native speakers were tested in pretest and posttest on the production of English vowel contrasts /i:/-/ɪ/, /æ/-/ʌ/, /ɑ:/-/ʌ/, /ɒ/-/ɔ:/ by means of a sentence reading and a spontaneous speech (picture description) task. Only one of the groups received 5 sessions of HVPT between pretest and posttest. Results showed that while the control group did not improve in posttest, the trained group improved significantly in both production tasks, with greater gains in reading.

Another question that has been raised is whether auditory training can actually improve speech processing in tasks which have more resemblance to “real-life” language processing conditions compared to the tasks usually used in pretest and posttest. For example, efficient speech perception requires listeners to filter out noise, as speech rarely occurs in silence. Moreover, negative effects of environmental signal distortion are greater for speech perception in a non-native than in the native language, and the size of this difference

correlates with proficiency in the non-native language (Bradlow & Alexander 2007; Zhang et al. 2014). Thus, if the effects of training are robust, it should improve learners' perception of a difficult L2 sound not only in silence, but in noisy conditions as well. This was tested by Lengeris & Hazan (2010) in a study on English vowel perception by Greek learners of English. Participants were divided into two groups, one did the pretest-posttest only, while the other one additionally received five training sessions. The pretest- posttest included identification of English vowels both in silence and in noise (multi-talker babble). Results showed that only the trained group improved in posttest and, crucially, this held for stimuli identification in quiet and in noise. This finding confirms that HPVPT can indeed enhance perception of difficult L2 sounds even under adverse listening conditions.

### **1.3.3 New methods – ecologically realistic environments**

In the past decade several methodological advances have been introduced in auditory training, especially aiming at creating more ecologically realistic environments for training. First, the classical HVPT has been administered at participants' homes through portable computer software or online servers, as opposed to in traditional well-controlled laboratory settings. For instance, in a number of studies (Iverson et al., 2005; Iverson & Evans, 2009; Lengeris & Hazan, 2010; Iverson et al. 2012), participants completed the training sessions on their own by using special software installed on their or a laboratory computer. Similarly, several free open-source software applications based on HVPT have also been developed to facilitate training experiment administration as well as to promote this technique in the language teaching community. For instance, the English Accent Coach (Thomson, 2012) is an application designed to improve English pronunciation by means of perception tasks involving elements of HVPT, such as high phonetic variability in the stimuli and corrective feedback. Similarly, Rauber et al. (2012) developed an application designed for speech perception testing and training. It involves identification and discrimination tasks, several languages (English, Portuguese, Spanish) and a possibility to use visual, auditory and audiovisual modes in the newer version. This software has been successfully used in training studies in laboratory settings (Rato & Rauber, 2015) and at home (Lengeris & Nikolaidis, 2014; Lengeris et al. 2018).

Moreover, in some studies training is being administered online (e.g., Motobashi-Saigo & Hardison, 2009; Huensch & Tremblay, 2015; Okuno & Hardison, 2016). Motobashi-Saigo & Hardison (2009) used audio-visual and audio-only Web-based training to enhance

the perception and production of Japanese geminates by English learners. Results showed that training administered online was effective. Moreover, in post-training interviews participants reported being motivated and satisfied by this type of training because of its flexibility and effectiveness. Authors note that although online training reduces the levels of control over the experiment, the user-friendly approach encourages participation and learning.

Finally, some studies used implicit training through game-based paradigms both to increase the ecological validity of the learning process and the motivation of participants (Lim & Holt, 2011; Liu & Zhang, 2016; Vlahou et al., 2019). For example, Lim & Holt (2011) designed a video game aimed at teaching Japanese learners to better discriminate between the English /ɪ/-/i/ contrast. Importantly, the purpose of the study was to find out if implicit unsupervised training could lead to improvements similar to the improvement induced by HVPT. The authors hypothesized that previous studies on implicit phonological learning had failed to reproduce the learning effects found in explicit training studies because they were based on unnatural tasks. In their paradigm, however, Lim & Holt did not ask participants to categorize sounds explicitly. Instead, the learners were encouraged to make associations between sounds and characters of the game. The goal of the game was to recognize and catch “good” aliens and “destroy” the bad ones. In this way the task was expected to induce learning that is more similar to learning under natural conditions: just like in real-life language processing, sounds in the game had a functional role and were strongly related with other perceptual and motor information (aliens had different shapes and movement patterns). Similarly, although the feedback was not explicit, better categorization led to higher scores in the game, which motivated participants to seek improvement. The results showed that this game-based paradigm induced learning and perceptual gains similar to those found in explicit training studies.

To summarize, research using HPVT has much advanced in the past decades. This paradigm has not only been tested for robustness, but it has also been adapted to more ecological testing conditions. Apart from the above mentioned “real-life” language elements tackled by training studies, such as speech perception in noise or accurate production in spontaneous speech, word recognition is an essential stage of naturalistic language processing. Thus, a robust auditory training paradigm should not only improve the perception of difficult sounds, but also improve the recognition of words containing these sounds. In section 3.3 of this thesis we will address this question by training French learners of English on the perception of the difficult sound /h/.

## 1.4 Outline of the following chapters

In Chapter 2 we examine the relationship between perception and production in L2, by focusing on Anglophones' processing of the French /u/-/y/ contrast. We present an experiment in which we use well-controlled tasks, allowing to obtain comparable measures of L2 perception and production. Furthermore, we compare perception and production accuracy across levels of processing. Finally, we investigate if good perception is a necessary prerequisite for good production, as suggested by one of the most influential L2 learning models, the Speech Learning Model (Flege, 1995).

In Chapter 3 we present three studies in which we further investigate the perception of non-native sounds across levels of processing by focusing on the perception of the English sound /h/ by French learners of English. In section 3.2 we first study whether perceptual difficulties with /h/ previously found in French learners of English at the prelexical level also persist at the lexical level. Furthermore, we examine if the asymmetry reported in the production of /h/ by French learners of English also persists in perception. Finally, based on the results, we discuss the causes of asymmetrical perception by reconsidering questions previously raised by L2 literature on lexical access: do asymmetries occur because of inaccurate perception, or rather, do they surface because the phonological representations of words in the mental lexicon are imprecise?

In section 3.3 we present an online phonetic training study, which looks at whether training French learners of English on the English sound /h/ at the prelexical level can enhance the perception of this sound both prelexically and lexically. We test participants both at the prelexical and lexical levels in pretest and posttest, using identification and lexical decision tasks. If in posttest participants improve on both tasks, this will suggest that auditory training can enhance both the prelexical and lexical levels of processing.

In section 3.4 we investigate the link between asymmetries at the prelexical vs. lexical levels. Specifically, by analyzing data from the training study (section 3.3) we look at whether asymmetries occur both at the prelexical and lexical levels and how they are related. We also investigate if asymmetries across levels of processing (if any), get modified by perceptual training. We end the section by comparing the findings on lexical asymmetries from the training study and from the experiment reported in section 3.2.

Finally, in Chapter 4 we revisit the findings of this thesis and discuss remaining questions and directions for further research.

# **Chapter 2: The relationship between perception and production: the processing of French /u/-/y/ by English natives**

## **2.1 Introduction**

This Chapter examines the relationship between perception and production in L2 phonological processing. Although a number of previous studies have addressed this question, their findings were often contradictory. We propose that one cause for these contradictions could lie in methodological issues when measuring perception and production accuracy. We will thus begin by reviewing previous literature on the perception-production link and discuss the methodological difficulties encountered in this research. We will further present an experiment testing the link between L2 perception and production, designed so as to address and overcome these main methodological problems. Furthermore, the paradigm chosen in this study will allow us to compare perception and production accuracy across levels of processing. Finally, additional analyses will be carried out to investigate if good perception is a necessary prerequisite for good production, as suggested by one of the most influential L2 learning models, the Speech Learning Model model (Flege, 1995).

## **2.2 On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u/-/y/ contrast**

This section constitutes the following manuscript: Melnik, G. A., Turnbull, R., Peperkamp, S. (submitted). On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u/-/y/ contrast. Submitted to: *Second Language Research*.

### **Abstract**

Previous studies have yielded contradictory results on the relationship between perception and production in L2 phonological processing. We reexamine the relationship between the two modalities both within and across processing levels, addressing several issues regarding methodology and statistical analyses. We focus on the perception and production of the

French contrast /u/-/y/ by proficient English-speaking late learners of French. In an experiment with a prelexical perception task (ABX discrimination) and both a prelexical and a lexical production task (pseudoword reading and picture naming), we observe a robust correlation between perception and production within but not across levels. Moreover, using a clustering analysis we provide evidence that good perception is a prerequisite for good production.

### **2.2.1 Introduction**

One of the difficulties for second language learners concerns phonological processing. It is well known that L2 sounds are hard both to perceive and to produce. Yet, L2 learners can learn to process such sounds, even though nativelike performance is very rarely achieved (for reviews, see Piske et al., 2001; Sebastián-Gallés, 2005). The extent to which L2 learners can acquire new sounds has been the topic of much research, as has been the question of the possible interaction of perception and production in this process. While a large literature has supported the idea that in the process of learning an L2 system, aptitude in perception precedes aptitude in production, other studies have shown the reverse effect. Furthermore, some studies question the very existence of a link between the perception and production of L2 sounds, as they have different underlying mechanisms. Finally, it is not clear whether perception and production are comparable across different levels of processing.

This decades-long debate on the perception-production link in L2 phonological processing remains of considerable importance both from a theoretical and a practical perspective, as it has consequences for models of speech processing on the one hand and for L2 teaching methods on the other hand. The aim of the present article is to shed new light on this issue by means of a study on the perception and production of the French vowel contrast /u/-/y/ by advanced English-speaking learners of French. We will start by reviewing previous work, and discuss methodological issues that might obscure the true relationship between perception and production in this type of research.

#### **2.2.1.1 Previous research**

A common assumption regarding L2 phonological processing is that learners cannot produce L2 sounds accurately without perceiving them well. For example, in the Speech Learning Model of Flege (1987, 1995), the production of L2 sounds depends on the perception of these sounds: L2 speakers can learn to produce a non-native sound only if they

have established *in perception* a new phonological category for it. Several experimental studies have yielded evidence in favor of this model. For instance, Flege (1993) found that experienced Taiwanese learners of English perceived the vowel duration cue to coda stop voicing in English as well as native speakers did, but failed to match the duration difference of native speakers in production. Focusing on beginning learners, Nagle (2018) examined the development over the course of one year of the perception and production of the Spanish /p/-/b/ contrast by native speakers of English. He found that improvement in perception preceded improvement in the production of Spanish-like VOT values for /p/ (but no relationship could be established with respect to the production of Spanish-like VOT values for /b/).<sup>1</sup> Some other studies, however, have demonstrated an effect opposite to the one predicted by the Speech Learning Model, namely that L2 speakers can have accurate production of a non-native contrast despite inaccurate perception. For example, Goto (1971) tested Japanese learners of English on their perception and production of English words containing /r/ and /l/ sounds, and found that even participants who achieved relatively high production accuracy still exhibited poor discrimination. Similar results were obtained by Sheldon & Strange (1982). Flege & Eeftink (1987) focused on Dutch learners of English, and observed a large increase in VOT during the production of English compared to Dutch voiceless stops but only by a small shift in the perceptual boundary between English voiced and voiceless stops. Bohn & Flege (1997), in a study of German L2 speakers' processing of the English vowel /æ/, also observed better production than perception. Other studies yet have obtained mixed effects, with more accurate perception for some sounds and more accurate production for others. For example, Hao & de Jong (2016) found that Korean learners of English show better perception than production of fricatives, but better production than perception of stops, suggesting that the L2 perception-production link is not monolithic.<sup>2</sup>

The link between perception and production has also been examined in training studies. In conformity with the Speech Learning Model, several of these studies show that specific perception training can result in improvement not only of the perception of the trained contrast but also of its production (Bradlow et al., 1997; Motohashi-Saigo & Hardison, 2009; Lengeris & Hazan, 2010; Wong, 2013, 2015; Huensch & Tremblay, 2015; Rato & Rauber, 2015; Lee & Lyster, 2016; Okuno & Hardison, 2016; see also the meta-analysis in Sakai & Moorman, 2018). However, while studies using production training are overall rarer, the inverse carry-over effect from production training to perception has been reported as well (Akahane et al., 1998; Kartushina, 2015). Moreover, an interference effect of production on perception training has also been observed: when participants overtly repeat the stimuli during



perception training, the effect of training on their post-test perception performance is disrupted (Baese-Berk & Samuel, 2016).

A third set of studies investigating the relationship between perception and production has focused on a possible correlation between the two modalities. Flege and colleagues thus conducted a series of studies on the perception and production of vowels and consonants in a variety of languages and with participants with various levels of L2 proficiency and a number of different L1s (Flege, 1993, 1999; Flege et al. 1997; Flege & Eeftink, 1988; Flege & Schmidt, 1995; Schmidt & Flege, 1995). In all of these studies, as well as more recent ones by other researchers (Baker & Trofimovich, 2006; Jia et al., 2006; Bettoni-Techio et al., 2007; Kluge et al., 2007; Hattori & Iverson, 2009, 2010; Zhang & Peng, 2017), a positive correlation between the two modalities was found, most often of modest size. Flege (1999) argued that while the correlation between perception and production might not be strong, it might also be underestimated in these studies due to methodological factors, such as the specific perception and production measures used. However, even when perception and production are correlated, they do not necessarily involve the same representations. For instance, in their study of the perception and production of the English /r/-/l/ distinction by Japanese learners, Hattori & Iverson (2010) found that production accuracy of the relevant acoustic cues does not correlate with perceptual sensitivity to these cues. Other studies, moreover, have failed to obtain a correlation between perception and production altogether. For instance, Peperkamp & Bouchon (2011) tested advanced French learners of English on the /i/-/ɪ/ contrast and found not even a hint of a correlation. Kartushina & Frauenfelder (2014) also found no correlation between the perception and production of French vowels by intermediate Spanish learners. Other studies have yet reported a correlation between perception and production for only some non-native sounds. For example, Levy (2009) and Levy & Law (2010) investigated the perception and production of three French vowel contrasts, /y/-/u/, /u/-/œ/ and /y/-/œ/, by three groups of American English learners of French, differing in L2 proficiency. Their results showed a correlation between perception and production across all proficiency groups for the /y/-/œ/ contrast, a correlation in all but the experienced learners for the /u/-/œ/ contrast, and no correlation in any of the groups for the /u/-/y/ contrast. Thus, it remains unclear which factors influence the strength and the very occurrence of a correlation between the two modalities.

To sum up, then, decades of research have not yielded a consensus concerning the relationship between perception and production in L2 speech sound processing. Some of this



lack of consensus may be attributed to methodological issues in these studies. We turn to these issues now.

### **2.2.1.2 Methodological issues**

The contradictory findings on the perception-production link might be explained to some extent by the methodological difficulty of assessing and comparing results from perception and production experiments (Levy & Law, 2010; Elvin et al., 2016). For one thing, results might differ even within a given modality, depending on the task. For instance, Mack (1989) compared the perception and production accuracy of early English-French bilinguals and English monolinguals on the English /d-t/ and /i-ɪ/ contrasts. In perception, she found that bilinguals performed differently from monolinguals in identification but not in discrimination. In a similar vein, Díaz et al. (2012) examined Dutch L2 learners' processing of the English /æ/-/ɛ/ contrast, and found that a larger performance gap between native and non-native listeners in lexical decision and word identification than in categorization. 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.

Furthermore, several tasks seem to involve both perception and production to some extent. On the one hand, as argued by Peperkamp & Bouchon (2011), certain perception tasks might be influenced by production, due to the automatic activation of a perception-production loop (Baddeley et al., 1984; Jacquemot & Scott, 2006). For instance, discrimination tasks require participants to retain stimuli in phonological short-term memory; provided the inter-stimulus interval (ISI) is not too short, this yields automatic covert rehearsal of the stimuli that are subsequently processed by the speech perception module. Similarly, identification leaves enough time for participants to subvocally rehearse the stimuli and process these covert productions before making a decision. On the other hand, production is sometimes assessed in an imitation or a repetition task (e.g., Flege & Eefting, 1988; Flege, MacKay & Meador, 1999; Levy & Law, 2010; Kartushina & Frauenfelder, 2014; Jia et al., 2016), both of which arguably contain a perception component. Hao & de Jong (2015) specifically raised the question of whether imitation is a better reflection of production or perception skills. They argue that although at first sight imitation seems to be a production task based on auditory prompts, it can also be viewed as a perception task with a verbal response. Looking at the performance of L2 learners in an imitation task compared to that in a reading task and an identification task, they found that the accuracy in imitation was not always constrained by accuracy in either the identification or the reading task. Thus, English learners of Mandarin

performed better on the imitation of Mandarin tones than in the identification and read-aloud tasks. However, Korean learners of English tested on English consonants showed a different pattern of results: their performance in imitation was less accurate than in read-aloud but – when the L2 sounds had a close counterpart in Korean – more accurate than in identification. The authors concluded that L2 imitation may not involve all the skills required by the perception and the production tasks and probably bypasses some aspect of phonological encoding.

Measuring accuracy in L2 remains problematic even after choosing the most appropriate tasks to test perception and production. Performance in perception typically depends not only on how well the target contrast is perceived, but also on factors such as cognitive control, memory, and attention. Adding a native control contrast provides an individual baseline for performance (e.g., Sebastián-Gallés & Baus, 2005; Peperkamp & Bouchon, 2011), but when an individual accuracy score needs to be computed – as is the case for correlation studies – the question arises as to how this baseline should be taken into account. For example, Peperkamp & Bouchon (2011) carried out linear regressions between the perception and production scores of bilingual participants and included the scores on the control condition in perception as a covariate. They also carried out an additional analysis where they used individual difference scores for perception, defined as the error rate on the experimental contrast minus that on the control contrast. An alternative way to take into account the performance on the control contrast is to include the native participants' data in the modeling, with native language entered as a fixed effect.

For production tasks, the problem consists in deciding what the dependent measure should be. One possibility is to obtain nativelikeness scores from judgments made by native speakers. But what should these judges listen to? Individual target sounds excised from recordings are often too short to be judged by native speakers, while larger portions might induce a judgment of the overall accent rather than of the target sound's accuracy. Even when asked to focus only on the target sound, the judges could be biased by the global accent of the L2 speaker, depending on their capacity to abstract away from it (cf. discussion in McCullough, 2013). In addition, providing full words as input might introduce a lexical bias in their judgments.

A different way of evaluating L2 production accuracy is by carrying out acoustic measurements of the recorded stimuli. A common assumption is that a greater acoustic distinctiveness of a non-native contrast implies a better command of L2 (Tsukada et al., 2005; Kartushina, 2014). The distinctiveness between two vowels is commonly measured as their

acoustic distance in the F1 x F2 vowel space. However, the choice of a measure of acoustic distinctiveness is not straightforward. Many previous studies used the Euclidian distance to estimate the distance between the centroids (means) of the two category distributions in the F1 x F2 acoustic space (e.g. Chandrasekaran, 2010, Tsukada et al., 2005, Lengeris, 2016). This method disregards both duration and formant dynamics, and has the further disadvantage that it does not take into account the shape of the distributions. That is, it ignores potentially relevant information like the category variance or overlap. Some recent studies have addressed this problem by using Mahalanobis distance, a unitless measure that captures the distance between a point and a distribution in terms of the number of standard deviations the point is from the distribution's mean (Mahalanobis, 1936). This metric can be used to estimate the distance between two vowel distributions by summing the individual distances between each exemplar of each category and the distribution of the other category (Kartushina & Frauenfelder, 2014; Renwick & Ladd, 2016). Finally, one more methodological aspect of acoustic measurements concerns the scale used to represent formant frequencies. A simple linear frequency scale does not reflect human perception accurately, as the frequency response of the human ear is somewhat *logarithmic* (Sawusch, 2005). Moreover, F2 has been shown to contribute more to the identification of vowels than F1 (Delattre et al., 1952). Transforming the Hertz scale into a psycho-acoustical scale such as the Bark or mel scale allows for a more accurate measurement of production accuracy.

Thus, whether evaluating perception or production data, one must take into account a range of issues and take non-trivial methodological decisions. This difficulty in assessing performance is even more striking for studies comparing perception and production, as the tasks used to assess each of them might not be of equivalent difficulty for L2 speakers. For instance, the task in one domain might be cognitively more demanding than the task in the other, as when different processing levels are involved. This is often the case, with perception being typically tested with a prelexical task and production with a lexical task. Similarly, target sounds are not always presented in the same phonetic contexts in the perception and production tasks, despite the fact that task difficulty can depend partly on the surrounding sounds (e.g., Strange et al., 2001; Levy & Law, 2010).

### **2.2.1.3 Current study**

We address some of the above-mentioned methodological issues in order to obtain more precise and comparable measures for perception and production accuracy, and further investigate the hypothesis that perception and production in L2 phonological processing are

correlated. Our case study concerns the perception and production of the French /u/-/y/ contrast (as in *pouce* ‘thumb’ - *puce* ‘flea’) by highly proficient English-speaking late learners of French. The contrast between the vowels /u/ and /y/ has been reported to be one of the most difficult ones for (American) English speakers to perceive (Levy & Strange, 2008) and produce (Levy & Law, 2010).

To assess production, we use both a prelexical and a lexical task. Since neither duration nor formant trajectories are important intrinsic aspects of the production of French oral vowels, we assess accuracy by measuring Mahalanobis distance between F1 and F2 midpoint measures in the Bark scale, comparing performance of the late learners to that of a control group of native French speakers. To assess perception, we use a prelexical task, and compare performance on the test contrast both to that of a series of control contrasts and to that of the control group of native French speakers. We also manipulate the ISI in this task in order to examine the role of the automatic activation of the perception-production loop when stimuli are coded in phonological short-term memory. That is, we use both a short and a long ISI, with only the latter allowing participants to subvocally rehearse the stimuli and process these covert productions before making a decision.

We evaluate the relation between perception and production both within and across processing levels. Specifically, using mixed-effects modeling, we compare performance on the prelexical perception task to performance in the prelexical production task (within-level comparison) on the one hand, and to that in the lexical production task (across-level comparison) on the other hand. We predict that perception correlates with production for the within-level comparison, especially for the long ISI condition, but not necessarily for the across-level comparison.

Additionally, we address the question of whether good perception is a necessary (but not necessarily sufficient) condition for good production, as stated by the Speech Learning Model (Flege 1987, 1995). In particular, we use a clustering algorithm to divide the late learners into relatively good and bad perceivers and relatively good and bad producers, based on their performance on each of the tasks. This method allows us to assess the relative performance of the individual learners in perception and production compared to the overall group performance, thus avoiding the problem of task comparability. Following the Speech Learning Model, we predict that more participants will fall within the clusters of good perceivers and bad producers than within the clusters of bad perceivers and good producers.

### 2.2.2 Methods

The experiment consisted of one perception task, i.e. ABX discrimination, and two production tasks, i.e. pseudoword reading and picture naming. In the ABX discrimination task, we tested participants' perception of the French /u/-/y/ contrast and compared it to their perception of a series of control contrasts, i.e. /a/-/i/, /a/-/e/, /o/-/i/, and /e/-/o/. In order to ensure that the task would be hard enough for our target group of highly proficient L2 learners, we used relatively long, trisyllabic stimuli and made the syllabic position of the experimental contrast vary across trials, such that participants' attention would not be drawn to one particular syllable position over the course of the experiment.

In the pseudoword reading task, we used the same pseudowords as those in the perception task, thus making it directly comparable to the perception task. As this task used both the same items and tapped into the same processing level as the perception task, it provided the strongest case for testing the hypothesis that perception and production are correlated in L2 speech sound processing. Finally, in the picture naming task, we used pictures of objects whose names contain /u/ or /y/ for the test items and pictures of objects whose names do not contain /u/ and /y/ for the filler items.

#### 2.2.2.1 Participants

Nineteen English-speaking late learners of French, sixteen women and three men aged between 20 and 35, participated. They were native speakers of American or British English who had started to learn French between the ages of 4 and 27 (mean: 12.9 years). They were all proficient speakers of French, and had been living in France for at least one year (mean: 4.58 years). A questionnaire based on the bilingualism dominance scale (Dunn & Fox Tree, 2009) was used to quantify language dominance. This questionnaire examines frequency and domains of use, age of acquisition, and the age at which they felt comfortable speaking each language. The resulting dominance score can range from -30 to +30, with 0 indicating perfect balance and a score lower than -5 or higher than +5 being interpreted as dominance in French or English respectively. Individual dominance scores for these participants ranged from +5 to +23 (mean: +17.6); thus, all participants were English-dominant, most of them substantially so. Participants also completed a questionnaire to self-evaluate their speaking, listening, reading, vocabulary and grammar skills in both languages, on scales from 1 to 10. For all aspects, participants scored themselves higher for English (mean: 9.8) than for French (mean: 7.1).

In addition, 11 native French speakers, eight women and three men aged between 20 and 29, participated as controls. None of the participants reported a history of speech or language problems. They were all paid a small fee for their participation.

### 2.2.2.2 Stimuli

For the ABX discrimination task, we created forty-eight pairs of trisyllabic French CVCVCV pseudowords differing only in a vowel (e.g. /vepuba/-/vepyba/) (the full list of stimuli is provided in Appendix, part A). For half of the pairs (test), the vowel contrast was /u/-/y/, for the other half (control), it was one of /a/-/i/, /a/-/e/, /o/-/i/, and /e/-/o/. The vowel contrast appeared in either the first, the second, or the third syllable. For the test contrast, the crucial vowels were preceded by an alveolar consonant (/t/, /d/, /n/) in half of the pairs and by a bilabial consonant (/p/, /b/, /m/) in the other half.<sup>3</sup> Three native speakers of French, two women and one man, recorded the stimuli in a soundproof booth, at 16 bits mono with a sampling rate of 44.1 kHz. The mean duration of the stimuli was 686 ms.

For the pseudoword reading task, we used the test pairs from the perception task, i.e. the ones containing the /u/-/y/ contrast. Stimuli were written in appropriate French orthography, e.g. *vépouba* for /vepuba/ and *vépyba* for /vepyba/.

For the picture naming task, we selected 120 color pictures of objects, the French names of which were likely to be familiar to all participants. Thirty of these names contained /u/, 30 /y/, and 60 neither of these vowels (the full list of stimuli is provided in Appendix, part B).<sup>4</sup> The lists were matched in terms of number of syllables and frequency.

### 2.2.2.3 Procedure

ABX discrimination Participants were presented in each trial with three trisyllabic items, the first two produced by the two female speakers and the third one by the male speaker.<sup>5</sup> Their task was to determine whether the last item (X) was identical to the first (A) or to the second one (B). There were 192 trials divided over four blocks. In each block, half of the trials featured the test /u/-/y/ contrast, the other half one of the control contrasts (/a/-/i/, /a/-/e/, /o/-/i/, or /e/-/o/). The identity of X and the correct response (A or B) were counterbalanced, and the trials were presented in a pseudo-random order, such that no more than three trials of the same type (test or control) or with the same correct response (A or B) would appear in a row. In each block the ISI was either 150 ms (henceforth: short ISI) or 1000 ms (long ISI). The ISI block types alternated. Half of the participants started with a short ISI

block, the other half with a long ISI block. Participants could take a short break in between blocks.

Each block started with a practice phase of five trials, during which participants received feedback as to whether their responses were correct. In the case of an incorrect response or no response within 2500 ms of the stimulus offset, the trial was repeated until the correct response was given. During the test phase, participant received no feedback and if they did not respond 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 stimulus.

Pseudoword reading The 48 items used in the test trials of the discrimination task (half containing /u/, the other half /y/) were embedded in a carrier sentence: *Je dis \_\_ deux fois*, /ʒœdi \_ døfwa/ "I say \_\_ twice". These sentences were presented on a computer screen in a pseudo-random order, such that no more than three sentences containing items with the same target vowel appeared in a row. Participants were asked to read them as naturally as possible, and to press a button to proceed from one sentence to the next.

Picture naming<sup>6</sup> The 120 pictures were presented one by one on the screen in a pseudo-random order, such that no more than three objects with the same target vowel in their name appeared in a row. Participants were asked to name the object they saw and to press a button to proceed to the next picture.

### **2.2.3 Results and discussion**

We first present the results for the production tasks. We then present together the results for perception and for the correlation between the two modalities, using a single regression model to analyze these aspects simultaneously.

#### **2.2.3.1 Production**

All recordings were checked for the absence of noise (e.g., coughs, sneezes, etc.), recording failures, and productions that differed from the target (i.e., names that did not correspond to the designated image in naming, and pseudowords produced with erroneous sounds in reading). A total of 1,5% of the recordings were discarded on this basis.

The waveform and the wideband spectrogram of the production data were visualized, and target vowels were segmented at zero crossings. After segmentation, the mean values of the first two formants (F1 and F2) at the acoustic midpoint of each token were automatically

estimated using Praat (Boersma & Weenink, 2016). These formant values were then Bark-transformed. Outliers more than 2.5 standard deviations from the by-talker by-vowel mean were discarded (3.9% of the datapoints). Vowel plot summaries for each participant are shown in Figure 2.1<sup>7</sup>.

