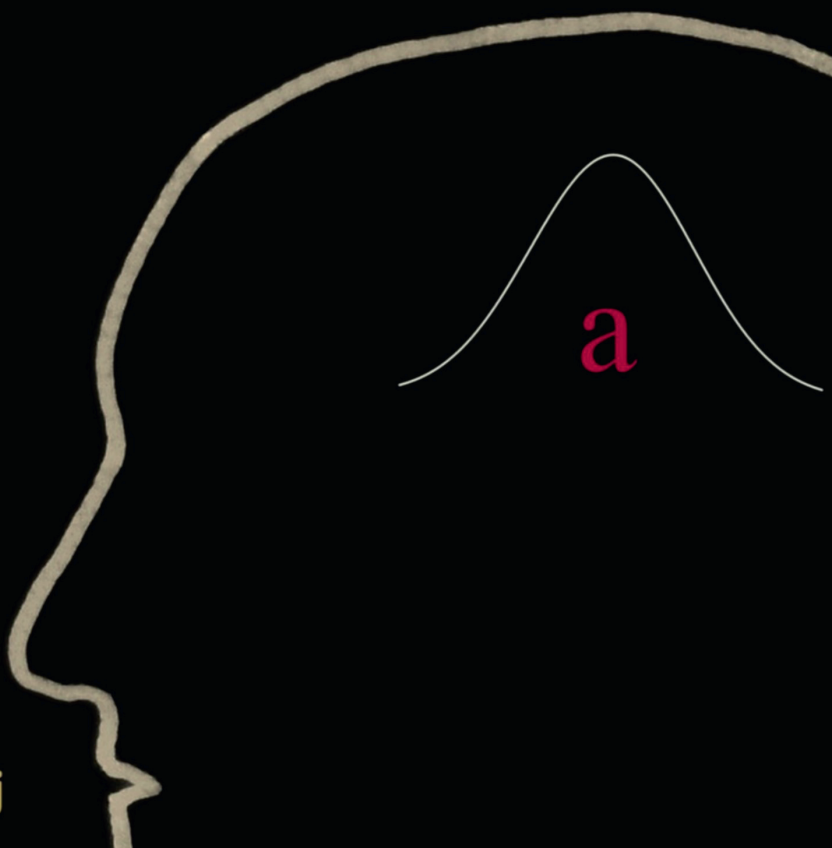
aa aaa a aa aaa

Distributional learning of vowel categories in infants and adults



Karin Wanrooij

Distributional learning of vowel categories in infants and adults

Karin Wanrooij

The research described in this thesis was performed at the Amsterdam Center for Language and Communication (ACLC) of the University of Amsterdam.

ISBN: 978-94-6259-489-0
NUR: 616
Author: Karin Wanrooij
Cover design: Matthijs Wanrooij
Printed by: Ipskamp Drukkers, Enschede, The Netherlands

Distributional learning of vowel categories in infants and adults

Academisch Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel
op donderdag 23 april 2015, te 14:00 uur
door

Karin Elisabeth Wanrooij

geboren te Rotterdam

Promotiecommissie

Promotor: prof. dr. P.P.G. Boersma (Universiteit van Amsterdam)
Co-promotor: dr. T.L. van Zuijen (Universiteit van Amsterdam)

Overige leden: prof. dr. M.T.C. Ernestus (Radboud Universiteit Nijmegen)
prof. dr. J.H. Hulstijn (Universiteit van Amsterdam)
dr. M. Huotilainen (Universiteit van Helsinki)
dr. J.E. Rispens (Universiteit van Amsterdam)
prof. dr. J.C. Schaeffer (Universiteit van Amsterdam)

Faculteit der Geesteswetenschappen

Contents

Dankwoord (acknowledgments)

Nog geen eeuw geleden verspeelde mijn grootvader het privilege om naar de universiteit te gaan, dat hem als oudste uit een doktersgezin met negen kinderen toegekend was. Hij vroeg zijn vader of hij niet in plaats van medicijnen zijn echte passie wis- en natuurkunde mocht gaan studeren. Het antwoord was duidelijk en meedogenloos: het voorrecht ging over op de volgende zoon in het rijtje. Mijn grootvader schreef op eigen kracht als niet-academicus zijn hele leven over de wis- en natuurkunde. Ik ben heel dankbaar dat ik eigen studiekeuzes heb kunnen maken en dit proefschrift heb mogen schrijven. Dat was niet gelukt zonder de herinnering aan mijn grootvader en de ondersteuning van mijn familie, vrienden, collega's en anderen.

Ik ben Paul Boersma dankbaar dat hij mij de kans heeft gegeven dit project te ondernemen. Vanaf het begin was er in onze "Vici-groep" een positieve sfeer van collegialiteit en constructieve kritiek, waarin iedereen zijn fouten mag maken en mensen elkaar graag helpen. Zo'n sfeer is niet vanzelfsprekend. Ook waardeer ik de vrijheid die ik heb gekregen om zelf een plan te trekken, en de medewerking om die plannen vervolgens uit te voeren. Ik denk daarbij bijvoorbeeld aan het EEG-lab, dat ik nodig had voor mijn project en dat we vervolgens opgezet hebben. Paul, heel veel dank daarvoor. Het was een feest om zo onderzoek te mogen doen.

Titia van Zuijen heeft mij wegwijs gemaakt in de wereld van het EEG-onderzoek. Zonder haar had ik de EEG-experimenten niet kunnen doen. Titia, het was fijn om jou als co-promotor te hebben en leuk om te ontdekken dat we veel gemeenschappelijke belangstellingen hebben. Verder stel ik het erg op prijs dat je mij geïntroduceerd hebt bij het lab in Helsinki, waar ik veel heb opgestoken.

Paola Escudero, jij bent degene die mij de wereld van de experimenten heeft ingetrokken. Ik was bij jou student-assistent. De eerste drie publicaties waarin ik auteur ben, heb ik met jou geschreven. Ook heb je me jouw data gegeven om hoofdstuk V van dit proefschrift te schrijven. Dat waardeer ik allemaal bijzonder.

Verder bedank ik Maartje Raijmakers en Titia Benders, twee co-auteurs die ik nog niet genoemd heb. Het was een groot plezier en inspirerend om met jullie samen te werken. Ik hoop dat we daar in de toekomst opnieuw kansen voor zullen krijgen!

Voor alle experimenten in dit proefschrift was onze technicus, Dirk Jan Vet, absoluut onmisbaar. Dirk, wat zet jij je in om alle experimenten tot een succes te maken! Voor technische vragen kan ik te pas en te onpas bij je binnenlopen. Je denkt mee en vooruit, zodat problemen waar wij (de onderzoekers) nog niet aan gedacht hebben, niet opduiken. Geweldig!

Johanna de Vos, Gisela Govaart, Marieke van den Heuvel en Marja Caverlé, jullie waren mijn trouwe student-assistenten, die met enthousiasme bij de experimenten geholpen hebben. Johanna, jij hebt ook je scriptie bij mij geschreven en daarmee een mooie bijdrage geleverd aan hoofdstuk IV. Allemaal, veel dank!

Wie ik ook erkentelijk ben zijn Sophie ter Schure, Caroline Junge, Janny Stapel, Marlene Meyer en vele anderen die baby-onderzoek doen en die ik heb leren kennen aan de Universiteit van Amsterdam, via de Baby Circle bijeenkomsten of via de Baby Brain and Cognition bijeenkomsten. Allemaal hebben jullie mij dingen geleerd over baby-onderzoek. Jullie openheid en bereidheid tot het delen van informatie apprecieer ik bijzonder.

Also, many thanks to my colleagues in Finland! Minna Huotilainen, Eino Partanen, Maria Mittag and many others, you were very kind in taking ample time to answer my questions and to show me the details of your EEG-research, in the lab for the methods and behind the computer for the analysis.

Natuurlijk bedank ik alle deelnemers aan de experimenten, met een bijzondere vermelding voor de baby's en hun ouders! De baby's waren stuk voor stuk schattig en voorbeeldige deelnemers (zoals te zien is aan de voorbeeldfoto's in hoofdstuk II).

Jan-Willem van Leussen en Katja Chládková, wat zijn jullie fijne kamergenoten! Margarita Gulian, ik mag jou vast toevoegen aan dit rijtje. Thuis krijg ik vaak te horen dat ik voor me uit prevel, maar jullie vonden alles best. Altijd bereid om te helpen of om gewoon een praatje te maken, en altijd attent. Dank!

Ten slotte kom ik weer terug bij mijn familie. Lieve Onno, Sterre en Mané, dank voor jullie niet-aflatende vertrouwen in en ook enthousiasme voor dit project. Vanaf het begin hebben jullie mij aangemoedigd om weer te gaan studeren en ook om aan dit project te beginnen. Datzelfde geldt voor mijn ouders, mijn schoonouders, mijn broer en andere familie. Er zat dus een heel ondersteuningsteam achter mij! Ik draag dit boek daarom aan mijn familie op.

Author contributions

I. General introduction

Karin Wanrooij

II. Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study.

Karin Wanrooij, Paul Boersma, and Titia L. Van Zuijen
Frontiers in Psychology - Language Sciences 2014, 5, article 77, 1-12, doi: 10.3389/fpsyg.2014.00077

KW posed the research question and designed the experiment, with valuable input from PB and TvZ. KW and PB did the stimulus generation. KW applied for approval with the Ethical Committee, recruited the infants, and ran the experiments. KW did the analysis of the sleep stages. KW, TvZ and PB did the EEG analysis. KW and PB did the statistical analysis. KW wrote the first version of the text. KW, PB and TvZ rewrote the text into its final version.

III. Distributional vowel training is less effective for adults than for infants: a study using the mismatch response.

Karin Wanrooij, Paul Boersma and Titia L. Van Zuijen
PLoS ONE 2014, 9(10), 1-13, doi: 10.1371/journal.pone.0109806.

KW, PB and TvZ designed the method for comparing infants and adults in their capacity for distributional learning. KW and PB made the stimuli. KW applied for approval with the Ethical Committee, recruited the participants, and ran the experiments. KW, PB and TvZ analyzed the data. KW wrote the first version of the text. KW, PB and TvZ rewrote the text into its final version.

Appendix to chapter III. Further exploring the ERP method for adult distributional training.

Karin Wanrooij

IV. Is distributional vowel training effective for Dutch adults? A behavioural control study.

Karin Wanrooij, Johanna de Vos and Paul Boersma
(to be submitted)

KW posed the research question and designed the experiment. KW generated the training stimuli. JdV wrote her Bachelor thesis on the experiments reported below, supervised by KW and PB (De Vos, 2012). For this thesis, JdV also recorded part of the test stimuli, recruited the participants and ran the experiments. KW and JdV analyzed the data. KW wrote the text in this thesis.

V. What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning.

Karin Wanrooij, Paola Escudero, and Maartje E.J. Raijmakers
Journal of Phonetics 2013, 41(5), 307-319, doi: 10.1016/j.wocn.2013.03.005

PE designed and supervised the experiment. KW ran part of the experiments. (Student assistants ran the other part of the experiments). MR performed the latent class regression analysis. KW did the remaining analyses. KW wrote the text, with valuable contributions from PE and MR.

VI. Distibutional training of speech sounds can be done with continuous distributions.

Karin Wanrooij and Paul Boersma
The Journal of the Acoustical Society of America 2013, 133, EL398–EL404, doi: 10.1121/1.4798618

KW posed the research question and designed the experiment. KW did the stimulus generation. Student assistants recruited the participants and ran the experiments, under KW's supervision. KW and PB analyzed the data. KW and PB wrote the text.

VII. Observed effects of "distributional learning" may not relate to the number of peaks. A test of "dispersion" as a confound.

Karin Wanrooij, Paul Boersma, and Titia Benders
(under review)

KW, PB and TB posed the research question, designed the experiment, and defined the training distributions and the alternative hypotheses for calculating the Bayes factors. KW generated the stimuli. Student assistants recruited the participants and ran the experiments, under KW's supervision. KW did the frequentist significance tests. PB calculated the Bayes factors in R. KW wrote the first version of the text. KW, PB and TB rewrote the text into its final version.

VIII. Neural correlates of distributional speech sound learning: a literature review.

Karin Wanrooij
(to be submitted)

IX. General Discussion

Karin Wanrooij

Funding

II. Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study.

Karin Wanrooij, Paul Boersma, and Titia L. Van Zuijen
Frontiers in Psychology - Language Sciences 2014, 5, article 77, 1-12, doi: 10.3389/fpsyg.2014.00077

This research was supported by grant 277.70.008 from the Netherlands Organization for Scientific Research (NWO) awarded to PB.

III. Distributional vowel training is less effective for adults than for infants: a study using the mismatch response.

Karin Wanrooij, Paul Boersma and Titia L. Van Zuijen
PLoS ONE 2014, 9(10), 1-13, doi: 10.1371/journal.pone.0109806.

This research was supported by grant 277.70.008 from the Netherlands Organization for Scientific Research (NWO) awarded to PB.

IV. Is distributional vowel training effective for Dutch adults? A behavioural control study.

Karin Wanrooij, Johanna de Vos and Paul Boersma
(to be submitted)

This research was supported by grant 277.70.008 from the Netherlands Organization for Scientific Research (NWO) awarded to PB.

V. What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning.

Karin Wanrooij, Paola Escudero, and Maartje E.J. Raijmakers
Journal of Phonetics 2013, 41(5), 307-319, doi: 10.1016/j.wocn.2013.03.005

This research was initiated and supported by grant 275.75.005 from the Netherlands Organization for Scientific Research (NWO) awarded to PE. Research assistants for participant recruitment and testing were also supported by NWO grant 016.024.018 awarded to Paul Boersma. KE's work was supported by NWO grant 277-70-008 awarded to Paul Boersma. MR's work was supported by NWO grant 452-06-008. PE's and MR's work was also supported by a grant from the priority program Brain & Cognition of the University of Amsterdam.

VI. Distibutional training of speech sounds can be done with continuous distributions.

Karin Wanrooij and Paul Boersma
The Journal of the Acoustical Society of America 2013, 133, EL398–EL404, doi: 10.1121/1.4798618

This research was supported by Grant No. 277.70.008 from the Netherlands Organization for Scientific Research (NWO) awarded to PB. Participant recruitment and testing for the Music and Discontinuous groups were also supported by NWO Grant No. 275.75.005 awarded to Paola Escudero.

VII. Observed effects of "distributional learning" may not relate to the number of peaks. A test of "dispersion" as a confound.

Karin Wanrooij, Paul Boersma, and Titia Benders
(under review)

This research was supported by grant 277.70.008 from the Netherlands Organization for Scientific Research (NWO) awarded to PB.

VIII. Neural correlates of distributional speech sound learning: a literature review.

Karin Wanrooij
(to be submitted)

This research was supported by grant 277.70.008 from the Netherlands Organization for Scientific Research (NWO) awarded to PB.

Chapter I

General introduction

1. Setting the stage

"Learning" is a fascinating topic of research. It has been studied at several ages, in several fields and from several angles. This thesis deals with learning to perceive the speech sounds of a language, both in infancy, when the speech sounds of the mother tongue must be mastered, and in adulthood, when speech sounds of new languages are learned. In addition, this thesis focuses on a particular learning mechanism that supposedly exists, namely *distributional learning*. Specifically, the topic of this thesis is *distributional learning of vowel categories in infants and adults*. What is meant precisely by "distributional learning" and "vowel categories" is explained in more detail below. Roughly, distributional learning is learning from plain exposure to sounds in the environment, i.e., perceptual learning that does not require pre-existing knowledge, feedback or social interaction. (Note that because the thesis concentrates on *perceptual* learning, it does not address how people learn to *pronounce* speech sounds). Vowel categories are a kind of speech sound categories, which are elements in the speech stream. Examples of vowel categories are the English vowels /ɛ/ (as in words like *pet*) and /æ/ (as in words like *pat*), or the Dutch vowels /aː/ (as in words like /maːn/, "moon") and /ɑ/ (as in words like /mɑn/, "man"). Infants who acquire their first language and older persons who acquire a new language need to learn that certain pronunciations in the speech stream belong to the same speech sound category, and that these pronunciations differ from other pronunciations that represent other speech sound categories.

In this introduction, I first explain the relevance of studying the acquisition of speech sound categories (section 1.1) and, in more detail than above, what is meant by "vowel categories" (section 1.2) and "distributional learning" (section 1.3). The explanation of these two concepts is partly repeated in the chapters of this thesis. Still, because the concepts are central to the thesis, it is important to include an explanation here in the Introduction. Section 1.4 then briefly states the aim of the thesis. Subsequently, section 2 describes what evidence for distributional learning of speech sound categories existed at the start of the project in 2009. Finally, section 3 explains how this previous evidence and linguistic theory inspired the research questions addressed in this thesis.

1.1. The relevance of studying the acquisition of speech sound categories

This thesis examines the perceptual acquisition via distributional learning of a type of speech sound categories (namely vowel categories). Knowledge of the acquisition of speech sound categories, both in infants and adults, is highly relevant. For infants, a proper acquisition of speech sound categories is crucial for infants' language development in general. Even though it is difficult to demonstrate causal relations between early speech perception and later language abilities, longitudinal research shows that the level of infants' speech sound perception in the second half of the first year of life predicts several later language abilities, "including the number of words produced, the degree of sentence complexity and the mean length of utterance" (Kuhl et al., 2008: 989; see also Kuhl et al., 2005), as well as word and phrase understanding (Tsao et al., 2004). Insight into speech sound acquisition can thus contribute to our understanding of first language acquisition in general, and enhance our ability to detect abnormal acquisition in an early phase of development. Studying adults' non-native speech sound acquisition is also relevant. For adults, the acquisition of certain non-native speech sound contrasts is notoriously hard, as evident in difficulties in perception (Polivanov, 1931 [translation 1974]; Flege, 1995) and production (Piske et al., 2001). Insight into the acquisition process and the problems that adults experience when learning these contrasts can help to improve training programs. If, as studies suggest (this is explained in section 2), distributional speech sound training can indeed be effective for adults already after only a few minutes of exposure, then such short-term distributional training seems an attractive alternative for the more common instruction programs for perceptual learning, which usually extend over days or even weeks.

1.2. A definition of "vowel categories"

Learners of a language must learn that certain elements in the speech stream belong to a certain speech sound category (e.g., /ɛ/ as in *pet*), and other elements to

another speech sound category (e.g., /æ/ as in *pat*).[1] This skill is not as simple as it may seem. It may seem that the speech sounds that we are familiar with are always pronounced the same. For instance, if a native speaker of English repeats the word *pet* ten times, English listeners will perceive the same vowel /ɛ/ ten times. In fact, however, each instance of a speech sound category differs from another instance of the same speech sound category in multiple acoustic dimensions. These acoustic differences can be measured in the speech signal. The reason why we do not *perceive* the differences between speech sound tokens of the same category, is that our brain has learned to ignore irrelevant auditory differences and to focus on the relevant auditory differences, i.e., the differences that cause a change in meaning (e.g., from *pet* to *pat*). That this skill is learned, is clear from the fact that the relevant and irrelevant differences between speech sound tokens are not the same across languages. A well-known example of a speech sound contrast that is relevant in one and not in another language, is the English contrast between /ɹ/ as in *rice* and /l/ as in *lice*, which is highly difficult to perceive for Japanese listeners (Goto, 1971; Miyawaki et al., 1975). It is difficult for these listeners because [ɹ]-like sounds and [l]-like sounds do not form separate words in Japanese, so that the difference is better ignored. This is indeed what Japanese listeners have learned to do: they perceive instances of both /ɹ/ and /l/ as the same sound, namely Japanese /ɾ/. A speech sound category thus reflects a group of speech sounds that we have learned to perceive as the same.

Let us now turn to a more technical explanation of a speech sound category. Differences between speech sound categories in a language are reflected in the different distributions of acoustic values of those categories (e.g., Lisker and Abramson, 1964; Newman et al., 2001; Lotto et al., 2004). For instance, if we plot the so-called first formant[2] (henceforth F1) values of numerous pronunciations of

1 In this thesis, the notation of speech sounds is based on the International Phonetic Alphabet (IPA). Also, I follow the practice of using square brackets [] for phonetic notations (reflecting actual pronunciations of speech sounds) and slashes // for phonemic notations (reflecting language-specific abstractions of speech sounds; here phonetic detail is ignored).

2 Formants are measurable frequency values that reflect the resonances of the sound wave in the vocal tract.

the British English vowel categories /æ/ and /ɛ/, we may obtain values similar to those indicated by the vertical lines in the top distribution of Figure I.1 (Hawkins and Midgley, 2005; for details see chapter II). It is evident that the F1 values for each of the two categories cluster around a particular value, the mean F1 value for that category (i.e., 10.44 ERB for /ɛ/ and 12.50 ERB for /æ/). In new measurements of instances of /ɛ/ or /æ/, the probability of finding an F1 value is highest around precisely these mean values. Hence the probability density curves (grey curves in the figure) display peaks here. The number of peaks in the probability density curves is a reflection of the number of categories. Thus, the two peaks hint at the presence of two English vowel categories.

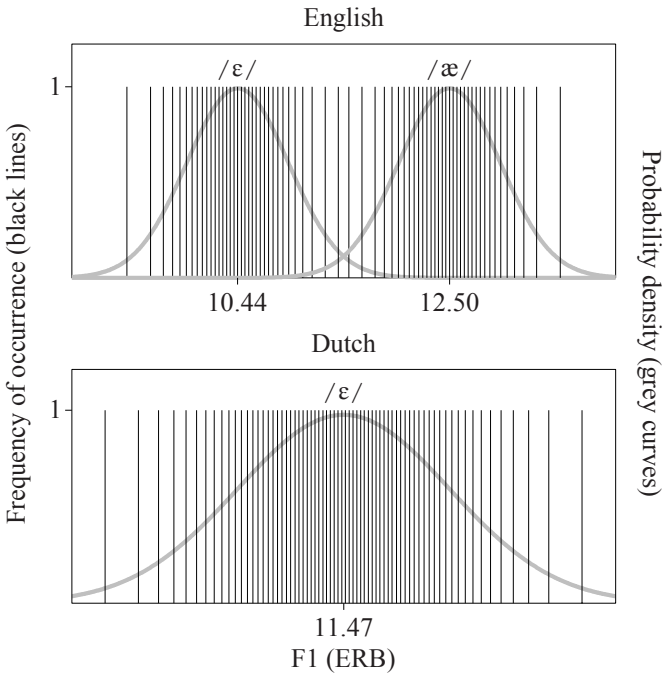


Figure I.1. Distributions of F1 values as hypothetically measured in vowels. Along the same range on the F1 continuum, English front vowels reflect a bimodal distribution (top), whereas Dutch front vowels reflect a unimodal distribution (bottom).

At this point it should be mentioned that for the sake of clarity, Figure I.1 presents a simplified, schematic version of real speech sound distributions. First, the figure shows the vowel distributions along only one acoustic dimension, the F1 value. In reality, speech sounds differ in more than one acoustic property, so that speech sound distributions are multi-dimensional. Apart from the F1 value, an important acoustic component that characterizes vowel categories is the second formant (F2) (Peterson and Barney, 1952). Second, the figure shows only a limited number of hypothetically measured values, which are distributed evenly around the mean. Due to many types of variations (e.g., due to the context of the speech sound token, due to the pitch, or due to the accent of the speaker) real distributions are less perfect.

So far in the technical explanation, I described how speech sound distributions appear in the environment, i.e., how they are shaped by speakers. We will now consider how such distributions are perceived by listeners. Not surprisingly, there appears to be a close relation between the distributions as pronounced by speakers and the speech sound categories as perceived by listeners: listeners tend to perceive speech sound tokens with acoustic values around each peak in the probability density functions as instances of the same speech sound category, and as different from instances around other peaks. This means that, in the example of Figure I.1, English listeners do not only *pronounce* instances of /ɛ/ with F1 values around 10.44 ERB, and instances of /æ/ with F1 values around 12.50 ERB, they also *perceive* such instances as /ɛ/ and /æ/, respectively. A specific "vowel category" can now be defined as a group of speech sound tokens (e.g., several instances of /ɛ/) that are pronounced and perceived as similar to one another in certain aspects (e.g., the F1 value), which differ for other speech sound tokens (e.g., for instances of /æ/).

The fact that some instances of /ɛ/ have F1 values that fall within the bounds of the category /æ/ (as visible in the overlap between the grey curves in Figure I.1, top), shows that it is difficult to define a vowel category sharply. This fuzziness of the boundary between categories is exacerbated by the focus on one acoustic characteristic only (in this case the F1 value). In fact, it is impossible to

define a vowel category on the basis of a single characteristic. This impossibility is illustrated by the existence of other vowels in the English vowel inventory than /ɛ/ and /æ/ along the same F1 continuum shown in Figure I.1, namely vowels such as /ʌ/ (as in *but*) and /ɒ/ (as the first vowel in *bottom*). These vowels differ from /ɛ/ and /æ/ mainly in having lower F2 values. Apart from F1 and F2, other acoustic characteristics, such as higher formants and duration, may also contribute to defining a vowel category. Properties that contribute less or not at all to defining a category, such as the fundamental frequency (F0)[3] of a vowel token in English, can vary more randomly between instances of the category than properties that contribute more to this definition (such as the F1 and F2 value). In sum, a category reflects a composite of properties, each of which can vary along a continuum, and this variation causes the boundaries between categories to be fuzzy.

The just-given definition of a "vowel category" is analogous to possible definitions of other (linguistic) categories, and was inspired by definitions of categories at a conceptual level as described by Rosch and colleagues (Rosch, 1973; Rosch and Mervis, 1975; Rosch et al., 1976). At a semantic word level, for instance, it is possible to define the category of "birds" as a *group* of animals that are *similar* to one another in certain aspects (e.g., in having feathers and being able to fly), which *differ* for other animals. Even though for semantic categories it may be more difficult to view the separate characteristics as a continuum of possible values, this is often possible. For instance, some birds are better equipped to fly (e.g., the arctic tern, which flies from pole to pole) than other ones (chickens only flutter for short distances), which are in turn better equipped to fly than even other species of birds (penguins and ostriches do not fly at all). That not all birds can fly shows that it is difficult to define a category on the basis of a single characteristic. Like vowel categories, semantic categories thus consist of *composites of properties* (other example characteristics of birds are "having a beak", and "laying eggs"), each of which can vary. The variation in each characteristic (running from being

3 The fundamental frequency is the rate of vocal fold vibration, and causes speech to be perceived at a certain pitch.

present to not being present at all) causes the boundaries between categories to be *fuzzy*.

Crucially, just as the division into categories at other levels of perception, the division into vowel categories has *functional relevance*: the categories contribute to the conveyance of different meanings. For instance, when appearing between /p/ and /t/ in English, /ɛ/ contributes to conveying the meaning of the word *pet*, while /æ/ contributes to conveying the meaning of the word *pat*. The difference between tokens of /ɛ/ and tokens of /æ/ is not functionally relevant across languages. For native speakers of Dutch, for example, the same instances of the English vowels /ɛ/ and /æ/ belong to a single vowel category, namely the Dutch /ɛ/ as in the Dutch word /pɛt/, meaning "cap". The Dutch distribution, with a single peak along the given F1 range, is illustrated at the bottom of Figure I.1. Such single-peaked distributions are called "unimodal". Two-peaked distributions, such as the one illustrated for English, are called "bimodal".

In sum, just as other types of categories, a vowel category can be viewed as a group of tokens that are similar to one another in certain aspects, and which differ in these aspects from tokens of other categories, in a functionally relevant way.

1.3. A definition of "distributional learning"

In this thesis, I presuppose that representations of vowel categories in the brain are largely acquired[4] on the basis of stimuli experienced in the environment, i.e., they are not hard-wired innately in the infant brain as properties that can be maintained or lost. For speech sound categories, this view diverges from that embraced in the seventies of the past century (which stressed the role of innate factors in establishing categorical speech sound production and perception; e.g., Chomsky

4 In this thesis, I ignore the distinction that is sometimes made between the terms "acquisition" and "learning". This distinction was introduced by Krashen (1981) to separate unconscious or subconscious "implicit learning" (termed "acquisition") from conscious "explicit learning" (termed "learning"). Distributional learning is a form of implicit learning (hence of "acquisition" rather than "learning" in Krashen's terms), since it is thought to ensue from mere exposure, without any explicit instruction or feedback (see section 1.3).

and Halle, 1968; Eimas et al., 1971), but is in line with the present dominant opinion, which has arisen on the basis of several computational, psychological and neurobiological findings since that time (see e.g., Guenther and Gjaja, 1996; Boersma, 1998; Karmiloff-Smith, 2006; Kuhl et al., 2008). The view that categories must be learned does not imply a denial of innately determined factors, which can also influence the acquisition of categories (see also chapter VIII).

If speech sound categories are learned, the question arises how. Distributional learning is possibly one of the learning mechanisms that contribute to this acquisition. It is thought to ensue from *simple exposure* to distributional patterns in the environment. Thus, distributional learning does not involve other types of learning that probably also play a role in speech sound acquisition, and which are based on pre-existing knowledge (Maye et al., 2002), feedback or social interaction (Kuhl et al., 2003). For instance, if an infant learned the English vowel categories /ɛ/ and /æ/ exclusively via distributional learning, then the infant would not need to have knowledge of words containing these vowels yet. With such knowledge, the infant could infer that tokens of /ɛ/ must probably represent a different vowel category than tokens of /æ/, because sounds like [pɛt] (*pet*) and [pæt] (*pat*) convey different meanings. Also, it would not be necessary for the infant to know how to pronounce the vowels, and the infant would not need feedback from or interaction with a caregiver explicitly teaching him or her the difference between the vowel categories. The infant would simply have to be exposed to the English language containing words with instantiations of /ɛ/ and /æ/. Thus, the idea of distributional learning is that, in the schematic example of Figure I.1, infants raised in English homes start creating two vowel categories, because they experience two groups of acoustic values (i.e., based on the bimodal distribution in Figure I.1) in the speech stream, while infants raised in Dutch homes create a single vowel category, because they experience one group of acoustic values (i.e., based on the unimodal distribution).

Distributional learning, which is also named “statistical learning” (e.g., Maye et al., 2008), has indeed been reported as a learning mechanism for the acquisition of speech sound categories (see section 2 in the Introduction). In

addition, distributional properties in the input have been shown to help infants learn several other aspects of language, including phonotactic patterns (e.g., Jusczyk et al., 1994), words (e.g., Saffran et al., 1996) and syntactic rules (e.g., Marcus et al., 1999). Moreover, statistical learning is not confined to language or to the auditory domain: it has been observed for non-linguistic auditory patterns and for visual patterns (reviews in Krogh et al., 2013; Lany and Saffran, 2013). Although it is not clear whether the neurobiological mechanisms behind these different manifestations of statistical learning are the same, simple exposure thus seems to affect not only speech sound perception, but perception in general. In the remainder of this thesis, the term "distributional learning" is used exclusively for distributional speech sound learning.

1.4. The aim

Bearing in mind the definitions of "a vowel category" (section 1.2) and "distributional learning" (section 1.3), the topic of this thesis can now be formulated as: learning to group vowel instances encountered in the environment into functional clusters ("vowel categories"), through plain exposure to their distributions ("distributional learning"). The main aim is to assess the role of such distributional learning in the acquisition of native (for infants) and non-native (for adults) vowel categories. The approach chosen to reach this aim is explained in section 3. Before turning to this section, let us first look at evidence for distributional speech sound learning in section 2.

2. Evidence for distributional learning of speech sound categories

There is evidence that distributional learning can indeed be a mechanism that contributes to speech sound learning, both for infants learning their first language, and for adults learning a new language. The evidence comes from observations during infants' natural language acquisition (section 2.1) and from psycholinguistic experiments with infant and adult participants (section 2.2).

2.1. Evidence from observations during natural language acquisition

A large body of research shows that infants' speech sound perception changes from universal (i.e., discrimination performance is the same across infants, irrespective of the native language that they experience) to language-specific (i.e., discrimination performance reflects the speech sound distributions of the native language) between 6 and 12 months of life (e.g., Werker and Tees, 1984; Kuhl et al., 1992; Cheour et al., 1998; for details, see chapter II). This developmental change, which is also called "perceptual reorganization" (Werker and Tees, 1984), is assumed to result mainly from exposure to native speech sound distributions, and thus from distributional learning, because it emerges before other ways of learning speech sound categories, such as noticing differences in word meaning and producing the differences in the categories, are fully effective (Stager and Werker, 1997; Maye et al., 2002; Bergelson and Swingley, 2012).

For adults who try to learn a non-native language, a similar developmental pattern (i.e., perceptual tuning that can be related to the length of exposure to the non-native language) cannot be observed readily (Escudero and Wanrooij, 2010; see also chapter III). This does not straightforwardly mean that distributional learning is not a mechanism in adults. The difficulty of clearly observing distributional learning during natural non-native language acquisition, may be due to the presence of many interfering factors that have been reported to play a role in non-native speech sound acquisition, such as the age of acquisition (Flege and MacKay, 2004), the nature of the native speech sound inventory (Polivanov, 1931 [translation 1974]), and the quality of the non-native language input (Moyer, 2009; see also the Discussion in chapter V). Thus, where for infants a developmental pattern in first-language speech sound acquisition can be coupled with distributional learning based on longitudinal observations during natural language acquisition, such a pattern is unclear for adults learning a new language. Experiments on distributional learning *in the lab*, however, have demonstrated effects of adult distributional learning more clearly. These experiments are discussed in the next section.

2.2. Evidence from psycholinguistic experiments

In laboratory settings, distributional learning has been observed not only in infants, but also in adults. In these experiments, exposure never lasts longer than a few minutes. Table I.1 lists the distributional training experiments known at the beginning of the current project in 2009. (It can be compared to Tables IX.2 and IX.3 in the Discussion, which give an overview of all experiments at the end of the project in 2014). As is visible in Table I.1, participants are usually exposed to either a bimodal or a unimodal distribution. The bimodal distribution reflects the speech sound contrast to be acquired; the unimodal distribution is representative of a single existing native speech sound category. A different control group than a unimodal control group was sometimes included. In these conditions, participants were exposed to "non-speech" or they did not receive any training at all. The latter condition is labelled "no training" in Table I.1.

After exposure, participants are always tested on how well they perceive a difference between the two speech sound categories in the contrast inherent in the bimodal distribution. In studies reporting a significant effect of distributional training, the bimodally trained participants are better at perceiving the difference between the two speech sound categories inherent in the bimodal distribution, than unimodally trained participants. Because participants do not receive any feedback, these effects can then be attributed to distributional learning.

Interestingly, Pons (2006) tested whether an effect of distributional training could also be elicited in adult rats, in a behavioural experiment based on Maye et al. (2002). The stimuli were the same as those in Maye et al. (2002; see Table I.1) and the procedure was as similar as possible, except for some obviously necessary adaptations to testing rats instead of humans. Also, exposure times were chosen to be substantially longer, i.e., eight sessions (with one session per day) of eight minutes each. Bimodally trained rats discriminated the tested contrast better than unimodally trained rats (with 31 rats included in the analysis, one excluded, and $p < 0.01$). In sum, behavioural experiments in the lab demonstrate that exposure to speech sound distributions can affect perception in human infants and adults and even in rats.

Table I.1. Studies on infant (top) and adult (bottom) distributional learning known in 2009. With participants' age and native language (L1), the non-native speech sound contrast in the bimodal training distributions (contrast), the duration of the training (Time, in minutes), the groups that were compared (bi = bimodal, uni = unimodal), the total number of participants in the combined groups mentioned in the groups column included in the analysis (N included), and additional participants tested (N excluded), and the *p*-value of the comparisons. Cf. Tables IX.2 and IX.3 in chapter IX.