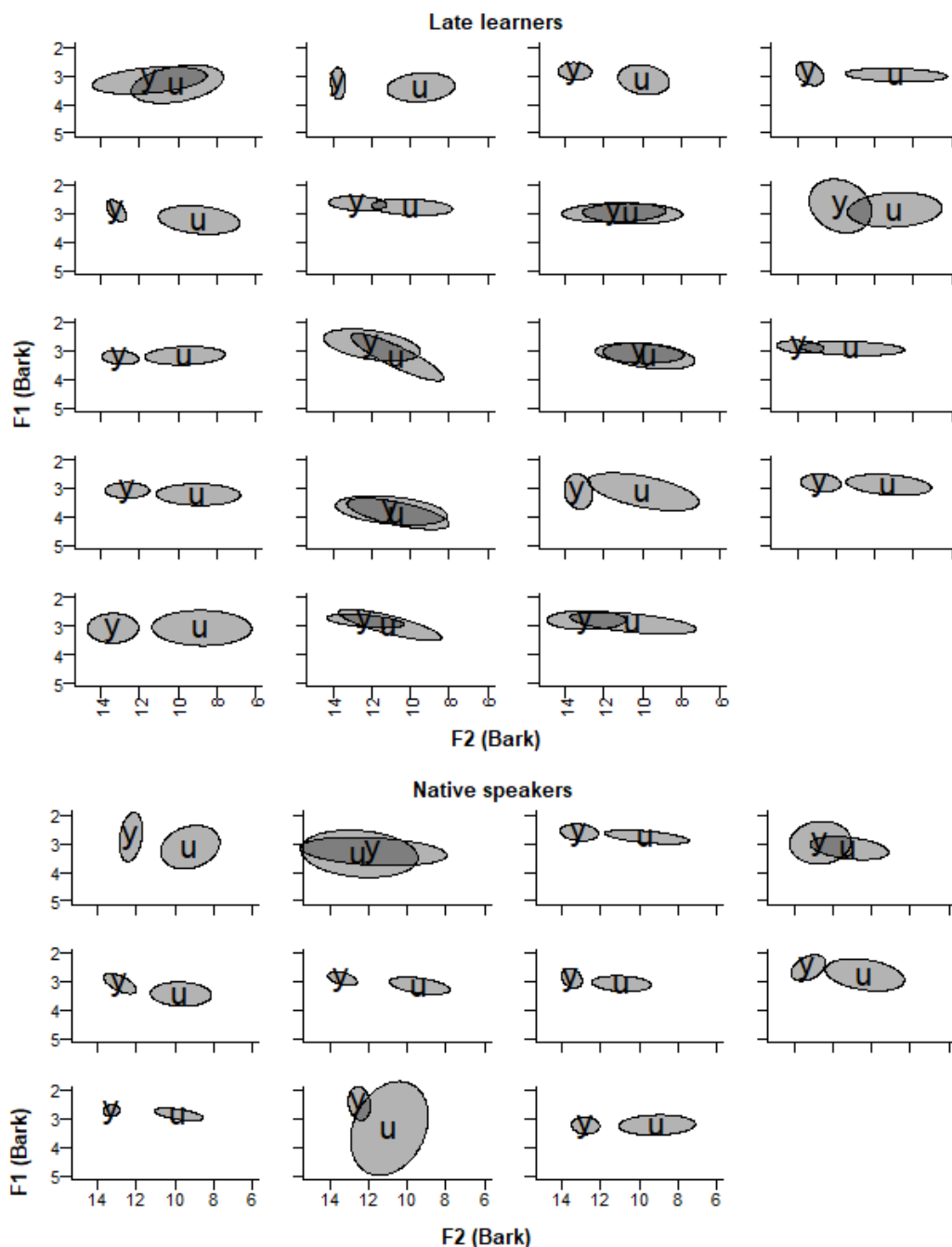


Figure 2.1. Bark-transformed first and second formant frequencies of /u/ and /y/ produced by 19 late learners (top) and 11 native speakers (bottom) in reading. Ellipses are centered on the mean, and their circumference represents one standard deviation.



In order to measure the acoustic distance between the /u/ and /y/ categories we used the Mahalanobis distance metric, which measures the number of standard deviations from a point to the mean of a distribution. For each vowel contrast for each participant, we computed the mean Mahalanobis distance between each token and the distribution of the other category. Thus, for every participant we obtained the mean Mahalanobis distance from each /u/ token to the entire /y/ category and from each /y/ token to the entire /u/ category. Finally, we summed these two distances to obtain an individual measure of the distance between the two categories. A larger distance is indicative of a better separation between the two vowels, and, by hypothesis, of a higher production accuracy. The mean individual distance scores are shown in Figure 2.2.

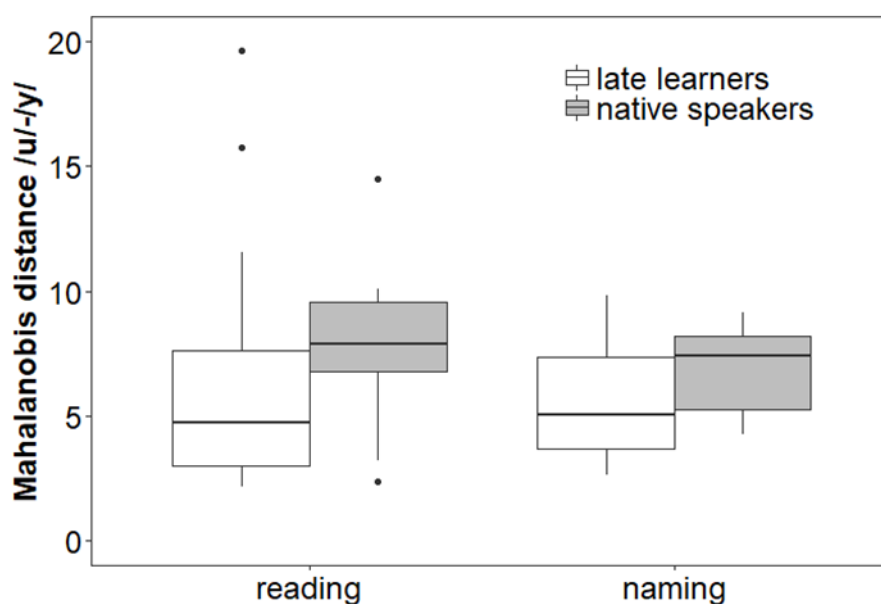


Figure 2.2. Boxplots of Mahalanobis distance scores between the categories /u/ and /y/ produced by late learners and native speakers in pseudoword reading and picture naming.

Mean Mahalanobis distance scores in the pseudoword reading task for late learners were not significantly different from scores for natives in a Welch t-test (learners: mean = 6.41, SD = 4.78; natives: mean = 7.9, SD = 3.32;  $t(26.90) = 1.00$ ,  $p > 0.5$ ,  $d = 0.34$ ). In picture naming, the difference between the mean production scores of the two groups was not

significant either (learners: mean = 5.57, SD = 2.22; natives: mean = 6.87, SD = 1.76;  $t(25.07) = 1.77$ ,  $p > 0.5$ ,  $d = 0.63$ ).

These findings are unexpected, although the numerical trends accord with our expectation that the French participants have more distinct /u/ and /y/ categories than the late learners. One explanation for the lack of a significant difference between the groups might be that the late learners are highly proficient and close to native-like. Their relatively high score on the bilingual dominance scale and their self-evaluations mentioned above, however, suggests otherwise. Alternatively, it might be attributed to differences in speech rate. In particular, if the native French speakers spoke faster than the late learners, this would have caused a reduction of their vowel space, i.e. the displacement of vowels towards the center of the acoustic F1xF2 space (Lindblom, 1963; Nadeu, 2014), and hence a reduced distance between their /u/ and /y/ categories. In order to test this hypothesis, we measured the duration of each target vowel produced by the participants, and carried out Welch t-tests to compare the duration of tokens of native speakers to those of late learners. In both reading and naming, tokens of /u/ and /y/ produced by native speakers were significantly shorter than those produced by late learners (reading<sub>natives</sub>: mean = 96 ms, SD = 44 ms; reading<sub>learners</sub>: mean = 125 ms, SD = 55 ms;  $t(24.84) = 3.44$ ,  $p = 0.002$ ,  $d = 1.23$ ; naming<sub>natives</sub>: mean = 88 ms, SD = 39 ms; naming<sub>learners</sub>: mean = 112 ms, SD = 51 ms;  $t(27.18) = 3.96$ ,  $p < 0.001$ ,  $d = 1.26$ ). Thus, the lack of a significant difference between the late learners and the native speakers might indeed be due to the native speakers' overall fluency, leading to a higher speech rate which likely caused their vowels to become more central overall.<sup>8</sup>

### **2.2.3.2 Perception and its correlation with production**

Figure 2.3 shows mean accuracy scores in the ABX task for the late learners and the native speakers, split by ISI condition (short vs long) and vowel contrast (/u/~y/ vs control). Scatter plots of the relationship between the ABX task and both production tasks can be found by following the link:

[http://cognitivetraining.epizy.com/Graphs/Graphs\\_supplementary\\_materials.pdf](http://cognitivetraining.epizy.com/Graphs/Graphs_supplementary_materials.pdf)

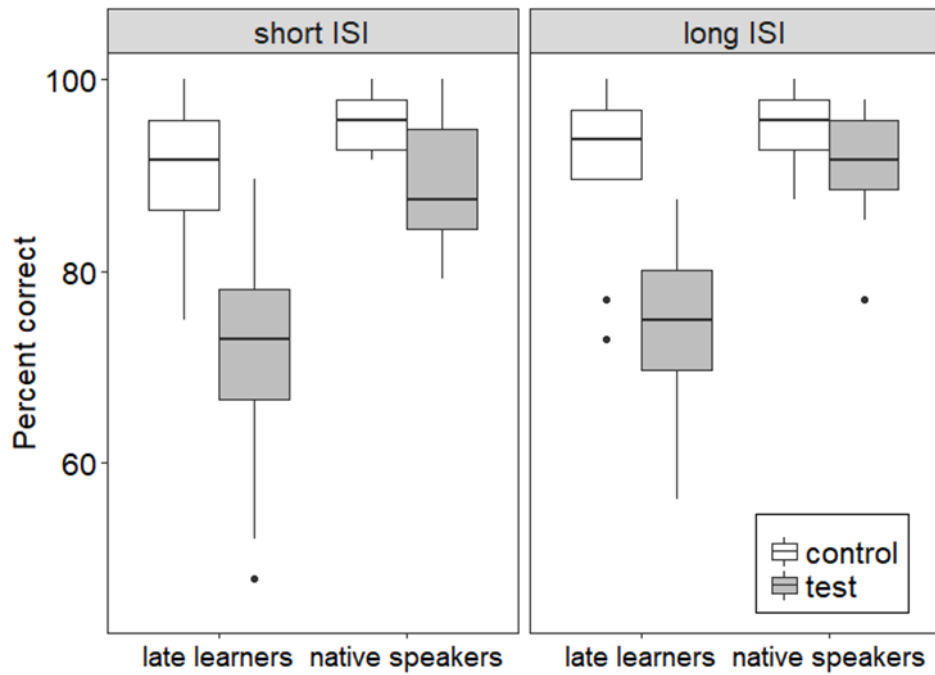


Figure 2.3. Boxplots of percent correct responses in the ABX discrimination task for late learners and native speakers in the short ISI (left panel) and long ISI (right panel) conditions.

We analyzed these data using logistic mixed effects regression modeling. Crucially, we included the production scores from the pseudoword reading and the picture naming tasks, respectively, as fixed effects in two separate models. For each model, a significant effect of production score would be evidence that perception and production are correlated. The R package lme4 (Bates et al., 2012) was used to carry out these analyses. Effect-size estimates were obtained using the MuMIn package (Barton, 2018).

Our procedure for model construction followed the stepwise algorithm outlined in Turnbull (2017). We started the analysis with a null model that included our binomial dependent variable (ABX task accuracy) and Participants and Items as random intercepts. The predictor variables that we tested for were Vowel Contrast (test vs. control (baseline)), Group (late learners (baseline) vs. native speakers), ISI (short (baseline) vs. long), and production score (either pseudoword reading score or picture naming score, depending on the model). At each step, we tested for each predictor variable not yet present in the model whether the model would improve if it was added. We evaluated each added effect using likelihood-ratio tests. At the end of each step, the effect with the lowest p-value below .05 was added to the model. We then repeated this process with the larger model with the remaining predictor

variables until no effects gave a significant model improvement. At each step, if main effects were retained in the model, we tested for an interaction between them.

In the model using pseudoword reading as a measure of production accuracy, the optimal model returned main effects of Reading Accuracy, Contrast, and Group, and an interaction between Contrast and Group (Table 2.1).

TABLE 2.1. Coefficients and log-likelihood comparisons for each retained fixed effect, in the model where the measure of production accuracy was performance in pseudoword reading.

variable	$\beta$	SE	z	$\chi^2$	DF	p
intercept	2.09	0.22	9.61	-	-	-
Contrast: test	-1.5	0.18	-8.26	102.88	2	<0.01
Group: native speakers	0.46	0.24	1.94	17.071	1	<0.01
Reading Accuracy	0.08	0.02	3.71	11.45	1	<0.01
Contrast: test $\times$ Group: native speakers	0.72	0.21	3.55	11.74	1	<0.01
Effect size (whole model) : $R^2_{\text{marginal}} = 0.14$ , $R^2_{\text{conditional}} = 0.26$						

Native French participants performed better than late learners (learners: mean = 82.3%, SD = 38.2%; natives: mean = 92.6%, SD = 26.2%) and performance was better on the control contrasts than on the /u/-/y/ test contrast (test: mean = 79.5%, SD = 40.3%; control: mean = 92.6%, SD = 26.1%), but the difference in performance between test and control contrasts was smaller for native speakers than for late learners. An effect of Reading Accuracy was observed, with higher production scores predicting higher perception accuracy. This means that performance in pseudoword reading was a good predictor of discrimination accuracy; hence, perception and production were correlated in both the late learners and the native speakers, and this correlation held for both the discrimination responses on the /u/-/y/ test contrast and those on the control contrasts.<sup>9</sup> Our prediction for the late learners was thus borne out. By contrast, the absence of an interaction between ISI and Reading Accuracy in the

final model indicates that, contrary to our prediction, this correlation was not modulated by ISI.

In the model using picture naming as a measure of production accuracy, two main effects and their interaction remained in the model: Contrast (test vs. control) and Group (late learners vs. native speakers) (Table 2.2). Crucially, Naming Accuracy was not retained in the final model. From this we infer that production in the picture naming task was not significantly correlated with perception in either the late learners or the native speakers of French, regardless of whether the discrimination responses concerned the test or the control contrasts.<sup>10</sup>

TABLE 2.2. Coefficients and log-likelihood comparisons for each retained fixed effect, in the model where the measure of production accuracy was performance in picture naming.

variable	$\beta$	SE	z	$\chi^2$	DF	p
intercept	2.61	0.19	13.92	-	-	-
Contrast: test	-1.50	0.18	-8.27	102.88	2	<0.01
Group: native speakers	0.59	0.27	2.21	17.07	1	<0.01
Contrast: test $\times$ Group: native speakers	0.72	0.20	3.55	11.74	1	<0.01
Effect size (whole model) : $R^2_{\text{marginal}} = 0.16$ , $R^2_{\text{conditional}} = 0.26$						

### 2.2.3.3 Clustering

In order to classify the late learners into relatively good and bad perceivers and relatively good and bad producers, we carried out separate clustering analyses on the discrimination, reading and naming data. We used non-hierarchical k-means clustering to group the participants into two groups according to their performance. We chose this technique rather than dividing the data into two equal groups based on a median split. Indeed, our aim is to infer groups based on similarity without requiring these groups to have the same size. For production, we entered two scores, i.e. the performance in pseudoword reading and the performance in picture naming. For perception, we entered one score, i.e. the mean of

performance on short and long ISIs in the test condition. Finally, based on the obtained clusters we superimposed the proficiency groups in the perception task and each of the production tasks, assigning the participants to one of the following groups: good perception and good production; good perception and bad production; bad perception and good production; bad perception and bad production (Table 2.3).

TABLE 2.3. Classification of late learners by their performance in the perception and production tasks.

		Production			
		Pseudoword Reading		Picture Naming	
		good	bad	good	bad
Perception					
good		8	6	6	8
bad		1	4	1	4

Regardless of the production task under consideration, we found that the majority of late learners belonged to one of three groups: those with good production and perception; those with bad production and perception; and those with bad production but good perception. Of interest is the fact that while several late learners belonged to the good perception / bad production group (6 in the reading task and 8 in the naming task), only one of them was assigned to the bad perception / good production group (in both reading and naming)<sup>11</sup>.

Thus, late learners who were good in production were also good in perception, although those who were good in perception were not necessarily good in production. These results are in accordance with the central claim of the Speech Learning Model (Flege 1987, 1995) that accurate perception of an L2 sound is a prerequisite for its correct production.

#### 2.2.4 General discussion

Investigating the relationship between perception and production in L2 phonological processing is all but straightforward. The specific tasks used to assess perception and production, the measure by which production accuracy is evaluated, the presence of a control contrast and/or group, and the statistical methods used to analyze the data all require non-

trivial choices to be made. Here, we focused on the French vowel contrast /u/-/y/, and tested proficient English-speaking L2 learners of French in a design aimed at obtaining more precise and comparable measures for perception and production accuracy. For perception we used a prelexical task, ABX discrimination task, with both the /u/-/y/ test contrast and a series of control vowel contrasts. For production, we used both a prelexical task, i.e. pseudoword reading (using the same items as those in ABX discrimination), and a lexical task, i.e. picture naming, and measured the Mahalanobis distance between /u/ and /y/ in the Bark scale to assess accuracy. In all tasks, we compared performance of the late learners to that of a control group of native French speakers. Using mixed-effects modeling, we found evidence for a correlation at the same processing level but not across levels. That is, prelexical ABX discrimination correlated with prelexical pseudoword reading but not with lexical picture naming. In addition, we tested whether good perception is a prerequisite for good production. Using a clustering algorithm, we found evidence that this is indeed the case. Before discussing the results regarding the relation between perception and production, we comment on the production and the perception results separately.

As to the production accuracy of late learners versus native speakers of French, we obtained mixed results. We expected late learners to produce /u/ and /y/ less accurately than native French speakers (with less distance between the two vowels in the acoustic space) and hence, to have less distinct /u/ and /y/ categories, as observed earlier with a repetition task by Levy & Law (2010). However, in both pseudoword reading and picture naming the difference between native speakers and late learners did not reach significance. A post-hoc analysis of token durations showed that compared to the late learners, the native French participants produced significantly shorter tokens of /u/ and /y/. This means that the distance between those vowels in French productions was likely reduced, as vowels typically become more central at increased speech rates. Thus, the lack of difference between the productions of native speakers and late learners could be explained by a difference in speech rate. This issue with the performance of control participants should be taken into account in further studies.

In perception, the late learners were less accurate than the native French speakers, and had more difficulty in perceiving the test contrast /u/-/y/ than the control contrasts. This reflects the strong effect of the listeners' native language on their phonological categorization. It is worth noting that /u/-/y/ is acoustically a smaller contrast than any of the control contrasts /a/-/i/, /a/-/e/, /o/-/i/, and /e/-/o/. Therefore, this contrast should be slightly harder to distinguish than the control contrasts, irrespective of the listener's language background. This is indeed what we found: the native speakers also performed less accurately on the /u/-/y/

contrast than on the control contrasts. Nevertheless, the native speakers' overall accuracy was high, and the significant difference between the native speakers and the late learners in the perception of test vs. control contrasts can thus be attributed to the late learners' difficulty to perceive an L2 contrast that does not exist in their native language. These results are consistent with the findings of Levy & Strange (2008), who examined the perception of French vowels by American English listeners with and without French language experience. In their study, both groups of American English listeners performed worse than French control participants. Moreover, for the experienced group the /u/-/y/ contrast was the most difficult one (the other test contrasts were /i/-/y/, /u/-/œ/, and /y/-/œ/).

We also examined the effect of different durations of ISI. The results revealed that there was no significant difference in the performance in the ABX task in the short versus long ISI condition. At first sight, this is contradictory to previous findings. For example, Werker & Logan (1985) tested native English speakers on their perception of a Hindi retroflex/dental contrast in an AX task, using three ISI conditions (1500 ms, 500 ms and 250 ms). They showed that the length of ISI affects performance differentially and argued that different ISIs tap different processing levels, as higher memory demands in the task trigger a higher processing level. Specifically, an ISI of 1500 ms would tap a phonological level, one of 500 ms a phonetic level, and one of 250 ms an auditory-acoustic level. However, the lack of an effect of ISI in our study is likely due to the fact that our task had high memory requirements even in the short ISI condition, and hence tapped a phonological processing level in both ISI conditions. That is, whereas Werker & Logan (1985) used short, monosyllabic stimuli, produced by a single speaker, we used long, trisyllabic, stimuli produced by three speakers, leading to substantially higher memory demands. Moreover, the ABX task we used is more demanding than their AX task, as participants have to attend to three instead of two items in each trial.

Turning now to the relationship between perception and production in the late learners, we obtained – as predicted – a correlation between discrimination and pseudoword reading. We consider this correlation to be reliable and robust: not only do the two tasks tap the same, prelexical, processing level, we also implemented them using the same items. Thus, we obtained comparable measures for assessing the participants' performance in the two modalities. This result contrasts with that of Levy & Law (2010), who also used the same items in perception and production but found no correlation. Their participants, though, had varying levels of French proficiency (from none to advanced), and their tasks were different: mapping of French vowels onto closest English one in perception, and non-word repetition in



production. Interestingly, the correlation observed in the present study held for both the discrimination responses on the test contrast and for those on the control contrasts; that is, /u/-/y/ production accuracy was predicted by discrimination accuracy of not only the same /u/-/y/ contrast but also different vowel contrasts. In other words, the correlation did not hinge upon the use of the same contrast, and, *a fortiori*, of the same set of test items. By contrast, we observed no correlation between discrimination and picture naming, suggesting that processing levels are to a certain extent independent, and that perception-production correlations can be restricted to specific levels.

The correlation between discrimination and reading was not moderated by ISI. We had chosen the ISIs such that the long but not the short ISI condition allowed for a complete activation of the perception-production loop (Baddeley et al., 1984; Jacquemot & Scott, 2006), i.e., for participants' automatic, subvocal rehearsal of the stimuli and their processing of these covert productions prior to decision making. The absence of a moderator effect of ISI indicates that the correlation between perception and production does not hinge upon the activation of participants' production module during the discrimination task, contrary to a suggestion by Peperkamp & Bouchon (2011). Of course, it is still possible that the perception-production loop plays a role during the process of L2 phonological learning and hence in the origin of the correlation in L2 learners (Nagle, 2018).

Finally, the clustering analysis showed that, overall, good perception is a prerequisite for good production, although good perception can occur in the absence of good production. This result is in accordance with the Speech Learning Model (Flege, 1987, 1995), according to which accurate production cannot be reached without accurate perception.

Our main result on the correlation between perception and production is in agreement with a number of previous studies (Flege, 1993; Flege et al., 1997, 1999; Flege & Schmidt, 1995; Schmidt & Flege, 1995; Jia et al., 2006; Bettoni-Techio et al., 2007; Kluge et al., 2007; Hattori & Iverson, 2009, 2010; Zhang & Peng, 2017). However, it contrasts with several others (Levy & Law, 2010; Peperkamp & Bouchon, 2011; Kartushina & Frauenfelder, 2014). Why some studies obtain a correlation and others do not remains an open question. In particular, it is not the case that all of the former and none of the latter used comparable tasks and stimuli across the two modalities, as we would expect based on our own results. Rather, the presence vs. absence of a correlation probably hinges on a host of factors, only some of which are methodological. For instance, there might be differences in the relation between perception and production according to the type of L2 sounds (consonants vs. vowels, or sounds that have a close L1 counterpart vs. those that do not (Bohn & Flege, 1997)), or the

general level of L2 proficiency (Levy, 2009; Levy & Law, 2010). Another factor that has often been suggested is L2 speakers' motivation - or lack thereof - to reduce their foreign accent (Sheldon, 1985; Mack, 1989; Bohn & Flege, 1997; Flege, 1999).

Similarly, the result of our clustering analysis that good perception is a prerequisite for good production seems to be in accordance with some previous studies (Flege, 1993; Nagle, 2018), but not with others (Goto, 1971; Sheldon & Strange, 1982; Flege & Eeftink, 1987; Bohn & Flege, 1997). However, in order to assess the issue of whether the ability to perceive L2 sounds develops before the ability to produce them, it can be more insightful to consider individual, not group performance. For instance, Flege (1993) observes that while in his study L2 learners' perception was overall better than their production, at the individual level participants were about equally divided between those having better perception and those having better production. Examining perception and production in beginning L2 learners over the course of one year, Nagle (2018) found that improvement in perception generally precedes improvement in production, but he likewise observed a lot of variability at the individual level. Goto (1971) and Sheldon & Strange (1982) also analyzed individual data; modulo the fact that they had few participants (11 and 6, respectively), their results do provide evidence for better production than perception in L2 phonological processing. This, then, clearly contrasts with the present data, where we used a clustering analysis to examine individual performance.

To conclude, using well-controlled experimental conditions, we provided robust evidence for a correlation between the perception and production of the French /u/-/y/ contrast by English advanced learners of French. The methodological framework we developed for studying the relationship between the two modalities can be used in further studies, focusing on other languages, other types of contrasts, and other profiles of L2 learners. Future research could also concentrate on lexical processing, and hence compare perception and production using lexical tasks; at least for the case of English learners' processing of the French /u/-/y/ contrast, we predict that a correlation will be obtained provided the same items are used for perception and production.

### 2.2.5 Notes

<sup>1</sup>Another study that argued for production lagging behind that of perception in beginning learners is that of Detey et al. (2014) and Detey & Racine (2015). They tested Japanese learners of French to both perceive and produce the French nasal vowels / $\tilde{a}$ /, / $\tilde{o}$ /, and / $\tilde{e}$ /. They found that the / $\tilde{a}$ /-/ $\tilde{e}$ / contrast was better perceived than the / $\tilde{a}$ /-/ $\tilde{o}$ / contrast, while in production there was no distinction among the three vowels. In the absence of a native control group and/or longitudinal data, though, the conclusion that perception preceded production seems unwarranted.

<sup>2</sup>We note here that similar effects have been observed, to some extent, in the L1 literature. Johnson et al. (1993) reported that some speakers of Californian English can reliably perceive but not produce the caught-cot distinction.

<sup>3</sup>This was motivated by the findings of Levy & Law (2010), who showed that American English speakers make more errors on discriminating pairs involving front vs. back rounded vowels (such as /u/ and /y/) in alveolar as opposed to bilabial contexts.

<sup>4</sup>Ideally, we would have used minimal pairs that differ in /u/ vs. /y/, but French does not have enough of them.

<sup>5</sup>The use of multiple speakers discourages participants from focusing on low-level acoustic details.

<sup>6</sup>As this task is the most engaging one, it was presented in between the two other ones.

<sup>7</sup>The plots were made using the R package phonR (McCloy, 2016).

<sup>8</sup>Note that this predicts a correlation between speech rate and category distinctiveness. A linear regression between Mahalanobis distance score and mean vowel duration performed on the native speakers' data, however, yielded no correlation (reading<sub>natives</sub>:  $r^2 = -0.04$ ,  $p > 0.1$ ; naming<sub>natives</sub>:  $r^2 = -0.04$ ,  $p > 0.1$ ). Possibly, the small sample size does not allow us to observe an existing correlation in this group. As to the late learners', their productions were overall long, making vowel reduction and hence a correlation unlikely to occur in this group to begin with.

<sup>9</sup>Note that if the correlation was restricted to the native speakers we would have observed an interaction between Reading Accuracy and Group. Similarly, if the correlation was restricted to the discrimination responses on the test contrast we would have observed an interaction between Reading Accuracy and Contrast.

<sup>10</sup>Note that if there was a correlation between the perception and production of the late learners we should have observed an interaction between Naming Accuracy and Group.

Similarly, if there was a correlation restricted to the discrimination responses on the test contrast we should have observed an interaction between Naming Accuracy and Contrast.

<sup>11</sup>For only two participants did performance differ in reading versus naming. In particular, they were relatively good in reading but relatively bad in naming. This could be explained by the difficulty of the naming task compared to the reading task. That is, these participants might have concentrated more on lexical retrieval than on the correct pronunciation in the naming task. Alternatively, it is possible that they had incorrect phonological representations of some of the words containing /u/ and /y/ and therefore pronounced them erroneously.