Study	Age	L1	Contrast	Time (min.)	Groups	N incl. (N excl.)	*p* value
Infants							
Maye et al., 2002	6 – 9 mths	English	/d/~/t/ [b]	2.3 [f]	- bi vs. uni	48 (12)	0.063
Maye et al., 2008	7 – 9 mths	English	/d/~/t/ [b] or	2.8	- bi vs. uni	97 (56)	0.001
			/g/~/k/ [b]		- bi vs. non-speech		0.001
Pons et al., 2006a [a]	6 mths	English	/ɛ/~/ɛː/	? [g]	- bi vs. uni	? [g]	"ns"
Pons et al., 2006b [a]	8 mths	English	/e/~/ɪ/	? [g]	- bi vs. uni	32 (? [g])	"ns"
Adults							
Maye & Gerken, 2000	18 – 41 yrs	English	/d/~/t/ [b]	9 [h]	- bi vs. uni	32	<0.05
Maye & Gerken, 2001	(Students)	English	/d/~/t/ [b]	9 [h]	- bi vs. uni	32	<0.01
	(Students)	English	/g/~/k/ [b]	9 [h]	- bi vs. uni	32	<0.05
Peperkamp et al., 2003	(Adults)	French	/ʁ/~/χ/ [c]	9	- bi1 vs. bi2 vs. uni [j]	60	>0.1 [k]
Shea et al., 2006	(Students)	English	/dæ/~/dˤɑ/ [d]	12 [i]	- bi vs. uni	32	<0.01 [l]
	(Adults)	Spanish					
Hayes-Harb, 2007	(Students)	English	/g/~/k/ [b]	9 [h]	- bi vs. uni	66	0.04
					- bi vs. no training		0.24
					- uni vs. no training		0.007
Gulian et al., 2007	16 – 60 yrs	Bulgarian	/ɑ/-/a/, /ɪ/-/i/ [e]	5	- bi vs. uni	40	0.029

a) Unpublished results presented in posters at conferences.
b) The contrast between voiced and voiceless unaspirated plosives (such as /d/ versus /t/ and /g/ versus /k/) is not phonemic in English, even though the orthography suggests that it is. The distinction only appears in allophonic contexts. English has a voicing contrast between "voiceless" unaspirated plosives (such as /d̥/ at the onset of the word "do" and /g̊/ at the onset of "game") and voiceless aspirated plosives (as /tʰ/ and /kʰ/ at the onset of "two" and "came" respectively).
c) The distinction between voiced and voiceless uvular fricatives is allophonic in French, not phonemic.
d) Participants were exposed to either a unimodal or a bimodal distribution based on either the consonant continuum /dV/~/dˤV/ (vowel kept constant) or the vowel continuum /Cæ/~/Cɑ/ (consonant kept constant). The consonant contrast represents the Arabic contrast between non-emphatic and emphatic (pharyngealized) alveolar plosives, which is accompanied by allophonic variation in the vowel /æ/. After the emphatic plosive, the second vowel formant is lowered, yielding /ɑ/. The vowel contrast is phonemic for English listeners, not for Spanish listeners.
e) Dutch vowel contrasts that are not phonemic in Bulgarian.
f) Training duration without fillers was around 1.5 minutes (deduced from the text as follows: 96 training stimuli * (465 ms + 500 ms inter-stimulus interval)).
g) ? = not reported
h) Half of the training stimuli consisted of fillers. The precise duration of exposure to the training stimuli (i.e., without the fillers) cannot be calculated from the article.
i) Training duration without fillers was 6 minutes.
j) There were two bimodal groups. In one group each VC-sequence in the training was coupled with a CV-syllable where the C agreed in voicing with the preceding C. In the other group the Cs did not agree in voicing.
k) The *p*-value represents the interaction between the test (post- vs. pre-test) and distribution (uni- vs. bimodal1 vs. bimodal2).
l) The *p*-value represents the interaction between the test (post- vs. pre-test) and distribution (uni- vs. bimodal) across language groups.

3. Research questions inspired by previous evidence and linguistic theory

The previous research on distributional learning (section 2) demonstrates that distributional learning of speech sound categories probably exists as a learning mechanism, and that this mechanism can be tapped in the lab already after a brief exposure duration. At the same time, this previous research as well as linguistic theories about distributional learning (which are touched upon in the sections below and which are explained in detail in chapter VIII) evoke many questions, among which the research questions addressed in this thesis and introduced in this section. These questions concern the replicability of distributional training experiments (section 3.1), the possibly changing role of distributional learning with age (section 3.2), potential differences in the effectiveness of distributional training between listener types within conditions (section 3.3), possible effects of manipulations of the training distributions (section 3.4), and neurobiological mechanisms of distributional learning (section 3.5).

3.1. Replicability of distributional training experiments

At first sight, Table I.1 presents a sound list of studies available in 2009, demonstrating that distributional learning is a mechanism that can be tapped after short exposure in the lab successfully, in both infants and adults. At the same time, a closer look at the table may temper such confidence, in particular for infants.

Specifically, the table shows that at the start of the project in 2009 there were only two published studies reporting infant distributional learning (Maye et al., 2002; 2008). These studies were from the same lab, used the same contrast (a voicing contrast), and tested infants from the same native language group (English) at approximately the same age (between 6 and 9 months). These similarities were intentional (i.e., the second study was designed to complement the first study), but spark curiosity as to whether distributional learning can be replicated with other contrasts and native-language groups, and with other age groups. Note that the other two infant studies in Table I.1 report unpublished null results, presented in posters at conferences (Pons et al., 2006a, who tested distributional learning of

vowel length distinctions in 6-month olds; Pons et al., 2006b, who tested distributional learning of vowel quality distinctions in 8-month olds). Even if null results cannot be interpreted as evidence against the occurrence of distributional learning, they do not provide clear evidence for it either. Unfortunately, null results tend to remain unpublished far more often than significant results, so that it was conceivable that more null results existed, at the start of the project in 2009. In view of the above, it was important to test *whether distributional learning can indeed be demonstrated as a mechanism in infants in a distributional training paradigm*. This thesis therefore includes a distributional training experiment with infants (chapter II).

For adults, previous research at the start of the project (see Table I.1 again) represented more diversity in the choice of the contrasts and the appropriate participant groups. However, the earlier adult studies showed a bias towards consonant contrasts (versus vowel contrasts), and towards contrasts containing speech sounds that occur in allophonic contexts in the native languages of the participants (versus contrasts that are neither phonemic nor allophonic). Another bias in the previous research, both in that with infants and in that with adults, is the exclusive use of behavioural paradigms (versus neurophysiological measurements). In view of these biases, it seemed important to examine *whether an effect of distributional training can be replicated with new speech sound contrasts for new participant groups (i.e., with other native languages), and with new research methods*. This thesis therefore presents distributional training experiments in which new contrasts are used with new participant groups, namely English vowels presented to listeners raised in Dutch homes (chapters II through IV), and Dutch vowels presented to listeners raised in Spanish homes (chapters V through VII). In addition, the thesis includes both behavioural (chapters IV through VII) and neurophysiological methods (chapters II and III).

3.2. The role of distributional learning with age

At the start of the project in 2009, there had not been any concrete investigation into the precise role of distributional learning in speech sound acquisition at different ages, a role that is possibly changing over time. The four studies on infant distributional learning known at the start of the project in 2009 (Table I.1) tested infants in the second half of the first year, i.e., at an age where infants are already beginning to show language-specific speech sound perception (section 2.1). The studies do not clarify a possible role of distributional learning in *achieving* such language-specific perception. Accordingly, it seemed relevant to ask *whether distributional learning can actually contribute to the development from universal to language-specific speech sound perception in the first year of life*, and thus to the acquisition of native-language speech sound categories. To this end, distributional learning had to be demonstrated at an age *before* the appearance of language-specific perception. Therefore, this thesis presents a distributional training experiment with 2-to-3-month olds (chapter II).

Further, at the start of the project in 2009, there was more evidence of distributional learning in the lab in adults than in infants (Table I.1). However, it was impossible to conclude on the basis of the evidence that the capacity for distributional learning was higher in adults than in infants: direct comparisons between the effect of distributional training in infants and that in adults had not been made, and experimental designs for infants and adults had been different (including longer training times for adults than for infants, as visible in Table I.1). Furthermore, in linguistic theories distributional learning tended to be viewed as a *more restricted* mechanism in adults than in infants (see chapter VIII). Therefore, if distributional learning was indeed a mechanism for learning speech sound categories, a relevant question was *whether the capacity for distributional learning is different in adulthood than in infancy, and consequently, whether the importance of distributional learning for the acquisition of native speech sound categories differs from that for the acquisition of non-native speech sound categories later in life*. To shed light on the issue, this thesis presents a first attempt to directly compare the effect of distributional training in infants to that in adults. The attempt

is based on the measurement of event-related potentials (ERPs) and the calculation of the "mismatch response" (MMR), in order to circumvent differences between the age groups in behavioural abilities (chapter III). In addition, a possible difference in the capacity for distributional learning between the age groups (infants versus adults) was probed by considering results from several subfields of neuroscience (chapter VIII).

3.3. Possible differences between listener types within conditions

The research questions presented above address the effect of distributional training on speech sound perception, for different types of participants *between* conditions (i.e., between bimodal and control conditions and between age groups). They do not address possibly different types of participants *within* conditions. Similarly, all previous distributional learning studies available at the start of the project (Table I.1) compared a group of bimodally trained participants to one or more control groups, irrespective of possible differences between participant types within each group. This is a valid approach in traditional experimental research. Predictions derived from linguistic theory (chapter VIII) also tend to apply to groups rather than to subgroups or to individuals, and thus tend to ignore potential differences among participants (e.g., Best, 1994; Flege, 1995)[5]. Recently, however, the interest in differences between participants within a condition is rising. A question in accordance with this trend is *whether exposure to speech sound distributions can affect types of listeners within conditions differently*. This thesis therefore includes a study examining this issue in adult native speakers of Spanish, who are trained on a Dutch vowel contrast that is difficult to perceive for these listeners (chapter V). Specifically, it was investigated whether it is possible to identify types of Spanish listeners, that each use different acoustic cues when perceiving Dutch vowels, and if so, whether such differential cue weightings influence what the listeners learn precisely during a subsequent distributional training.

5 Authors sometimes mention that individual differences play a role (e.g., Best, 1994), but these differences are seldom accounted for in the theory (an exception is Escudero, 2005).

3.4. Possible effects of manipulations of the distributions

Previous research on distributional learning available at the start of the project in 2009 focused on determining whether distributional learning is a mechanism at all, and not on how the distributional learning mechanism (if it exists) could be influenced by manipulations of the training distributions. Typically, the differential number of peaks in the distributions (namely two peaks in the bimodal distribution versus either one peak in the unimodal distribution or an undefined number of peaks in non-distributional training) was viewed as the main determinant of the observed distributional training effects. In other words, attention had focused on the *means* (i.e., the peaks) of speech sound distributions, and no attention had been paid to a possible influence of *measures of dispersion*, and to a possible influence of *variability* in the presented speech sound tokens. These issues are addressed in this thesis, as explained below.

Natural speech sound distributions vary in measures of dispersion. For instance, distributions in infant-directed speech (IDS) appear to be "enhanced" as compared to distributions in adult-directed speech (ADS): the means of each speech sound category are spaced at a larger acoustic distance from one another, thereby also stretching the range of probable acoustic values (Kuhl et al., 1997). Such enhancement can also be observed in foreigner-directed speech (Uther et al., 2007) and in "clear speech", a speech style that is used in, for example, noisy environments (Smiljanić and Bradlow, 2009). Enhancement can reduce the overlap between speech sound distributions, and can thus improve the discriminability of the speech sounds involved. Indeed, there are several indications that enhancement is related to better speech sound discriminability (for infants: Liu et al., 2003; in clear speech: Smiljanić and Bradlow, 2009; in computer models: De Boer and Kuhl, 2003). Also, Kuhl and colleagues posit that enhancement in IDS is an important driving force enabling infants to create language-specific speech sound categories (Kuhl et al., 2008; chapter VIII). A relevant question that logically follows from the just-given observations is *whether enhancement of bimodal distributions in a distributional training experiment can benefit participants' ability to learn speech sound categories*. Therefore, following Escudero et al.

(2011), this thesis compares the effects of exposure to enhanced versus non-enhanced bimodal training distributions on adult learners' categorization of tokens representative of the two speech sound categories in the bimodal distribution (chapter V).

Distributions in IDS are not only enhanced as compared to those in ADS, they also contain a larger "variety of instances" (Kuhl et al., 1997: 685; Kuhl, 2000). The presence of various different instances of speech sound categories supposedly helps infants to create the categories, because it allows them to detect relevant similarities and differences between the instances (Kuhl et al., 1997). Presenting a large variety of instances has also been hypothesized to benefit speech sound learning in adults (Jamieson and Morosan, 1986). Accordingly, such "high variability" has been implemented in many experiments in which adults received speech sound training, for instance by including multiple tokens pronounced by multiple speakers (Logan et al., 1991; Lively et al., 1993; Bradlow et al., 1997) or by creating a large number of acoustically different synthetic stimuli (Jamieson and Morosan, 1986). Although high- and low-variability training were usually not compared in a direct statistical comparison, and although the difference between the two was not straightforwardly significant in the few cases when this was done (McCandliss et al., 2002; Jamieson and Morosan, 1989), the studies using high variability in their training stimuli generally report improvement in adults' classification or discrimination of speech sounds representative of the trained contrast (Logan et al., 1991; Lively et al., 1993; Bradlow et al., 1997).

Notably, all previous research on distributional learning available at the start of the project in 2009, used training distributions with relatively low variability, namely 8-step "discontinuous" distributions. Such distributions are created by dividing the acoustic continuum in only eight steps and by repeating the stimuli at each step in certain proportions (for a more detailed explanation see chapter VI). Although usually for each step more than one speech sound token was created (for example on the basis of different pronunciations), variability was highly reduced by the discontinuity and the repetition of tokens. A relevant question that logically follows from the above is *whether adding variability to the*

training stimuli can benefit distributional speech sound learning. Therefore, this thesis presents an experiment in which the effect on speech sound perception of discontinuous distributions is compared to that of "continuous" distributions (chapter VI). These distributions contain a large number of acoustically different tokens (e.g., 900 in chapters II and III), each of which is presented only once (i.e., there is no token repetition). Continuous distributions, which are closer to natural distributions, and which are thus more ecologically valid, were also used in other experiments (chapters II, III, and VII).

3.5. Neurobiological mechanisms of distributional learning

In linguistic theory, distributional learning is viewed as a low-level, bottom-up mechanism, i.e., a mechanism that only involves the lower levels of representation in the brain (low-level), and which is entirely driven by the external stimulus (bottom-up) and thus not by internal knowledge (see Chapter VIII for a detailed explanation). However, linguistic theory contains very few references to concrete neuroscientific evidence for such a bottom-up, low-level mechanism. Accordingly, a relevant question is *whether it is possible to pinpoint concrete neurobiological processes in the brain that could represent or affect distributional learning*. This thesis gives a literature review of possible neural correlates of distributional learning, as found in diverse subfields of neuroscience (chapter VIII).

3.6. Overview

In sum, this thesis examines the role of distributional learning in the acquisition of vowel categories in infants and in adults. This is done on the basis of neurophysiological (chapters II and III) and behavioural experiments (chapters IV through VII) and on the basis of a literature review of possible neurobiological underpinnings of the mechanism (chapter VIII). Table I.2 presents an overview of the five research topics and the related questions, as inspired by previous

experimental and theoretical research (as explained in sections 3.1 through 3.5). The table also mentions the chapters in which the questions are addressed.

Table I.2. Five research topics, with the related questions and corresponding chapters.

Topic	Research question	Chapter
1. Replicability of distributional training experiments	Is a distributional training effect replicable in infants?	II
	Can an effect of distributional training be replicated with other speech sound contrasts appropriate for other adult populations (as defined by the native language) than previously tested?	English vowels for Dutch learners: II, III, IV[a] Dutch vowels for Spanish learners: V, VI, VII[a]
	Can an effect of distributional training be replicated with new research methods?	Neurophysiology: II, III Behaviour: IV, V, VI, VII
2. The role of distributional learning with age (infants vs. adults)	Can distributional learning contribute to the emergence of language-specific speech sound perception in infancy?	II
	Is the capacity for distributional learning different in adulthood than in infancy?	III
3. Possible differences between listener types within conditions	Can exposure to speech sound distributions affect types of listeners within conditions differently?	V
4. Possible effects of manipulations of the distributions	Can enhancement of bimodal training distributions benefit speech sound learning?	V
	Can adding variability to distributional training stimuli benefit speech sound learning?	VI
5. Neurobiological mechanisms of distributional learning	Is it possible to pinpoint concrete neurobiological processes in the brain that could represent or affect distributional learning?	VIII

a) The table does not show the research questions for chapters IV and VII: these chapters describe control experiments that were added in the course of the project (see the respective chapters and the Discussion in chapter IX).

Chapter II

Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study

Karin Wanrooij, Paul Boersma, and Titia L. Van Zuijen
Frontiers in Psychology - Language Sciences 2014, 5, article 77, 1-12
doi: 10.3389/fpsyg.2014.00077

Data available from the public figshare database:
http://dx.doi.org/10.6084/m9.figshare.1157812

Abstract

An important mechanism for learning speech sounds in the first year of life is "distributional learning," i.e., learning by simply listening to the frequency distributions of the speech sounds in the environment. In the lab, fast distributional learning has been reported for infants in the second half of the first year; the present study examined whether it can also be demonstrated at a much younger age, long before the onset of language-specific speech perception (which roughly emerges between 6 and 12 months). To investigate this, Dutch infants aged 2 to 3 months were presented with either a unimodal or a bimodal vowel distribution based on the English /æ/~/ɛ/ contrast, for only 12 minutes. Subsequently, mismatch responses (MMRs) were measured in an oddball paradigm, where one half of the infants in each group heard a representative [æ] as the standard and a representative [ɛ] as the deviant, and the other half heard the same reversed. The results (from the combined MMRs during wakefulness and active sleep) disclosed a larger MMR, implying better discrimination of [æ] and [ɛ], for bimodally than unimodally trained infants, thus extending an effect of distributional training found in previous behavioural research to a much younger age when speech perception is still universal rather than language-specific, and to a new method (using event-related potentials). Moreover, the analysis revealed a robust interaction between the distribution (unimodal vs. bimodal) and the identity of the standard stimulus ([æ] vs. [ɛ]), which provides evidence for an interplay between a perceptual asymmetry and distributional learning. The outcomes show that distributional learning can affect vowel perception already in the first months of life.

The pictures in Figures II.2 and II.3 were not part of the publication in *Frontiers in Psychology*.

1. Introduction

Distributional learning, i.e., learning by simply being exposed to the frequency distributions of stimuli in the environment, may be one of the mechanisms by which infants start to acquire the phonemes of their language (Lacerda, 1995; Guenther and Gjaja, 1996). *Fast* distributional learning of speech sounds after just a few minutes of exposure in the lab has been observed in infants in the second half of the first year (e.g., Maye et al., 2008). This study investigates whether such fast distributional learning can also take place in very young infants, i.e., 2-to-3-month olds. This is relevant if we want to establish that the distributional learning mechanism is in place early enough to be able to contribute to the transition from universal to language-specific speech perception, which becomes apparent in infants' speech sound discrimination from around 6 months of age (e.g., Werker and Tees, 1984/2002; Polka and Werker, 1994), or perhaps even from 4 months (Yeung et al., 2013).

In the first year of life, infants' speech sound perception has been observed to change from universal to language-specific. Specifically, in the course of this transition discrimination performance is enhanced for native speech sound contrasts (Cheour et al., 1998b; Kuhl et al., 2006; Tsao et al., 2006), and reduced for non-native contrasts that are irrelevant in the native language (Werker and Tees, 1984/2002; Kuhl et al., 1992; Tsushima et al., 1994; Polka and Werker, 1994; Best et al., 1995; Bosch and Sebastián-Gallés, 2003; Kuhl et al., 2006; Tsao et al., 2006). In general, language-specific speech sound discrimination emerges between 4 and 6 months for tones (i.e., in tonal languages; Cheng et al., 2013; Yeung et al., 2013), around 6 months for vowels (Kuhl et al., 1992; Polka and Werker, 1994; Bosch and Sebastián-Gallés, 2003), and between 8 and 12 months for consonants (Werker and Tees, 1984/2002; Tsushima et al., 1994; Best et al., 1995; Kuhl et al., 2006; Tsao et al., 2006), although language-specific discrimination of difficult contrasts may develop later (e.g., Cheour et al., 1998b; Polka et al., 2001; Sundara et al., 2006).

One of the mechanisms that has been hypothesized to contribute to the emergence of language-specific speech perception is distributional learning

(Lacerda, 1995; Guenther and Gjaja, 1996). The existence of this mechanism has indeed been supported by observations in the lab. In particular, fast distributional learning has been demonstrated most reliably in 8-month olds by Maye et al. (2008; $p < 0.001$), and (nearly) significantly in 6-to-8-month olds by Maye et al. (2002; $p = 0.063$), in 10-to-11-month olds by Yoshida et al. (2010; $p = 0.036$ for one of the experiments), and in 11-month olds by Capel et al. (2011; $p = 0.053$), although null results were found in 10-to-11-month olds by Yoshida et al. (2010; for two experiments) and ambiguous results were found in 5-month olds by Cristià et al. (2011; $p > 0.16$ for the main effect, but $p = 0.007$ for an interaction effect).

If distributional learning indeed contributes to the acquisition of language-specific perception, and discriminational evidence for the latter starts being observed from 4 or 6 months on, fast distributional learning can be expected to be detectable in even younger infants. This expectation is supported by neuroscientific research. Cortical layers involved in top-down processing (e.g., Kral and Eggermont, 2007) become anatomically available in humans from around 4 to 5 months of age (Moore and Guan, 2001; Moore, 2002; Moore and Linthicum, 2007), which suggests that speech perception before 4 months relies mainly on bottom-up processing. The distributional learning mechanism, which supposedly does not require top-down processing (Guenther and Gjaja, 1996), should therefore at this early age be relatively unimpeded by learning mechanisms that require top-down influence from higher-level (e.g., lexical) representations.

We therefore performed a fast distributional learning experiment with infants aged 2 to 3 months. Specifically, we presented Dutch infants of this age with speech sounds from an acoustic continuum encompassing the British-English vowel contrast /æ/~/ɛ/; this is a contrast that does not exist in Dutch, and which Dutch adults find difficult to master (e.g., Schouten, 1975; Weber and Cutler, 2004; Broersma, 2005; Escudero et al., 2008). These vowels differ in their first formant (F1), as illustrated in Figure II.1, where the F1 values are given in ERB (Equivalent Rectangular Bandwidth; see section 2.3 for details). In our experiment, one half of the infants were exposed to a unimodal distribution (Figure II.1, grey), i.e., to a large number of different vowel tokens whose F1 values center around

11.47 ERB, which is phonetically halfway between English [ɛ] and [æ], and the other half of the infants were exposed to a bimodal distribution (Figure II.1, black), i.e., to a large number of vowel tokens whose F1 values center around 10.44 and 12.50 ERB, which are F1 values typical of English [ɛ] and [æ], respectively. The bimodal distribution thus suggests the existence of a contrast between /æ/ and /ɛ/ (as would be appropriate for learners of English), while the unimodal distribution does not suggest a contrast between the two vowels (as would be appropriate for learners of Dutch). Immediately after the training we tested how well the infants discriminated an open variant of English [ɛ], i.e., a vowel with an F1 of 10.78 ERB, and a closed variant of English [æ], i.e., a vowel with an F1 of 12.16 ERB, both visible in Figure II.1. If distributional learning occurred, bimodally trained infants should discriminate them better than unimodally trained infants.

Discrimination ability after training had to be measured with a method appropriate for young infants. All previous research on infant or adult distributional learning employed behavioural measures, which for infants always meant looking time. Since suitable behavioural responses are difficult to obtain from 2-to-3-month olds, we instead measured an automatic brain response, namely

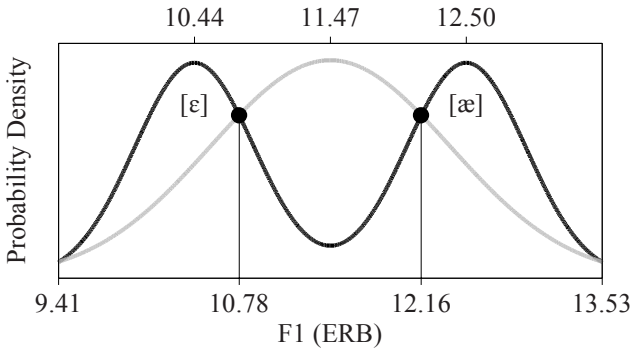


Figure II.1. Unimodal (grey curve) and bimodal (black curve) training distributions of the first vowel formant (F1). The values of the test stimuli lie at the intersections of the two distributions.

the mismatch response (MMR; e.g., Näätänen et al., 1978). In contrast to behavioural measurements, which require the infant's cooperation and attention (Cheour et al., 2000, p. 6), the MMR is elicited even in the absence of voluntary attention to the stimuli (e.g., Schröger, 1997; Näätänen and Winkler, 1999), and can be measured even when the infant is asleep (Friederici et al., 2002; Martynova et al., 2003). The MMR has been shown to reflect behavioural discrimination in adults (for a review, see Näätänen et al., 2007), and has been used successfully before to demonstrate vowel discrimination in infants of 3 months and younger (e.g., Cheour-Luhtanen et al., 1995; Cheour et al., 1998a; Martynova et al., 2003; Kujala et al., 2004; Shafer et al., 2011; Partanen et al., 2013). The MMR can be elicited in an oddball paradigm (e.g., Näätänen, 1992), where a series of "standard" stimuli (e.g., [ɛ] tokens) is interspersed infrequently with "deviant" stimuli (e.g., [æ] tokens). If the auditory perception system detects that deviants differ from standards, it will process the two kinds of stimuli in different ways, which can be reflected in the event-related potentials (ERPs). The MMR can be computed as the difference between the ERP elicited by the deviants and the ERP elicited by the standards.

When measuring MMRs to speech sounds in an oddball paradigm, it can make a difference whether one or the other stimulus of a pair is chosen as a standard. Possible asymmetries in participants' perception can exist, which can make discrimination easier if one particular stimulus (e.g., [æ]) is the standard than if the contrasting stimulus (e.g., [ɛ]) is the standard. Perceptual biases have been reported for several speech sounds (Pisoni, 1977; Aslin and Pisoni, 1980; Polka and Bohn, 1996, 2003) and seem especially strong in young infants (Pons et al., 2012). For vowels one relevant perceptual bias is a peripherality-related asymmetry: when hearing a more peripheral vowel after a more central vowel (i.e., in a two-dimensional acoustic space defined by the first and second vowel formants) discrimination is easier than when hearing the same vowels in the opposite order (e.g., Polka and Bohn, 1996, 2003; Pons et al., 2012). This would predict that in our oddball paradigm discrimination may be easier if [ɛ] is the standard stimulus than if [æ] is the standard. Further, the "natural referent vowel"

hypothesis (Polka and Bohn, 2011) predicts that this perceptual bias will vanish or grow fainter for native contrasts and will remain or grow stronger for non-native contrasts. This would predict that if fast distributional training already leads to some sort of vowel category formation, unimodally trained infants, for whom the contrast /æ/~/ɛ/ is new ("non-native"), will show a perceptual asymmetry, whereas the bias will not be clear in bimodally trained infants, for whom the contrast is experienced during training ("native"). Other perceptual biases can be expected on the basis of hypotheses involving underspecification (Lahiri and Reetz, 2010), according to which a featurally underspecified phoneme will mismatch with a preceding specified phoneme, but the reverse order will not lead to a mismatch. This would predict that if [æ] is specified for the feature [low] and [ɛ] is not, discrimination may be easier if [æ] is the standard stimulus than if [ɛ] is the standard. To accommodate the main and interaction effects of any perceptual biases, we counterbalanced the identity of the standard ([æ] or [ɛ]) across the infants and included it as a factor in the analysis.

In sum, the aim of the current study was to investigate whether 2-to-3-month old infants already show fast distributional learning, by training Dutch infants of this age on either a unimodal or a bimodal distribution of the English vowel contrast /æ/~/ɛ/, and then testing in an ERP oddball paradigm how well they discriminate [æ] from [ɛ]. If the distributional learning mechanism exists, it is expected that bimodally trained infants, who hear a distribution that suggests the existence of a contrast between /æ/ and /ɛ/, discriminate [æ] and [ɛ] better, and thus have a larger MMR amplitude, than unimodally trained infants, who hear a distribution that does not suggest a contrast between /æ/ and /ɛ/.

2. Materials and methods

2.1. Participants

The 32 infants (11 girls) accepted for the study met the following criteria. The language spoken at home had to be Dutch only. The infant had to be healthy and had to have passed the Dutch otoacoustic emissions test for newborns. Birth weight

had to be normal (each infant weighed over 2500 g). The Apgar score had to be 8 or higher 10 minutes after birth. The gestational age at birth had to be between 37 and 42 weeks, and the post-natal age from birth to time of testing between 8 and 12 weeks. Finally, we excluded infants born with complications, but accepted infants delivered by Caesarean section. The study protocol was approved by the Ethical Committee of the Faculty of Social and Behavioural Sciences at the University of Amsterdam. Parents signed informed consent forms.

2.2. Design

All infants listened to a training distribution and performed a subsequent discrimination test. During the training, half of the infants heard a bimodal distribution, with peaks around [æ] and around [ɛ], and the other half a unimodal distribution, with a single broad peak between [æ] and [ɛ]. During the test, half of the infants in each distributional training group listened to standard [æ] and deviant [ɛ], and the other half to standard [ɛ] and deviant [æ]. Thus, based on Distribution Type (unimodal vs. bimodal) and Standard Vowel ([æ] vs. [ɛ]) the 32 infants were assigned to four "groups," namely Unimodal [æ], Unimodal [ɛ], Bimodal [æ], and Bimodal [ɛ], each consisting of eight infants. Apart from balancing the sexes, assignment to the groups was random.

After separating the data into non-quiet sleep (non-QS) and quiet sleep (QS) data (section 2.5) and applying a criterion for a sufficient number of valid responses (section 2.6), we could include the non-QS data of 22 infants in the non-QS dataset, and the QS data of 21 infants in the QS dataset (12 infants contributed to both datasets, 19 to one dataset, and one to no dataset). In the non-QS dataset the number of contributing infants was five in Unimodal [æ], six in Unimodal [ɛ], six in Bimodal [æ], and five in Bimodal [ɛ]. In the QS dataset the number of contributing infants was six in Unimodal [æ], four in Unimodal [ɛ], five in Bimodal [æ], and six in Bimodal [ɛ].

To sum up, the experimental design for measuring the effect of distributional training had Distribution Type (unimodal vs. bimodal) and Standard

Vowel ([æ] vs. [ɛ]) as between-subject factors, and the MMR amplitude as the dependent variable, to be determined separately for the QS and the non-QS dataset.

2.3. Stimuli

Test and training stimuli were made with the Klatt synthesizer in the computer program Praat (Boersma and Weenink, 2010) and varied only in the values for the first and second formants, F1 and F2 (see sections 2.3.1 and 2.3.2). The duration of each stimulus was kept at 100 ms (e.g., Cheour-Luhtanen et al., 1995; Cheour et al., 1998a; Cheour et al., 2002b) including rise and fall times of 5 ms. The fundamental frequency contour fell from 150 to 112.5 Hz, which represents a male voice (e.g., Cheour-Luhtanen et al., 1995; Cheour et al., 1998a; Cheour et al., 2002b; Martynova et al., 2003). The source signal was filtered with eight additional formants (F3 through F10). The values for F3, F4, and F5, which were 2400, 3400,

Figure II.2. Infant participating in the experiment.

and 4050 Hz respectively, were extracted from American-English vowels representing /æ/ and /ɛ/ in the TIMIT database (Lamel et al., 1986), while those for F6 through F10 were calculated as the previous formant plus 1000 Hz (e.g., F6 = F5 + 1000 Hz). Similarly, the bandwidth values for the first four bandwidths, which were 80, 160, 360, and 530 Hz, respectively, were based on the TIMIT database, while an additional six bandwidths were calculated as the corresponding formant divided by 8.5 (e.g., bandwidth 5 = F5/8.5). Each stimulus was made equally loud, to avoid possible confounds in the ERPs based on intensity differences (Näätänen et al., 1989; Sokolov et al., 2002). The stimuli were played (during training and test) at around 70 dB SPL, measured at about one meter from the two loudspeakers, where the infant was lying.

2.3.1. In the training

The unimodal and bimodal training distributions were created in the manner reported in chapter VI. In contrast with previous research, which typically employed only eight different stimulus values, each of which was repeated multiple times during training, this method uses more ecologically valid continuous training distributions, where all presented stimuli are acoustically different. Each of the two distributions thus consisted of 900 unique vowels and had an identical range of F1 and F2 values: 9.41 to 13.53 ERB for F1 and 21.05 to 18.31 ERB for F2 (see also Figure II.1). These ranges were based on values for F1 and F2 as reported by Hawkins and Midgley (2005). Specifically, we took the reported F1 and F2 values of /æ/ and /ɛ/, each pronounced four times by five male speakers of British English in the age group 35–40 years, and converted the hertz values to ERB. Hawkins and Midgley's mean F1 and F2 were 12.51 ERB and 18.94 ERB, respectively for /æ/, and 10.43 ERB and 20.42 ERB for /ɛ/. Because in the current study the stimuli were produced by one synthetic speaker, a single-speaker standard deviation, for F1 and F2 separately, was calculated as the mean of the five speakers' standard deviations for the vowel /ɛ/. The standard deviations were 0.51 ERB for F1 and 0.32 ERB for F2. The edges of the F1 and F2 ranges,

mentioned above, were determined to lie two standard deviations from the mean F1 and F2 values of the vowels; for instance, the lower edge of the F1 continuum lay at 10.43 - 2 × 0.51 = 9.41 ERB. Note that in going from /ɛ/ to /æ/ the F1 rises, while F2 declines.

The shape of the distributions was defined in accordance with earlier distributional learning studies in that the ratio of the least to most frequent stimuli was about 1 to 4 (e.g., Maye et al., 2002, 2008). As illustrated in Figure II.1, the unimodal mean lay exactly in the middle of the range of F1 (or F2) values and precisely in between the two bimodal means, which lay at 25 and 75% of the range, for both F1 and F2. This led to the mean F1 and F2 values listed in Table II.1, which are quite close to those reported for /æ/ and /ɛ/ by Hawkins and Midgley (2005; see above in this section). The unimodal and bimodal distributions consisted of one and two Gaussian peaks, respectively, with standard deviations equal to 22 and 11% of the range, respectively. On the basis of these distributions, the F1 and F2 values for the 1800 training vowels were determined by a procedure described in chapter VI, which approximates the intended probability densities of Figure II.1 optimally. The order of presentation of the 900 stimuli in the training was randomized separately for each infant. The inter-stimulus interval (the silent interval between the end of a stimulus token and the start of the next token) was 707 ms.

Table II.1. F1 and F2 values (in ERB): means in the unimodal and bimodal training distributions, and values of the two test stimuli.

	Bimodal /ɛ/	**Test stimulus 1**	**Unimodal**	**Test stimulus 2**	**Bimodal /æ/**
F1	10.44	10.78	11.47	12.16	12.50
F2	20.37	20.14	19.68	19.22	18.99

2.3.2. In the test

In the test phase, infants were presented with two different stimuli, i.e., a standard and a deviant, repeated at most 2200 and 300 times respectively, depending on the infant (see section 2.4). Thus, deviants were presented at a rate of 12%. The F1 and F2 values of the test stimuli (Table II.1) were determined by computing the intersections (circles in Figure II.1) between the unimodal and bimodal distributions. In this way, the two groups of listeners came to the test phase with equal prior exposure (during training) to sounds in the region of the test stimuli, so that any difference between the groups observed in the test could not be attributed to differences in familiarity with the test stimuli. As during training the inter-stimulus interval in the test was 707 ms. In the test, minimally three standards (10 at the start of the test) appeared before each deviant. Apart from this constraint, the presentation of standards and deviants was randomized separately for each infant.

2.4. Procedure

Before training, the EEG cap with electrodes was placed on the infant's head (Figures II.2 and II.3). During training and testing, infants were lying on the caregiver's lap or in an infant seat beside the caregiver, in a sound-shielded room. Caregivers could watch a silent movie. Researchers in the adjacent room could hear caregiver and infant via loudspeakers, and observe them through a window. Researcher and caregiver did not know and could not consciously detect whether the distribution that was played during the training was unimodal or bimodal. The infant's behaviour was monitored and documented. Notes on behaviour included the documentation of open or closed eyes, movement, fussiness, and pauses. Caregivers were asked not to interact with the infant, unless necessary to keep the infant quiet. In this case, recording was paused or (if it happened in the last minutes of the test) stopped. Excluding pauses, the training always lasted 12.1 minutes (900 training stimuli) and the test lasted between 29.7 and 33.6 minutes (between 2208 and 2500 test stimuli).

2.5. Coding sleep stages

A factor that has to be considered when measuring MMRs is that during the relatively long experimental duration (viz., in the current experiment over 30 minutes, as compared to less than 10 minutes in behavioural distributional learning experiments) young infants tend to fall asleep (see also e.g., Friederici et al., 2002; He et al., 2009). It was therefore important to take a possible influence of sleep stages on MMR measurements into account. Infant sleep stages are usually divided into quiet sleep (QS), active sleep (AS), and wakefulness. Although some studies have not found any differences in neonates' MMR amplitudes between different sleep stages (e.g., Martynova et al., 2003), there are two arguments to analyse data obtained in QS separately from data obtained during wakefulness for 2-to-3-month olds. First, for 2-month olds Friedrich et al. (2004) report a significantly larger positive MMR in QS than during wakefulness, as well as a preceding small negative MMR in wakefulness that was absent in QS. Second, sleep stages and the related EEG-patterns develop quickly into adult-like patterns

Figure II.3. Infants participating in the experiment.

already in the first 3 months of life (e.g., Crowell et al., 1982; Kahn et al., 1996; Graven and Browne, 2008), and the adult MMR during wakefulness differs from that during sleep, particularly during the successor of QS, non-rapid eye movement (NREM) sleep, where the response tends to disappear (e.g., Loewy et al., 1996; Loewy et al., 2000). In sum, there is at least some evidence that for 2-to-3-month olds the MMR in QS is different from that during wakefulness.