## 2.2.6 Appendix

**A: Stimuli used in the ABX discrimination and the pseudoword reading task. Note that the latter only contained the test items.**

Test stimuli

### Pseudowords with /u/

boulipa	/bulipa/
pougamon	/pugamõ/
kabouzin	/kabuzẽ/
vépouba	/vepuba/
tigobou	/tigobu/
nimapou	/nimapu/
boutafi	/butafi/
pouzidé	/puzide/
tapoudi	/tapudi/
méboufa	/mebufa/
vossipou	/vosipu/
rajébou	/ʁaʒebu/
touréna	/tuʁenã/
doumiko	/dumiko/
katoudin	/katudẽ/
tidouza	/tiduza/
fitadou	/fitadu/
béjitou	/beʒitu/
toupaki	/tupaki/
douféni	/dufeni/
latoumé	/latume/
pédouvi	/peduvi/
romatou	/ʁomatu/
késsidou	/kessidu/

### Pseudowords with /y/

bulipa	/bylipa/
pugamon	/pygamõ/
kabuzin	/kabyzẽ/
vépuba	/vepyba/
tigobu	/tigoby/
nimapu	/nimapy/
butafi	/bytafi/
puzidé	/pyzidé/
tapudi	/tapydi/
mébyfa	/mébyfa/
vossipy	/vossipy/
rajébu	/ʁaʒeby/
turéna	/tyʁenã/
dymiko	/dymiko/
katudin	/katydẽ/
tidyza	/tidyza/
fitady	/fitady/
béjity	/beʒity/
typaki	/typaki/
dyféni	/dyfeni/
latyme	/latyme/
pédyvi	/pedyvi/
romaty	/ʁomaty/
késsidy	/kessidy/

### Control stimuli

koubado /kubado/  
loudapon /ludapõ/  
passoutin /pasutẽ/  
métoussi /metusi/  
mélibou /melibu/  
voquadou /vokadu/  
joumélo /zumelo/  
goufané /gufane/  
térouna /teɾuna/  
danoupo /danupo/  
nidapou /nidapu/  
ponatou /ponatu/  
lumaro /lymaɾo/  
pufadin /pyfadẽ/  
manussin /manysẽ/  
té luna /telyna/  
fédabu /fedaby/  
ritadu /ɾitady/  
kutalo /kytalo/  
bussiné /bysine/  
gobuza /gobyza/  
nadyvé /nadyve/  
jokatu /ʒokaty/  
fopassu /fopasy/

koubédo /kubedo/  
loudipon /ludipõ/  
péssoutin /pesutẽ/  
motoussi /motusi/  
molibou /molibu/  
viquadou /vikadu/  
jouméli /zumeli/  
goufano /gufano/  
térouni /teɾuni/  
danoupi /danupi/  
nidépou /nidepu/  
ponitou /ponitu/  
lumiro /lymiɾo/  
pufédin /pyfedẽ/  
ménussin /menysẽ/  
té luni /telyni/  
fodabu /fodaby/  
ritédu /ɾitedy/  
kutali /kytali/  
bussino /bysino/  
guibuza /gibyza/  
naduvo /nadyvo/  
jikatu /ʒikaty/  
fopissu /fopisy/

## B: stimuli used in the naming task

### /u/-words

ours	/uʁs/
fourchette	/fuʁʃɛt/
couronne	/kuʁɔn/
mouche	/muʃ/
coussin	/kusɛ̃/
bouchon	/buʃɔ̃/
chou	/ʃu/
coude	/kud/
goutte	/gut/
moustache	/mustaʃ/
poule	/pul/
trousse	/tʁus/
genou	/ʒənu/
poupée	/pupe/
yaourt	/jauʁt/
bouteille	/butej/
tatouage	/tatwaʒ/
bougie	/buʒi/
cou	/ku/
couteau	/kuto/
kangourou	/kɑ̃guʁu/
moustique	/mustik/
roue	/ʁu/
ampoule	/ɑ̃pul/
journal	/ʒuʁnal/
bambou	/bɑ̃bu/
bouton	/butɔ̃/
douche	/duʃ/
loupe	/lup/
pouce	/pus/

### /y/-words

sucré	/sykʁ/
prune	/pʁyn/
pendule	/pɑ̃dyl/
lunettes	/lynɛt/
légume	/legym/
fumée	/fyme/
cube	/kyb/
chaussure	/ʃosyʁ/
cactus	/kaktys/
voiture	/vwaityʁ/
statue	/staty/
plume	/plym/
peluche	/pəlyʃ/
luge	/lyʒ/
jupe	/ʒyp/
écureuil	/ekyʁœj/
confiture	/kɔ̃fityʁ/
ceinture	/sɛ̃tyʁ/
bus	/bys/
tortue	/toʁty/
autruche	/otʁyʃ/
allumette	/alymɛt/
turban	/tyʁbɑ̃/
pull	/pyl/
perruque	/peʁyk/
nuage	/nyaʒ/
fusil	/fyzi/
cure-dents	/kyʁdɑ̃/
capuche	/kapyɛ̃/
bulle	/byl/

## Fillers

vache	/vaʃ/	cœur	/kœʁ/
télé	/tele/	cinq	/sɛ̃k/
stylo	/stilo/	chat	/ʃa/
roi	/ʁwa/	cartes	/kaʁt/
pont	/pɔ̃/	cadeau	/kado/
piano	/pjano/	banane	/banan/
oignon	/ɔɲɔ̃/	bague	/bag/
médecin	/medsɛ̃/	ange	/ɑ̃ʒ/
lampe	/lɑ̃p/	verre	/vɛʁ/
fenêtre	/fənɛtʁ/	valise	/valiz/
escalier	/ɛskalje/	téléphone	/telefon/
doigt	/dwa/	table	/tabl/
citron	/sitʁɔ̃/	soleil	/solej/
chien	/ʃjɛ̃/	porte	/poʁt/
chaise	/ʃɛz/	plage	/plaz/
carotte	/kaʁot/	palmier	/palmje/
bateau	/bato/	montre	/mɔ̃tʁ/
arbre	/aʁbʁ/	livres	/livʁ/
crayon	/kʁɛjɔ̃/	gâteau	/gato/
train	/tʁɛ̃/	vélo	/velo/
tasse	/tas/	écharpe	/ɛʃaʁp/
souris	/suri/	clé	/kle/
renard	/renaʁ/	chocolat	/ʃokola/
pomme	/pom/	chapeau	/ʃapo/
pantalon	/pɑ̃talɔ̃/	canapé	/kanape/
moto	/moto/	bébé	/bebe/
maison	/mɛzɔ̃/	ballon	/balɔ̃/
glace	/glas/	avion	/avjɔ̃/
étoile	/etwal/	baguette	/bagɛt/
église	/egliz/	assiette	/asjet/



## 2.2.8 References

- Akahane-Yamada, R., McDermott, E., Adachi, T., Kawahara, H., Pruitt, J. S. (1998). Computer-based second language production training by using spectrographic representation and HMM-based speech recognition scores. In: *Proceedings of International Conference on Spoken Language Processing* (pp. 1–4).
- Baddeley, A., Lewis, V., Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A*, 36(2), 233–252.  
<http://doi.org/10.1080/14640748408402157>
- Baker, W., Trofimovich, P. (2005). Interaction of native and second language vowel system(s) in early and late bilinguals. *Language and Speech*, 48, 1-27.  
<https://doi.org/10.1177/00238309050480010101>
- Baese-Berk, M.M., Samuel, A. G. (2016). Listeners beware: speech production may be bad for learning speech sounds. *Journal of Memory and Language*, 89, 23–36.  
<https://doi.org/10.1016/j.jml.2015.10.008>
- Barton, K. (2018). MuMIn: Multi-model inference. R package version 1.42.1. [Computer Software]
- Bates, D. M., Maechler, M., Bolker, B., Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. [Computer Software]
- Bettoni-Techio, M., Rauber, A.S, Koerich, R.D. (2007). Perception and production of alveolar stops by Brazilian Portuguese learners of English. In: *Proceedings of Interspeech* (pp. 2293-2296).
- Boersma, P., and Weenink, D. (2013). “Praat: doing phonetics by computer” [Computer program]. <http://www.praat.org/>
- Bohn, O. S., Flege, J. E. (1997). Perception and production of a new vowel category by adult second language learners. In: J. Allen & J. Leather (Eds), *Second-language speech: Structure and process*, (pp. 53–73), New York: Mouton de Gruyter.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4), 2299–310. <http://doi.org/10.1121/1.418276>
- Chandrasekaran, B., Sampath, P. D., Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, 128(1), 456–465. <http://doi.org/10.1121/1.3445785>

- Delattre, P., Liberman, M., Cooper, F. S., Gerstman, L. (1952). An experimental study of the acoustic determinants of vowel colour: observations on one- and two- formant vowels synthesized from spectrographic patterns. *Word*, 8(3), 192–210.  
<http://doi.org/10.1080/00437956.1952.11659431>
- Detey, S., Racine, I. (2015). Does perception precede production in the initial stage of French nasal vowel quality acquisition by Japanese learners? A corpus-based discrimination experiment. In: *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Detey, S., Racine, I., Eychenne, J., Kawaguchi, Y. (2014). Corpus-based L2 phonological data and semi-automatic perceptual analysis: The case of nasal vowels produced by beginner Japanese learners of French. In: *Proceedings of Interspeech* (pp.539–543).
- Díaz, B., Mitterer, H., Broersma, M., Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, 22(6), 680–689.  
<http://doi.org/10.1177/0023830916647079>
- Dunn, A. L., Fox Tree, J. E. (2009). A quick, gradient Bilingual Dominance Scale\*. *Bilingualism: Language and Cognition*, 12(3), 273–289.  
<http://doi.org/10.1017/S1366728909990113>
- Elvin, J., Williams, D., Escudero, P. (2016). The relationship between perception and production of Brazilian Portuguese vowels in European Spanish monolinguals. *Loquens*, 3(2), 2386-2637. <http://doi.org/http://oadoi.org/10.3989/loquens.2016.031>
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, 15(1), 47–65.
- Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America*, 93(3), 1589–1608.  
<http://doi.org/10.1121/1.406818>
- Flege, J. E. (1995). Second language speech learning: theory, findings, and problems, In: W. Strange (Eds.), (pp. 233-277). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, Timonium, MD: York Press.  
<http://doi.org/10.1111/j.1600-0404.1995.tb01710.x>
- Flege, J. E. (1999). The relation between L2 production and perception. In: *Proceeding of the 14th International Congress of Phonetic Sciences* (pp. 1273–1276).

- Flege, J., Bohn, O., Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470.  
<http://dx.doi.org/10.1006/jpho.1997.0052>
- Flege, J. E., Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, 15, 67–83.
- Flege, J. E., Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America*, 83(2), 729–740.  
<http://doi.org/10.1121/1.396115>
- Flege, J. E., MacKay, I. R., Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, 106(5), 2973–2987. <http://doi.org/10.1121/1.428116>
- Flege, J.E., Schmidt, A.M. (1995). Native speakers of Spanish show rate-dependent processing of English stop consonants. *Phonetica*, 52, 90–111.  
<http://doi.org/10.1159/000262062>
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R”. *Neuropsychologia*, 9(3), 317–323.  
[http://doi.org/10.1016/0028-3932\(71\)90027-3](http://doi.org/10.1016/0028-3932(71)90027-3)
- Hao, Y. C., de Jong, K. (2015). Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics*, 54, 151–168.  
<http://doi.org/10.1016/j.wocn.2015.10.003>
- Hattori, K., Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America*, 125, 469–479.  
<http://doi:10.1121/1.3021295>
- Hattori, K., Iverson, P. (2010). Examination of the relationship between L2 perception and production: an investigation of English /r/-/l/ perception and production by adult Japanese speakers. In: *Interspeech Workshop on Second Language Studies: Acquisition, Learning, Education and Technology*.
- 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. <http://doi.org/10.1016/j.wocn.2015.06.007>

- Jacquemot, C., Scott, S.K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Science*, 10, 480–486.  
<http://dx.doi.org/10.1016/j.tics.2006.09.002>
- Jia, G., Strange, W., Collado, Y. Wu, J., Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America*, 119, 1118–30.  
<https://doi.org/10.1121/1.2151806>
- Johnson, K., Ladefoged, P. & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, 94, 701–714.  
<https://doi.org/10.1121/1.406887>
- Kartushina, N., Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, 5, 1–17.  
<http://doi.org/10.3389/fpsyg.2014.01246>
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, 138(2), 817–832. <http://doi.org/10.1121/1.4926561>
- Kluge, D. C., Rauber, A. S., Reis, M. S., and Bion, R. A. H. (2007). The relationship between the perception and production of English nasal codas by brazilian learners of English. In: *Proceedings of Interspeech* (pp. 2297–2300).
- Lee, A. H., Lyster, R. (2016). Effects of different types of corrective feedback on receptive skills in a second language: A Speech Perception Training Study. *Language Learning*, 66(4), 809–833.  
<http://doi.org/10.1111/lang.12167>
- Lengeris, A., Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, 128(6), 3757–3768.  
<http://doi.org/10.1121/1.3506351>
- Lengeris, A. (2016). Comparison of perception-production vowel spaces for speakers of Standard Modern Greek and two regional dialects. *The Journal of the Acoustical Society of America*. 140(4), EL314–EL319.  
<http://doi.org/10.1121/1.4964397>

- Levy, E. S. (2009). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America*, 125(2), 1138–1152.  
<http://doi.org/10.1121/1.3050256>
- Levy, E. S., Law, F. F. (2010). Production of French vowels by American-English learners of French: language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America*, 128(3), 1290–1305.  
<http://doi.org/10.1121/1.3466879>
- Levy, E. S., Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, 36(1), 141–157.  
<http://doi.org/10.1016/j.wocn.2007.03.001>
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35, 1773–1781.  
<http://doi.org/10.1121/1.2142410>
- Mack, M. (1989). Consonant and vowel perception and production: early English-French bilinguals and English monolinguals. *Perception & Psychophysics*, 46(2), 187–200.  
<http://doi.org/10.3758/BF03204982>
- Mahalanobis, Prasanta C. (1936). On the generalized distance in statistics. In: *Proceedings of the National Institute of Sciences of India* 2, pp. (49–55).
- McCloy, D. R. (2016). phonR: tools for phoneticians and phonologists. R package version 1.0-7. [Computer Software]
- McCullough, E.A., (2013). Acoustic Correlates of Perceived Foreign Accent in Non Native English, PhD dissertation, Ohio State University.
- Motohashi-Saigo, M., Hardison, D. M. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Technology*, 13(2), 29–47.
- Nadeu, M. (2014). Stress- and speech rate-induced vowel quality variation in Catalan and Spanish. *Journal of Phonetics*, 46(1), 1–22.  
<http://doi.org/10.1016/j.wocn.2014.05.003>
- Okuno, T., & Hardison, D. M. (2016). Perception-production link in L2 Japanese vowel duration: training with technology. *Language Learning & Technology*, 20(2), 61–80.
- Peperkamp, S., Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. In: *Proceedings of Interspeech* (pp.161–164).
- Piske, T., MacKay, I. R., Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29(2), 191–215.

- <http://doi.org/10.1006/jpho.2001.0134>
- Rato, A., Rauber, A. (2015). The effects of perceptual training on the production of English vowel contrasts by Portuguese learners. In: *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Renwick, M. E. L., Ladd, D. R. (2016). Phonetic distinctiveness vs. lexical contrastiveness in non-robust phonemic contrasts. *Laboratory Phonology*, 7(19), 1–29.  
<http://doi.org/10.5334/labphon.17>
- Sadakata, M., McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America*, 134, 1324–1335.  
[http://doi: 10.1121/1.4812767](http://doi:10.1121/1.4812767)
- Sakai, M., Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, 39(1), 187–224.  
<http://doi:10.1017/S0142716417000418>
- Sawusch, J. R (2005). Acoustic analysis and synthesis of speech. In: D. B. Pisoni and R. E. Remez (Eds), (pp. 7–27), *The Handbook of Speech Perception*, Malden, MA: Blackwell.
- Sebastián-Gallés, N. (2005). Cross-language speech perception. In: D.B. Pisoni, & R.E. Remez (Eds.), (pp. 546–566). *The Handbook of Speech Perception*, Malden, MA: Blackwell.
- Sebastián-Gallés, N., Baus, C. (2005). On the relationship between perception and production in L2 categories. In: A. Cutler (Ed), (pp. 279–292), *Twenty-first century psycholinguistics: Four cornerstones*, New York: Erlbaum.
- Sheldon, A., Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3(3), 243–261.  
<http://doi.org/10.1017/S0142716400001417>
- Schmidt, A.M., Flege, J.E. (1995). Effects of speaking rate changes on native and non-native production. *Phonetica*, 52, 41–54.
- Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A. T., Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *The Journal of the Acoustical Society of America*, 109(4), 1691–1704. <http://doi.org/10.1121/1.1353594>

- Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, 33(3), 263–290.  
<http://doi.org/10.1016/j.wocn.2004.10.002>
- Turnbull, R. (2017). The role of predictability in intonational variability. *Language and Speech*, 60(1), 123–153.  
<http://doi.org/10.1177/0023830916647079>
- Werker, J. F., Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37(1), 35–44.  
<http://doi.org/10.3758/BF03207136>
- Wong, J. W. S. (2013). The effects of perceptual and or productive training on the perception and production of English vowels /I/ and /i:/ by Cantonese ESL learners. In: *Proceedings of Interspeech* (pp.2113–2117).
- Wong, J. W. S. (2015). Comparing the perceptual training effects on the perception and production of English high-front and high-back vowel contrasts by Cantonese ESL Learners. In: *Proceedings of the 18th International Congress of Phonetic Sciences* (pp.107-111).
- Zhang, K., Peng, G. (2017). The relationship between the perception and production of non-native tones. In: *Proceeding of Interspeech* (pp.1799-1803).

## **2.4 Conclusion**

In this Chapter we discussed the complex methodological issues one must face when examining and comparing speech perception and production. We proposed a well-controlled experimental paradigm aimed at solving at least the most problematic of these issues. Using this paradigm, we showed that there is a link between perception and production in the phonological processing of French /u/-/y/ by English natives. However, we found evidence that perception and production are related only when tested at the same level of processing. Finally, additional analyses revealed that good perception is a prerequisite for good production, in accordance with one of the most influential L2 learning models, the Speech Learning Model (Flege, 1995).



# **Chapter 3: Non-native sound perception across levels of processing: the perception of English /h/ by French natives**

## **3.1 Introduction**

As discussed in the previous Chapter, L2 phonological processing might have different degrees of difficulty at the prelexical vs the lexical level. In this Chapter we will further investigate the differences of performance across levels of processing, by focusing on L2 perception and its development. We will examine the perception of the English sound /h/ by French learners of English. This sound does not exist in French and it has been reported to cause perceptual difficulties at the prelexical level, as French learners confuse it with silence (Mah et al., 2016). In production, there is evidence that French learners of English delete /h/ in some cases and insert it in the wrong position in others, although deletions happen much more often than insertions (Janda & Auger, 1992).

In part 3.2 we examine whether perceptual problems with /h/ previously found in French learners of English at the prelexical level also persist at the lexical level. We also explore if the asymmetry reported in production (more deletions than insertions) is also observed in perception. If an asymmetry is found, our data could bring additional evidence to the ongoing discussion on the source of such asymmetrical lexical access in second language learners: namely, on the hypotheses of imprecise lexical representations versus inaccurate phonetic perception.

In part 3.3 we present an online phonetic training study, which looks at whether training on the English sound /h/ at the prelexical level can enhance the perception of this sound both prelexically and lexically. Moreover, we retest the participants in a retention test four months after posttest to examine whether the learning effects, if any, are robust. An improvement after training in both prelexical and lexical perception would suggest that phonetic training can effectively enhance word recognition, pointing to a particular link between the development of perception at prelexical and lexical levels during learning.

Finally, in part 3.4 we explore the link between asymmetries at the prelexical and lexical levels. First, we focus on data from the training study (section 3.3) and investigate if asymmetries are found at pretest at both levels of processing and whether training induces changes in these asymmetries in posttest. Second, we compare the findings from the training study with the results from the Melnik & Peperkamp (2019) study, reported in part 3.2. We discuss the results in light of models of speech perception.

### **3.2 Perceptual deletion and asymmetric lexical access in second language learners – Melnik & Peperkamp (2019)**

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



# Perceptual deletion and asymmetric lexical access in second language learners

Gerda Ana Melnik<sup>a)</sup> and Sharon Peperkamp

Laboratoire de Sciences Cognitives et Psycholinguistique (Ecole Normale Supérieure - Université Paris Sciences et Lettres, Ecole des Hautes Etudes en Sciences Sociales, Centre National de la Recherche Scientifique), Département d'Etudes Cognitives, 29 rue d'Ulm, 75005 Paris, France  
[gerda.ana.melnik@ens.fr](mailto:gerda.ana.melnik@ens.fr), [sharon.peperkamp@ens.fr](mailto:sharon.peperkamp@ens.fr)

**Abstract:** French listeners have difficulty perceiving /h/, and as L2 speakers they frequently omit /h/ in English words. This study investigated their processing of English /h/-initial words. Participants performed a lexical decision task on words and nonwords, where the nonwords were created from /h/-initial and vowel-initial words by removing or adding /h/, respectively. The results mirrored the production pattern, with more misses on /h/-initial words (e.g., holiday) and more false alarms on vowel-initial nonwords (e.g., usband). These results are interpreted in light of previous research on asymmetries in L2 lexical access.

© 2019 Acoustical Society of America

[AL]

Date Received: October 8, 2018      Date Accepted: December 11, 2018

## 1. Introduction

The phonology of our native language shapes our perception of foreign sounds, often resulting in perceptual distortions. In bilingual speakers such perceptual distortions have a detrimental effect on lexical processing in the non-native language. This was first shown by Pallier *et al.* (2001), who focused on early Spanish-Catalan bilinguals' processing of the Catalan /e/-/ɛ/ contrast. Using a lexical decision task with long-term repetition priming, they found that for these bilinguals, minimal pairs of words based on the /e/-/ɛ/ contrast prime one another; hence, they are treated as homophones. Likewise, in a standard lexical decision task, Spanish-Catalan bilinguals were found to have difficulty rejecting nonwords that are created from real words containing /e/ or /ɛ/ by replacing this vowel with its counterpart (Sebastián-Gallés and Baus, 2005). Similar effects in L2 lexical processing have been reported in studies using these and other paradigms (cross-modal priming, word identification, eye-tracking, semantic relatedness judgment), and focusing on a variety of bilingual populations with various levels of proficiency, including Dutch-English (Weber and Cutler, 2004; Escudero *et al.*, 2008; Broersma and Cutler, 2011; Díaz *et al.*, 2012), Japanese-English (Cutler *et al.*, 2006; Ota *et al.*, 2009), English-Japanese (Darcy *et al.*, 2013), Cantonese-Mandarin (Zhang *et al.*, 2012), and English-German (Barrios *et al.*, 2016).

Many of the above-mentioned studies have reported asymmetries, such that performance is better on one member of the L2 contrast than on the other one (Weber and Cutler, 2004; Escudero *et al.*, 2008; Cutler *et al.*, 2006; Broersma and Cutler, 2011; Díaz *et al.*, 2012; Darcy *et al.*, 2013; Zhang *et al.*, 2012; Barrios *et al.*, 2016). For instance, Weber and Cutler (2004) examined Dutch-English bilinguals using eye-tracking with a visual world paradigm. Given the target *panda* (containing the vowel /æ/, which does not exist in Dutch), a distractor picture of a pencil (containing the vowel /e/) received longer fixations than unrelated distractors; the same did not hold when a picture of a panda was presented as a distractor for target *pencil*. Weber and Cutler (2004) argued that the /e/-/æ/ contrast is maintained in lexical representations but that whichever member is heard in the input, only words containing /e/ are initially activated, due to inaccurate phonetic perception. Thus, /e/ is said to be the active category, presumably because Dutch has a vowel /e/ but not /æ/. By contrast, Darcy *et al.* (2013) argued that perception is accurate but that the phonological representations of words containing the difficult category are—as they call it—fuzzy, i.e., imprecise. They tested American English learners of Japanese and German with a lexical decision task, and observed asymmetric performance in the processing of the Japanese contrast between singleton and geminate consonants and the German contrast between front

<sup>a)</sup> Author to whom correspondence should be addressed.

and back rounded vowels. That is, in both cases performance was better on *words* with the category that *has* an English equivalent (Japanese: singleton consonant; German: back rounded vowel) and on *nonwords* with the category that *has no* English equivalent (Japanese: geminate consonant; German: front rounded vowel). Thus, for Japanese, there were more hits on *akeru* “open” than on *kippu* “ticket,” and more false alarms on *\*kipu* than on *\*akkeru*; for German, there were more hits on *Honig* /hɔnɪç/ “honey” than on *König* /kɔnɪç/ “king,” and more false alarms on *\*Konig* /kɔnɪç/ than on *\*Hönig* /hɔnɪç/.

While asymmetries seem widespread, in some previous studies they were absent or observed only in certain experiments or conditions (Escudero *et al.*, 2008; Broersma and Cutler, 2011). Moreover, asymmetries have also been observed in control native speakers, and in some of these cases the same asymmetries were absent or smaller in the bilingual speakers (Sebastián-Gallés and Baus, 2005; Díaz *et al.*, 2012). Thus, both the presence of bilingual-specific asymmetries in word recognition and the cognitive locus of such asymmetries in either inaccurate phonetic perception and/or imprecise lexical representations remain largely open questions.

All of the studies mentioned so far concern contrasts between two sounds in L2 that are difficult to distinguish. Here, we are interested in a case in which an L2 sound is not confused with another one, but rather, with its absence. Specifically, French lacks /h/, and native French speakers have difficulty perceiving the contrast between the presence vs absence of /h/ in English stimuli (Mah *et al.*, 2016). Of course, in a French–English bilingual’s daily life, the perception of the presence vs absence of /h/ is completely asymmetric: French speakers have difficulty perceiving /h/ when it is present but do not hallucinate it when it is absent. This asymmetry is reflected in their English production. That is, while they occasionally insert /h/ (e.g., *hafter* instead of *after*), they much more frequently omit /h/ (*eadache* instead of *headache*) (Janda and Auger, 1992).<sup>1</sup> Given the inherent asymmetry of the contrast between the presence vs absence of /h/, we predict asymmetric lexical processing of English /h/-initial and vowel-initial words in French listeners. We also predict that in native English listeners, this asymmetry is—if anything—smaller.

Previous work has revealed difficulty in the processing of /h/ at the lexical level in French listeners: in an electroencephalography study by White *et al.* (2017), French–English bilinguals performed a semantic classification task on words and nonwords, where the nonwords were created from /h/- and vowel-initial words by removing or adding /h/, respectively. Crucially, they were not informed that the items contained nonwords as well as real words. Low-proficiency bilinguals failed to show an N400 nonword effect; thus, they processed the nonwords as if they were real words. As /h/-initial and vowel-initial nonwords were not analyzed separately, the question of a possible asymmetry was not addressed. Here, we report on an experiment with French–English bilinguals, using a lexical decision task in which, as in White *et al.* (2017), nonwords differ from real words in the presence or absence of /h/. We predict, first, that nonwords such as *usband* and *hoffer* will activate the representations of the real words *husband* and *officer*, respectively, and thus will cause false alarms. Moreover, given that French speakers produce *usband* more often than *hoffer* (and are exposed to it more often in the speech of other French speakers), we predict higher false alarm rates on the former than on the latter. Finally, for the same reason, we also predict more misses on *husband* than on *officer*, despite relatively good performance on real words.

## 2. Methods

### 2.1 Participants

Forty-one French intermediate to advanced learners of English participated. There were 20 women and 21 men, aged between 18 and 34 (mean: 25.6), who had started learning English at school. They filled in a questionnaire to self-evaluate their speaking, listening, reading, vocabulary, and grammar skills in English and French, on a scale from 1 to 10. The overall mean score was 7.4 [standard deviation (SD)=1.2] for English and 9.8 (SD=0.1) for French. In addition, 26 native speakers of American, British, or Canadian English, 13 women and 13 men, aged between 23 and 34 (mean: 26.8), were tested as controls. None of the participants reported a history of speech or language problems. They were paid a small fee for their participation.

### 2.2 Stimuli

We selected 80 English test words, 40 starting with /h/ (e.g., *husband*) and 40 with a vowel (e.g., *officer*). They consisted of nouns, verbs, and adjectives, and contained between two and four syllables. The /h/-initial and vowel-initial words did not differ in



mean frequency in the Subtlex database (Brysbaert and New, 2009) or in mean number of syllables (both  $t < 1$ ). They also did not differ in mean familiarity, as rated by a group of 45 adult French learners of English who did not participate in the experiment ( $t = 1.0$ ,  $p > 0.1$ ). It was not possible to match for both frequency and cognate status; about one-quarter of the h-initial and half of the vowel-initial items had a French cognate, but none of these cognates were (near-)homophones.

For each word, we created a paired nonword by deleting or adding /h/ at the beginning (e.g., *husband* → *usband*, *officer* → *hofficer*). In addition, we selected 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. For each control word, we created a paired nonword by replacing, deleting, or inserting one phoneme other than /h/.