Sleep stages for each infant were determined on the basis of the infant's behaviour and the EEG. Stages in the EEG were coded in accordance with the AASM manual (Iber et al., 2007) and, because the manual's age granularity is not precise enough to deduct recommendations for 2-to-3-month olds specifically, specifications for approximately the same age group from Crowell et al. (1982) and Niedermeyer (2005). Specifically, the stage was coded as "QS" when the infant's eyes were closed and the EEG contained frequent spindles (i.e., more or less sinusoidal waves of 12 to 14 Hz, clearly distinguishable from background activity, and lasting at least 0.5 s; see also Rodenbeck et al., 2007) or apparent slow waves (with or without spindles) coming after parts with abundant spindling. The stage was coded as "AS" when the infant's eyes were closed and the EEG featured transient muscle movements and low-amplitude mixed frequency activity. Finally, the stage was coded as "awake" when the eyes were open. When unequivocal identification was not possible (i.e., the eyes were closed but the EEG did not suggest QS or AS), the state was coded as "indeterminate sleep" (IS). A change of stage was not coded if the relevant changes in EEG and behaviour lasted for less than 30 s (Iber et al., 2007).

It turned out that none of the infants stayed awake during recording. On average, they spent 13% of test time awake, 47% in QS, 1% in AS and 39% in IS. There were no significant differences in the time spent in each sleep stage between the four groups (four independent-samples Kruskal–Wallis tests, one for each sleep state, all p-values > 0.74).

For all subsequent analysis, we combined the three non-QS sleep stages (AS, IS, wakefulness) and labelled them together as "non-QS" (cf. Weber et al., 2004). As for AS, only three infants were in this stage for a short while (accounting

for less than 2% of test time in any group), which is not surprising in the light of the rare AS onsets at 3 months of age and the relatively late expected start of AS after sleep onset as compared to the total test duration (Ellingson and Peters, 1980; Crowell et al., 1982); moreover, no reliable differences have been reported between MMRs during AS and MMRs during wakefulness in newborns (e.g., Cheour et al., 1998a; Kushnerenko, 2003). As for IS, we suspected that the infant was either well awake or drowsy, even though the eyes were closed, because the EEG in IS looked similar to that during wakefulness and did not contain any visual sign of QS. After combining the three non-QS variants, the sleep stages ended up being nearly equally divided between QS (47% of the time) and non-QS (53%).

2.6. ERP recording and analysis

The EEG was recorded with a 32-channel Biosemi Active Two system (Biosemi Instrumentation BV, Amsterdam, The Netherlands) at a sampling rate of 8 kHz. Beside the 32 electrodes in the cap, two external electrodes were placed on the mastoids. After recording, the EEG was downsampled to 512 Hz (with Biosemi Decimator 86). Subsequent analysis was done in the computer program Praat (Boersma and Weenink, 2010). First, the EEG was tagged for sleep stages (see section 2.5). Then the EEG in each of the 32 channels was referenced to the mastoids (i.e., the average of the two mastoid channels was subtracted from each channel), "detrended" (i.e., a line was subtracted so that beginning and end of the channel signal were zero) and filtered (Hann-shaped frequency-domain, i.e., zero-phase, filter: pass-band 1–25 Hz, low width 0.5 – high width 12.5 Hz).

The subsequent analysis was done for QS and non-QS data separately, as follows. The EEG was segmented into epochs (32-channel ERP waveforms) of 760 ms duration (from 110 ms before to 650 ms after stimulus onset), for standard and deviant stimuli separately. For each epoch, a baseline correction was performed in each channel by subtracting from each (1-channel) ERP waveform the mean of the waveform in the 110 ms before stimulus onset. If after this an epoch (i.e., a 32-

channel ERP waveform) still contained a peak below –150 μV or above +150 μV in one or more channels, the whole epoch was deemed invalid and rejected from further analysis. If after this fewer than 75 deviant epochs remained, the infant was rejected from the dataset for the relevant sleep stage. For each remaining infant, the standard and deviant responses were averaged separately, so as to obtain a mean standard ERP and a mean deviant ERP for each electrode. The infant's 32-channel MMR waveform was obtained by subtracting the standard ERP from the deviant ERP.

2.7. MMR analysis

In order to be able to submit the MMR measurements to statistical analysis, each infant's MMR waveform was reduced to a small set of MMR amplitude values (see below in this section). To achieve this reduction, it was necessary to decide what electrodes and what time window(s) to include in the analysis. The literature that uses infant MMR analysis varies in these decisions and, relatedly, also in the reported results on where on the scalp the MMR was found and when the response occurred (see below in this section). In addition, the literature reports different polarities for the infant MMR (see below in this section). Thus, whereas the adult MMR is invariably a negative deflection (hence usually called a mismatch negativity, or MMN) that usually occurs between 150 and 250 ms after change onset, and is strongest at frontocentral electrodes (when the mastoids or the nose is used as a reference; for a review, see Näätänen et al., 2007), the infant MMR is much less defined in terms of what its polarity is, and when it occurs where on the scalp. We now explain our decisions on how these three aspects of the MMR waveform enter in our analysis.

As for the polarity of the infant MMR, it is sometimes reported as negative (e.g., Cheour-Luhtanen et al., 1995; Cheour et al., 1998b), sometimes as positive (e.g., Dehaene-Lambertz and Baillet, 1998; Dehaene-Lambertz, 2000; Carral et al., 2005), and sometimes as both negative and positive (e.g., Morr et al., 2002; Friederici et al., 2002; Friedrich et al., 2004). Regarding the variation in

observed MMR polarities for infants across studies, we include both negative and positive values of individual infant's MMR amplitudes in our analysis.

As for the location of interest on the scalp, some previous research selected only frontal electrodes (e.g., Morr et al., 2002) or frontal and central electrodes (e.g., Cheour et al., 1998b; Morr et al., 2002). When more posterior electrodes were included a significant infant MMR was sometimes reported only at frontal or frontocentral electrodes (Cheour-Luhtanen et al., 1995; Friederici et al., 2002; Friedrich et al., 2004), and sometimes also in more posterior areas (Cheour et al., 2002a; Van Leeuwen et al., 2008; He et al., 2009). As there is therefore some evidence that the infant MMR can be measured beyond frontocentral electrodes, our analysis includes not only six frontocentral electrodes (Fz, F3, F4, Cz, C3, C4), but also two temporal electrodes (T7 and T8); parietal and occipital electrodes were not included, because some infants had been lying on these electrodes. Following Cheour et al. (1998b), Morr et al. (2002) and Friedrich et al. (2004) we include the eight electrodes in the main analysis as a within-subject factor.

As for the chosen time window, the previous literature on infant MMR used various windows for vowels (e.g., 0–500 ms after stimulus onset in Cheour-Luhtanen et al., 1995; 200–500 ms in Cheour et al., 1998a) and various windows for 2- or 3-month olds (e.g., 0–1000 ms in Friederici et al., 2002; 200–600 ms in Friedrich et al., 2004; 100–450 ms and 550–900 ms in He et al., 2009). The only publication on vowels with infants in our age range (3-month olds: Cheour et al., 2002b) used a window from 150 to 400 ms. Regarding the reported variation, and because control of the Type I error rate dictates that analysis windows be chosen before the ERP results are seen, we had to choose in advance a window that includes at least the possible times at which the MMR can occur, namely a window running from 100 to 500 ms. In order to submit this window to an analysis of variance (ANOVA), we divide it into eight consecutive time bins of 50 ms each (Cheour-Luhtanen et al., 1995; He et al., 2009), and compute the average amplitude of the difference waveform in each bin as our measurement variable. To conclude, each infant's MMR waveform is reduced to only 64 (8 time bins × 8 channels) MMR amplitude values.

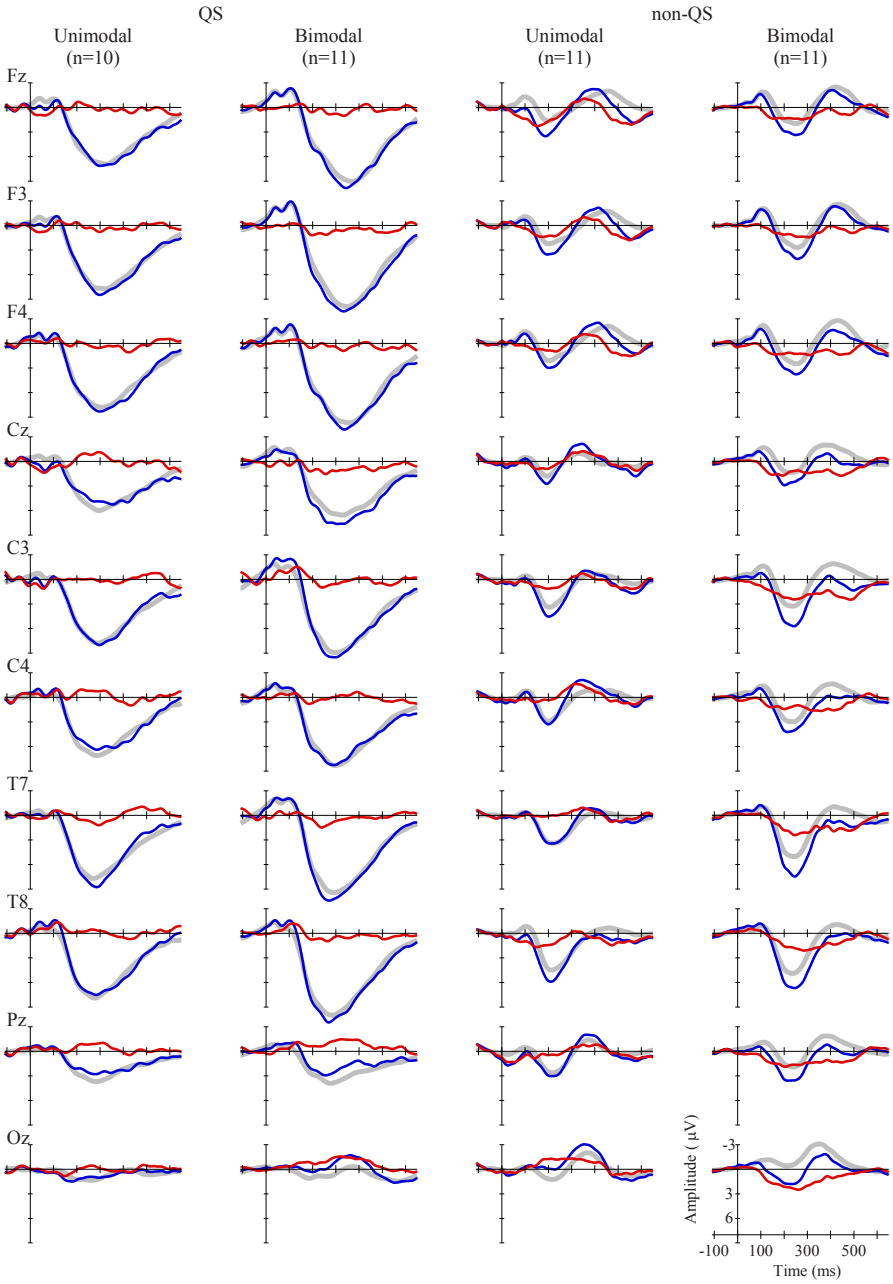
2.8. Statistical analysis

To test whether there is a difference between unimodally and bimodally trained infants, while controlling for differences in the presented standard, we subjected the QS and non-QS datasets separately to an ANOVA with a mixed design (between-subject factors and repeated measures). The MMR amplitude was the dependent variable, Time Bin (100–150, 150–200, 200–250, 250–300, 300–350, 350–400, 400–450, and 450–500 ms) and Electrode (Fz, F3, F4, Cz, C3, C4, T7, and T8) were within-subject factors, and Distribution Type (unimodal vs. bimodal) and Standard Vowel ([æ] vs. [ɛ]) were between-subject factors. The design also included all possible interactions between the factors, up to the fourth order. To compensate for the double chance of finding results (separate QS and non-QS analyses) all tests employ a conservative α level of 0.025.

3. Results

The grand average waveforms for each Distribution Type (unimodal vs. bimodal) pooled over the two levels of the factor Standard Vowel are presented in Figure II.4, for 10 electrodes. In line with previous research on 2-to-3-month olds, the standard and deviant ERPs contained prominent slow positive waves (e.g., Friederici et al., 2002; Morr et al., 2002; Carral et al., 2005; Shafer et al., 2011), and the ERPs in the QS data appeared large compared to those in the non-QS data (e.g., for 2-month olds: Friederici et al., 2002; for newborns: Pihko et al., 2004; Sambeth et al., 2009; but see Cheour et al., 2002a, for conflicting results).

Figure II.4 (opposite page). Grand average standard (grey, thick curves), deviant (blue, thin curves) and MMR (red, thin curves) waveforms, at 10 electrodes (see rows), for unimodally and bimodally trained infants in QS (left two columns) and non-QS (right two columns).


QS
non-QS
Unimodal
(n=10)
Bimodal
(n=11)
Unimodal
(n=11)
Bimodal
(n=11)
Fz
F3
F4
Cz
C3
C4
T7
T8
Pz
Oz
Amplitude (µV)
-3
3
6
-100
100
300
500
Time (ms)

For the QS data, the ANOVA on the MMR amplitude yielded significant results neither for the research question (main effect of Distribution Type: p = 0.88), nor for any other main effect (Standard Vowel: p = 0.23; Electrode: F < 1; Time Bin: F < 1), nor for any of the 11 interactions (all p -values > 0.07).

For the non-QS data, the ANOVA revealed a positive grand mean (+0.84 μV), with a 97.5% confidence interval (CI) that does not include zero (+0.35 ~ +1.33 μV), implying that on average Dutch 2-to-3-month old infants can discriminate the test vowels, and that vowel discrimination in these infants is reflected in a *positive* MMR. Regarding our specific research question, the analysis showed a main effect of Distribution Type (mean difference = +1.06 μV, CI = +0.08 ~ +2.04 μV, F [1,18] = 7.03, p = 0.016, η_p^2 = 0.28): across electrodes and time windows the bimodally trained infants had a higher positive MMR (+1.37 μV, CI = +0.68 ~ +2.06 μV) than the unimodally trained infants (+0.31 μV, CI = –0.38 ~ +1.00 μV), indicating that Dutch 2-to-3-month olds' neural discrimination of [æ] and [ɛ] is better after bimodal than after unimodal training.

As for factors not directly pertaining to our research question, there was no effect of Standard Vowel (p = 0.98), so that we cannot state with confidence that one of the two combinations of standard and deviant vowel yields a higher MMR amplitude (and thus better neural discrimination) than the other combination. Further, the analysis showed no main effects of Time Bin (F [7ε, 126ε, ε = 0.334] = 1.37, Greenhouse–Geisser corrected p = 0.27) or Electrode (F < 1). Thus, there was no support for a more positive or more negative MMR in any specific time window as compared to other ones within 100 and 500 ms, and at any specific electrode as compared to other ones among the frontocentral and temporal electrodes. Interestingly, we found a highly significant interaction effect between Distribution Type and Standard Vowel [F(1,18) = 20.22, p = 0.0003, η_p^2 = 0.53], which shows that the attested difference between unimodally and bimodally trained Dutch 2-to-3-month olds differs depending on the standard that they hear in the oddball test (see section 3.1).

3.1. Exploratory results for the four groups

To examine the responses of the four non-QS groups separately, we pooled the MMR amplitudes across electrodes and time bins in view of the lack of significant differences herein (see section 3). Figure II.5 shows the pooled MMR waveforms per group, and Table II.2 lists the corresponding averaged MMR amplitudes. The amplitude differed from zero significantly only for the Bimodal [ɛ] group (p = 0.004, uncorrected for multiple comparisons) implying that bimodally trained Dutch 2-to-3-month olds who are tested with standard [ɛ] and deviant [æ] can hear the difference between the two vowels.

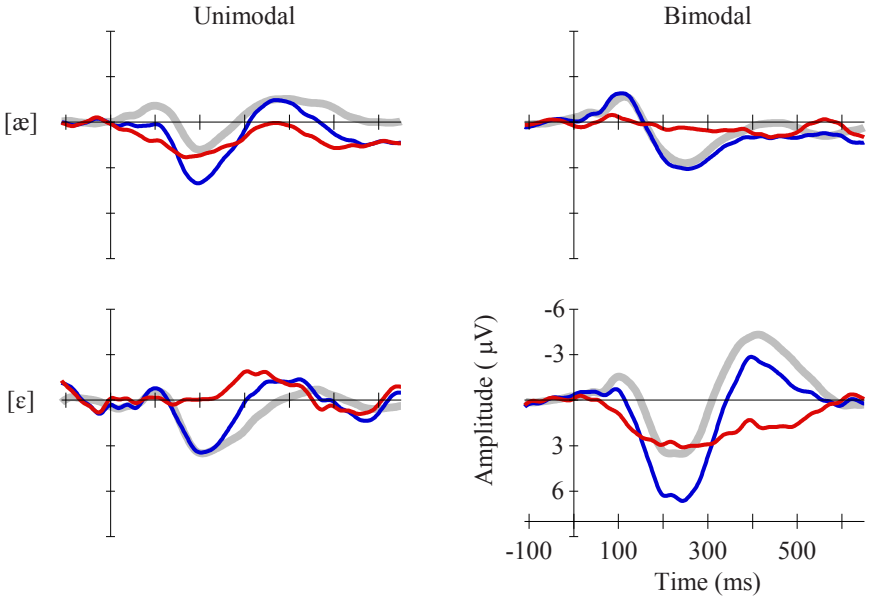


Figure II.5. Standards (grey, thick curves), deviants (blue, thin curves) and MMRs (red, thin curves) in non-QS, pooled across eight electrodes, per group (Unimodal [æ] top left vs. Bimodal [æ] top right, Unimodal [ɛ] bottom left vs. Bimodal [ɛ] bottom right).

Table II.2. Mean MMR amplitudes (in μV) between 100 and 500 ms across eight electrodes per subgroup for non-QS data, with within-group standard deviations (SD; between parentheses) and 97.5% confidence intervals.