The test and control items were divided into four lists.<sup>2</sup> Each list 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. No list contained both members of a given word–nonword pair. Finally, two additional words and two additional nonwords, none involving /h/, were selected for a practice phase.

All items were recorded by a male native speaker of American English in a soundproof booth, at 16 bits mono with a sampling rate of 44.1 kHz.

### 2.3 Procedure

Participants were tested individually in a soundproof booth. They were randomly assigned to one of the four lists. Hence, participants heard only one of the members of each word–nonword pair in both the test and control condition. They performed an 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.

### 3. Results

Prior to analysis, we inspected the performance of English participants on words and nonwords separately to detect outlier items. Items were discarded if their rate of correct responses was three SDs below the mean of the corresponding items (words or nonwords). Two real words (*advise*, *holly*) and three nonwords (*essintial*, *acclord*, *advin-ture*) were thus discarded.

Mean accuracy for French and English participants on /h/-type items (e.g., *husband*, *usband*), vowel-type items (e.g., *officer*, *hofficer*) and control items, separated by lexiconality, are shown in Table 1.

We started with a global analysis of accuracy, including both test and control items. As French participants had a strong bias for yes-responses (shown by their low accuracy scores on test nonwords), we used the  $A'$  statistic. This statistic provides a non-parametric, unbiased, index of sensitivity (here: to the difference between words and nonwords), with 0.5 indicating chance performance and 1.0 perfect performance. The data from one participant in the French group were excluded, as this participant’s

Table 1. Mean accuracy on test and control items for French learners of English and native English speakers (standard errors in parentheses).

	Words			Nonwords		
	h-type (husband)	vowel-type (officer)	control	h-type (usband)	vowel-type (hofficer)	control
French	88.11 (5.12)	92.07 (4.28)	95.86 (3.15)	32.98 (7.44)	47.14 (7.90)	82.44 (6.02)
English	92.31 (5.24)	97.67 (2.96)	97.71 (2.93)	93.70 (4.77)	94.14 (4.61)	97.29 (3.19)

Table 2. Mean  $A'$  scores on test and control items for French learners of English and native English speakers (standard deviations in parentheses).

	test	control
French	0.72 (0.20)	0.94 (0.04)
English	0.97 (0.03)	0.99 (0.01)

$A'$  score was more than three SDs below the group mean. Mean  $A'$  scores for the French and English groups are shown in Table 2.

An analysis of variance by participant with the factors Group (French vs English), Condition (test vs control), and Item List (1 vs 2 vs 3 vs 4) revealed main effects of Group [ $F(1,58)=42.3$ ,  $p<0.001$ ] and Condition [ $F(1,58)=54.6$ ,  $p<0.001$ ], as well as a Group  $\times$  Condition interaction [ $F(1,64)=28.8$ ,  $p<0.001$ ]. Restricted analyses revealed that the interaction was due to the fact that although there was an effect of Condition in both language groups, with better performance on control than on test items, this effect was larger in the French group [French:  $F(1,39)=53.3$ ,  $p<0.001$ ; English:  $F(1,25)=10.8$ ,  $p<0.01$ ]. As there was no effect of nor an interaction with the counterbalancing factor Item List, this factor was omitted in further analyses.

Next, we focused on the test items. The  $A'$  statistic cannot be used to test for possible asymmetries in the performance on h-type vs vowel-type items in words vs nonwords. Rather, a linear mixed effects model was run on the accuracy scores, with Group (French vs English), Lexicality (word vs nonword), and Item Type (/h/ vs vowel) as contrast-coded fixed factors, and intercepts for participants and items as random factors.  $P$ -values were obtained by likelihood ratio tests of the full model against the model without the effect or interaction in question. All three main effects were significant, with higher accuracy by English than by French participants [ $\beta=2.21$ , standard error (SE)=0.33,  $\chi^2(1)=37.4$ ,  $p<0.0001$ ], on words than on nonwords [ $\beta=1.81$ , SE=0.18,  $\chi^2(1)=103.8$ ,  $p<0.0001$ ] and on vowel-type than on h-type items [ $\beta=0.65$ , SE=0.20,  $\chi^2(1)=10.3$ ,  $p=0.001$ ]. There was also a Lexicality  $\times$  Group interaction [ $\beta=2.76$ , SE=0.35,  $\chi^2(1)=43.0$ ,  $p<0.0001$ ], due to the fact that French learners performed better on words than on nonwords whereas there was no difference for English speakers. Crucially, the triple interaction between Group, Lexicality, and Item Type was also significant [ $\beta=1.40$ , SE=0.70,  $\chi^2(1)=4.17$ ,  $p=0.04$ ]. This triple interaction was examined in two mixed effects models on the French and English data, respectively, with Lexicality and Item Type as contrast-coded fixed factors, and intercepts for participants and items as random factors. For the French data, Cognate Status was also added as a contrast-coded fixed factor. For French learners there were main effects of Lexicality [ $\beta=3.18$ , SE=0.18,  $\chi^2(1)=508.8$ ,  $p<0.0001$ ] and Item Type [ $\beta=0.71$ , SE=0.18,  $\chi^2(1)=13.9$ ,  $p<0.0001$ ] but no interaction; performance was better on words than on nonwords and better on vowel-type items than on /h/-type items. There was no effect of Cognate Status ( $p>0.1$ ). For native English speakers, there was no effect of either Lexicality or Item Type but an interaction between these factors [ $\beta=1.28$ , SE=0.65,  $\chi^2(1)=4.07$ ,  $p=0.04$ ]. For real words, they were more accurate on vowel-type than on /h/-type items, while for nonwords they showed no such difference.

#### 4. Discussion

We used a lexical decision task to test whether French learners of English encounter difficulty in processing /h/ at the lexical level, and whether they show the same asymmetry as observed in production (i.e., more incorrect deletions than insertions of /h/; Janda and Auger, 1992). While both the French learners and native English control participants performed better on control than on test items, this difference was much larger for the French learners. Thus, we found that French learners of English indeed have difficulty in the processing of /h/ at the lexical level, in agreement with White *et al.* (2017).<sup>3</sup> Furthermore, French learners made more errors on /h/-type items than on vowel-type items, both in words and nonwords, while native English speakers showed the same difference, but of a smaller magnitude and only in real words. Thus, as predicted, we observed an asymmetry in French learners that parallels the one observed in production. That is, just like these learners typically produce more /h/-initial words without the /h/ (e.g., *usband* instead of *husband*) than that they produce vowel-initial words with an added initial /h/ (e.g., *hofficer* instead of *officer*), they show more false alarms on *usband*-type than *hofficer*-type nonwords and more misses on *husband*- than on *officer*-type real words.

Similar asymmetries were reported in two case studies by Darcy *et al.* (2013), who also used a lexical decision task. They argued that these asymmetries are due to what they called fuzzy, i.e., imprecise, lexical representations of words containing the difficult category (i.e., the one without a native equivalent), and not to problems in phonetic perception. Based on data from a visual world paradigm, however, Weber and Cutler (2004) and Cutler *et al.* (2006) argued for the opposite view, namely, that asymmetries in L2 learners indicate inaccurate phonetic perception of the difficult category combined with accurate lexical representations. These two views are incompatible yet congruent with the respective datasets from lexical decision and the visual world paradigm, respectively, that they were meant to account for. To see why, first consider the present lexical decision data (the data of Darcy *et al.* show the same pattern): Lexical representations of /h/-initial words being less precise than those of vowel-initial words leads French listeners to be less accurate on h-type items (*husband/usband*) than on vowel-type items (*officer/hofficer*). By contrast, if the asymmetries were exclusively due to inaccurate perception of /h/, French learners should have performed worse on *hofficer*-type than on *usband*-type nonwords, contrary to fact. That is, failing to perceive the /h/ in *hofficer* should have led them to generally accept it as a real word, whereas they should have had no particular difficulty rejecting *usband* given their correct lexical representation of the word *husband*. (For real words, the prediction would be correct: that is, more errors on *husband*-type than on *officer*-type words, since a failure to perceive /h/ should make it more difficult to accept the former than the latter.)

As to the data from the visual world paradigm, Weber and Cutler (2004) and Cutler *et al.* (2006) found that whichever member of a difficult L2 contrast is heard, only words containing the category that has a native equivalent are initially activated. For instance, Dutch learners of English, whose native language has /ɛ/ but not /æ/, were found to briefly activate the word *pencil* upon hearing *panda* but not the reverse, supposedly due to misperception of the /æ/ of *panda* as the /ɛ/ of *pencil*. If perception were accurate, but the lexical representation of words containing /æ/ imprecise, the expected pattern of results would be different; there should be no brief activation of *pencil* upon hearing *panda*, but both *panda* and *pencil* should initially be activated upon hearing *pencil*.

To conclude, there are two contrasting accounts of asymmetric lexical access in second language learners: imprecise lexical representations versus inaccurate phonetic perception. Our data are in accordance with the former: they suggest that intermediate to advanced French learners—in addition to the difficulty they may have with the perception of /h/—have imprecise lexical representations of /h/-initial words. More comparative research with different contrasts, tasks, and populations is needed to further examine asymmetries in L2 lexical processing. As such asymmetries can also occur in native speakers (Sebastián-Gallés and Baus, 2005; Díaz *et al.*, 2012), it would be important to always compare the learners' performance to that of native speakers, such as to clearly identify asymmetries that are specific to L2 processing.

### Acknowledgments

This research was supported by a Ph.D. fellowship from Ecole normale supérieure to G.A.M. and by grants from the Agence Nationale de la Recherche (Grant Nos. ANR-17-EURE-0017 and ANR-17-CE28-0007-01). We would like to thank Jeremy Kuhn for recording the stimuli.

### References and links

<sup>1</sup>As argued by Janda and Auger (1992), insertion is a sign of hypercorrection, showing that French speakers have metalinguistic knowledge about /h/ in English.

<sup>2</sup>We used four rather than two counterbalanced lists, such that the stimuli can be used in a pre- and a post-test of a planned training study.

<sup>3</sup>In White *et al.* (2017) this processing problem was observed only in learners who, based on their overall pronunciation, were rated as having low proficiency. It is difficult to compare proficiency across the two groups, as the classification of our participants as intermediate to advanced was based on their self-evaluation of a range of skills. Note also that the participants in White *et al.* lived and were tested in Montreal, a French–English bilingual city; they thus received larger amounts of native English input than our participants, who lived and were tested in Paris.

Barrios, S. L., Namyst, A. M., Lau, E. F., Feldman, N. H., and Idsardi, W. J. (2016). "Establishing new mappings between familiar phones: Neural and behavioral evidence for early automatic processing of nonnative contrasts," *Front. Psychol.* 7, 1–16.

Broersma, M., and Cutler, A. (2011). "Competition dynamics of second-language listening," *Q. J. Exp. Psychol.* 64(1), 74–95.



- Brysbart, M., and 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," *Behav. Res. Methods* **41**(4), 977–990.
- Cutler, A., Weber, A., and Otake, T. (2006). "Asymmetric mapping from phonetic to lexical representations in second-language listening," *J. Phonetics* **34**(2), 269–284.
- Darcy, I., Daidone, D., and Kojima, C. (2013). "Asymmetric lexical access and fuzzy lexical representations in second language learners," *Mental Lexicon* **8**(3), 372–420.
- Díaz, B., Mitterer, H., Broersma, M., and Sebastián-Gallés, N. (2012). "Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access," *Learn. Individual Diff.* **22**(6), 680–689.
- Escudero, P., Hayes-Harb, R., and Mitterer, H. (2008). "Novel second-language words and asymmetric lexical access," *J. Phonetics* **36**, 345–360.
- Janda, R. D., and Auger, J. (1992). "Quantitative evidence, qualitative hypercorrection, sociolinguistic variables—and French speakers' headaches with English h/Ø," *Lang. Commun.* **12**(3/4), 195–236.
- Mah, J., Goad, H., and Steinhauer, K. (2016). "Using event-related brain potentials to assess perceptibility: The case of French speakers and English [h]," *Front. Psychol.* **7**, 1–14.
- Ota, M., Hartsuiker, R. J., and Haywood, S. L. (2009). "The KEY to the ROCK: Near-homophony in nonnative visual word recognition," *Cognition* **111**(2), 263–269.
- Pallier, C., Colomé, A., and Sebastián-Gallés, N. (2001). "The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries," *Psychol. Sci.* **12**(6), 445–449.
- Sebastián-Gallés, N., and Baus, C. (2005). "On the relationship between perception and production in L2 categories," in *Twenty-first Century Psycholinguistics: Four Cornerstones*, edited by A. Cutler (Erlbaum, New York), pp. 279–292.
- Weber, A., and Cutler, A. (2004). "Lexical competition in non-native spoken-word recognition," *J. Memory Lang.* **50**(1), 1–25.
- White, E. J., Titone, D., Genesee, F., and Steinhauer, K. (2017). "Phonological processing in late second language learners: The effects of proficiency and task," *Bilingualism: Lang. Cognition* **20**, 162–183.
- Zhang, X., Samuel, A. G., and Liu, S. (2012). "The perception and representation of segmental and prosodic Mandarin contrasts in native speakers of Cantonese," *J. Memory Lang.* **66**(3), 438–457.



### 3.3 The effect of phonetic training on L2 word recognition

This section constitutes the following manuscript: Melnik, G. A., Peperkamp, S. (Submitted). High Variability Phonetic training enhances L2 lexical processing: evidence from training French learners of English online. Submitted to: *Journal of Phonetics*.

A condensed version of this paper will appear in the *Proceedings of CogSci 2019* and is included as Appendix B of this thesis.

#### Abstract

High-Variability Phonetic Training (HVPT) has been shown to be effective in improving the perception of even the hardest second-language (L2) contrasts. However, little is known as to whether such training can improve phonological processing at the lexical level. The present study tested whether this type of training also improves word recognition. Adult proficient French late learners of English completed eight online sessions of HVPT on the perception of English word-initial /h/. This sound does not exist in French and has been shown to be difficult to process by French listeners both on the prelexical (Mah, Goad & Steinhauer, 2016) and the lexical level (Anonymous, 2019). In pretest and posttest participants were administered both a prelexical identification task and a lexical decision task. The results demonstrate that after training the learners' accuracy had improved in both tasks. Moreover, four months after posttest, participants completed a retention test, showing that the positive effects of phonetic training did not reduce in either task. This is the first evidence that even short training results in gains not only in prelexical perception, but also in word recognition.

#### 3.3.1 Introduction

A common assumption in second language research is that producing and perceiving non-native speech sounds can be problematic (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 /ɪ/; consequently, Japanese listeners fail to perceive these sounds as different. Yet, auditory training on Japanese learners has proved successful (e.g., McCandliss et al. 2002; Iverson et al. 2005; Zhang et al., 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). HVPT uses multiple natural exemplars of the target sounds in a variety of phonetic environments. This variability enhances the process of building novel phonological categories. Importantly, perceptual training involves immediate corrective feedback that provides information to participants about their performance and promotes rapid learning by driving the learner's attention to the relevant phonetic cues of the sounds to be learned (Homa & Cultice, 1984; Logan, Lively & Pisoni, 1991). The effectiveness of this technique has been shown in many studies in a variety of languages, using several target contrasts and structures, including vowels (Carlet & Cebrian, 2014; Lee & Lyster, 2016), consonants (Kim & Hazan, 2010; Shinohara & Iverson, 2018), tones (Wang et al. 1999; Wang et al., 2003), and syllable structure (Huensch & Tremblay, 2015). Moreover, both high- and low-proficiency speakers benefit from HVPT (Iverson, Pinet & Evans, 2012), and HVPT generalizes to new tokens and new speakers (Lively et al., 1994; Okuno & Hardison, 2016). Finally, it gives rise to long-term retention of the new categories (Lively et al., 1994), and it helps to improve L2 production (for a review, see Sakai & Moorman, 2018).

Several studies investigated at which level of processing the benefits of HVPT occur. For instance, Sadakata & McQueen (2013) studied whether auditory training affects the

formation of phonetic categories or, rather, the sensitivity to pre-categorical acoustic information. They showed that more accurate perception following HVPT was due to enhanced and more robust abstract sound categories and not to improved acoustic sensitivity. Shinohara & Iverson (2018) further tested if combining different training methods leads to improvements of different underlying processes. However, they found that both identification and discrimination training enhanced L2 perception in a similar way (both methods of training improved learners' identification, auditory discrimination, and category discrimination accuracy). Finally, Iverson & Evans (2009) found that despite improvements in the identification accuracy following training, participants did not change their judgments of what are the best exemplars of the non-native sounds. They therefore suggested that training improves and automatizes the learners' capacity

to apply their existing category knowledge to natural variable speech without necessarily changing the use of relevant acoustic cues.

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 in 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 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 require higher memory demands and 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 (Sebastián-Gallés & Baus, 2005; Díaz et al., 2012).

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

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

real words (Anonymous, 2019). Thus, they have difficulty not only in perceiving the contrast between /h/ and silence, but also in distinguishing between words and nonwords that differ only in the presence vs. absence of /h/.

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

In the current study we trained French learners on the perception of English /h/ in a pretest–training–posttest design. 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. In 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 in the controls words. 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 non-words (false alarms) compared to the control items, as previously shown in (Anonymous, 2019). In posttest, both tasks were repeated with the same stimuli, but the identification task was supplemented by trials with novel items, such as to test for generalization. Four months after the posttest, participants returned to the lab for a long-term retention test, which was identical to the posttest.

Training was administered online, and 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, 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 (Lively et al., 1994; Lee & Lyster, 2016). 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.

### **3.3.2 Methods**

#### **3.3.2.1 Pretest-Posttest-Generalization: Identification in nonwords**

##### **3.3.2.1.1 Stimuli**

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

An additional 30 pairs of nonwords (10 monosyllabic, 10 disyllabic and 10 trisyllabic, containing the 10 vowels mentioned above) were selected to test for generalization at the end of the posttest.

Half of the pairs were recorded by a male, and the other half by a female native of American English. Table 3.1 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.

TABLE 3.1. 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>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>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)

\*:  $p < .05$

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

### 3.3.2.1.2 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/. There were 194 trials divided over two blocks. Trials were presented in a semi-random order such that no more than four trials of the same type (vowel-initial or /h/-initial) and no more than three trials recorded by the same speaker appeared in a row.

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

At the end of the posttest only, participants performed the same task in 60 trials with the 30 additional nonword pairs.

### **3.3.2.2 Pretest-Posttest: Lexical decision**

#### **3.3.2.2.1 Stimuli**

The stimuli were the same as in Anonymous (2019). They consisted of 80 English test words, 40 starting with /h/ (e.g., husband) and 40 with a vowel (e.g. officer), recorded by the same male American English speaker who recorded stimuli for the identification task. They consisted of nouns, verbs and adjectives, and contained between two and four syllables. The /h/-initial and the vowel-initial words did not differ in mean frequency in the Subtlex database (Brysbaert & New, 2009) or in mean number of syllables (both  $t < 1$ ). They also did not differ in mean familiarity, as rated by a separate group of 45 adult French learners of English in a rating questionnaire ( $t = 1.0$ ,  $p > 0.1$ ). Table 3.1 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* - *\*usband*, *officer* - *\*hofficer*). In addition, there were 240 English control words (nouns, verbs and adjectives), none of which starting with /h/. They were matched for mean

frequency and mean number of syllables with the test words. Each control word was paired with a nonword created by replacing, deleting or inserting one phoneme other than /h/.

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

#### 3.3.2.2.2 Procedure

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

### **3.3.2.3 Training: Identification in real words**

#### **3.3.2.3.1 Stimuli**

We selected 59 minimal pairs of real words differing in the presence or absence of an initial /h/. Given the limited number of such minimal pairs, we used both frequent words (e.g. *hair-air*) and infrequent ones (e.g. *hosier-osier*). However, word frequency was not considered to have an impact, as the task used in training did not involve 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 3.1 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.

#### **3.3.2.3.2 Procedure**

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

All training sessions were run at the participants' homes through internet. The online training sessions were designed using the JsPsych library (de Leeuw, 2015) in JavaScript. Before each 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 was 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.

#### **3.3.2.4 Participants**

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

### **3.3.3 Results and discussion**

#### **3.3.3.1 Pretest-Posttest-Generalization: Identification**

Prior to analysis, we discarded trials with no response or time-out (2.4 % of the data). Figure 3.1 displays the identification accuracy of participants in pretest, posttest and generalization. As the identification task is a signal detection task, we used the A' statistic, which provides a non-parametric, unbiased, index of sensitivity, with 0.5 indicating chance performance and 1.0 perfect performance. A repeated-measures ANOVA by participant with the factor Session (Pretest vs. Posttest vs. Generalization), revealed a main effect of Session ( $F(2,46) = 26.75, p < .001$ ), with the accuracy improving from an average A' score of 0.74 in pretest to 0.86 in posttest and 0.86 in generalization. Bonferroni-adjusted pairwise t-tests revealed that there was a significant difference between pretest and posttest ( $p < .01$ ), as well as between pretest and generalization ( $p < .01$ ). There was no difference between the performance in the posttest and in the generalization ( $p = .82$ ).

We also examined if performance was influenced by the acoustic properties of the stimuli produced by each speaker. Recall from Table 3.1 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 vs. 0.40;  $t(79.65) = 7.83, p < 0.001$ ). Performance on stimuli produced by each of the two speakers, however, did not differ ( $\text{Mean}_{\text{male\_speaker1}} = 76.03\%$  correct,  $\text{Mean}_{\text{fem\_speaker2}} = 76.47\%$  correct,  $p > 0.1$ ).

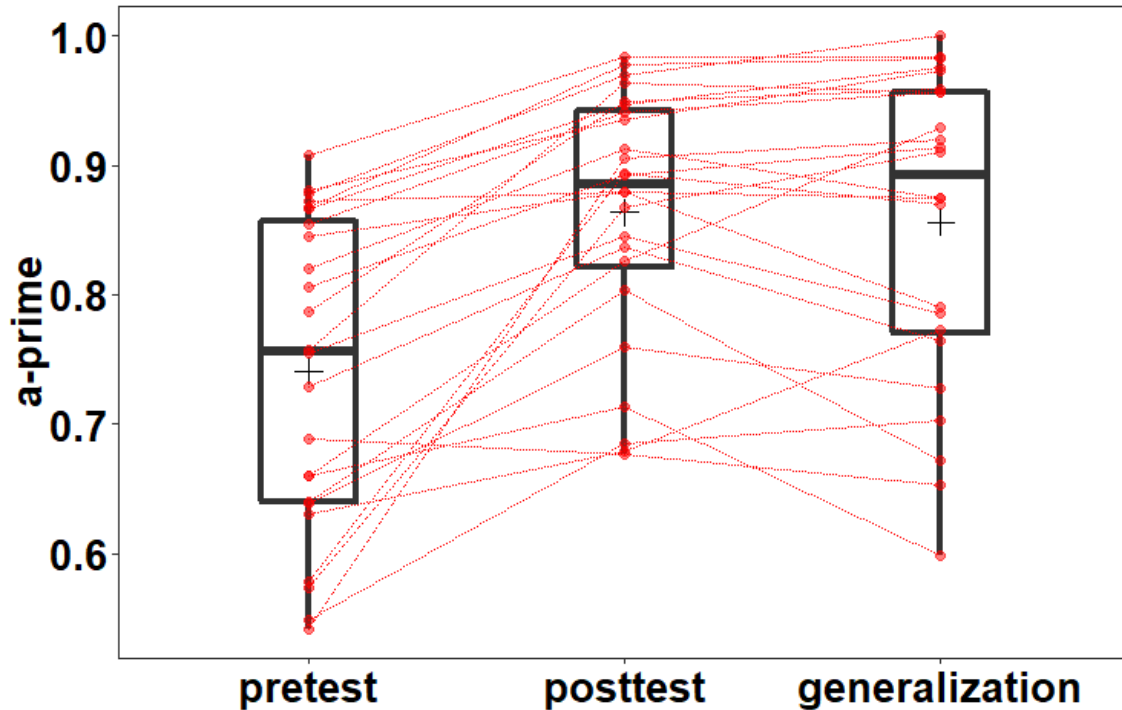


Figure 3.1. Boxplots of A' scores in the identification task in pretest, posttest, and generalization. The red dots represent individual participants' scores; the lines link each participant's scores in the three parts of the tests. The black cross marks indicate mean A' scores.

### 3.3.3.2 Retention: Identification

For the 21 participants who returned to the laboratory four months after posttest, a repeated-measures ANOVA by participant with the factor Session (Posttest vs. 4-months delayed posttest) revealed an effect ( $F(1,20) = 9.25, p < 0.01$ ), 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 change after a period of four months.

### 3.3.3.3 Pretest-Posttest: Lexical Decision

Prior to analysis, we discarded responses with no response or time-out (1.3 % of the data). Figure 3.2 displays the participants' accuracy on the test and control items in pretest and posttest. As the participants had a strong bias for "yes"-responses (shown by their low accuracy

scores on test nonwords), we used the A' statistic, as in the analysis of performance in the identification task.

We carried out a repeated-measures ANOVA by participant with the factors Session (pretest vs. posttest), Condition (test vs. control) and Lists (AC vs. BD), as well as an interaction between Session and Condition. We found main effects of Session ( $F(1, 23) = 39.36, p < .001$ ) and Condition ( $F(1, 23) = 73.93, p < .001$ ), and a Session  $\times$  Condition interaction ( $F(1, 23) = 30.87, p < .001$ ). 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 ( $r = 0.41, p = 0.04$ ).

We also examined if performance was influenced by the acoustic properties of word and nonword stimuli. Recall from Table 3.1 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,  $t(75.38) = 4.00, p < 0.0001$ ; Mean ratio: words = 0.52, nonwords = 0.56,  $t(76.32) = 2.33, p = 0.02$ ). 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.

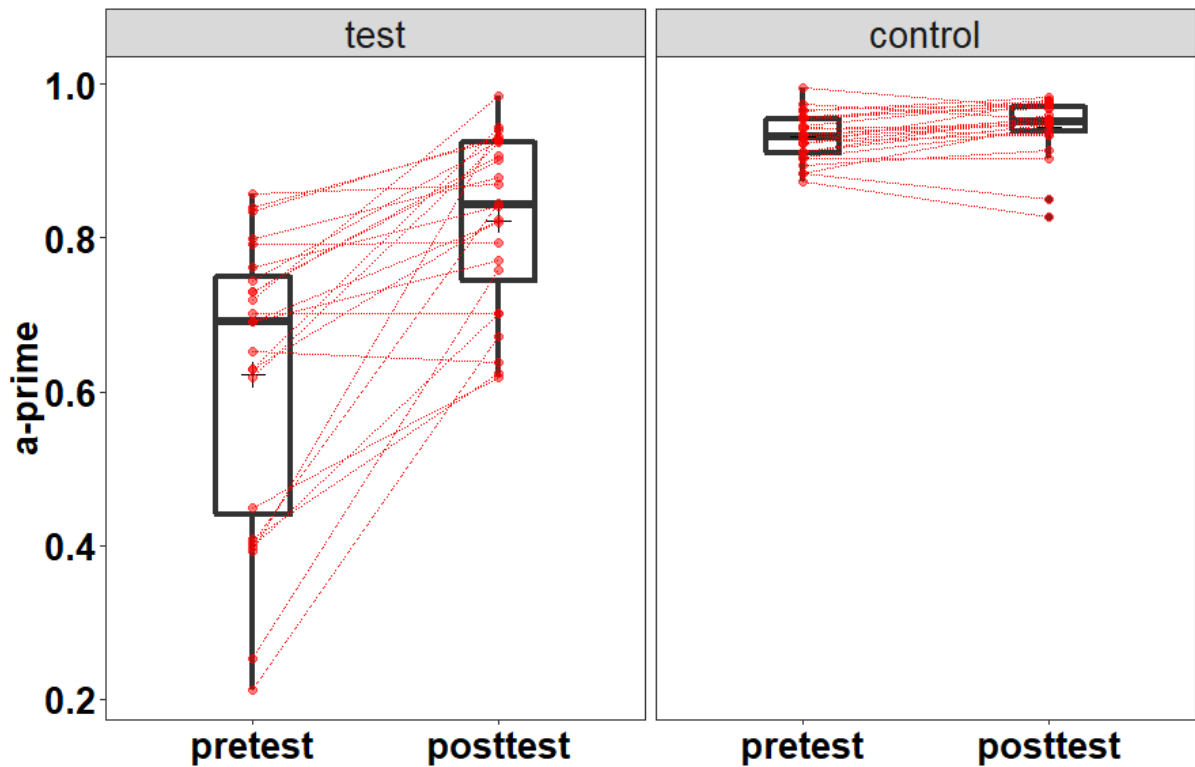


Figure 3.2. Boxplots of A' scores in the lexical decision task in pretest and posttest for test (left panel) and control (right panel) conditions. The red dots represent individual participants' scores; the lines link each participant's scores in both sessions. The black cross marks indicate mean A' scores.