Distribution Type	Standard Vowel	N	Mean (SD)	Confidence Interval	*t*	*p*
Unimodal	**[ɛ]**	6	–0.59 (0.86)	–1.71 to +0.52	–1.69	0.153
Unimodal	**[æ]**	5	+1.21 (1.23)	–0.71 to +3.14	+2.20	0.092
Bimodal	**[ɛ]**	5	+2.26 (0.83)	+0.97 to +3.55	+6.12	0.004
Bimodal	**[æ]**	6	+0.48 (0.80)	–0.55 to +1.50	+1.46	0.203

Significance is tested against zero in four one-sample *t*-tests (without correction for multiple testing).

The individual group's MMR amplitudes presented in Table II.2 are visualized in Figure II.6. The interaction between Distribution Type and Standard Vowel, which was found in the main ANOVA for the non-QS data (see section 3), is clearly visible. We did the four relevant group comparisons, assuming equal variances for all groups (as in the ANOVA): Bimodal [ɛ] vs. Unimodal [ɛ], Bimodal [æ] vs. Unimodal [æ], Bimodal [ɛ] vs. Bimodal [æ] and Unimodal [æ] vs. Unimodal [ɛ] (technically, this was done via *post hoc* comparisons using Fisher's Least Significant Difference in SPSS). The Bimodal [ɛ] group's response was reliably more positive than that of the Unimodal [ɛ] group (see the arc numbered 1 and the black line in Figure II.6; uncorrected $p = 0.00008$); this indicates that when the standard in the oddball paradigm is [ɛ] and the deviant is [æ], bimodally trained Dutch 2-to-3-month olds show better neural discrimination than unimodally trained infants. The difference between Bimodal [æ] and Unimodal [æ] was not significant ($p = 0.21$); thus, when the standard is [æ] and the deviant [ɛ], unimodally trained infants do not necessarily have higher response amplitudes. The Bimodal [ɛ] group's response was greater than that of the Bimodal [æ] group (the arc numbered 2 in Figure II.6; $p = 0.005$), suggesting that neural discrimination is easier for bimodally trained Dutch 2-to-3-month olds when the standard is [ɛ] and the deviant

is [æ] than when standard and deviant are reversed. Conversely, the Unimodal [æ] group's response was more positive than that of the Unimodal [ɛ] group's response (the arc numbered 3 in Figure II.6; $p = 0.005$), which suggests that neural discrimination is easier for unimodally trained Dutch 2-to-3-month olds when the standard is [æ] and the deviant is [ɛ] than when standard and deviant are reversed.

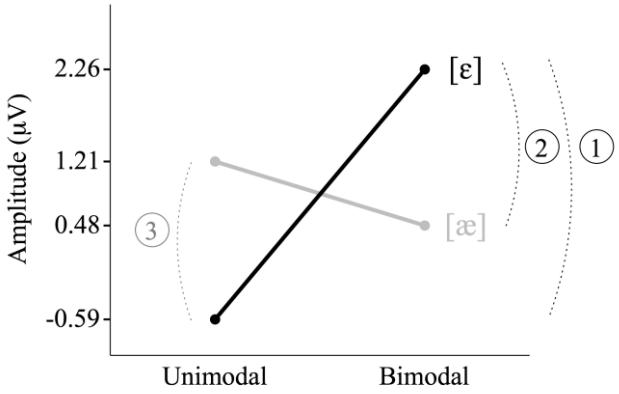


Figure II.6. Three post hoc significant differences in MMR amplitude between the four subgroups. Unimodal [ɛ] (left black), Unimodal [æ] (left grey), Bimodal [ɛ] (right black), Bimodal [æ] (right grey). Note: Among the four amplitudes, only the one for the Bimodal [ɛ] group differed from zero significantly.

4. Discussion

The present study provides the first evidence for fast distributional learning in very young infants. The specific research question was whether Dutch 2-to-3-month old infants show larger mismatch responses, hence presumably better discrimination, of English [ɛ] and [æ] after bimodal than after unimodal training. This was answered in the affirmative with a *p*-value of 0.016. The age of 2 to 3 months is early enough for the distributional learning mechanism to be able to play a role in

the transition from universal to language-specific speech perception, which has been observed to take place from 4 to 12 months.

This outcome extends previous research in two ways. First, fast distributional learning has now been attested at widely different ages, namely at 2 to 3 months (the present study), between 6 and 11 months (Maye et al., 2002, 2008; Yoshida et al., 2010; Capel et al., 2011), and in adults (Maye and Gerken, 2000, 2001; Gulian et al., 2007; Hayes-Harb, 2007; Escudero et al., 2011; Chapters V and VI). One can now hypothesize that the mechanism is available throughout life and can contribute to first and second language acquisition. Second, the ERP method has now been added to the set of methods by which distributional learning can be demonstrated. We needed the ERP method because of the young age of our participants, but this technique might have the general advantage over behavioural methods that it does not require the participant's attention and that it taps the response process at a time when the response is still little influenced by the myriads of factors that contribute to the behavioural part of the response. An assessment of the general usefulness of the ERP technique, especially in comparison with behavioural techniques, has to await replication with more age groups, larger sample sizes and more phonological contrasts.

The ERP method potentially yields information on the scalp distribution and the timing of the responses. Our results, however, do not allow us to determine any precise scalp location or timing. This indeterminacy is not uncommon in studies on infant MMRs (see section 2.7), and may be due to the more pronounced shapes of sulci and gyri in adults than in infants (Hill et al., 2010) and to the larger variability in MMR timing among infants than among adults (e.g., Kushnerenko, 2003). More location- or time-specific results can be expected at later ages.

This study detected an interaction between the type of distribution (bimodal vs. unimodal) in the training and the identity of the standard vowel ([ɛ] vs. [æ]) in the test ($p < 0.001$); *post hoc* exploration suggested that bimodally trained infants discriminated better if the standard was [ɛ] and unimodally trained infants discriminated better if the standard was [æ]. This confirms none of the three

predictions that we derived from previous literature in the Introduction: the peripherality-related asymmetry predicted on the basis of Polka and Bohn (1996), namely that the MMR should be larger if the standard is [ɛ], was not found (main effect of Standard Vowel: $p = 0.98$); a prediction indirectly derived from the "natural referent vowel" hypothesis (Polka and Bohn, 2011), namely that the peripherality-related asymmetry should occur only in the unimodal group, was contradicted by our detection of the asymmetry in the bimodal group and the opposite asymmetry in the unimodal group; a prediction derived from the "featural underspecified lexicon" model (Lahiri and Reetz, 2010), namely that the MMR should be larger if the standard is [æ], was not confirmed (main effect of Standard Vowel: $p = 0.98$). None of the hypotheses in the literature predicted the asymmetry that we did find, and we cannot speculate on it before many more ERP results on asymmetries have been collected.

Given the effect of distributional training in the young infants tested, the question arises what the mechanism is: is there an enhanced discrimination in the bimodally trained infants (acquired distinctiveness), or is there a reduced discrimination in the unimodally trained infants (acquired similarity), or both? We cannot answer this question on the basis of our results, because time constraints prevented us from testing the infants' perception before training. Also, a pre-test would have been an additional distributional training and could therefore have distorted the intended training distributions. Although to our knowledge MMRs for 2-to-3-month olds in response to similar small differences between vowels as between our test vowels (i.e., 1.38 ERB in F1 and 0.92 ERB in F2) have not been examined before, the acoustic difference between the test vowels was well above the discrimination threshold reported for 8-week old infants as measured behaviourally by high-amplitude sucking (Swoboda et al., 1976, 1978). On the other hand, the vowels in those studies were different, had different durations and were presented with different inter-stimulus intervals than in the current study, so that we cannot be certain that our 2-to-3-month olds discriminated the test vowels before training. Similarly, we cannot say if a potential perceptual ease of listening to the order [ɛ] – [æ] strengthened the effect of distributional learning for the

bimodally trained infants and/or if a potential perceptual difficulty of listening to the opposite order weakened this effect.

One may wonder why the training–test paradigm works at all. After all, the test phase presents a (shrunk) bimodal distribution to the infants, and it can be expected that they continue to learn during the test, which lasts quite a bit longer (30 minutes) than what we call the "training" (12 minutes). The persistent influence of the training is possibly related to the much larger variability during training (900 different stimuli) than during the test (2 different stimuli). From other training paradigms it is known that a large variability in training stimuli can facilitate learning and could be instrumental in category formation (e.g., Lively et al., 1993). Future research is necessary to examine the persistence of short-term distributional learning over time.

With regard to the methodology of testing 2-to-3-month olds, the results highlight the importance of documenting sleep stages and analysing QS data separately from non-QS data. In QS the MMR did not emerge, which is in line with the disappearance of the MMN in adult NREM sleep, and with the development of infant QS into an adult-like NREM in the first 3 months of life (see section 2.5), but in contrast to the lack of differences in the MMR between sleep stages in newborns (Martynova et al., 2003), and, for 2-month olds, to the *larger* MMR in QS than during wakefulness in Friedrich et al. (2004) and to the robust MMR in QS in Van Leeuwen et al. (2008). The many differences between these infant studies and the current study (if not simply due to chance) make it difficult to pinpoint the cause of this discrepancy. One difference from the studies mentioned is that the current study tested perception *after short-term training*. Thus, it may be that training effects were not yet sufficiently encoded in neural activation patterns to surface in QS. Alternatively, if infants who were in QS during the test, had already been in QS during the training, learning may have been hampered in QS as compared to non-QS.

We conclude that 2-to-3-month olds are sensitive to distributions of speech sounds in the environment. This is earlier than what has been shown in previous experiments with fast distributional learning, and earlier than the onset of

language-specific speech perception. A linguistic interpretation of these results is that at 2 months of age infants already have a mechanism in place that can support the acquisition of phonological categories.

Chapter III

Distributional vowel training is less effective for adults than for infants: a study using the mismatch response

Karin Wanrooij, Paul Boersma and Titia L. Van Zuijen
PLoS ONE 2014, 9(10), 1-13, doi: 10.1371/journal.pone.0109806

Data available from the public figshare database:
http://dx.doi.org/10.6084/m9.figshare.1157804

Abstract

Distributional learning of speech sounds (i.e., learning from simple exposure to frequency distributions of speech sounds in the environment) has been observed in the lab repeatedly in both infants and adults. The current study is the first attempt to examine whether the capacity for using the mechanism is different in adults than in infants. To this end, a previous event-related potential study that had shown distributional learning of the English vowel contrast /æ/~/ɛ/ in 2-to-3-month old Dutch infants was repeated with Dutch adults. Specifically, the adults were exposed to either a bimodal distribution that suggested the existence of the two vowels (as appropriate in English), or to a unimodal distribution that did not (as appropriate in Dutch). After exposure the participants were tested on their discrimination of a representative [æ] and a representative [ɛ], in an oddball paradigm for measuring mismatch responses (MMRs). Bimodally trained adults did not have a significantly larger MMR amplitude, and hence did not show significantly better neural discrimination of the test vowels, than unimodally trained adults. A direct comparison between the normalized MMR amplitudes of the adults with those of the previously tested infants showed that within a reasonable range of normalization parameters, the bimodal advantage is reliably smaller in adults than in infants, indicating that distributional learning is a weaker mechanism for learning speech sounds in adults (if it exists in that group at all) than in infants.

1. Introduction

"Distributional learning" is learning from simple exposure to the frequency distributions of stimuli in the environment (Lacerda, 1995; Guenther and Gjaja, 1996). It is assumed to be an important mechanism by which infants start to learn the phonemes of their native language (e.g., Werker and Tees, 1984). In the lab, where exposure to speech sound distributions lasts only a few minutes, the mechanism has been reported not only for infants, but also for adults who try to master difficult speech sound contrasts of a second language (section 1.1).

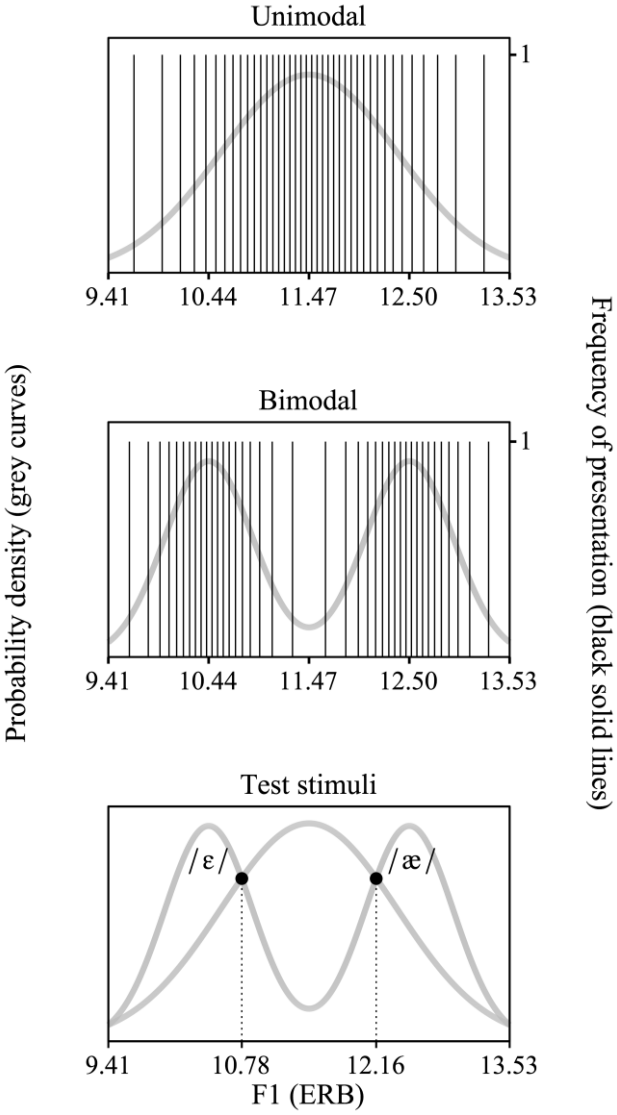
Some of the previous research suggests that the capacity for distributional learning of speech sounds is smaller in adults than in infants (section 1.2), while other research implies that this capacity remains fairly robust in adulthood (section 1.3). Here we present the first attempt to *directly* compare adults and infants in their capacity for distributional learning of speech sounds. For this, a recent distributional learning experiment with infants (chapter II) was repeated with adults, and the effect of distributional training on the adults' neural auditory discrimination performance was compared to that of the infants.

1.1. Distributional learning

The concept of distributional learning can be illustrated best with an example. The chosen example is relevant in the current study, where we use distributions encompassing the same speech sound contrast, namely the English vowel contrast /æ/~/ɛ/ as in *bat* vs. *bet*. In Southern British English (SBE) the vowels in this contrast differ primarily in the first and second formants (F1 and F2). Specifically, /æ/ has a higher F1 and a lower F2 than /ɛ/ (Hawkins and Midgley, 2005). For the sake of clarity, we focus on the F1 values only. When hypothetically measuring the F1 values in many tokens of SBE /æ/ and /ɛ/ (mixed), it can be observed that the values are grouped around two values, one for the mean of /æ/ and one for the mean of /ɛ/. This is illustrated in the middle graph of Figure III.1. Each vertical line indicates an F1 value. The curve shows the underlying probability density function. Because the function has two peaks, the distribution is called bimodal.

In Dutch the contrast /æ/~/ɛ/ is not phonemic, and Dutch listeners show difficulty in mastering it (e.g., Schouten, 1975; Weber and Cutler, 2004; Broersma, 2005; Escudero et al., 2008). This is probably because Dutch has the single vowel /ɛ/ (as in the Dutch word *pet*, "cap") in roughly the region of the F1-by-F2 vowel space occupied by SBE /æ/~/ɛ/ (Adank et al., 2004; Van Leussen et al., 2011). When hypothetically measuring the F1 values in many tokens of Dutch /ɛ/, the values cluster around a single value, which is the mean F1 of Dutch /ɛ/ (top graph of Figure III.1). Consequently, the underlying probability density function (the curve) has only one peak and is thus unimodal. Distributional learning reflects the idea that the language-specific distributions cause English listeners to experience two vowels in this region of the vowel space, and Dutch listeners one vowel.

Figure III.1 (opposite page). Distributions of first formant (F1) values (in ERB). The unimodal (top) and bimodal (middle) distributions represent the Dutch vowel /ɛ/ and the English vowel contrast /ɛ/~/æ/, respectively. Each solid vertical line indicates a vowel token with a specific F1 value. Each vowel token was presented only once (i.e., the height of the vertical lines is 1). The grey curves are the underlying probability density functions. When creating training distributions, the acoustic values of the test stimuli can be calculated by computing the intersections (black discs, bottom) of the unimodal and bimodal distributions.


Unimodal
1
9.41
10.44
11.47
12.50
13.53
Bimodal
1
9.41
10.44
11.47
12.50
13.53
Test stimuli
/ɛ/
/æ/
9.41
10.78
12.16
13.53
F1 (ERB)
Probability density (grey curves)
Frequency of presentation (black solid lines)

The existence of distributional learning has been demonstrated in the lab, where exposure to speech sound distributions takes just a few minutes. In a typical distributional learning experiment, participants (e.g., the Dutch infants in chapter II) are exposed to either a bimodal distribution of speech sounds representing a contrast to be acquired (e.g., the SBE contrast /æ/~/ɛ/, as for one group of infants in chapter II) or to a unimodal distribution that represents a single native speech sound (e.g., Dutch /ɛ/, as for a second group of infants in chapter II). After exposure, participants are tested on their discrimination or identification of two tokens that were represented equally in both distributions during training (e.g., for the infants in chapter II: an [æ] and an [ɛ], as illustrated by the black discs in the bottom graph of Figure III.1). If distributional training is effective, bimodally trained listeners should discriminate or identify the two test stimuli better than unimodally trained listeners, because the bimodal distribution is expected to make listeners experience the test stimuli as belonging to different speech sound categories and the unimodal distribution is expected to make them experience these stimuli as being representatives of a single speech sound category. Indeed, several studies report such an effect of distributional training, both studies with infants (including chapter II, and Maye et al., 2002, 2008; Yoshida et al., 2010; Capel et al., 2011), and studies with adults (Maye and Gerken, 2000, 2001; Gulian et al., 2007; Hayes-Harb, 2007; Escudero et al., 2011; chapters V and VI).

1.2. Previous research with plosive distributions

Only one set of studies has examined distributional learning of the same speech sound contrast in adults (Maye and Gerken, 2000, 2001; Hayes-Harb, 2007) *and* infants (Maye et al., 2002, 2008; Yoshida et al., 2010), namely the voicing contrast between the "voiced" plosive in the English word *day* and a voiceless unaspirated plosive similar to that in the English word *stay*, with participants from English homes. The overall results suggest in a weak manner (namely, by comparing multiple degrees of significance, which does not constitute a valid statistical test) that distributional learning, which was observed in both adults and infants, might

have a smaller scope in the former than in the latter group. Specifically, for infants, exposure to a bimodal distribution of the voicing contrast at one place of articulation (e.g., a distribution of [d]~[t]) turned out to enhance discrimination of the same contrast at *another* place of articulation (e.g., between [g] and [k]) (Maye et al., 2008), whereas for adults the parallel results were not significant (Maye and Gerken, 2001). Also, Yoshida et al. (2010) argue that the capability to learn from exposure to a speech sound distribution may weaken with age already within the first year of life. Two groups of 10-to-11-month olds in this study did not improve discrimination significantly after a 2.3-minute bimodal training (which is the same duration as used earlier for the younger infants, who were reported to exhibit distributional learning; Maye et al., 2002, 2008). After a longer training (4.6 minutes) an additional group of 10-to-11-month olds did exhibit significantly improved discrimination (a direct comparison between the three groups was not reported). Exposure duration in the adult studies was chosen to be even longer (9 minutes) (Maye and Gerken, 2000, 2001; Hayes-Harb, 2007).

In sum, on the basis of this set of studies (i.e., those using plosive distributions), one might hypothesize that distributional learning is a less prominent mechanism in adults than in infants. Unfortunately, the method differed between the adult and infant studies in several aspects (including the actual stimuli, the procedure and, as just mentioned, the training duration). Moreover, as said above, neither adults and infants, nor older infants and younger infants, nor groups exposed to different training durations, were compared with a direct statistical test. Consequently, the studies in this set cannot really be interpreted as providing evidence for a declining prominence of distributional learning with age. Also, the contrast used in this set was a voicing distinction in plosives, for which the distributional learning mechanism may be very different from the distributional learning of vowels, which we investigate in the current study (section 1.4).

1.3. Previous research with vowel distributions

A second set of studies on distributional learning used vowel distributions, as we do in the present study, and also includes both studies with adults (Gulian et al., 2007; Escudero et al., 2011; chapters V and VI) and a study with infants (chapter II). The results demonstrate that an effect of distributional training *can* be measured in adults after short exposure (5 minutes in Gulian et al., 2007; less than 2 minutes in Escudero et al., 2011, and in chapters V and VI), thus suggesting that the capacity for distributional learning can remain rather robust in adulthood. Unfortunately, the vowel contrasts used for the adults (Dutch /ɑ/~/aː/ and /ɪ/~/i/ for Bulgarian learners in Gulian et al., 2007; Dutch /ɑ/~/aː/ for Spanish learners in Escudero et al., 2011, and in chapters V and VI) do not match those for the infants (SBE /æ/~/ɛ/ for Dutch infants in chapter II), and test procedures differed between the adult and infant studies. Consequently, it is not clear how the observed effects of distributional training in adults relate to those in infants.

1.4. The objective of the current study

As explained above (sections 1.2 and 1.3), previous research implies conflicting conclusions about the capacity for distributional learning in adults as compared to that in infants. On the one hand, this capacity may decline with age (section 1.2). On the other hand, the capacity for distributional learning seems robust regardless of age, as it is measurable in a fast distributional training paradigm in both infancy and adulthood (section 1.3). The purpose of the current study was to shed light on the effect of age on the capacity for distributional learning. Specifically, the aim was to directly compare adults' capacity for distributional learning to that of infants, and thus to determine the relative importance of the mechanism for learning speech sounds in adulthood, when speech sounds of new languages are learned, versus that in infancy, when the speech sounds of the native language are learned. In order to examine whether adults have a smaller capacity for distributional learning than infants, we first repeated a recent study that demonstrated an effect of distributional training of SBE /æ/~/ɛ/ in Dutch infants

aged 2 to 3 months (chapter II), with Dutch adults. Subsequently, we aimed to determine whether any observed effect of distributional training in the adults was smaller than the corresponding effect observed in the infants in chapter II.

1.5. Comparing distributional learning in infants and adults

In any comparison between participant groups, it is important to use the same method, i.e., the exact same training, with the same duration, and the same method for testing discrimination after training for all participant groups. A method that can be used for both infants and adults to test discrimination after distributional training is the measurement of the mismatch response (MMR), a brain response that can be calculated from event-related potentials (ERPs). The MMR has been related to behavioural discrimination in adults (for a review see Näätänen et al., 2007) and has been used widely to test discrimination in newborns (e.g., Cheour-Luhtanen et al., 1995; Partanen et al., 2013), older infants (e.g., Cheour et al., 1997; Van Zuijen et al., 2013), children (e.g., Kraus et al., 1999; Shafer et al., 2011) and adults (e.g., Näätänen et al., 1997; Winkler et al., 1999). The MMR has also been used to compare speech sound discrimination in infants versus adults (Pang et al., 1998).

The MMR can be recorded in an oddball paradigm (Näätänen, 1992), in which infrequent "deviant" stimuli (e.g., [æ]) appear randomly in a train of "standard" stimuli (e.g., [ɛ] tokens). If the auditory system signals a difference between the standards and the deviants, it will generate different brain responses (ERPs) to the two kinds of stimuli. This difference between the ERP to the deviants and that to the standards is the MMR. Larger perceived differences between standard and deviant stimuli have been related to larger MMR amplitudes, not only in adults (Näätänen et al., 1997; Aaltonen et al., 1997), but also in children (Kraus et al., 1999) and in one-year old infants (Cheour et al., 1998b).

The cause of the MMR method being suitable for infants and adults alike is that the MMR reflects automatic auditory processing, which occurs before participants can pay conscious attention to the stimuli (Näätänen and Winkler,

1999), and which is elicited even if participants do not attend to the stimuli at all (Näätänen et al., 1978; Näätänen, 1992; Schröger, 1997). Consequently, the response does not depend on a behavioural task, which young infants cannot perform. The MMR thus allows for minimizing methodological differences between testing infants and testing adults on their discrimination performance.

When comparing the MMR of infants and adults, a point of concern is that the infant and adult MMR may not reflect the same neural processes: the underlying ERPs have a very different morphology in infants than in adults, which is probably partly due to structural differences (i.e., the size and anatomical structure of the brain and skull), and partly to representational differences (i.e., linguistic representations are likely to be either absent or immature in infants). Notice, however, that as the MMR is a difference wave (see above in this section), part of the differences between infant and adult ERPs is removed by the subtraction. Nevertheless, in order to compensate for differences between infant and adult MMRs that cannot be avoided by using the same method and by subtracting ERPs, some kind of normalization has to be performed that scales the MMR amplitudes prior to statistical analysis (section 2.7). Normalization between infant and adult MMRs was applied before (Pang et al., 1998), albeit without a specification of the exact normalization method.

In order to facilitate the comparison of the effect of distributional training between infants and adults, the present study: (1) minimizes methodological differences by measuring the MMR in adults, as was done for the infants in chapter II, and (2) normalizes the MMR amplitudes before statistical analysis.

In sum, the present study first examines whether distributional training of SBE /æ/~/ɛ/ is effective for Dutch adults, by repeating an experiment that demonstrated an effect of such training in Dutch 2-to-3-month-old infants (chapter II). Specifically, we expose the Dutch adults to either a bimodal or a unimodal distribution encompassing /æ/~/ɛ/, and then test their discrimination of a representative [ɛ] and [æ] by recording the MMR in an oddball paradigm. On the basis of earlier reported effects of distributional training in adults (discussed in sections 1.2 and 1.3), it is expected that the bimodally trained participants will

discriminate the test vowels better, and will thus have a larger MMR amplitude, than the unimodally trained listeners. Secondly, we examine whether the difference in the normalized MMR amplitude between bimodally and unimodally trained participants is indeed smaller in the adults than in the infants in chapter II.

2. Method

Below we first describe the method for determining whether distributional vowel training is effective for Dutch adults. This method is identical to that used in the previous infant study (chapter II), except where stated otherwise. The final section (section 2.7) explains our approach to normalizing the MMR amplitudes across infants and adults.

2.1. Design

All adults received a pre-test, a training and a post-test. Because the infants in chapter II did not do a pre-test, the pre-test data will not be discussed in this paper. The reason for not doing a pre-test with infants was that such a test could be an additional distributional training that distorts the intended training distributions (section 4 in chapter II); since there is strong evidence that adults do not learn in "passive" tests (i.e., where they do not have to perform a specific task and can ignore the presented stimuli, as was the case in the present experiment; e.g., Keuroghlian and Knudsen, 2007), a pre-test was included for the adults to permit later comparisons with other studies on distributional training of adults (Gulian et al., 2007; Escudero et al., 2011; chapters V and VI).

During training, participants listened to either a unimodal or a bimodal distribution of vowels encompassing /æ/~/ɛ/ (section 2.4). Distribution Type (unimodal vs. bimodal) was included as the main between-subject factor in the statistical analysis.

In the post-test, the MMR was recorded in an oddball paradigm (Näätänen, 1992) to assess discrimination of a representative [æ] and [ɛ] (section 2.4). Half of

the participants in each training group listened to standard [æ] and deviant [ɛ] in the test, and the other half to standard [ɛ] and deviant [æ]. This was done in view of possible asymmetries in participants' perception. For instance, an asymmetry predicted by Polka and Bohn (1996, 2003) can make discrimination easier if relatively central vowels (in the two-dimensional vowel space defined by F1 and F2; e.g., [ɛ] as compared to [æ]) are presented *before* relatively peripheral vowels (e.g., [æ] as compared to [ɛ]) than if they are presented in the reverse order. Conversely, an asymmetry predicted on the basis of the "featurally underspecified lexicon" theory by Lahiri and Reetz (2010) can make discrimination easier if a vowel specified for the phonological feature [low] (i.e. /æ/) is followed by a vowel not specified for that feature (i.e. /ɛ/), than if they are presented in the reverse order. To control for such potential biases, Standard Vowel ([ɛ] vs. [æ]) was included as a between-subject factor in the statistical analysis.

Thus, the statistical analysis of the effect of distributional training on adults' discrimination performance had the MMR amplitude as the dependent variable, and Distribution Type (unimodal vs. bimodal) and Standard Vowel ([æ] vs. [ɛ]) as between-subject factors.

2.2. Participants

Participants were native speakers of Dutch that had been raised monolingually, had not lived abroad during childhood, and had never passed more than four weeks in countries where English is the national language. Forty-four participants were tested, of whom 5 were excluded from analysis (see section 2.5). On the basis of the factors Distribution Type and Standard Vowel (section 2.1), the remaining 39 participants belonged to one of four "groups", namely Unimodal [æ] (n=9), Unimodal [ɛ] (n=10), Bimodal [æ] (n=9) or Bimodal [ɛ] (n=11). Apart from balancing the sexes (there were 2 or 3 men in each of these groups), the assignment to these groups was random. The Unimodal group thus contained 19 participants (mean age 22 years, range 18 to 28 years) and the Bimodal group 20 participants (mean age 22 years, range 18 to 30 years). In the infant study (chapter II), the

relevant analysis had been based on a smaller number of participants, namely 11 infants in the Unimodal and Bimodal groups each.

2.3. Ethics statement

The Ethical Committee of the Faculty of Social and Behavioural Sciences at the University of Amsterdam approved the study protocol. Participants were recruited via posters and flyers distributed at the University of Amsterdam and at public places in Amsterdam. Each participant received an information brochure before coming to the lab. The participant signed an informed consent form before the experiment and was paid 20 euros.

2.4. Stimuli and procedure

The stimuli used in the training and in the test were created with the Klatt synthesizer in Praat (Boersma and Weenink, 2014). All had the same duration (100 ms, including a rise and fall time of 5 ms), fundamental frequency (F0) contour (150 to 112.5 Hz), intensity (70 dB) and third through tenth formants (F3 = 2400 Hz, F4 = 3400 Hz, F5 = 4050 Hz, F6 through F10: previous formant plus 1000 Hz). The stimuli varied in F1 and F2 (see below). All stimuli were played at 70 dB SPL, measured at about one meter from two loudspeakers, where the participant was sitting. The inter-stimulus interval in the training and the tests was 707 ms. Total experimental time was 45.7 minutes (i.e., 12.1 minutes for the training and 16.8 minutes for each test).

2.4.1. Training

The unimodal (Figure III.1, top) and bimodal (Figure III.1, middle) training distributions each consisted of 900 acoustically different vowels, of which the values of the varying parameters (F1 and F2) reflected a probability density function that approximated a continuous distribution. The distributions were made

as described in chapter VI. Both distributions had identical ranges of F1 and F2 values, based on values reported in Hawkins and Midgley (2005): 9.41 to 13.53 ERB (Equivalent Rectangular Bandwidth) for F1 and 21.05 to 18.31 ERB for F2 (for details see chapter II). The 1800 F1 and F2 values were calculated on the basis of these defined ranges for F1 and F2 and the defined shapes of the distributions, which were based on earlier distributional learning studies (see chapter II for details). The unimodal and bimodal mean F1 and F2 values, i.e., the values represented by the peaks of the Gaussian curves in Figure III.1, were 11.47 and 19.68 ERB respectively for the unimodal mean, 10.44 and 20.37 ERB for the bimodal mean representing /ɛ/, and 12.50 and 18.99 ERB for the bimodal mean representing /æ/. The presentation of the stimuli was randomized per listener. Participants were instructed to relax and listen to the vowels carefully. Because the exposure time was longer than in previous behavioural studies on adult distributional learning (namely more than 12 minutes versus 9 minutes in Maye and Gerken, 2000, 2001 and Hayes-Harb, 2007; 5 minutes in Gulian et al., 2007; and less than 2 minutes in Escudero et al., 2011, and in chapters V and VI), there was the risk that participants would fail to pay attention to the vowels during the whole training. This had to be avoided because there is extensive evidence that in contrast to infant listeners, adult listeners do not learn if they do not pay attention to the task (Keuroghlian and Knudsen, 2007). Therefore, in order to help participants to keep their attention on the training vowels, they were not only asked to listen carefully, but also to indicate after training how many different vowels they had perceived. The inclusion of a task to keep participants' attention to the training vowels is not uncommon in studies on adult distributional training (Maye and Gerken, 2000; Hayes-Harb, 2007).

2.4.2. Test

The F1 and F2 values of the standard stimulus and the deviant stimulus in the post-test were defined by the intersections of the unimodal and bimodal F1 and F2 distributions (the black discs in Figure III.1, bottom). These intersections represent

the values that have been trained equally intensively in both distributions. The F1 and F2 values were 10.78 and 20.14 ERB respectively for the stimulus representative of [ɛ] and 12.16 and 19.22 ERB respectively for the stimulus representative of [æ]. Half of the participants heard [ɛ] as the standard and [æ] as the deviant, and the other half heard the opposite pattern. The post-test contained 1100 standard tokens and 150 (i.e., 12%) deviant tokens, which is half the numbers presented to the infants in chapter II. This was done because we expected less noisy data for the adults. Besides the constraint that minimally three standards (ten at the start of the test) had to appear before each deviant, the presentation of standards and deviants was randomized per participant. Participants watched a silent movie during recording.

2.5. ERP recording and analysis

The ERP recording and analysis were similar to those in chapter II. The EEG was recorded with a 64-channel Biosemi Active Two system (Biosemi Instrumentation BV, Amsterdam, The Netherlands). In addition to the 64 electrodes in the cap, reference electrodes were placed on the mastoid processes and the nose. (The nose reference was not used. It was recorded to permit later comparisons with studies that use the nose as a reference). Also, one electrode was placed to the left of the left eye and one to the right of the right eye in order to track horizontal eye movements, and two electrodes were placed above and below the right eye respectively to monitor vertical eye movements. The sampling rate was 8 kHz, which was downsampled to 512 Hz after recording (Biosemi Decimator 86). The subsequent analyses were performed in Praat (Boersma and Weenink, 2014). The EEG in each channel was referenced to the mastoids (i.e., the mean of the two mastoid signals was subtracted from each of the 64 channel signals), detrended (i.e., a straight line was subtracted from each channel signal in such a way that its beginning and end became zero) and filtered with a zero-phase pass-band filter between 1 and 25 Hz (implemented in the frequency domain; Hann-shaped smoothing 0.5 Hz at the low edge, 12.5 Hz at the high edge). We then extracted

from the EEG signal a large number of 500-ms epochs, namely one for each stimulus token. Each epoch started 110 ms before the onset of the stimulus and ended 390 ms after it. Subsequently, we performed a baseline correction on each epoch by subtracting from each of its channels the mean in that channel of the 110 ms before the onset of the stimulus. Subsequently, we removed all epochs that contained a voltage below –75 μV or above +75 μV in one or more of its channels. In this way, we obtained a set of standard epochs and a set of deviant epochs; if the number of deviant epochs was below 100 for a certain participant, we excluded all of this participant's data from further analysis (this happened for five of the 44 participants).

The data of each remaining participant was simplified in the following way. By averaging over all (at most 1100) standard epochs, we computed the participant's "mean standard ERP", which is a 500-ms 64-channel ERP whose average over the first 110 ms is 0. Similarly, we computed the participant's "mean deviant ERP" by averaging over all (100 to 150) deviant epochs. Finally, we computed the participant's 64-channel MMR waveform by subtracting the mean standard ERP from the mean deviant ERP.

In this way, ERPs were recorded and analysed similarly to those of the Dutch infants in chapter II. The differences, which reflect adaptations to the measurement of adult as opposed to infant MMRs, were a larger number of electrodes (64 vs. 32), shorter epochs (500 ms vs. 760 ms; see section 2.6) and more stringent norms for artefact rejection (±75 μV vs. ±150 μV) and for the minimum number of deviants (100 vs. 75).

2.6. MMR analysis

Numerous studies have established the *adult* MMR as a negativity (as reflected in the name "mismatch negativity" or MMN; Näätänen et al., 1978) occurring predominantly at fronto(central) electrodes (when the chosen reference is the nose or the mastoids) in a time frame between roughly 150 and 250 ms after change onset (for a review, see Näätänen et al., 2007). In many studies, the analysis is

confined to the midline frontal electrode Fz (e.g., Näätänen et al., 1997; Winkler et al., 1999; Pakarinen et al., 2009), because the MMN tends to be prominent there (Näätänen et al., 2007).

In line with these properties of the MMN, we performed the following steps for each of the four groups, i.e. for each combination of Distribution Type (i.e., uni- and bimodal) and Standard Vowel ([æ] and [ɛ]). We first determined the group's 64-channel waveform by averaging the MMR waveforms of the group's participants, and then determined the "group latency" as the time of the most negative voltage occurring in this average waveform in the Fz channel between 150 and 250 ms after stimulus onset. Then, we defined a 50-ms "group window" of analysis, starting 25 ms before and ending 25 ms after the group latency. Subsequently, we determined each participant's "MMR amplitude" at Fz by time-averaging the participant's MMR waveform at Fz over this window. In this way we reduced the MMR waveform for each participant to one relevant number only.