#### 3.3.3.4 Retention: Lexical Decision

To examine retention of the training improvements after four months in the 21 participants who returned for the retention test, we carried out a repeated-measures ANOVA by participant with the factors Session (posttest vs. 4-months delayed posttest), Condition (test vs. control), as well as an interaction between Session and Condition. We found a main effect of Condition ( $F(1, 20) = 31.06, p < 0.001$ ;  $\text{mean}_{\text{test}} = 0.81, \text{mean}_{\text{control}} = 0.94$ ) but no effect of Session ( $F(1, 20) < 1$ ), and no interaction ( $F(1, 20) < 1$ ). Thus, overall performance in the lexical decision task did not significantly reduce 4 months after the immediate posttest.

### 3.3.4 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 intermediate level 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 it with silence (Mah et al., 2016). Eight sessions of High-Variability Phonetic training were administered to the participants online. They were tested in pretest, posttest and a retention test by means of an identification and a lexical decision task.

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

Finally, results from the retention test showed that the positive effects of training did not decrease 4 months after the posttest. This suggests that learning induced by auditory training is robust both at the prelexical and the lexical levels of processing. In identification there was a quite surprising improvement of participants (although not large, A-prime difference of 0.03) in

the retention test compared to the posttest. One explanation for this could be 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, it is possible, that their perception of /h/ further continued improving due to subsequent exposure to English outside the laboratory. High variability phonetic training and naturalistic learning have been shown to provide different types of improvements and thus are thought to be complementary (Iverson & Evans, 2009; Iverson et al., 2012).

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

Hisagi and colleagues (2010) investigated the role of focused attention needed in processing difficult L2 sounds. They tested American-English listeners on a Japanese vowel duration contrast and found that directing the attention of participants to the non-native contrast led to higher MMNs (which indicate automatic pre-attentive discrimination) than when their attention was directed away to a visual oddball task.



According to the Automatic Selective Perception model, the processing of simple tasks becomes more automatic and nativelike as proficiency grows. Thus, in our study, training possibly rendered the prelexical processing more efficient, thus allowing participants to allocate more cognitive resources to the lexical level of processing. A similar finding on the benefit of phonetic training on the automatization of phonetic processes was reported in a study on the perception of L2 speech in noise (Lengeris & Hazan, 2010). Adverse listening conditions such as a high signal-to-noise ratios (SNRs) have been shown to involve increased cognitive load (Pichora-Fuller et al., 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 et al., 2010). Lengeris & Hazan (2010) showed that HVPT in quiet improves the perception of difficult L2 sounds in noise (multi-talker babble). Finally, Shinohara & Iverson (2018) found that auditory training also improved production of the English /ɪ/-/I/ contrast by Japanese learners and suggested that training might induce automaticity of phonetic processes which in turn allows the speakers to pronounce the correct acoustic contrasts when producing speech.

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

To conclude, we showed that even short online HVPT can improve both prelexical and lexical processing of a difficult L2 sound. 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.

### 3.3.5 References

- Brysbart, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977-90. <https://doi.org/10.3758/BRM.41.4.977>
- Carlet, A. & Cebrian, J. (2014). Training Catalan speakers to identify L2 consonants and vowels: A short-term high variability training study. In: *Proceedings of the 7th International Symposium on the Acquisition of Second Language Speech* (pp. 85-98). Concordia University Working Papers in Applied Linguistics.
- Cooper, A., & Wang, Y. (2011). The influence of tonal awareness and musical experience on tone word learning. In: *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 512-515). Hong Kong, SAR, China.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1), 1-12. <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(6), 680-689. <https://doi.org/10.1016/j.lindif.2012.05.005>
- Hisagi, M., Shafer, V. L., Strange, W., Sussman, E. S. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: Behavioral and electrophysiological measures. *Brain Research*, 1360, 89-105. <https://doi.org/10.1016/j.brainres.2010.08.092>
- 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(3), 1045.  
[https://doi.org/10.1044/1092-4388\(2012/12-0024\)](https://doi.org/10.1044/1092-4388(2012/12-0024))
- 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 Journal of the Acoustical Society of America*, 118(5), 3267–3278.  
<https://doi.org/10.1121/1.2062307>
- Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, 126, 866–877. <https://doi.org/10.1121/1.3148196>
- Iverson, P., Pinet, M., Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(1), 145-160.  
<https://doi.org/10.1017/S0142716411000300>
- Kim, Y. H. & Hazan, V. (2010). Individual variability in the perceptual learning of L2 speech sounds and its cognitive correlates. In: *Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech* (pp. 251-256). Poznań, Poland.
- Lecumberri, M. L. G., Cooke, M., Cutler, A. (2010). Non-native speech perception in adverse conditions: a review. *Speech Communication*, 52(11-12), 864-886.  
<https://doi.org/10.1016/j.specom.2010.08.014>
- Lee, A. H. & Lyster, R. (2016). Effects of Different Types of Corrective Feedback on Receptive Skills in a Second Language: A Speech Perception Training Study. *Language Learning*, 66(4), 809-833.  
<https://doi.org/10.1111/lang.12167>
- Lengeris, A. & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *Journal of the Acoustical Society of America*, 128(6), 3757-3768.  
<https://doi.org/10.1121/1.3506351>

- Lively, S. E., Pisoni, D. B., Yamada, R. A., Tokhura, Y., Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, 96(4), 2076-2087.  
<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(2), 89–108. <https://doi.org/10.3758/CABN.2.2.89>
- Okuno, T. & Hardison, D. M. (2016). Perception-production link in L2 Japanese vowel duration: Training with technology. *Language Learning & Technology*, 20(2), 61-80.
- Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64(3), 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(6), 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(1), 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(2), 191-215.  
<https://doi.org/10.1006/jpho.2001.0134>

- Pisoni, D. & Luce, P. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, 25(1-2), 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. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America*, 134(2), 1324–1335.  
<https://doi.org/10.1121/1.4812767>
- Sakai, M. & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, 39(1), 187-224. <https://doi.org/10.1017/S0142716417000418>
- Sebastián-Gallés, N. (2005). Cross-language speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception*. Oxford: Blackwell.
- Shinohara, Y., Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. *Journal of Phonetics*, 66, 242-251.  
<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>
- 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), 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 perception. *Perception & Psychophysics*, 37(1), 35–44.  
<https://doi.org/10.3758/BF03207136>

- Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, 75(6), 1866–1878.  
<https://doi.org/10.1121/1.39098>
- Yamada, J. (1991). The discrimination learning of the liquids /r/ and /l/ by Japanese speakers. *Journal of Psycholinguistic Research*, 20(1), 31-46.  
<https://doi.org/10.1007/BF01076918>
- Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., Kawakatsu, M., Tohkura, Y., Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *NeuroImage*, 46(1), 226–240.  
<http://doi.org/10.1016/j.neuroimage.2009.01.02>

### **3.4 The relationship between prelexical and lexical asymmetries in L2 perception**

#### **3.4.1 Introduction**

As discussed in the introduction of the thesis (section 1.2), much research reported perceptual asymmetries occurring either at the prelexical or the lexical level of processing. At the prelexical level, a number of studies showed that in vowel contrast discrimination, infants tend to have better performance when the pair of vowels is presented in one direction compared to the opposite direction, such as /y/ changing to /u/ versus /u/ changing to /y/ (for a review, see Polka & Bohn, 2003; Polka & Bohn, 2011). Further evidence showed that infants from different language backgrounds perform more accurately when the vowel changes from a more central to a more peripheral one in the F1-F2-F3 vowel space. Based on these findings, Polka & Bohn (2011) framed the Natural Referent Vowel model (NRV). According to this model, a universal bias to prefer peripheral vowels is observed in young infants, as such vowels act as reference or perceptual anchors (Polka & Bohn, 2003) in the development of the native vowel inventory. However, as the perceptual system gets attuned to the L1, the perceptual biases favoring peripheral vowels can be over-ridden in order to optimize native language perception. Thus, asymmetries fade for native categories and subsequently might disappear in adult native speakers. However, the bias for peripheral vowels might remain even in adulthood for non-native contrasts that are not used in the L1.

The Native Language Magnet Theory (Kuhl, 1991; Kuhl et al., 2008), on the other hand, suggests that asymmetries in L2 sound perception are driven by L1 sounds acting as perceptual magnets and attracting more prototypical L2 sounds. Thus, both L2 vowels and consonants that are close to the L1 prototypes in the perceptual space will be attracted to them, making those L2 sounds less distinct. Asymmetries occur when discriminating sounds which fall within an L1 category and which differ in level of prototypicality: the more similar an L2 sound is to the L1 prototype, the greater the native language magnet effect and therefore the greater the difficulty to distinguish such a sound from the L1 prototype.

Despite the recent developments of both theories aimed at explaining data on asymmetries in a variety of languages and age groups, none of them looked at asymmetries occurring at the lexical level. Similarly, no study compared asymmetries at different levels of

processing. As discussed in part 1.2.3, a separate branch of research, that of L2 phonological acquisition, also investigated perceptual asymmetries. Nevertheless, this literature focused almost exclusively on asymmetries in lexical processing, showing that in lexical tasks L2 learners are often more accurate on one member of an L2 contrast than on the other (Weber and Cutler, 2004; Escudero et al., 2008; Cutler et al., 2006; Broersma and Cutler, 2011; Diaz et al., 2012; Zhang et al., 2012; Darcy et al., 2013). As these studies did not look at prelexical asymmetries, they did not address the possibility that lexical asymmetries might be caused or at least influenced by asymmetries in prelexical perception.

TABLE 3.2. Summary of the literature review on prelexical and lexical asymmetries described in section 1.2.3. The sign “>” indicates which of the two sounds in the contrast is perceived better.

STUDY	LEVEL	ASYMMETRY	DIRECTION
<b>Japanese</b>			
Hattori & Iverson (2009)	prelexical	yes	/ɪ/ > /I/
Hattori & Iverson (2010)	prelexical	no in identification yes in discrimination F3	– /ɪ/ > /I/
Guion & Flege (2000)	prelexical	yes	/ɪ/ > /I/
Bradlow et al. (1997)	prelexical	yes, and smaller in post- than in pretest	/ɪ/ > /I/
Hazan et al. (2005)	prelexical	no (neither in pre- nor posttest)	– –
Cutler et al. (2006)	lexical	yes	/I/ > /ɪ/
<b>Dutch</b>			
Cutler et al. (2004)	prelexical	yes	/ɛ/ > /æ/
Escudero et al. (2012)	prelexical	yes	/ɛ/ > /æ/
Weber & Cutler (2004)	lexical	yes	/ɛ/ > /æ/
Broersma & Cutler (2011)	lexical	no	–
Diaz et al. (2012)	lexical	yes	/æ/ > /ɛ/

In Chapter 1 (section 1.2.3) we reviewed literature that found asymmetries at the prelexical level of processing and compared it to the results from other studies on the same



contrasts which looked at the lexical level of processing. More precisely, we focused on the perceptual problems of Japanese speakers with the English /ɪ/-/I/ contrast and the difficulties of Dutch speakers with the English /æ/-/ɛ/ contrast (see Table 3.2 for summary). The review showed a rather complex image, where for the same contrast asymmetries were found by certain studies but not others. Moreover, when such asymmetries were observed at both the prelexical and lexical levels, their direction did not always match. Similarly, only in certain cases could the direction of such asymmetries be explained by either of the two theories (NRV vs. NLM); in others, neither one of them gave accurate predictions. However, conclusions on prelexical and lexical asymmetries in perception should be drawn carefully from a comparison of studies that used different tasks and different groups of participants.

In order to obtain more robust results on the link between prelexical and lexical perceptual asymmetries, we reanalyzed the data from the training study (section 3.3) and compared them to the data from the Melnik & Peperkamp (2019) study on lexical asymmetries (section 3.2). Importantly, the data from the training study allow us to investigate asymmetries at both the prelexical and lexical levels as we tested the same participants on two tasks tapping into different levels of processing but involving the same difficult L2 contrast. Namely, we collected data on the prelexical perception and lexical perception of the English /h/ by French learners of English, in an identification and a lexical decision task, respectively. At the prelexical level, participants heard /h/- and vowel-initial nonwords and had to decide whether the nonwords started with the sound /h/ or not. Although Mah et al. (2016) already showed that French learners of English encounter difficulties when perceiving /h/-initial vs. vowel-initial stimuli, the asymmetry in prelexical perception of this contrast has not yet been tested. Concerning the perception at the lexical level of processing, we used words and nonwords in a lexical decision task, where the test nonwords were created from /h/-initial and vowel-initial words by removing or adding /h/, respectively. In section 3.2 we already documented a perceptual asymmetry in L2 lexical processing that was observed in a different group of French learners of English, using the same stimuli and the same experimental paradigm. Results revealed that participants performed worse on h-type items (*happy* - *\*appy*) compared to vowel-type items (*officer* - *\*hofficer*). We will thus retest this asymmetry on data from the training study and compare the findings with results from the prelexical task. Furthermore, the comparison of data from pretest and posttest can provide evidence on the possible changes in asymmetries induced by training. Finally, we

will compare the findings of the Melnik & Peperkamp (2019) study (section 3.2) with findings from the training study to investigate if the lexical asymmetry in the former study corresponds to the lexical asymmetry (if any) in the later one.

### **3.4.2 The phonetic properties of /h/ and predictions of existing models**

As described previously, French learners of English have great difficulty with producing (Janda & Auger, 1992) and perceiving (LaCharité & Prévost 1999) the English phoneme /h/ which does not exist in French. This sound is neither assimilated to, nor substituted by a French phoneme. In contrast to most cases in non-native sound acquisition, /h/ does not form a contrast with another sound. Instead, it is a contrast between a sound and silence.

Importantly, /h/ has particular phonetic properties that could hinder its successful perception. Phonetically, /h/ differs from other consonants in how it is produced. Although /h/ is considered to be a glottal fricative, it is sometimes suggested that /h/ has no manner features as, unlike in production of other fricatives, no turbulence is created during the production of /h/ (the air passes rather freely through the narrowed glottis and no narrowing occurs at the supralaryngeal level) (Ladefoged & Maddieson, 1996). Moreover, some researchers propose that [h] lacks the place of articulation feature, as it is produced differently depending on the properties of the following vowel (Schluter et al., 2016). Thus, although phonologically it plays a role of a consonant, phonetically [h] could be considered as a voiceless vowel with the quality of the voiced vowel that follows (Roach, 2009). These phonetic characteristics make the /h/ acoustically low in salience and thus perceptually weak.

Is, then, the difficulty with /h/ that French learners of English encounter acoustically based? In order to test this Mah et al. (2016) carried out an EEG experiment on the prelexical perception of /h/ by French learners of English and native English speakers. This experiment was based on the findings of Werker and Tees (1984), showing that adults discriminate non-native segmental contrasts more accurately when they do not identify the stimuli as linguistic data, as under such conditions the influence of the native language phonology can be diminished. Thus, Mah et al. used two conditions: a linguistic one, in which the stimuli were syllables [ʌm] ‘um,’ [hʌm] ‘hum,’ and [θʌm] ‘thumb’ (the authors chose [θ] as a distracter, as it is a low-intensity fricative also absent from French), and a nonlinguistic one, where stimuli consisted of fricative noise bursts corresponding to [f], [hf], and [θf] ([f] was used as another fricative of low- intensity

that exists in both French and English). They hypothesized that if the problem with /h/ is acoustically based, no MMN (Mismatch negativity) will be observed in learners when hearing stimuli involving /h/ in either of the conditions. If, however, the MMN is observed in the non-linguistic condition but not in the linguistic one, the difficulty with /h/ stems from bad phonological representations and not from mere acoustic non-salience. The results of the study are in line with the second hypothesis, namely, that French learners of English are able to perceive /h/ on the acoustic level but have problems at the phonological level. More precisely, in the non-linguistic condition French learners of English exhibited MMN responses which were similar to the MMN of natives, showing that they are able to perceive the acoustic properties of /h/ if the influence of the L1 is diminished. However, the learners showed no MMN in the linguistic condition when /h/ was involved, which points to the influence on the L1 phonological system whenever the stimuli are presented as linguistic data. Although this finding suggests that the perceptual problems with the sound /h/ are caused by the phonological grammar of the learners, and not by the low acoustic salience of this sound, Mah et al. (2016) did not test whether the perceptual difficulty of /h/ leads to asymmetries.

What would the predictions of both NRV and NLM be, concerning the perception of /h/ by French learners of English? As its name indicates, the Natural Referent Vowel model has been developed to explain perceptual bias in vowel perception (Polka & Bohn; 2011). However, Nam & Polka (2016) made a first attempt to adapt this model to consonant perception. According to them, the consonant that is more salient in a contrast, will act as a referent and bias perception. Although the case of /h/ is rather unusual, as /h/ is a consonant (which has phonetic properties of a vowel) and it contrasts with silence, not another consonant, the NRV does provide predictions concerning the perception of /h/. If we consider the salience of the sound to be the driving factor for perceptual bias, then, given that /h/ is more salient than silence, /h/-initial items should be better perceived than items where the /h/ is omitted (vowel-initial items).

From the perspective of the Native Language Magnet Theory (Kuhl, 1991; Kuhl et al., 2008), L2 sounds that are less prototypical of the L1 sounds will be easier to discriminate. As /h/ is not a prototypical consonant of French, the NLM would predict that it will be perceived rather well. Even if /h/ is considered as a devoiced version of the following vowel, its production is still very different from the prototypical way French vowels are produced. Thus, even if /h/ should be compared to vowels, it still would be considered non-prototypical and thus, according to NRV, it

should be perceived quite accurately. However, it is not clear what the predictions should be for items where the /h/ is omitted.

### 3.4.3 Results from the training study

#### 3.4.3.1 Prelexical level

A linear mixed effects model was run on the accuracy scores in the identification tasks in pretest and posttest, with contrast-coded fixed factors Session (pretest vs. posttest) and Item Type (/h/ vs. vowel) as well as the interaction between them. Intercepts for participants and items were entered as random factors (Table 3.3). P-values were obtained by likelihood ratio tests of the full model against the model without the effect or interaction in question. There was a main effect of Session, with better performance in posttest (mean = 80.04 %, SE = 8.16) than in pretest (mean = 67.37 %, SE = 9.57), but no effect of Item Type (mean<sub>h</sub> = 75.40 %, SE = 8.79) (mean<sub>vowel</sub> = 72.13 %, SE = 9.15). There was a significant interaction between Session and Item Type. Post-hoc paired t-tests showed that the interaction was due to the fact that in pretest participants performed better on h-type items than on vowel-type items ( $t(23) = 2.50$ ,  $p = 0.02$ ) whereas there was no such difference in posttest ( $p > 0.5$ ). The mean accuracy scores for /h/-initial and vowel initial nonwords in the identification task in pretest in posttest are shown in Figure 3.3.

TABLE 3.3. Coefficients and log-likelihood comparisons for each fixed effect in the model for identification.

variable	B	SE	Z	$\chi^2$	df	p
intercept	1.24	0.14	9.04	-	-	-
Session	0.78	0.05	14.70	220.25	1	<0.001
Item Type	-0.14	0.12	-1.15	1.30	1	>0.1
Session $\times$ Item Type	0.54	0.11	5.08	25.58	1	<0.001

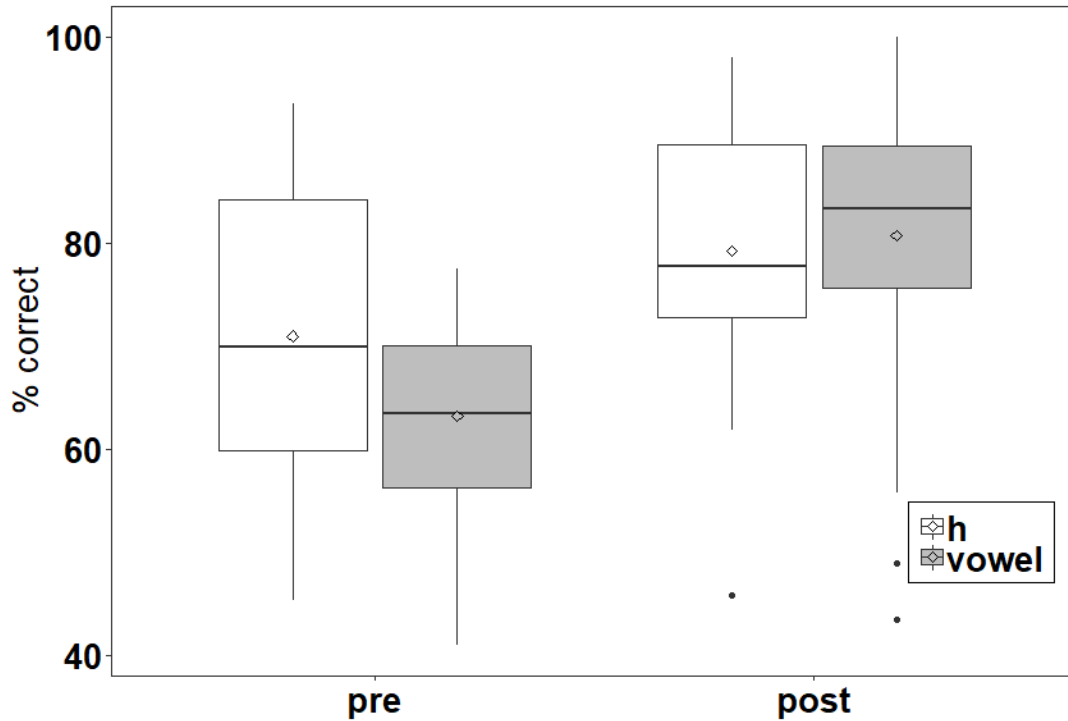


Figure 3.3. Boxplots of accuracy (percent correct) on /h/-initial and vowel-initial stimuli in the identification task at pretest and posttest.

### 3.4.3.2 Lexical level

A linear mixed effects model was run on the accuracy scores, with Session (pretest vs. posttest), Lexicality (word vs. nonword), and Item Type (/h/ vs. vowel) as contrast-coded fixed factors, and intercepts for participants and items as random factors (Table 3.4). P-values were obtained by likelihood ratio tests of the full model against the model without the effect or interaction in question. All three main effects were significant, with higher accuracy in posttest (mean = 74.05 %, SE = 8.95) than in pretest (mean = 57.07 %, SE = 10.11), on words (mean = 84.23 %, SE = 7.44) than on nonwords (mean = 46.76 %, SE = 10.19) and on vowel-type (mean = 70.01 %, SE = 9.36) than on h-type items (mean = 61.48 %, SE = 9.94). There was also a Lexicality x Session interaction, due to the fact that the difference between words and nonwords was larger in pretest than in posttest. The triple interaction between Session, Lexicality, and Item Type was also significant. This triple interaction was examined in two mixed effects models on

the pretest and posttest data, respectively, with Lexicality and Item Type as contrast-coded fixed factors, and intercepts for participants and items as random factors. In pretest, there was a main effect of Lexicality, with better performance on words than on nonwords ( $\beta = 2.47$ ,  $SE = 0.18$ ,  $Z = 13.96$ ,  $\chi^2(1) = 258.20$ ,  $p < 0.001$ ), but no effect of Item Type ( $\beta = -0.13$ ,  $SE = 0.19$ ,  $Z = -0.71$ ,  $\chi^2(1) = 0.50$ ,  $p = 0.48$ ), nor an interaction between the factors ( $\beta = -0.71$ ,  $SE = 0.39$ ,  $Z = -1.80$ ,  $\chi^2(1) = 3.09$ ,  $p = 0.08$ ). In posttest, there was a main effect of Lexicality ( $\beta = 1.57$ ,  $SE = 0.18$ ,  $Z = 8.76$ ,  $\chi^2(1) = 84.29$ ,  $p < 0.001$ ) and Item Type ( $\beta = -0.72$ ,  $SE = 0.22$ ,  $Z = -3.25$ ,  $\chi^2(1) = 9.93$ ,  $p = 0.002$ ), as well as an interaction between these factors ( $\beta = 1.73$ ,  $SE = 0.35$ ,  $Z = 4.90$ ,  $\chi^2(1) = 23.89$ ,  $p < 0.001$ ). For nonwords, participants were more accurate on vowel-type than on /h/-type items, while for words they showed no such difference. The mean accuracy scores for /h/-type and vowel-type words and nonwords in the lexical decision task, both in pretest and posttest, are shown in Figure 3.4.

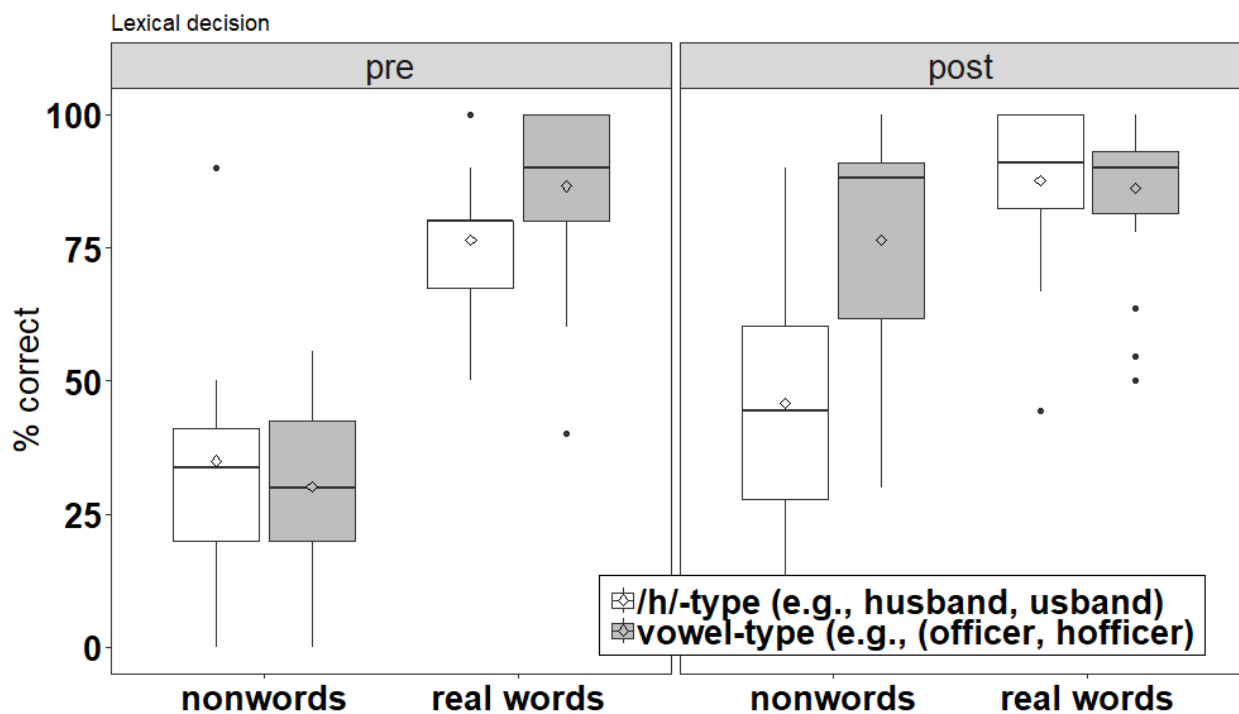


Figure 3.4. Boxplots of accuracy on the test items as a function of Lexicality (word vs. Nonword) and Item Type (/h/-type vs vowel-type) in pretest (left panel) and posttest (right panel).

TABLE 3.4. Coefficients and log-likelihood comparisons for each fixed effect in the model for lexical decision.

Variable	$\beta$	SE	Z	$\chi^2$	df	p
Intercept	0.86	0.13	6.50	-	-	-
Session	0.87	0.16	5.59	27.33	1	<0.001
Underlying Sound	-0.42	0.14	-2.96	8.44	1	<0.01
Lexicality	2.01	0.12	16.14	306.62	1	<0.001
Session $\times$ Item Type	-0.55	0.29	-1.91	3.69	1	>0.05
Session $\times$ Lexicality	2.42	0.52	-3.98	15.72	1	<0.001
Lexicality $\times$ Item Type	0.47	0.26	1.81	3.29	1	>0.05
Session $\times$ Item Type $\times$ Lexicality	2.42	0.52	4.70	21.36	1	<0.001

### 3.4.4 Discussion

We looked at the link between asymmetries at the prelexical vs. lexical levels. At the prelexical level we found a perceptual asymmetry in pretest, where participants performed more accurately on /h/-initial items than on vowel-initial items. In posttest, however, this asymmetry disappeared, as the accuracy of participants increased on both item types, but the improvement was larger for vowel-initial items. The initial asymmetry is in accordance with what could be predicted by the Natural Referent Model. Namely, the /h/-initial items were more salient than the ones without /h/. Thus, perception was better for the stimuli containing /h/. The observation that the perceptual asymmetry decreased and disappeared with training is in accordance to the findings of Bradlow et al., (1997). In this study they trained Japanese participants and showed that the asymmetry between the perception of English /ɹ/ and /l/ decreased after training, due to the larger improvement on /l/. This suggests that, contrary to the findings of Kriengwatananaa & Escudero (2017) (see section 1.2.1), experience with an L2 can induce changes in perceptual asymmetries.

Another explanation for the prelexical asymmetry found in our study is related to (the) task particularities. The perception of participants was tested by means of an identification task, where participants had to answer for each item whether it started with /h/ or not. It might be the case that such a formulation of the question biases participants into excessively concentrating on

the sound /h/ and “hallucinate” this sound even in items in which it was not present. (Usually, identification involves a contrast between two sounds. In such cases the task of the participants is to say, for instance, whether they heard /ɪ/ or /l/, which means that their attention is equally driven to both options.) Moreover, it is possible that this bias was reinforced by the poor perception of /h/: the uncertainty about what was heard together with the will to be accurate, led to more “yes” responses. For instance, John & Cardoso (2008) showed that in production, participants epenthesize /h/ metalinguistically, in order to imitate what they think is a nativelike and thus more prestigious production. We could hypothesize that the improved perception of /h/ after training leads to more confident symmetrical responses.

At the lexical level, in pretest there was no asymmetry between /h/-type and vowel-type items. However, in posttest a significant asymmetry was observed in nonwords, where participants performed better on items starting with /h/ (e.g., *\*hofficer*) than on items starting with a vowel (e.g., *\*usband*). Thus, it seems that the asymmetries in prelexical and lexical levels do not correspond. That is, in pretest there was an asymmetry at the prelexical but not at the level level, whereas in posttest there was an asymmetry at the lexical but not at the prelexical level. This could be explained in two ways. First, it could be the case that asymmetries exist at prelexical and lexical levels independently and should be explained by different underlying processes. Thus, on the one hand, prelexical asymmetries would occur because of certain perceptual bias, but this bias would be attenuated by more complex processing in tasks such as word recognition. On the other hand, asymmetries at the lexical level would occur because of some higher order processing constraints and would have no effect on prelexical perception as they would occur later in the processing. Conversely, it is possible that the asymmetry at the prelexical level found in this experiment was an artefact caused by the specifics of the task. Thus there might be an independent lexical asymmetry and no prelexical asymmetry et all, at least in the case of the perception of /h/.

At the lexical level, we can notice that the asymmetry in posttest occurred because of the very clear change in performance on nonwords starting with /h/ (e.g., *\*hofficer*). That is, following training, participants became much better at rejecting this type of nonwords. In addition, there is a numerical (although not significant) trend showing that they also became more accurate at correctly recognizing real words starting with /h/. This suggests that by better perceiving the sound /h/ prelexically, participants (also) became stronger in noticing the /h/ both



in nonwords created by adding an /h/ and in real words starting with /h/. Why then did the performance hardly change for nonwords starting with a vowel (e.g., *\*usband*)? An unlikely possibility is that similarly to the prelexical level in pretest, participants “hallucinated” an /h/ at the beginning of the nonword (they perceived *husband* when hearing *\*usband*) and thus accepted it as a real word. Indeed, the results at the prelexical level in posttest indicates that following training, the asymmetry between the perception of stimuli with and without /h/ disappeared and participants performed rather well on both types of items (average 80% correct). An alternative explanation is that French learners of English have inaccurate phonological representations of English words containing /h/. More specifically, because learners misperceive /h/ (at least until a certain stage of proficiency), the representations in the mental lexicon created for words starting with /h/ are imprecise. This would be the reason why participants tended to accept both words (such as *husband*) and nonwords (such as *\*usband*) as real English words. This problem persisted even in posttest, suggesting that short auditory training and improvements in perception were not sufficient to change the phonological representations of words. It is possible, though, that longer training might induce such changes, as we observed a non-significant trend for improvement from pretest (35 % correct) to posttest (44 % correct) for these nonwords.

To sum up, the asymmetry at the lexical level was caused by an improvement in prelexical perception, which enhanced the performance on nonwords starting with /h/ (*\*hofficer*), but not the performance on nonwords without /h/ (*\*usband*). This points to the fact that problems with both types of nonwords in pretest had different sources: a perceptual difficulty with *\*hofficer*-type nonwords and imprecise phonological representations for *\*usband*-type nonwords. While the first one can be easily improved by training, the other one seems to be more problematic.

Thus, although prelexical perception influenced the asymmetry at the lexical level, it is not the case that a prelexical bias is the source of this lexical asymmetry. It seems instead that if a prelexical asymmetry exists at all (it might be an artefact caused by the task), it has no major influence on the perception at the lexical level. This is in line with the reviewed literature in Chapter 1, where asymmetries seemed to occur rather independently at the prelexical and lexical levels. Moreover, the direction of such asymmetries did not seem to strictly follow the predictions of none of the two theories (NRV vs. NLM). Further research is therefore needed to clarify the nature of prelexical and lexical asymmetries in perception. From the methodological

point of view, if an asymmetry was found using a discrimination (ABX) task to test prelexical perception, this would ensure that no bias is induced by the task itself.

### **3.4.5 Comparing results from lexical decision in the Melnik & Peperkamp (2019) study and the training study**

For simplicity reasons, in this part we will refer to the Melnik & Peperkamp (2019) study on lexical asymmetries as “experiment 1” and to the training experiment, as “experiment 2”. In this section we provide a broad comparison of the findings of the two studies. We found an asymmetry going in the same direction for nonwords in experiment 1 and the posttest of experiment 2 (mean accuracy scores for the /h/-type and vowel-type words and nonwords in the lexical decision task in experiment 1 are shown in Figure 3.5). Why does the pattern of asymmetries in experiment 1 resemble more the performance of learners after training than the pattern of performance before training in experiment 2? One possibility could be that the proficiency of participants in experiment 1 was overall higher than that of participants in the training experiment and thus was closer to their level after training. Although we could not perform a median split to divide participants into higher and lower proficiency groups because their number was too small ( $N = 24$ ), the scores in the self-evaluation questionnaire point in this direction. Namely, learners in experiment 1 had a mean score of 7.4/10 ( $SD = 1.2$ ), while it was of 6.4/10 ( $SD = 1.6$ ) in experiment 2. Note also that this is in line with the A-prime scores obtained by the groups in lexical decision. For test items, the learners in experiment 1 had a mean A-prime score of 0.72, but learners in the pretest of experiment 2 scored only 0.62. However, after training these participants reached an A-prime score of 0.82 in posttest.

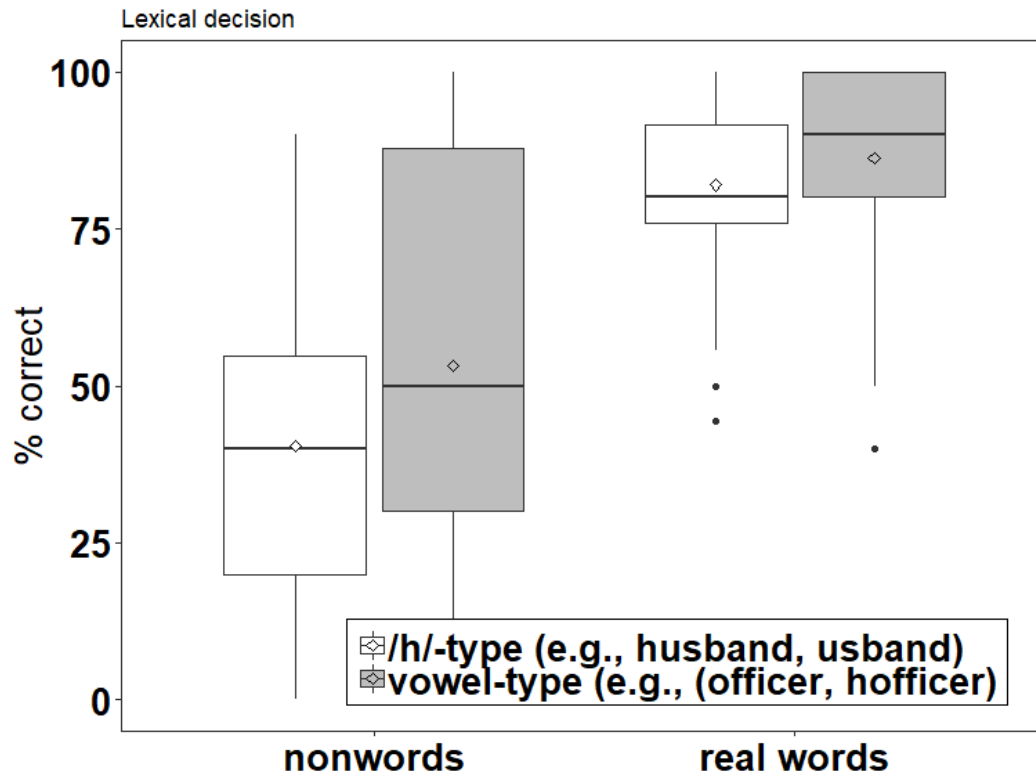


Figure 3.5. Boxplots of accuracy on the test items as a function of Lexicality (word vs. Nonword) and Item Type (/h/-type vs vowel-type) in the lexical decision task in the Melnik & Peperkamp (2019) study.

This would mean that in accordance with what was said earlier, the asymmetry occurs with experience, either induced by natural learning or by auditory training. More precisely, till some point in learning there would be no asymmetry or only a small one because both the perception of /h/ and the representations of words containing this sound would be imprecise. This is reflected by the performance of participants in the pretest of experiment 2. However, with the improving perception of /h/, an asymmetry at the lexical level occurs as better perception enhances the performance on nonwords starting with /h/ (\**hofficer*), but not the performance on nonwords without /h/ (\**usband*). The latter involves a wrongly encoded representation of the word *husband* which leads to accepting both the word and the nonword. This pattern is reflected by the asymmetry in experiment 1 and in the posttest of experiment 2. We could assume that the asymmetry gets larger with experience: participants from experiment 1 who have an A-prime score of 0.72 in test items show a much smaller asymmetry than participants of experiment 2 in

posttest, who reached an A-prime score of 0.82. What remains unclear is whether with even more time and experience, the representations of words also become more precise, leading to the reduction of the asymmetry. As noted in the previous subsection, we observed a non-significant trend for improvement from pretest (35 % correct) to posttest (44 % correct) for the *\*usband-* type nonwords, which might be an indication that more learning can induce such changes. Note, though, that we can only draw tentative conclusions from this comparison of the Melnik & Peperkamp (2019) and the training study, as we did not carry out additional statistical analyses between the data of the two studies.

### 3.5 Conclusion

In this Chapter we examined L2 perception and its development across levels of processing. We focused on the perception of the English sound /h/ by French learners of English. In part 3.2 we showed that perceptual problems with /h/ previously found in French learners of English at the prelexical level also persist at the lexical level. Results also showed that the perception of /h/ at the lexical level is asymmetrical, and this asymmetry follows the one reported in production (more deletions than insertions). These findings were discussed in light of two contrasting explanations on the source of such asymmetrical lexical access in second language learners: that of imprecise lexical representations versus inaccurate phonetic perception. We propose that our data goes into the direction of the first of these explanations.

In part 3.3 we carried out an online phonetic training study, based on the findings of the section 3.2. We looked at whether training on the English sound /h/ at the prelexical level can enhance the perception of this sound both prelexically and lexically. We found that training was indeed effective on improving the perception of /h/ at both levels of processing. Moreover, this improvement did not reduce four months after posttest, as shown by the results of the retention test. We suggest that the finding that phonetic training can effectively enhance word recognition, points to a sequential order in learning across levels of processing. Specifically, we suggest that the automatization of prelexical perception allows to allocate more cognitive resources to the processing at the lexical level, which in turn facilitates its improvement.

Finally, in part 3.4 we explored the link between asymmetries at the prelexical and lexical levels. First, we reanalyzed the data from the training study (3.3) and investigated if asymmetries are found at pretest at both levels of processing and whether training induces changes in these asymmetries in posttest. We found that prelexical and lexical asymmetries were independent. However, training seemed to have an impact on their development. Second, we compared these findings with the results from the Melnik & Peperkamp (2019) study on perceptual asymmetries in the lexical perception of /h/, reported in part 3.2. We found that the lexical asymmetry in the Melnik & Peperkamp study and the posttest of the training study go into the same direction. We suggest that the asymmetry at the lexical level occurs with experience, either induced by natural learning or auditory training. This asymmetry is caused by improvements in prelexical perception which precede changes in the phonological representations of words.

## Chapter 4: General discussion

By studying the development of L2 speech perception and production in adulthood, this thesis inscribes itself within the theoretical framework of the Speech Learning Model (Flege, 1995) and the Perceptual Assimilation Model (Best, 1995). Specifically, we focus on second language phonological processing and its development across modalities and levels of processing (Fig. 4.1). On the one hand, we explored the relationship between perception and production in L2 and the strength of this relationship when tested at different levels of processing. On the other hand, we considered the link between prelexical and lexical perception by focusing on perceptual asymmetries and on the effects of prelexical training on the perception at the lexical level. Here we will revisit the results reported in the previous chapters and discuss remaining questions and directions for further research.

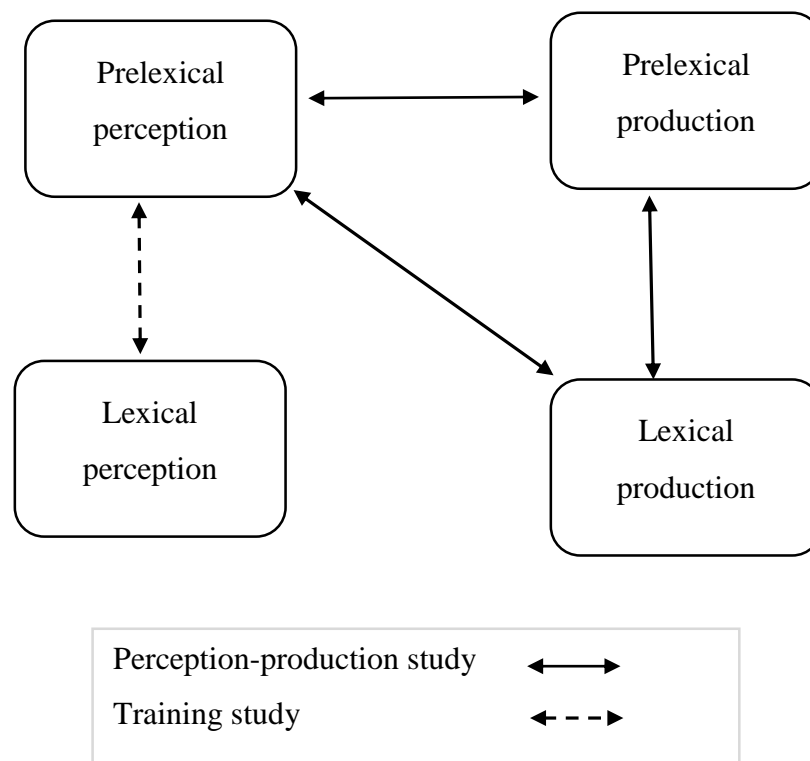


Figure 4.1. Schematic representation of links between modalities across levels of processing studied in this thesis. The arrow type corresponds to the study that looked at each specific link.

## **4.1 The relationship between perception and production**

### **4.1.1 Summary of empirical work**

The main research question in Chapter 2 was whether there is a relationship between perception and production in L2 phonological processing and if so, whether it holds across levels of processing. We addressed important methodological issues encountered in previous literature and designed an experiment aimed at obtaining more precise and comparable measures for perception and production accuracy. We tested proficient English learners of French and a control group of French natives on the French contrast /u/-/y/ in an ABX discrimination task (prelexical perception) and two production tasks, i.e., nonword reading (prelexical level) and naming (lexical level). We found that there is a relationship between the perception and production of the contrast /u/-/y/, but only when tested at the same level of processing.

A complementary question raised in this section (2.2.3.3) addressed one of the main claims of the Speech Learning Model (Flege, 1995), namely, that accurate production of an L2 contrast cannot be achieved unless this contrast is perceived accurately. We carried out a clustering analysis on the results of the perception and production tasks and found evidence that good perception is indeed a prerequisite for good production.

### **4.1.2 Further questions**

In the introduction of Chapter 2, we discussed a range of methodological decisions one must make to assess perception and production, and crucially, to be able to compare the performance in both modalities. More precisely, we considered the following methodological concerns: the tasks used to assess perception and production, the measure by which production accuracy is evaluated, the presence of a control contrast and/or group, and the statistical methods used to analyze the data. Although we addressed most of these issues when creating our experimental paradigm, the production tasks we chose could not take into account the speech rate differences which occurred between groups. Namely, native French participants spoke faster than the English learners of French in both production tasks. Based on previous literature on the relationship between speech rate and vowel reduction (Lindblom, 1963; Nadeu, 2014) we hypothesized that the faster speech of French natives resulted in the reduction of vowels they produced. We could not directly test whether the speech rate of participants was correlated with

the average F1-F2 frequencies of their productions because we did not have a sufficient number of participants to carry out such analyses. Thus, further studies could include more participants and test this hypothesis.

Most importantly, though, further experimental research on production should control for differences in speech rate between natives and learners or even between individuals of the same language background. A simple way to monitor speech rate in a task such as nonword reading is by using a metronome (Mitterer & Ernestus, 2008; Motobashi-Saigo & Hardison, 2009; Kittredge & Dell, 2016). This method would allow us to pace the productions of participants according to a predefined absolute speech rate. However, in a naming task, where stimuli differ in length and familiarity, a more ecological solution to control for speech rate would be to take into account the relative speech rate of each participant. This could be achieved by introducing speech rate conditions in production for both learners and natives (Schmidt & Flege, 1995; Nadeu, 2014). For instance, Schmidt & Flege (1995) used a Magnitude production task in which participants first read the stimulus phrase and then were instructed to practice saying it at what they considered to be their normal speaking rate. They were then asked to practice saying it at a rate that was twice as fast, and finally at a speaking rate that was half as fast as their normal speaking rate. After this practice phase, they were recorded saying the stimulus phrase several times in all three speaking rates. The slow rate productions were discarded, as they only served as a control. Thus, comparing speech at a slow rate and a faster rate using this or a comparable task in both learners and natives would allow us to take into account the possible reduction processes that happen in faster speech. In this way, data from native speakers could be used as a more robust baseline for comparison with data from learners.

Turning to the relationship between perception and production, we showed that these two modalities are correlated, but only when tested at the same, prelexical level of processing. Future research should further test whether perception and production are related at the lexical level of processing. This would require adding a lexical perception task, such as lexical decision. If perception and production develop in parallel within each level of processing, a correlation should be found between tasks that tap into the same level of processing, such as nonword reading and ABX discrimination, as well as between naming and lexical decision tasks. Moreover, testing several groups of proficiency in such a paradigm would allow us to shed more light on the speed of the development of perception and production within each level of



processing. Furthermore, this would allow us to understand better whether perception must attain a certain level of accuracy at each level of processing before improvements in production in the corresponding level could be achieved.

## 4.2 Non-native sound perception across levels of processing

### 4.2.1 Summary of empirical work

In Chapter 3 we further examined the phonological processing of difficult L2 sounds, by focusing on the perception and its development at prelexical and lexical levels. We looked at the perception of the English sound /h/ by French learners of English. In section 3.2 (Melnik & Peperkamp, 2019) we tested whether perceptual difficulties with /h/ previously reported in French learners of English at the prelexical level also persist at the lexical level. We also explored if the asymmetry found in production, where French learners of English delete /h/ more often than they insert it, is also observed in perception. French learners of English and a group of native English speakers were administered a lexical decision task on words and nonwords, where the test nonwords were created from /h/-initial and vowel-initial words by removing or adding /h/, respectively. We found that French learners of English do have difficulties in processing /h/ at the lexical level, as they tended to accept /h/- and vowel-initial nonwords as real words. Moreover, an asymmetry was observed in their performance, such that French learners made more errors on /h/-initial words (e.g., *husband*) and the nonwords derived from these words (e.g., *usband*) than on vowel-initial words (e.g., *officer*) and nonwords derived from them (e.g., *hofficer*) (native English speakers showed the same difference, but of a smaller magnitude and only in real words). This finding sheds new light on the ongoing discussion in the literature on the source of such asymmetries, previously also found in other languages. More specifically, our data suggests that the perceptual asymmetry is caused by inaccurate representations of /h/-initial words in the lexicon, and not by mere inaccurate perception.

Based on these results, in section 3.3 we carried out an online phonetic training study, which investigated whether training on the English sound /h/ at the prelexical level can enhance the perception of this sound both prelexically and lexically. In pretest, posttest and retention test we used the same lexical decision task and the same stimuli as in the study in 3.2, as well as an identification task, the latter aimed at assessing the perception of /h/ at the prelexical level. Participants received eight sessions of online High Variability Phonetic training. Results revealed that following training participants improved not only in identification accuracy, but also in their performance on the lexical decision task, suggesting that phonetic training can effectively enhance word recognition. Moreover, we retested participants in a retention test four months

after posttest and found that the effects of training did not diminish, compared to posttest, in either of the tasks.

Finally, in part 3.4 we explored the link between asymmetries at the prelexical and lexical levels by reanalyzing the data from the training study (3.3) and comparing them to the data from the Melnik & Peperkamp (2019) study (3.2). We first focused on the training study and examined if at pretest asymmetries could be found at both levels of processing and whether training induced changes in these asymmetries. We found that prelexical and lexical asymmetries were quite independent both in pretest and posttest, although training did influence the asymmetries within each level of processing. Second, we compared these findings with the results from the Melnik & Peperkamp (2019) study on perceptual asymmetries in the lexical processing of /h/ (part 3.2). This comparison showed that asymmetries in this first experiment follow the same patterns as the asymmetries found in the posttest of the training study. This was interpreted as an indication that asymmetries in perception develop with changes in proficiency.

#### **4.2.2 Further questions**

##### Asymmetries

Previous literature on asymmetries in perception looked either at prelexical asymmetries or at asymmetries at the lexical level, but they never compared the two. However, if perceptual asymmetries occur while perceiving speech sounds, this might have an influence on how we recognize words containing these sounds. More precisely, as prelexical perception is closely linked to perception at the lexical level, one could expect the prelexical asymmetries to influence the asymmetries at the lexical level. In this thesis we made a first attempt to compare perceptual asymmetries across levels of processing, using the same participants. The results of the training study revealed a rather complex picture, as we found both prelexical and lexical asymmetries, but the lexical asymmetry could not be explained by the prelexical one. More precisely, in pretest we found a prelexical but no lexical asymmetry, while in posttest we conversely found a lexical but no prelexical asymmetry. We suggested that the asymmetry at the lexical level in posttest was caused by an improvement in prelexical perception following training. This would lead to the conclusion that prelexical and lexical asymmetries are independent. Thus, on the one hand, prelexical asymmetries would occur because of certain perceptual bias, but this bias would be attenuated by more complex processing in tasks such as word recognition. On the other hand,

asymmetries at the lexical level would occur because of some higher order processing constraints and would have no effect on prelexical perception as they would occur later in the processing

However, it is possible that the prelexical asymmetry we observed in the perception of /h/ was due to task particularities. Namely, we tested participants using an identification task, in which they had to answer for each item whether it started with /h/ or not. It is possible that this formulation of the task biased participants into excessively concentrating on the sound /h/ and “hallucinating” this sound even in items in which it was absent. This might have resulted in more “yes”-responses, regardless of whether the /h/ was present or not. Thus, our finding that in pretest participants performed worse on vowel-initial stimuli compared to /h/-initial ones could be an artefact of the task, and not evidence of a robust prelexical asymmetry.

Further research could thus first explore if there is a robust prelexical asymmetry in the perception of /h/ by French learners of English, by using a different prelexical task than identification, for instance, ABX discrimination. If a prelexical asymmetry similar to the one observed in our study is found, this would be an indication that the prelexical perception of /h/ by French learners of English is indeed asymmetrical. Moreover, such a result would confirm our proposal that prelexical and lexical asymmetries are independent. If, however, no prelexical asymmetry is found, this would lead to the conclusion that a lexical asymmetry can occur independently in cases where a prelexical asymmetry is absent. It would be then of great interest to examine a different language, where strong prelexical asymmetries have been reported. More specifically, it is possible that in cases where a strong prelexical asymmetry is found, it would influence perception at the lexical level. For instance, the prelexical perception of English /ɪ/-/I/ by Japanese learners has been shown to be asymmetrical by many studies (Bradlow et al., 1997; Guion & Flege, 2000; Hattori & Iverson, 2009). Thus, the same Japanese learners of English should be tested at the lexical and prelexical levels to investigate if robust prelexical asymmetries in L2 processing have an influence on lexical processing.

### Training

Our study showed that phonetic training can improve word recognition, and this positive effect does not reduce four months after training is completed. This first evidence that such transfer from prelexical to lexical level is possible raises some questions about the nature of the changes training can induce. The analyses of the asymmetries in pretest and posttest revealed

that the improvements in the lexical decision task reported in posttest were due to improvements of the prelexical perception of /h/, but not to changes of representations of words containing /h/ in the mental lexicon of learners. That is, an asymmetry at the lexical level occurred in posttest, as better perception enhanced the performance on nonwords starting with /h/ (*\*hofficer*), but not the performance on nonwords without /h/ (*\*usband*), which involve a wrongly encoded representation of the word *husband*. Nevertheless, we observed a trend for improvement for these *\*usband*-type nonwords, suggesting that longer training might induce changes in the mental representations of words containing /h/. Importantly, the almost perfect one-to-one mapping in English of the grapheme <h> onto the phoneme /h/ might help in the process of updating the phonological representations of words. That is, once the prelexical perception of /h/ improves following training, learners might be able to update their mental representations of known /h/-initial words by mapping the percept to the orthographic form of the word. In a study on English word learning by Dutch learners of English, Escudero et al. (2008) showed that the availability of the mapping between spelled forms and phonological forms can help in establishing a lexical contrast which can then be used in auditory word recognition. Therefore, the explicit knowledge of the spelling of the word *husband*, for example, could help the learners in linking the percept of /hʌzbənd/ to the previously imprecise representation /Xʌzbənd/, where /X/ was not well defined. It remains unclear, however, what is the minimal intensity and length of training that could induce robust changes in the mental lexicon. It is true that although the training we administered in this experiment was eight sessions long, which is common for training studies (Wang et al., 1999; Huensch & Tremblay, 2015; Okuno & Hardison, 2016), each individual session was rather short (10-20 min), compared to the 40- or 50-min long sessions in certain studies (Logan et al., 1991; Hazan et al., 2005; Sadakata & McQueen, 2014). Thus, in order to understand if phonetic training can trigger further changes in the mental representations of words, learners should be tested on longer or more intensive training.

It is also possible, though, that much more time and practice is needed to integrate the improvements in prelexical perception of a difficult L2 sound, so that changes in the phonological representations of words containing this sound would occur. Several previous studies looked at whether phonological representations in the mental lexicon change at all with mastery in L2. For instance, Imai et al. (2005) compared groups of different proficiency levels. They tested native English speakers and two groups of Spanish learners of English with different

proficiency on lexical tasks involving English words produced by a native English and a native Spanish speaker (accented speech). The authors hypothesized that phonological representations in L2 are initially affected by the L1 sound system, but that as L2 learning proceeds, these representations become more attuned to the L2 and thus more similar to those of native speakers. Thus, proficient speakers were expected to perform better on native English stimuli, while low proficiency speakers – better on Spanish-accented stimuli. They found that the performance of high proficiency learners on native English-spoken stimuli was indeed closer to that of native speakers' than to the performance of low-proficiency learners. This indicated that the lexical representations of the high-proficiency group were more similar to those of the native English group, thus suggesting that lexical representations in L2 improve with higher levels of proficiency in the language. In a more recent study, Cook & Gor (2015) tested advanced learners of Russian and compared the lexical representations they have of familiar and less familiar words in Russian. In a lexical decision with auditory priming task they showed that when targets are preceded by familiar primes with overlapping onsets, they are processed slower – this is an inhibitory effect previously found in native speakers. Conversely, when the prime with overlapping onset is a less known word, it produces a facilitation effect, commonly found in L2 speakers. The authors conclude that as L2 proficiency and word familiarity increases, the lexical representations become more precise, thus leading to more nativelike processing. Overall, these studies indicate that as proficiency increases, the performance on word recognition improves as well, suggesting that the mental representations of L2 words become more precise. As these studies tested participants only at a specific time point in learning, and there are no longitudinal studies on the topic, training studies could provide important evidence on the pace and dynamics of changes in representations of words in L2.

Another useful addition that could be made in further training studies is the inclusion of a control group. This group would do the pretest, posttest and retention test without undergoing training. This would allow us to have an additional proof of the effectiveness of training, as we would expect the control group to show less improvement at posttest and the retention test, compared to the training group. However, many recent training studies did not include a control group (Sadakata & McQueen, 2014; Tamminen et al., 2015; Leong et al., 2018; Shinohara & Iverson, 2018), and those who did, usually had a small number of participants (13 test and 9

control in Lengeris & Nicolaidis (2014); 12 test and 12 control in Huensch & Tremblay (2015); 15 test and 9 control in Lengeris (2018)). One of the explanations for the lack of a control group or for the inclusion of very few participants, is that the inclusion of additional participants is often costly and problematic, for two reasons.

First, the control group should be matched for level of proficiency to the test group, which is often complicated. More specifically, the common proficiency levels usually used, such as the A1-C2 levels proposed by the Common European Framework of Reference for Languages, do not always reflect the proficiency of learners at the phonological level. That is, such systems of reference usually focus on reading, speaking, writing and comprehension skills, without evaluating the pronunciation or perceptual abilities of the learners. Thus, when it comes to perception or production, there might be great interindividual differences within learners who are assigned to the same level of proficiency. For instance, for our training study we recruited participants with a B2 level of English. In order to avoid ceiling performance or insufficient knowledge of English vocabulary, we screened the participants based on their performance in the perception tasks. 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. Note though, that of the 51 participants who did the pretest, only 25 satisfied these criteria. Thus, from all of the B2 level participants we recruited, only half had a comparable level in the perception of L2 sounds. This points to the fact that besides identifying a necessary L2 level for participants in a commonly acknowledged system of proficiency referencing, screening learners for their proficiency in L2 phonology is a crucial step in order to obtain homogeneous groups of participants.

The second problem is that in order to include a control group, researchers will have to test a rather large number of participants who will not know if they will be able to go through the whole training experiment. This in turn can create problems for the recruitment as participants might lack motivation to come to the laboratory for a screening test if they are not guaranteed with the possibility to undergo training. Such practical issues, though, could probably be diminished if the whole experiment could be done online. We showed that online training can be an efficient alternative to training in the laboratory. Moreover, participants reported being motivated by doing the training at home. In a similar way, we would expect that doing the pretest

online could encourage learners to participate and at least partly diminish the difficulty of recruitment.



### 4.3 Insights from perception on the development of production

Although experiments described in Chapter 2 and Chapter 3 focused on different languages and used different tasks, we will make an attempt to discuss the general trends observed in these studies and the insights on L2 phonological acquisition they provide. More specifically, while in Chapter 2 we report on a study where learners were tested on their production of L2 sounds at the prelexical and lexical levels, in Chapter 3 we provide evidence on L2 perception at both levels of processing. What can we conclude from these studies on the changes that happen during learning in both modalities and across levels of processing?

Starting from production, in Chapter 2 we used non-hierarchical k-means clustering to group the participants according to their performance in the perception task and each of the production tasks, separately. The results revealed that more participants were good in production when tested on the prelexical nonword reading task ( $N = 9$ ) than when tested on the lexical naming task ( $N = 7$ ). Although, the number of participants is too small to draw robust conclusions from this difference, this might be an indication that learners first master production at the prelexical level and can only then achieve accuracy in production at the lexical level. We can interpret this observation in light of the findings from the training study, where improvements in prelexical perception led to improvements in lexical perception. We proposed that phonological learning happen in a bottom-up sequential order, where improvements in prelexical processing might give rise to changes at the lexical level. This is in accordance with the Automatic Selective Perception model (Strange, 2011), which proposes that L2 phonological processing is less automatic than phonological processing in L1. Thus, while learners might achieve relatively high accuracy in relatively simple prelexical tasks, where they can exclusively focus their attention on crucial phonetic cues, the same performance level might not be obtained in tasks requiring the processing of more complex stimuli and attention to other information, such as word meaning.

However, according to this model, the processing of simple tasks becomes more automatic and nativelike as proficiency grows and this in turn allows to allocate more attention and other cognitive resources to higher level processing. Thus, results from the study reported in section 3.3 suggest that training rendered the prelexical processing more efficient and led to enhanced performance in word recognition. In a similar way, accurate production at the

prelexical level could be a prerequisite for accurate production in a naming task, which requires retrieving the meaning and the phonological form of the word from the mental lexicon, and thus is cognitively more demanding. To our knowledge, there are no studies on prelexical production training, which would improve prelexical and lexical production of L2 learners. Shinohara & Iverson (2018), though, showed that phonetic perception training led to improvements in production at the lexical level (tested in word and passage reading tasks). The authors suggested that training might induce automaticity of phonetic processes which in turn allows the speakers to pronounce the correct acoustic contrasts when producing speech. Further studies could explicitly investigate how production at the prelexical and lexical levels are linked. If perception and production follow the same path in development, improvements at the prelexical level should facilitate improvements at the lexical one.

# **Appendix A: The role of domain-general cognitive capacities in L2 phonological learning**

## **A.1 Introduction**

Second language literature has provided ample evidence that learning outcomes are not equal for all learners (for a review, see Dörnyei, 2014). As a consequence, the effectiveness of phonetic training depends not only on the methodology used, but also on the capacities of the individual speakers being trained (Bradlow et al., 1997; Colantoni & Steele, 2008; Lengeris & Hazan, 2010; Shinohara & Iverson, 2018). First, learning success depends on learners' pre-training sensitivity to various phonetic cues. For example, Perrachione et al. (2011) trained native American English speakers on Mandarin tone contrasts. Before training they measured participants' baseline ability to perceive tone contrasts. Based on their performance participants were grouped into high and low aptitude groups. Participants were trained on either low variability (one speaker) or high variability (several speakers) training. The posttest results revealed that only the participants with high aptitude benefited from high variability training, while those with low aptitude benefited more from training with low variability. A subsequent study by Sadakata and McQueen (2014) further confirmed that the effectiveness of different training paradigms depends on individual perceptual abilities.

Various other cognitive factors might also cause individual differences in the success of learning. For instance, working memory has been shown to play a role in phonological processing and word learning (Baddeley, 1998), and working memory training enhances listeners' speech perception in noise (Ingvalson et al., 2015). Furthermore, L2 phonological processing correlates with working memory in L2 (Wen, 2016). For instance, Darcy et al. (2015) tested Korean learners of English on three phonological processing tasks and showed that individuals with a higher working memory capacity develop more native-like phonological processing in L2. Other studies have demonstrated that inhibitory skill is closely linked to proficiency in a foreign language. For example, Linck et al. (2012) demonstrated that having better inhibitory control reduces switch costs when switching between the native and foreign

language. Moreover, Darcy et al. (2016) found that Spanish learners of English and English learners of Spanish who had higher inhibitory skill perceived L2 sounds more accurately. Finally, a third domain-general cognitive capacity, attention control, has been shown to be an important predictor of learning success (Segalowitz & Frenkiel-Fishman, 2005). Regarding L2 phonology, Darcy et al. (2014) showed that faster attention switching could enhance the processing of phonologically relevant acoustic information in the L2 input and may lead to more accurate L2 speech perception.

While the above-mentioned studies on perceptual sensitivity tested participants both in single-session experiments and in longer training studies, the studies on domain-general cognitive capacities report only single-session experiments, where participants are tested at one time-point during foreign language learning. For exploratory purposes we included to our training study three cognitive tasks to evaluate some of the domain-general cognitive capacities in the learners and to thus provide evidence from the same learners at two time points in learning: pretest and posttest. Specifically, we studied to what extent working memory, inhibition and selective attention influence the effectiveness of training. To test this, three classical psychology tasks were administered at the end of the posttest: the Digit span (short-term memory) task, the Flanker (attention control) task and the Simon (inhibition) task.

## **A.2 Methods**

Three cognitive tasks were administered in posttest after the language tasks. They were taken from the PEBL (Psychology Experiment Building Language) software (Mueller, 2011). The instructions for each task were translated from English to French. The three tasks taken together lasted 20 min.

Short-term memory was tested by means of the Digit Span task. Participants were visually presented with strings of numbers and were asked to recall and type down the sequence correctly. The first list contained three digits and the length of the sequences increased with the trials. There were two lists for each list length. If participants recalled at least one of the two lists correctly, they were presented with a list of the next length. The longest possible list was ten-

digits-long. The Span score for each participant corresponded to the maximum sequence length each participant succeeded to remember.

Attention control was measure by means of the Flanker's task. In each trial participants saw on the screen a target arrow surrounded by four flanker stimuli. The task of the participants was to indicate the direction of the target arrow by pressing as quickly as possible the corresponding right or left arrow key on the keyboard. The flanker stimuli (the stimuli surrounding the target arrow) were directed either to the same direction as the target (congruent flankers), either to the opposite direction (incongruent flankers), or to neither (neutral flankers). Reaction times were recorded for each trial. The Flanker score for each participant was calculated by subtracting the mean RT for congruent trials from the mean RT for the incongruent trials.

Inhibitory control was evaluated with the Simon's task. In each trial participants saw a red or a blue circle that appeared at varying positions on the screen. Participants were asked to press the right arrow key if the circle was red and the left arrow key if it was blue. The goal of the task was to answer to the color of the stimulus, while ignoring the position (the Simon effect occurs as participants tend to have longer RTs when the side of the circle does not match the button press associated with the shape (incongruent trials) than when the side and the button press match (congruent trials)). The Simon score for each participant was calculated by subtracting the mean RT for congruent trials from the mean RT for the incongruent ones.

### A.3 Results

TABLE A.1. Digit Span, Flanker's and Simon's score for each of the twenty-four participants.

Participant	Digit Span	Flanker's score	Simon's score
nb1	7	63.1	27.767
nb2	7	34.4	10.734
nb3	8	36.2	-12.133
nb4	8	47.55	68.2
nb5	8	65.875	-35.767
nb6	8	17.6	19.8
nb7	7	62.225	40.033
nb8	10	42.975	22.266
nb9	6	38.9	2.633
nb10	5	24.025	21.2
nb11	7	48.4	-43.834
nb12	7	37.475	43.734
nb13	8	21.3	47
nb14	7	48.375	-14.632
nb15	8	57.275	18.1
nb16	6	25.025	45.234
nb17	9	37.45	38.967
nb18	7	32.075	39.6
nb19	9	90.25	18.8
nb20	8	53.925	16.367
nb21	7	83.9	157.234
nb22	6	48.25	27.2
nb23	10	50.575	12.467
nb24	9	46.175	9.4

Table A.1 presents the individual scores for each of the three cognitive tasks. We examined whether pre- or post-training performance on perception tests (lexical decision and identification) was correlated with any of the cognitive skills. We included in the analyses an additional measure of improvement (gain), obtained by subtracting the accuracy in pretest from accuracy in posttest, separately for the identification and the lexical decision task. Pearson correlations revealed that the only significant correlation that included a cognitive score was obtained between the Digit Span score and Gain in Identification (see the correlation matrix in Table A.2 and Figure A.1),  $r = 0.64$ ,  $p < 0.001$ . This suggests that participants with higher short-term memory capacity improved more in the identification task after training. Note, though, that

these analyses were exploratory and the number of participants tested in our study is likely to be too small to capture more fine-grained relationships.

TABLE A.2. Correlation coefficients (Pearson's  $r$ ) showing significant relations ( $p < 0.05$ , uncorrected for multiple comparison) between the scores in cognitive tasks and scores in perception tasks at pretest and posttest; n.s. = non-significant; ID = Identification; LEX = Lexical decision; Pre = Pretest; Post = Posttest; Gain = Posttest – Pretest. The dark grey rectangle frames the correlations that include cognitive scores.

	Span	Flanker	Simon	Pre ID	Post ID	Pre LEX	Post LEX	Gain ID	Gain LEX
Span		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	<b>0.64</b>	n.s.
Flanker			n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Simon				n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Pre ID					0.65	n.s.	n.s.	-0.6	n.s.
Post ID						n.s.	0.61	n.s.	n.s.
Pre LEX							0.51	n.s.	- 0.8
Post LEX								n.s.	n.s.
Gain ID									0.41
Gain LEX									

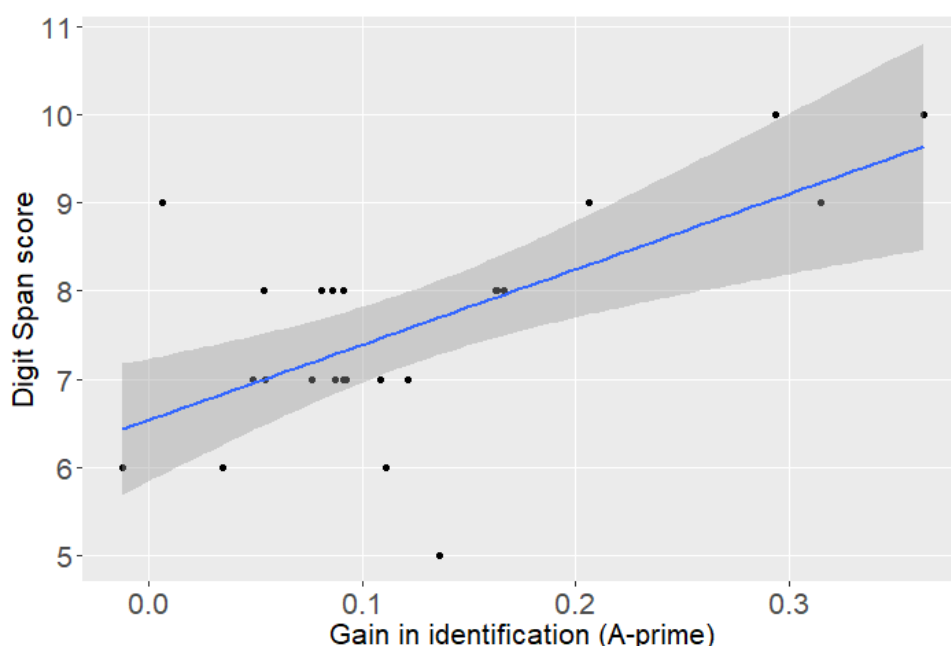


Figure A.1. Correlation between the Gain in identification accuracy and the Digit Span score.

## A.4 Discussion

The current section focused on the relationship between domain-general cognitive capacities and phonological processing in L2 as measured by means of identification (prelexical level) and lexical decision (lexical level) tasks. More specifically, we examined if short-term memory, attention and inhibition control correlate with the accuracy in pretest, posttest as well as with the perceptual gains induced by training. We found that only one cognitive measure was linked to perceptual accuracy, namely, there was a significant correlation between the Digit Span score and the amount of perceptual gain obtained in the identification task following training.

The result that short-term memory is related to identification accuracy is in agreement with previous findings. Having a larger short-term memory capacity has been suggested to allow to retain longer sound sequences in memory which can be useful for language learning and processing (Jacquemot & Scott, 2006; Pierce et al. 2017). Moreover, in first language acquisition, young children with a higher short-term memory capacity have been shown to have larger vocabulary sizes compared to other children of the same age (Baddeley et al., 1998). Concerning adult L2 learning, Aliaga-García et al. (2011) proposed that having larger short-term



memory allows to better attend to the acoustic cues of the foreign language which are not used contrastively in one's native language. Thus, participants with better short-term memory should benefit more from the High-Variability training paradigm, in which they are driven to attend to specific acoustic cues. This is in accordance with our results, where larger memory capacity possibly allowed participants to retain better and more of the relevant acoustic information during training, which resulted in more accurate perception and better identification scores. However, the observation that the Digit Span score did not correlate with identification accuracy in pretest or posttest could be a suggestion that not all learners make optimal use of their short-term memory capacity in naturalistic learning conditions. That is, some of the participants who had the highest memory scores and the best gains in identification accuracy following training performed rather poorly in pretest. Thus, one could wonder why the pretest scores of such participants were so low and why they did not use their memory capacity for learning prior to this training experiment? One explanation could be that because of the abundance of information and the higher processing requirements in naturalistic learning conditions, the learners use their memory capacity to deal with this information flow and have less opportunity to focus on specific phonetic cues. Conversely, phonetic training has been shown to have additional advantages to natural exposure (Iverson et al., 2012), which could at least partly be explained by the possibility to allocate more short-term memory to relevant phonetic cues.

On the other hand, the finding that the short-term memory score did not correlate with the performance on the lexical decision task is in accordance with the results of Darcy et al. (2015) who tested L2 learners on three phonological tasks (ABX, sequence repetition and lexical decision) and three memory tasks (digit, non-word span task, and sentence recall task). They found that memory measures correlated significantly with the first two phonological tasks, but not with the lexical decision task. Because the gains in prelexical and lexical tasks were correlated, it is possible, though, that there is some relationship between short-term memory and phonological processing at the lexical level, which our analyses failed to show because of lack of statistical power.

Turning to the absence of correlations between attention and inhibition control and perception accuracy, it is somewhat surprising that we found no relationship between phonological measures and these cognitive skills, although such links have been reported in

previous studies (Darcy et al., 2014; Darcy et al., 2015). One explanation could be that the tasks used in our experiment might have not been sensitive enough to measure the capacities evaluated. Another explanation for the lack of correlations is that the number of participants (N = 24) in our study was rather small, thus making hard to seize some less strong relationships between variables. Further research should train and test more participants to obtain more robust correlations or to show the lack thereof.

In conclusion, these findings suggest that learners who have higher memory capacity can benefit more from short High Variability Phonetic training, at least at the prelexical level of processing. This raises the possibility to predict the success of phonetic training from short-term memory tests. Further studies should use more sensitive measures of attention and inhibition control to examine the role of these capacities in the phonological learning of L2 during training.

# Appendix B: Online Phonetic Training Improves L2 Word

## Recognition – Melnik & Peperkamp (to appear, 2019)

This section constitutes the following manuscript: Melnik, G. A., Peperkamp, S. (to appear, 2019). Online Phonetic Training Improves L2 Word Recognition. *Proceedings of the 41nd Annual Meeting of the Cognitive Science Society*.

### Online Phonetic Training Improves L2 Word Recognition

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

Sharon Peperkamp  
(sharon.peperkamp@ens.fr)

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

#### Abstract

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

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

#### Introduction

It is well known that producing and perceiving non-native speech sounds can be very challenging (for reviews, see Piske, MacKay & Flege, 2001; Sebastián-Gallés, 2005). In the realm of perception, much research has shown that with auditory training, the difficulty of perceiving even the hardest non-native sounds can be reduced. The most common training paradigm used to improve second language (L2)

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

Although the effectiveness of HVPT is well studied, most previous work focused exclusively on prelexical perception, using identification or discrimination tasks. The difficulty with the perception of L2 sounds, though, is paralleled by less efficient lexical processing (e.g., Pallier, Colomé &

Sebastián-Gallés, 2001; Weber & Cutler, 2004). Thus, truly successful training should also enhance performance at the lexical level. While prelexical processing only involves a phonetic analysis, lexical processing is more complex as it additionally requires mapping the incoming speech signal onto phonological representations stored in memory, and the performance gap between native and non-native listeners in L2 speech perception increases as the tasks have greater lexical involvement (Díaz et al., 2012).

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

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

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

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

Training was administered on-line, and consisted of eight sessions of an identification task using minimal pairs of real words (such as *air-hair*), with corrective feedback.<sup>2</sup> We expected the training to enhance performance in the identification task at posttest, thus replicating the findings of previous studies on the effectiveness of HVPT in improving phonetic perception of L2 sounds. Moreover, if the effect of training extends to lexical processing, performance in lexical decision should likewise improve with training.

## Method

### Pretest-Posttest-Generalization: Identification

#### Stimuli

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

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

#### Procedure

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

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

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

<sup>2</sup> Training can be done either with nonwords (e.g., Yamada, 1991) or with real words (e.g., Logan et al., 1991). Here, we chose to use real words because repeated exposure to a large

number of nonwords during training might have induced a bias to excessively accepting nonwords in the lexical decision task in pre- and posttest.

## Pretest-Posttest: Lexical decision

### Stimuli

The stimuli were the same as in Melnik & Peperkamp (2019). They consisted of 80 English test words, 40 starting with /h/ (e.g., *husband*) and 40 with a vowel (e.g. *officer*), recorded by the same male American English speaker who recorded stimuli for the identification task. They consisted of nouns, verbs and adjectives, and contained between two and four syllables. The /h/-initial and the vowel-initial words did not differ in mean frequency in the Subtlex database (Brysbaert & New, 2009) or in mean number of syllables (both  $t < 1$ ).<sup>3</sup>

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

The test and control minimal pairs were divided into two equal groups, one for pretest and one for posttest, respecting the matching in terms of frequency and number of syllables. The pretest stimuli were further divided into two counterbalancing lists: list A and list B. Each of them contained only one member of each pretest minimal pair. For instance, if the word *husband* was in list A, its nonword counterpart *usband* was in list B. The posttest stimuli were divided into lists C and D following the same principle. Thus, no list contained both members of a given word-nonword pair. Each of the four lists contained 10 /h/-initial and 10 vowel-initial words, 10 /h/-initial and 10 vowel-initial nonwords, as well as 60 control words and 60 control nonwords. Finally, for a practice phase there were two additional words and two additional nonwords, none involving /h/.

### Procedure

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

The procedure was identical to that in Melnik & Peperkamp (2019): Participants performed a speeded auditory lexical decision task. In each trial they heard a word or a nonword and had to answer if the item was an English word. They were instructed to use their dominant hand for “yes”- and their non-dominant hand for “no”-responses on a

button box. There were 160 trials divided over two blocks, each containing the same number of test and control stimuli. Trials were presented in a semi-random order such that between one to three control trials appeared between two experimental ones, and that no more than four trials of the same type (word or nonword) appeared in a row.

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

## Training: Identification

### Stimuli

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

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

### Procedure

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

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

<sup>3</sup> The familiarity of these words was evaluated by a separate group of 45 adult French learners of English in an online

rating questionnaire. The /h/- and vowel-initial words that were chosen for the experiment did not differ in mean familiarity ( $t = 1.0$ ,  $p > 0.1$ ).



day in between two sessions. Thus, the whole course of training was completed in eight to fifteen days.

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

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

## Participants

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

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

## Results

### Pretest-Posttest-Generalization: Identification

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

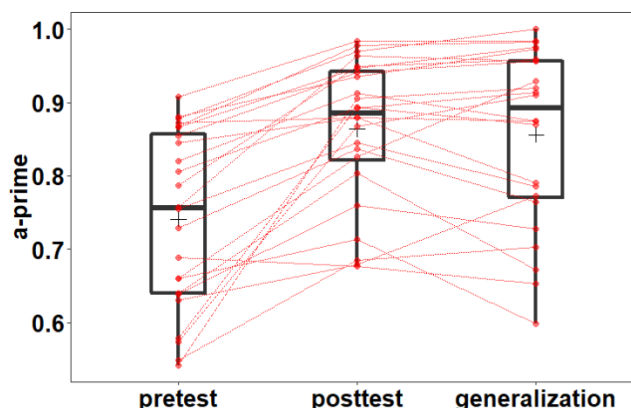


Figure 1. Boxplots of A' scores in the identification task in pretest, posttest, and generalization. The red dots represent individual participants; the lines link each participant's performance in the three sessions. The black cross marks indicate mean A' scores in each session.

### Pretest-Posttest: Lexical Decision

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

We carried out a repeated measures ANOVA by participant with the factors Session (pretest vs. posttest), Condition (test vs. control) and Lists (AC vs. BD), as well as an interaction between Session and Condition. We found main effects of Session ( $F(1, 23) = 39.36$ ,  $p < .001$ ) and Condition ( $F(1, 23) = 73.93$ ,  $p < .001$ ), and a Session X Condition interaction ( $F(1, 23) = 30.87$ ,  $p < .001$ ). Pairwise t-tests revealed that the interaction was due to the fact that in control items, the effect of Session was not significant, while in test items, there was a significant difference between

pretest and posttest ( $p < .001$ ), with the accuracy improving from an average  $A'$  score of 0.62 in pretest to 0.82 in posttest. There was no effect of the counterbalancing factor Lists.

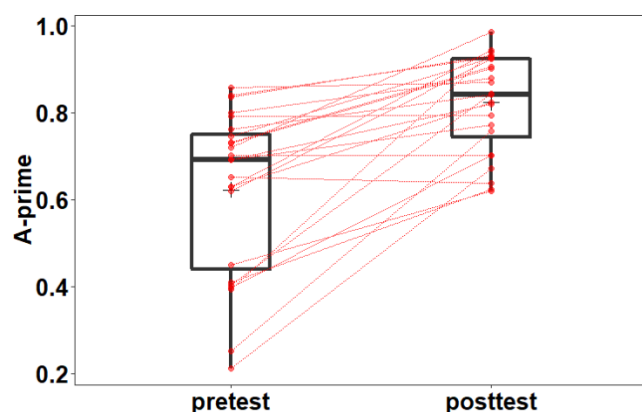


Figure 2. Boxplots of  $A'$  scores in the lexical decision task in pretest and posttest. The red dots represent individual participants; the lines link each participant's performance in both sessions. The black cross marks indicate mean  $A'$  scores in each session.

## Discussion

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

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

participants encountered with the test items was caused by the presence of /h/ and not by a lack of English vocabulary. Importantly, this difficulty was clearly reduced after training, as in posttest participants made less errors on the test items with /h/ than in pretest, while their performance did not change on control items.

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

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

On the practical side, the current findings could have implications for language teaching. The above-mentioned aspects of speech processing – lexical perception and perception of speech in noise – are inherent elements of “real life” language processing. The fact that they can be improved

by relatively short HVPT is encouraging. Moreover, our training was administered online and not in a well-controlled laboratory setting; it can thus easily complement traditional language teaching methodologies. Finally, we note that participants of our study reported that being trained on real words was very motivating, as they had the occasion not only to enhance their perception but to learn new words as well.

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

## Acknowledgments

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

## References

- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977-90.
- Carlet, A. & Cebrian, J. (2014). Training Catalan speakers to identify L2 consonants and vowels: A short-term high variability training study. *Proceedings of the 7th International Symposium on the Acquisition of Second Language Speech* (pp. 85-98). Concordia University Working Papers in Applied Linguistics.
- Cooper, A., & Wang, Y. (2011). The influence of tonal awareness and musical experience on tone word learning. *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 512-515). Hong Kong, SAR, China.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1), 1-12.
- Díaz, B., Mitterer, H., Broersma, M., Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, 22(6), 680-689.
- Homa, D., & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology*, 10, 83-94.
- Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics*, 52, 105-120.
- Ingvalson, E. M., Barr, A. M., & Wong, P. C. (2013). Poorer Phonetic Perceivers Show Greater Benefit in Phonetic-Phonological Speech Learning. *Journal of Speech Language and Hearing Research*, 56(3), 1045.
- Iverson, P., Pinet, M., Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(1), 145-160.
- Kim, Y. H. & Hazan, V. (2010). Individual variability in the perceptual learning of L2 speech sounds and its cognitive correlates. *Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech* (pp. 251-256). Poznań, Poland.
- Lecumberri, M. L. G., Cooke, M., Cutler, A. (2010). Non-native speech perception in adverse conditions: a review. *Speech Communication*, 52(11-12), 864-886.
- Lee, A. H. & Lyster, R. (2016). Effects of Different Types of Corrective Feedback on Receptive Skills in a Second Language: A Speech Perception Training Study. *Language Learning*, 66(4), 809-833.
- Lengeris, A. & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *Journal of the Acoustical Society of America*, 128(6), 3757-3768.
- Lively, S. E., Pisoni, D. B., Yamada, R. A., Tokhura, Y., Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, 96(4), 2076-2087.
- Logan, J.S., Lively, S.E., Pisoni, D.B., (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *Journal of the Acoustical Society of America*, 89, 874-886.
- Mah, J., Goad, H., Steinhauer, K. (2016). Using event-related brain potentials to assess perceptibility: The case of French speakers and English [h]. *Frontiers in Psychology*, 7, 1-14.
- Melnik, G. A., Peperkamp, S. (2019). Perceptual deletion and asymmetric lexical access in second language learners. *Journal of the Acoustical Society of America*, 145(1), EL13-EL18.
- Okuno, T. & Hardison, D. M. (2016). Perception-production link in L2 Japanese vowel duration: Training with technology. *Language Learning & Technology*, 20(2), 61-80.
- Pallier, C., Colomé, A., Sebastián-Gallés, N. (2001). The Influence of Native-Language Phonology on Lexical Access: Exemplar-Based Versus Abstract Lexical Entries. *Psychological Science*, 12(6), 445-449.
- Piske, T., Mackay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29(2), 191-215.



- Pisoni, D. & Luce, P. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, 25(1-2), 21-52.
- Sakai, M. & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, 39(1), 187-224.
- Sebastián-Gallés, N. (2005). Cross-language speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception*. Oxford: Blackwell.
- Shinohara, Y., Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/-/l/. *Journal of Phonetics*, 66, 242-251.
- Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39, 456-466.
- Wang, Y., Jongman, A., Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, 113, 1033-1043.
- Wang, Y., Spence, M., Jongman, A., Sereno, J. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, 106, 3649-3658.
- Weber, A., Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1), 1-25.
- Yamada, J. (1991). The discrimination learning of the liquids /r/ and /l/ by Japanese speakers. *Journal of Psycholinguistic Research*, 20(1), 31-46.

## References

- Aliaga-García, C., Mora, J. C., & Cerviño-Povedano, E. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznań Studies in Contemporary Linguistics*, 47(1), 1. <https://doi.org/10.2478/psicl-2011-0002>
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1), 158–173. <https://doi.org/10.1037/0033-295X.105.1.158>
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross- Language Research*. Timonium, MD: York Press.
- Best, Catherine T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 345–360. <https://doi.org/10.1037/0096-1523.14.3.345>
- Birdsong, D. (2006). Age and Second Language Acquisition and Processing: A Selective Overview. *Language Learning*, 56(1), 9–49. <https://doi.org/10.1111/j.1467-9922.2006.00353.x>
- Bradlow, A R, Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4), 2299–2310. <https://doi.org/10.1121/1.418276>
- Bradlow, Ann R, & Alexander, J. a. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, 121(4), 2339–2349. <https://doi.org/10.1121/1.2642103>
- Brannen, K. (2002). The Role of Perception in Differential Substitution. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, 47(1–2), 1–46. <https://doi.org/10.1017/s0008413100018004>
- Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology*, 64(1), 74–95. <https://doi.org/10.1080/17470218.2010.499174>

- Carlet, A., & Cebrian, J. (2014). Training Catalan Speakers to Identify L2 Consonants and Vowels: A Short-Term High Variability Training Study. *Concordia Working Papers in Applied Linguistics*, 5, 85–98.
- Cho, M.-H. (2008). Asymmetries in the perception and production of the English incomplete off-gliding diphthongs by Korean speakers. *Studies in Phonetics, Phonology, and Morphology*, 14(3), 483–499. <https://doi.org/10.17959/sppm.2008.14.3.483>
- Colantoni, L., & Steele, J. (2008). Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics*, 29(3), 489–534. <https://doi.org/10.1017/S0142716408080223>
- Cook, S. V., & Gor, K. (2015). Lexical access in L2: Representational deficit or processing constraint? *The Mental Lexicon*, 10(2), 247–270. <https://doi.org/10.1075/ml.10.2.04coo>
- Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, 34(2), 269–284. <https://doi.org/10.1016/j.wocn.2005.06.002>
- Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, 116(6), 3668–3678. <https://doi.org/10.1121/1.1810292>
- Dar, M., Keren-Portnoy, T., & Vihman, M. (2018). An order effect in English infants' discrimination of an Urdu affricate contrast. *Journal of Phonetics*, 67, 49–64. <https://doi.org/10.1016/j.wocn.2017.12.002>
- Darcy, I., Daidone, D., & Chisato, K. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, 8(3), 372–420. <https://doi.org/10.1075/ml.8.3.06dar>
- Darcy, I., Mora, J., & Daidone, D. (2014). Attention Control and Inhibition Influence Phonological Development in a Second Language. *International Symposium on the Acquisition of Second Language Speech*, 115–129. Retrieved from [http://doe.concordia.ca/copad/documents/10\\_Darcy\\_Mora\\_Daidone\\_Vol5.pdf](http://doe.concordia.ca/copad/documents/10_Darcy_Mora_Daidone_Vol5.pdf)
- Darcy, I., Mora, J. C., & Daidone, D. (2016). The Role of Inhibitory Control in Second Language Phonological Processing. *Language Learning*, 66(4), 741–773. <https://doi.org/10.1111/lang.12161>

- Darcy, I., Park, H., & Yang, C.-L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, 40, 63–72.  
<https://doi.org/10.1016/j.lindif.2015.04.005>
- Davidson, L., & Shaw, J. a. (2012). Sources of illusion in consonant cluster perception. *Journal of Phonetics*, 40(2), 234–248. <https://doi.org/10.1016/j.wocn.2011.11.005>
- Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, 22(6), 680–689.  
<https://doi.org/10.1016/j.lindif.2012.05.005>
- Dörnyei, Z. (2014). *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. New York: Routledge. <https://doi.org/10.4324/9781410613349>
- Dufour, S., Nguyen, N. (2008). L'influence de la langue maternelle dans la perception de la parole. *Travaux Interdisciplinaires sur la Parole et le Langage*, 27, 39-49.
- Dupoux, E., & Pallier, C. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes*, 16(5–6), 491–505.  
<https://doi.org/10.1080/01690960143000191>
- Dupoux, Emmanuel, Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1568–1578. <https://doi.org/10.1037/0096-1523.25.6.1568>
- Dupoux, Emmanuel, Pallier, C., Sebastian, N., & Mehler, J. (1997). A Destressing “Deafness” in French? *Journal of Memory and Language*, 36(3), 406–421.  
<https://doi.org/10.1006/jmla.1996.2500>
- Dupoux, Emmanuel, Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress “deafness”: The case of French learners of Spanish. *Cognition*, 106(2), 682–706.  
<https://doi.org/10.1016/j.cognition.2007.04.001>
- Durvasula, K., & Kahng, J. (2016). Phonological Alternations Modulate Illusory Vowels in Perceptual Epenthesis. *Phonology*, (32(3)), 385–416.  
<https://doi.org/10.1017/S0952675715000263>

- Eimas, P. D., Miller, I. L., & Iusczyk, P. W. (1987). On infant speech perception and the acquisition of language. In S. Hamad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 161-195). Cambridge University Press: New York.
- Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, 36(2), 345–360.  
<https://doi.org/10.1016/j.wocn.2007.11.002>
- Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics*, 40(2), 280–288.  
<https://doi.org/10.1016/j.wocn.2011.11.004>
- Flege, J., Bohn, O., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *J. Phonetics*, 25, 437–470. Retrieved from  
<http://dx.doi.org/10.1006/jpho.1997.0052>
- Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language : evidence for the effect of equivalence classification. *Journal of Phonetics*, 15(1), 47–65. Retrieved from <http://www.citeulike.org/group/278/article/520469>
- Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America*, 93(3), 1589–1608.  
<https://doi.org/10.1121/1.406818>
- Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross- language research* (pp. 229–273). Timonium, MD: York Press. <https://doi.org/10.1111/j.1600-0404.1995.tb01710.x>
- Flege, J. E. ., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, 106(5), 2973–2987. <https://doi.org/10.1121/1.428116>
- Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech*, 38(1), 25–55.  
<https://doi.org/10.1177/002383099503800102>
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R.” *Neuropsychologia*, 9(3), 317–323. [https://doi.org/10.1016/0028-3932\(71\)90027-3](https://doi.org/10.1016/0028-3932(71)90027-3)

- Guevara-Rukoz, A., Lin, I., Morii, M., Minagawa, Y., Dupoux, E., & Peperkamp, S. (2017). Which epenthetic vowel? Phonetic categories versus acoustic detail in perceptual vowel epenthesis. *The Journal of the Acoustical Society of America*, 142(2), EL211–EL217. <https://doi.org/10.1121/1.4998138>
- Guion, S. G., & Flege, J. E. (2000). An investigation of current models of second language speech perception : The case of Japanese adults ' perception of English consonants, 107(5), 2711–2724.
- Hallé, P. A., & Best, C. T. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *The Journal of the Acoustical Society of America*, 121(5), 2899–2914. <https://doi.org/10.1121/1.2534656>
- Hancin-Bhatt, B. J. (1994a). Phonological transfer in second language perception and production. University of Illinois: Unpublished doctoral dissertation.
- Hattori, K., & Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America*, 125(1), 469. <https://doi.org/10.1121/1.3021295>
- Hattori, K., & Iverson, P. (2010). Examination of the relationship between L2 perception and production: an investigation of English/r/-/l/perception and production by adult Japanese speakers. In *Interspeech Satellite Workshop on Second Language Studies* (pp. 3–6). Retrieved from [http://www.gavo.t.u-tokyo.ac.jp/L2WS2010/papers/L2WS2010\\_P2-04.pdf](http://www.gavo.t.u-tokyo.ac.jp/L2WS2010/papers/L2WS2010_P2-04.pdf)
- Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, 24(1), 5–33. <https://doi.org/10.1177/0267658307082980>
- Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47(3), 360–378. <https://doi.org/10.1016/j.specom.2005.04.007>
- 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>
- Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and

- Spanish listeners. *The Journal of the Acoustical Society of America*, 117(2), 896–907.  
<https://doi.org/10.1121/1.1823291>
- Ingvalson, E. M., Dhar, S., Wong, P. C. M., & Liu, H. (2015). Working memory training to improve speech perception in noise across languages. *The Journal of the Acoustical Society of America*, 137(6), 3477–3486. <https://doi.org/10.1121/1.4921601>
- Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, 126(2), 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 Journal of the Acoustical Society of America*, 118(5), 3267–3278.  
<https://doi.org/10.1121/1.2062307>
- IVERSON, P., PINET, M., & EVANS, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(01), 145–160.  
<https://doi.org/10.1017/S0142716411000300>
- Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences*, 10(11), 480–486.  
<https://doi.org/10.1016/j.tics.2006.09.002>
- Janda, R. D., & Auger, J. (1992). Quantitative Evidence, Qualitative Hypercorrection, Sociolinguistic Variables - and French Speakers' 'eadhaches with English h/0. *Language & Communication*, 12(3/4), 195–236.
- John, P., & Cardoso, W. (2008). Francophone ESL Learners and [h]-Epenthesis. *Concordia Working Papers in Applied Linguistics*, 1(II), 76–97.
- Kang, Y. (2011). Loanword Phonology. In M. van Oostendorp, C. J. Ewen, E. V. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (Vol. 4, pp. 2258–2281). Hoboken, N.J.: Wiley-Blackwell. <https://doi.org/10.3758/s13428-014-0458-y>
- Kim, Y. H., & Hazan, V. (2010). Individual variability in the perceptual learning of L2 speech sounds and its cognitive correlates. In *New Sounds 2010. Sixth International Symposium on the Acquisition of Second Language Speech* (pp. 251–256). Retrieved from [http://ifa.amu.edu.pl/newsounds/files/proceedings/proceedings\\_quotable\\_version.pdf](http://ifa.amu.edu.pl/newsounds/files/proceedings/proceedings_quotable_version.pdf)



- Kittredge, A. K., & Dell, G. S. (2016). Learning to speak by listening: Transfer of phonotactics from perception to production. *Journal of Memory and Language*, 89, 8–22.  
<https://doi.org/10.1016/j.jml.2015.08.001>
- Kriengwatana, B. P., & Escudero, P. (2017). Directional Asymmetries in Vowel Perception of Adult Nonnative Listeners Do Not Change Over Time With Language Experience. *Journal of Speech, Language, and Hearing Research*, 60(4), 1088–1093.  
[https://doi.org/10.1044/2016\\_jslhr-h-16-0050](https://doi.org/10.1044/2016_jslhr-h-16-0050)
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 979–1000. <https://doi.org/10.1098/rstb.2007.2154>
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2). <https://doi.org/10.1111/j.1467-7687.2006.00468.x>
- Lacharite, D., & Prévost, P. (1999). Le rôle de la langue maternelle et de l’enseignement dans l’acquisition des segments de l’anglais langue seconde par des apprenant francophones.
- Ladefoged, P., Maddieson, I. (1996). *The Sounds of the World’s Languages*. Blackwell; Malden, MA.
- Lee, A. H., & Lyster, R. (2016). Effects of Different Types of Corrective Feedback on Receptive Skills in a Second Language: A Speech Perception Training Study. *Language Learning*, 66(4), 809–833. <https://doi.org/10.1111/lang.12167>
- Lengeris, A. (2018). Computer-based auditory training improves second-language vowel production in spontaneous speech. *JASA EL*, 144. <https://doi.org/10.1093/bjc/azq038>
- Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *Journal of the Acoustical Society of America*, 128(6), 3757–3768.  
<https://doi.org/10.1121/1.3506351>
- Lengeris, A., & Nicolaidis, K. (2014). Effect of phonetic training on the perception of English consonants by Greek speakers in quiet and noise conditions. *The Journal of the Acoustical Society of America*, 136(4), 2107–2107. <https://doi.org/10.1121/1.4899580>
- Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley



- Lenneberg, E. H. (1969). On Explaining Language. *Science*, 164(3880), 635–643.  
<https://doi.org/10.1126/science.164.3880.635>
- Leong, C. X. R., Price, J. M., Pitchford, N. J., & van Heuven, W. J. B. (2018). High variability phonetic training in adaptive adverse conditions is rapid, effective, and sustained. *PLOS ONE*, 13(10), e0204888. <https://doi.org/10.1371/journal.pone.0204888>
- Lewis, P., & Roux, J. C. (1999). XHOSA L2 SPEECH INTELLIGIBILITY: EJECTIVES , IMPLOSIVES AND CLICKS. In *PROCEEDINGS OF THE INTERNATIONAL CONGRESS OF PHONETIC SCIENCES* (pp. 1149–1152). San Francisco.
- Lim, S., & Holt, L. L. (2011). Learning Foreign Sounds in an Alien World: Videogame Training Improves Non-Native Speech Categorization. *Cognitive Science*, 35(7), 1390–1405.  
<https://doi.org/10.1111/j.1551-6709.2011.01192.x>
- Linck, J. A., Schwieter, J. W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism*, 15(3), 651–662.  
<https://doi.org/10.1017/S136672891100054X>
- Liu, J., & Zhang, J. (2016). The effects of talker variability and variances on incidental learning of lexical tones of Asian Languages and Literatures , University of Minnesota , USA 2 . Incidental Learning — Video Game. *Proceedings of the 5th International Symposium on Tonal Aspects of Languages (TAL 2016)*, (May), 23–27.
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories, 94, 1242–1255.
- Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, 96(4), 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. *The Journal of the Acoustical Society of America*, 89(2), 874–886. <https://doi.org/10.1121/1.1894649>
- Lyster, R., & Izquierdo, J. (2009). Prompts versus recasts in dyadic interaction. *Language Learning*, 59, 453–498.  
[doi:10.1111/j.1467-9922.2009.00512.x](https://doi.org/10.1111/j.1467-9922.2009.00512.x)

- 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(OCT), 1–14. <https://doi.org/10.3389/fpsyg.2016.01469>
- Masuda, H., & Arai, T. (2008). Perception and production of consonant clusters in Japanese-English bilingual and Japanese monolingual speakers. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 1988–1991.
- Melnik, G. A., & Peperkamp, S. (2019). Perceptual deletion and asymmetric lexical access in second language learners. *The Journal of the Acoustical Society of America*, 145(1), EL13–EL18. <https://doi.org/10.1121/1.5085648>
- Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, 109(1), 168–173. <https://doi.org/10.1016/j.cognition.2008.08.002>
- Motobashi-Saigo, M., & Hardison, D. M. (2009). Acquisition of L2 Japanese geminates training with waveform displays. *Language Learning & Technology*, 13(2), 29–47. Retrieved from <http://llt.msu.edu/vol13num2/motohashisaigohardison.pdf>
- Nadeu, M. (2014). Stress- and speech rate-induced vowel quality variation in Catalan and Spanish. *Journal of Phonetics*, 46(1), 1–22. <https://doi.org/10.1016/j.wocn.2014.05.003>
- Nam, Y., & Polka, L. (2016). The phonetic landscape in infant consonant perception is an uneven terrain. *Cognition*, 155, 57–66. <https://doi.org/10.1016/j.cognition.2016.06.005>
- Okuno, T., & Hardison, D. M. (2016). Perception-Production Link in L2 Japanese Vowel Duration: Training With Technology. *Language Learning & Technology*, 20(2), 61–80. Retrieved from <http://llt.msu.edu/issues/june2016/okunohardison.pdf>
- Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition*, 111(2), 263–269. <https://doi.org/10.1016/j.cognition.2008.12.007>
- Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The Influence of Native-Language Phonology on Lexical Access: Exemplar-Based Versus Abstract Lexical Entries. *Psychological Science*, 12(6), 445–449. <https://doi.org/10.1111/1467-9280.00383>

- Patricia K. Kuhl. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93–107.
- Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2018). Advanced Second Language Learners’ Perception of Lexical Tone Contrasts. *Studies in Second Language Acquisition*, (May), 1–28. <https://doi.org/10.1017/s0272263117000444>
- Peperkamp, S., Vendelin, I., & Nakamura, K. (2008). On the perceptual origin of loanword adaptations: experimental evidence from Japanese. *Phonology*, 25(1), 129–164. <https://doi.org/10.1017/S0952675708001425>
- Pierce, L. J., Genesee, F., Delcenserie, A., & Morgan, G. (2017). Variations in phonological working memory: Linking early language experiences and language learning outcomes. *Applied Psycholinguistics*, 38(6), 1265–1300. <https://doi.org/10.1017/S0142716417000236>
- Piske, T., MacKay, I. R. a., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, 29(2), 191–215. <https://doi.org/10.1006/jpho.2001.0134>
- Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, 25(1–2), 21–52. [https://doi.org/10.1016/0010-0277\(87\)90003-5](https://doi.org/10.1016/0010-0277(87)90003-5)
- Polka, L., & Bohn, O. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America*, 100(1), 577–592. <https://doi.org/10.1121/1.415884>
- Polka, L., & Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41(1), 221–231. [https://doi.org/10.1016/S0167-6393\(02\)00105-X](https://doi.org/10.1016/S0167-6393(02)00105-X)
- Polka, L., & Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, 39(4), 467–478. <https://doi.org/10.1016/j.wocn.2010.08.007>
- Rauber, A., Rato, A., Kluge, D. and Santos, G. R. (2012). TP (Version. 3.1) [Software]. Available from: [http://www.worken.com.br/tp\\_regfree.php/](http://www.worken.com.br/tp_regfree.php/)
- Rato, A., & Rauber, A. (2015). The Effects of Perceptual Training on the Production of English vowel contrasts by Portuguese Learners. In *Proceedings of the 18th International*

- Congress of Phonetic Sciences*. Retrieved from <http://www.icphs2015.info/pdfs/Papers/ICPHS0656.pdf>
- Riney, T. J., & Flege, J. E. (1998). Changes Over Time in Global Foreign Accent and Liquid Identifiability and Accuracy. *Studies in Second Language Acquisition*, 20(02), 213–243. <https://doi.org/10.1017/s0272263198002058>
- Roach, P. (2009). *English Phonetics and Phonology*.
- Rojczyk, A. (2010). Forming New Vowel Categories in Second Language Speech: The Case of Polish Learners' Production of English /I/ and /e/. *Research in Language*, 8(0), 1–13. <https://doi.org/10.2478/v10015-010-0008-3>
- Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, 5(NOV), 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(1), 187–224. <https://doi.org/10.1017/S0142716417000418>
- Schmidt, A. M., & Flege, J. E. (1995). Effects of Speaking Rate Changes on Native and Nonnative Speech Production. *Phonetica*, 52(1), 41–54. <https://doi.org/10.1159/000262028>
- Schwartz, J.L., Abry, C., Boe, L. J., Ménard, L., Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion–Focalisation Theory. *Speech Communication*, 45, 425-434.
- Scovel, T. (1988). A time to speak: A psycholinguistic inquiry into the critical period for human speech. Rowley, MA: Newbury House.
- Sebastián-Gallés, N., & Baus, C. (2005). On the relationship between perception and production in L2 categories. In *Twenty-first century Psycholinguistics: Four cornerstones*.
- Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: attention shifting and second-language proficiency. *Memory & Cognition*, 33(4), 644–653. <https://doi.org/10.3758/BF03195331>

- 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>
- Simmonds, A. J., Wise, R. J. S., Dhanjal, N. S., & Leech, R. (2011a). A comparison of sensory-motor activity during speech in first and second languages. *Journal of Neurophysiology*, 106(1), 470–478. <https://doi.org/10.1152/jn.00343.2011>
- Simmonds, A. J., Wise, R. J. S., & Leech, R. (2011b). Two tongues, one brain: Imaging bilingual speech production. *Frontiers in Psychology*, 2(JUL), 1–13.  
<https://doi.org/10.3389/fpsyg.2011.00166>
- Strange, Winifred. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39(4), 456–466.  
<https://doi.org/10.1016/j.wocn.2010.09.001>
- Strange, Winifred, & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, 36(2), 131–145.  
<https://doi.org/10.3758/BF03202673>
- Takagi, N. (1993). Perception of American English /r/ and /l/ by adult Japanese learners of English : A unified view. Unpublished Dissertation, University of California, Irvine, Irvine, CA.
- 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(1), 23–29. <https://doi.org/10.1016/j.ijpsycho.2015.04.020>
- Thomson, R. I. (2012). Improving L2 Listeners' Perception of English Vowels: A Computer-Mediated Approach. *Language Learning*, 62(4), 1231–1258.  
<https://doi.org/10.1111/j.1467-9922.2012.00724.x>
- Trubetzkoy, N. (1939). *Principles of phonology*. Paris: Klincksieck.
- Tsushima, T., Shiraki, S., Yoshida, K., & Sasaki, M. (2003). On stimulus order effects in discrimination of nonnative consonant contrasts. *Acoustical Science and Technology*, 24(6), 410–412. <https://doi.org/10.1250/ast.24.410>
- Tsushima, T., Shiraki, S., Yoshida, K., & Sasaki, M. (2005). Stimulus order effects in discrimination of a nonnative consonant contrast, English /b-v/, by Japanese listeners in

- the AX discrimination procedure. Unpublished paper presented at ‘First Acoustical Society of America Workshop on L2 Speech Learning’, Vancouver, Canada.
- Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, 71(1), 4–21.  
<https://doi.org/10.1159/000356237>
- Vlahou, E., Seitz, A. R., & Kopčo, N. (2019). Nonnative implicit phonetic training in multiple reverberant environments. *Attention, Perception, & Psychophysics*.  
<https://doi.org/10.3758/s13414-019-01680-0>
- Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, 113(2), 1033–1043. <https://doi.org/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>
- Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, 36(2), 250–267.  
<https://doi.org/10.1016/j.wocn.2007.06.004>
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1), 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 perception. *Perception & Psychophysics*, 37(1), 35–44.  
<https://doi.org/10.3758/BF03207136>
- Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, 75(6), 1866–1878.  
<https://doi.org/10.1121/1.390988>
- White, E. J., Titone, D., Genesee, F., & Steinhauer, K. (2017). Phonological processing in late second language learners: The effects of proficiency and task. *Bilingualism: Language and Cognition*, 20(1), 162–183. <https://doi.org/10.1017/S1366728915000620>

- Yun, S. (2014). The Role of Acoustic Cues in Nonnative Cluster Repairs. In R. E. Santana-LaBarge (Ed.), *Proceedings of the 31st West Coast Conference on Formal Linguistics* (pp. 514–523). Somerville: Cascadilla Proceedings Project.
- Zhang, J., Xie, L., Li, Y., Chatterjee, M., & Ding, N. (2014). How noise and language proficiency influence speech recognition by individual non-native listeners. *PLoS ONE*, 9(11). <https://doi.org/10.1371/journal.pone.0113386>
- Zhang, X., Samuel, A. G., & Liu, S. (2012). The perception and representation of segmental and prosodic Mandarin contrasts in native speakers of Cantonese. *Journal of Memory and Language*, 66(3), 438–457. <https://doi.org/10.1016/j.jml.2011.12.006>



## RÉSUMÉ

L'apprentissage d'une langue étrangère nécessite une quantité considérable de temps et d'efforts. Les apprenants doivent faire face à de nombreux défis dans cet apprentissage, dont le traitement des sons qui n'existent pas dans leur langue maternelle. La différence entre les propriétés de la langue maternelle et de la langue étrangère entraîne des distorsions dans la perception et un accent dans la production des sons non-natifs. De plus, ces difficultés persistent à tous les niveaux de traitement, car les problèmes de perception et de production d'un son influencent le traitement des mots contenant ces sons. Heureusement, la capacité à percevoir et à produire les sons de la L2 (langue seconde) s'améliore progressivement. Cette thèse porte sur le traitement phonologique de la L2 et son développement à travers les modalités (perception vs. production) et les niveaux de traitement (niveau prélexical vs. lexical).

Dans la première partie de la thèse, nous étudions la relation entre la perception et la production en L2. Les résultats des études précédentes ont souvent été contradictoires et nous suggérons que plusieurs limitations méthodologiques aient pu y créer des confusions. Nous avons donc pris en compte ces limitations méthodologiques et nous avons développé un paradigme expérimental afin de tester la perception et la production du contraste français /u/-/y/ par des apprenants anglophones. Nous avons utilisé des tâches qui visent le traitement prélexical et lexical afin d'examiner si le lien entre les deux modalités, s'il en existe un, est maintenu à travers les niveaux de traitement. Les résultats ont montré que la perception et la production sont corrélées, mais uniquement au niveau prélexical. De plus, nous avons trouvé que le développement de la perception précède celui de la production car il faut d'abord bien percevoir un son non-natif afin de le produire correctement.

Dans la deuxième partie, nous avons poursuivi l'étude du traitement phonologique à travers les niveaux de traitement en nous concentrant sur la perception du son anglais /h/ par des apprenants francophones. Nous avons d'abord examiné si les difficultés à percevoir ce son précédemment signalées au niveau prélexical posaient également problème au niveau lexical. De plus, nous avons examiné si l'asymétrie observée dans la production (les francophones omettent le /h/ plus souvent qu'il ne l'insèrent) était présente dans la perception. Les résultats ont révélé que les apprenants francophones ont du mal à percevoir des mots et des non-mots contenant le /h/. De plus, une performance asymétrique a été observée. Nous avons interprété ceci comme une indication que les représentations phonologiques des mots anglais contenant le /h/ sont imprécises chez les apprenants francophones. Dans un second temps, nous avons examiné si un entraînement phonétique pouvait améliorer la perception du /h/ non seulement au niveau prélexical, mais également au niveau lexical. Nous avons démontré que l'entraînement phonétique améliorait la perception du /h/ dans les deux niveaux de traitement. De plus, cet effet positif a été maintenu quatre mois après l'entraînement. Enfin, nous avons examiné si les asymétries dans la perception du /h/ au niveau lexical pouvaient s'expliquer par des asymétries au niveau prélexical. Un tel lien n'a cependant pas été observé dans les résultats.

Dans l'ensemble, cette thèse démontre que les mécanismes sous-jacents au traitement de la parole en L2 sont complexes et dynamiques, et influencent ainsi la perception et la production tant à travers les modalités qu'à travers les niveaux de traitement. Enfin, des pistes pour les recherches futures, qui permettraient d'explorer davantage les liens entre ces éléments du traitement phonologique, sont proposées. Cela mènerait à une compréhension plus approfondie des processus impliqués dans l'acquisition de la L2.

## MOTS CLÉS

Psycholinguistique; phonologie; acquisition de la langue seconde; perception de la parole; production; niveaux de traitement; reconnaissance des mots; asymétries perceptives; entraînement phonétique

## ABSTRACT

Learning a foreign language (L2) is a difficult task, requiring considerable amounts of time and effort. One of the challenges learners must face is the processing of sounds that do not exist or are not used contrastively in their native language. The mismatch between the properties of the native language and the foreign one leads to distortions in the perception of non-native sounds and to foreign accent in their production. Moreover, these difficulties persist across levels of processing as problems in prelexical L2 sound perception and production influence the processing of words containing these sounds. Fortunately, with growing proficiency the abilities to perceive and produce L2 sounds gradually improve, although they might never attain native-like levels. This thesis focuses on L2 phonological processing and its development across modalities (perception vs. production) and across levels of processing (prelexical vs. lexical).

In the first part of the thesis, we investigate the relationship between perception and production in L2. Previous literature has provided contradictory evidence as to whether perception and production develop in parallel. We hypothesized that several methodological limitations could have brought confounds in some of these previous studies. We therefore designed an experiment that addressed these methodological issues and tested proficient English learners of French on their perception and production of the French contrast /u/-/y/ that does not exist in English. We included tasks that tap into both prelexical and lexical levels of processing in order to examine whether the link between the two modalities, if any, holds across levels of processing. Results showed that perception and production were correlated, but only when tested with tasks that tap into the same level of processing. We next explored if the developments in one modality precede developments in the other and found that good perception is indeed a prerequisite for good production.

In the second part of the thesis, we continue to investigate the phonological processing of L2 across levels by focusing on the perception of the English sound /h/ by intermediate to proficient French learners of English. We first studied if the poor perception of this sound previously reported at the prelexical level also causes problems at the lexical level. We also looked at whether asymmetries found in production (i.e. more deletions than insertions) are reflected in perception. The results revealed that French learners of English have difficulty in perceiving /h/-initial words and non-words at the lexical level. Moreover, an asymmetry was indeed observed in their performance, which was interpreted as an indication that French learners of English have imprecise phonological representations of /h/-initial but not of vowel-initial words. Second, we carried out a training study to test if phonetic training could improve the perception of /h/ not only at the prelexical, but at the lexical level as well. We found that the High Phonetic Variability training did improve the perception of /h/ both at the prelexical and lexical levels, and that this positive effect was retained four months after training. Finally, we examined if asymmetries in the perception of /h/ at the lexical level could be explained by asymmetries at the prelexical level. The results revealed no such relationship.

Overall, this thesis demonstrates the complex and dynamic nature of the mechanisms underlying non-native speech processing and its development during learning both across modalities and across levels of processing. We discuss how future research could further explore the links between these elements of the phonological processing apparatus to get a better understanding of L2 acquisition.

## KEYWORDS

Psycholinguistics; phonology; second language acquisition; speech perception; speech production; levels of processing; word recognition; perceptual asymmetries; phonetic training