It should be noted that for the infants in chapter II the MMR amplitude had been computed somewhat differently due to the larger uncertainty about the location on the scalp and the timing of the MMR for infants than for adults (for a discussion, see chapter II). Because of the uncertainty as to scalp location, the infant response was not analysed at Fz only, but at eight different electrodes, ranging in scalp position from frontal to central and temporal (parietal and occipital electrodes were not used because several infants had been lying on these electrodes), and Electrode was included as a within-subject factor in the statistical analysis. In view of the uncertainty pertaining to the timing, the infant response was analysed across eight 50-ms windows between 100 and 500 ms after stimulus onset, and Time Bin was included as a within-subject factor in the statistical analysis. After observing that all effects involving Electrode or Time Bin were insignificant, the infant MMR amplitudes were pooled across electrodes and time bins, thus reducing them to one number for each participant only, reflecting the mean MMR amplitude in a 50-ms window between 100 and 500 ms after stimulus onset, and across electrodes.

In sum, the adult MMR amplitude was the mean amplitude at Fz in one data-dependent 50-ms window determined between 150 and 250 ms after stimulus onset, and the infant MMR amplitude was the mean amplitude averaged across eight electrodes and all eight 50-ms windows between 100 and 500 ms after stimulus onset.

2.7. Comparing infant and adult MMRs: normalization

Even after minimizing methodological differences between testing infants and testing adults, it was possible that the MMR amplitudes (as computed in the previous section) still incorporated differences between the age groups that do not pertain to neural discrimination. In an attempt to filter out these residual differences, we examined whether a quantifiable relation between infant and adult MMR amplitudes could be deduced from previous literature. Because the difference between the test vowels [æ] and [ɛ] can be termed a difference in vowel quality, we looked for pairs of adult and infant studies in which MMRs in response to the *same* vowel quality differences were recorded. Table III.1 presents the MMR amplitudes in the pairs of studies found in the literature.

When aiming to quantify the relation between adult and infant MMRs, the first issue to be addressed is a potential *polarity* difference, as the table shows for [i:]~[e:]. As mentioned above (section 2.6), adult MMRs are commonly negative. Infant MMRs differ in polarity across studies. In some studies they are negative (as in many studies in Table III.1), in other studies positive (e.g., Dehaene-Lambertz and Baillet, 1998; Dehaene-Lambertz, 2000; Carral et al., 2005; Partanen et al., 2013), and in still other studies both negative and positive MMR components are reported (e.g., Morr et al., 2002; Friederici et al., 2002; Friedrich et al., 2004). To accommodate polarity differences between infant and adult MMRs, we consider from now on the *absolute values* of the mean MMR amplitudes in Table III.1.

The second issue to be addressed in a comparison of adult and infant MMRs is the *size* of the MMR. If we collapse all MMR amplitudes listed in Table III.1 per vowel (regardless of factors such as age and sleep stage) and then average

over the five vowel contrasts, we obtain an adult average of 2.98 μV and an infant average of 2.54 μV. Based on these numbers, infant MMRs become comparable to adult MMRs if they are multiplied by a scaling factor of 1.18. We could be more precise and restrict ourselves to studies where the vowels are matched and where two factors that may influence the MMR amplitude, namely age (Shafer et al., 2011) and sleep stage (Friedrich et al., 2004), are taken the same for the infants as in chapter II. In that case only three comparisons between infant and adult MMR amplitudes are left in Table III.1, namely those where the infants were 3 months old and were awake. The absolute MMR amplitudes in these studies were 4.0 μV (Cheour et al., 1997) or 3.1 μV (Cheour et al., 1998a) for infants versus 4.5 μV (Aaltonen et al., 1987) for adults, and 2.0 μV (Cheour et al., 1997) for infants versus 3.3 μV (Aaltonen et al., 1987) for adults, which would lead to a scaling factor of 1.41.

Another factor that can affect the MMR amplitude is the offset-to-onset inter-stimulus interval (Cheour et al., 2002a). If this inter-stimulus interval is required to be the same in the infant study as in the adult study, only one comparison mentioned in the table is left: 3.5 μV (Pakarinen et al., 2009) vs. 1.7 μV (Partanen et al., 2013). This would yield a (too unreliable) scaling factor of 2.05.

As the scaling factors thus determined are based on a very small sample of studies, the analyses below will include a *range* of scaling factors for the infant MMR amplitudes rather than just one or two. In addition, because the polarity of the MMR in the infants in chapter II was positive and a negative polarity is expected for the adults, we will multiply the adult MMR amplitudes by –1 before comparing them to the MMR amplitudes of the infants in chapter II.

Table III.1. Adult and infant studies in which MMRs to the same vowel pairs differing in quality ([standard]~[deviant]) were recorded. The MMR amplitude (MMR; in microvolts) is listed for both the adults and the infants. For the infants, age (in months) and sleep stage are also shown.

Vowel stimuli	**Adults**		**Infants**			
	Study	**MMR**	**Study**	**Age**	**Sleep stage**	**MMR**
[y]~[i]	Aaltonen et al., 1987	–4.5[a]	Cheour-Luhtanen et al., 1995	0	quiet sleep	–1.3[b]
	Aaltonen et al., 1987	–4.5[a]	Cheour et al., 1997	3	awake	–4.0[c]
	Aaltonen et al., 1987	–4.5[a]	Cheour et al., 1998a	0	quiet sleep	–1.7[d]
	Aaltonen et al., 1987	–4.5[a]	Cheour et al., 1998a	3	awake	–3.1[d]
[y]~[y/i]	Aaltonen et al., 1987	–3.3[a]	Cheour et al., 1997	3	awake	–2.0[c]
[e]~[ø]	Näätänen et al., 1997	–1.6[e]	Cheour et al., 1998b	6	awake	–4.5[f]
[e]~[o] (adults), [o]~[e] (infants)	Näätänen et al., 1997	–2.0[e]	Martynova et al., 2003	0	active sleep	–1.8[g]
	Näätänen et al., 1997	–2.0[e]	Martynova et al., 2003	0	quiet sleep	–2.1[g]
[i:]~[e:]	Pakarinen et al., 2009	–3.5[h]	Partanen et al., 2013	0	several	+1.7[h]

Notes: see opposite page.

a) Amplitudes calculated from the amplitudes mentioned for the "Ignore condition" at Fz ([45]: p.202)
b) Amplitude calculated from the amplitudes at F3 and F4 between 200 and 300 ms after stimulus onset.
c) Amplitude at C4 (peak observed between 200 and 300 ms).
d) Amplitude at F4.
e) Amplitude at Fz inferred from graph (Näätänen et al., 1997: Figure 4a on page 434).
f) Amplitudes at Cz inferred from graph (Cheour et al., 1998b: Figure 3 on page 353).
g) Amplitudes averaged across Fp1, Fp2, C3 and C4, and across MMN (measured between 100 and 300 ms) and LDN (measured between 300 and 500 ms).
h) Only the MMRs obtained in an oddball paradigm (the MMRs obtained in a multifeature paradigm are not included). At Fz in Pakarinen et al. (2009). At F3 and F4 in Partanen et al. (2013).

3. Results

3.1. Descriptives

3.1.1. Grand average waveforms

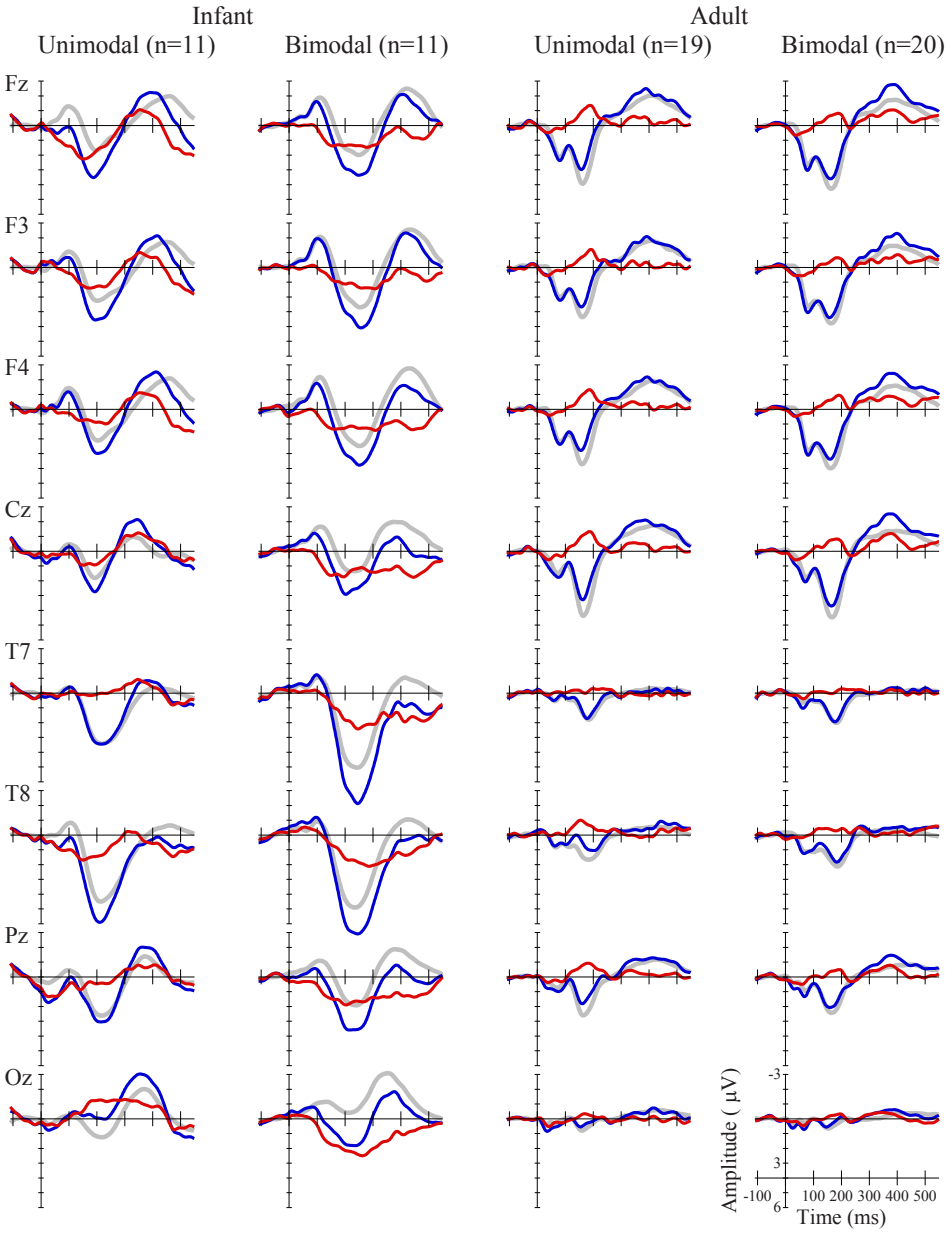
Figure III.2 shows the grand average standard, deviant and MMR waveforms of the adults in the current study (right) and, for comparison, of the infants in chapter II (left), at eight electrodes, for each Distribution Type (unimodal vs. bimodal) pooled over Standard Vowel. The figure confirms the negative polarity and the expected latency and fronto(central) scalp distribution of the adult MMN (section 2.6): the red curve, which is the MMR waveform, deviates in the negative direction (notice that negative polarities are plotted upwards) from the baseline between 150 and 250 ms, and seems to do so more at frontocentral sites then elsewhere. The figure also confirms that the infant MMR contains less pronounced peaks (Kushnerenko et al., 2002) and that its scalp distribution is less defined than in adults (e.g., Cheour et al., 2002a, see also chapter II). Also, in accordance with

several previous studies (e.g., Dehaene-Lambertz and Baillet, 1998; Dehaene-Lambertz, 2000; Carral et al., 2005; Partanen et al., 2013), the polarity of the infant MMR is positive.

3.1.2. Scalp distributions

Figure III.3 depicts the scalp distributions, which were made in Praat (Boersma and Weenink, 2014), for the unimodally (top) and bimodally (bottom) trained adults in the current study (right) and, for comparison, for the infants in chapter II (left). The adult distributions were measured between 167 and 217 ms after stimulus onset, i.e., in a 50-ms window around the average MMR latency (i.e., the time of the most negative voltage occurring in the grand average waveform at Fz between 150 and 250 ms), which was at 192 ms. The infant distributions were measured between 100 and 500 ms after stimulus onset (section 2.6). Just as the grand average waveforms in Figure III.2, the topographies of the MMR in Figure III.3 illustrate the adult negative polarity (always blue, never red) and frontocentral distribution (darkest blue at frontocentral sites). For the infants, the positive polarity (red) and less specified distribution (darkest colours are spread over the scalp) are clearest for the bimodally trained infants. The MMR was not significantly different from zero for the unimodally trained infants (details are provided in section 3.2).

Figure III.2 (opposite page). Grand average waveforms. Standard (grey, thick curves), deviant (blue, thin curves) and MMR (red, thin curves), at eight electrodes (see rows), for the unimodally and bimodally trained infants in chapter II (the two columns on the left) and adults in the current study (the two columns on the right).


Infant
Adult
Unimodal (n=11)
Bimodal (n=11)
Unimodal (n=19)
Bimodal (n=20)
Fz
F3
F4
Cz
T7
T8
Pz
Oz
Amplitude (μV)
-3
3
6
-100
100 200 300 400 500
Time (ms)

3.1.3. MMR amplitudes

The MMR amplitude in the overall window where the response was expected (i.e., between 150 and 250 ms after stimulus onset; see Method 6) was significantly negative for both the bimodally trained adults (mean = –0.45 μV, 95% confidence interval [henceforth CI] = –0.95 ~ –0.05 μV, $t[19] = -1.89$, $p = 0.037$) and the unimodally trained adults (–0.80 μV, 95% CI = –1.39 ~ –0.20 μV, $t[18] = -2.82$, $p = 0.006$), thus suggesting that both groups discriminated the two test vowels to some extent.

Subsequently, for each adult participant the MMR amplitude was calculated at Fz in a 50-ms window around the MMR latency for the participant's group (see Method 6). This group latency was 193 ms for Unimodal [æ], 196 ms for Bimodal [æ], and 189 ms for Unimodal [ɛ] and Bimodal [ɛ]. The MMR amplitudes, averaged over the participants per Distribution Type and Standard Vowel, are presented in Table III.2, together with their standard deviations and confidence intervals. For comparison, the corresponding numbers of the infant MMR amplitudes (see Method 6) are also shown.

In chapter II, no significant difference had been observed between the *infant* MMR amplitudes at frontal, central and temporal electrodes (Fz, F3, F4, Cz, C3, C4, T7, T8). To further explore the frontocentral scalp distribution observed in the *adult* grand average waveforms and scalp topographies, we performed an analysis of variance (ANOVA) with Electrode (the same eight electrodes as for the infants) as a within-subject factor. The effect of Electrode was significant ($F[7\varepsilon, 266\varepsilon, \varepsilon = 0.504] = 9.94$, Greenhouse–Geisser corrected $p < 0.001$). The amplitude at T7 (mean = –0.19 μV) was significantly less negative ("smaller") than the amplitudes at all frontal and central electrodes (mean at Fz = –0.91 μV, mean at Cz = –0.90 μV, mean at F3 = –0.77 μV, mean at F4 = –0.93 μV, mean at C3 = –0.85 μV, mean at C4 = –0.84 μV; all ps ≤ 0.002), and not significantly different from the amplitude at T8 (mean = –0.50 μV, $p = 0.80$). These results are in line with a predominantly frontocentral distribution of the adult MMN.

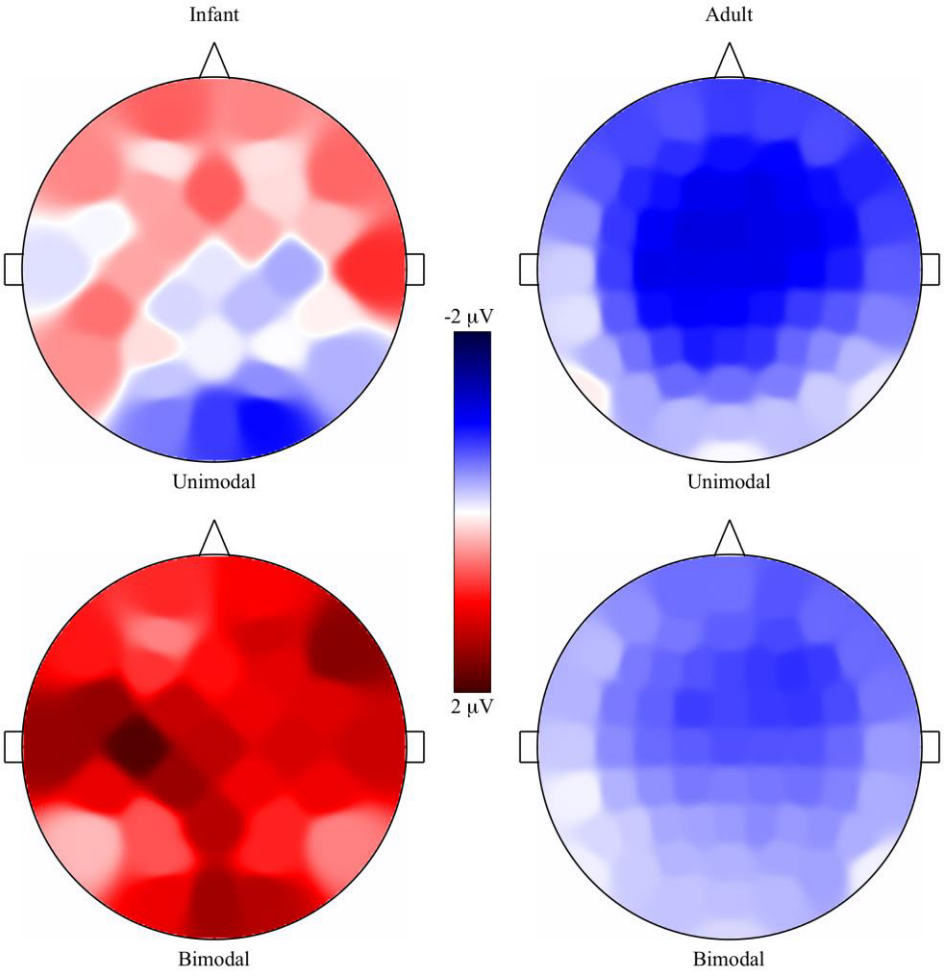


Figure III.3. MMR scalp distributions. Unimodally (top) and bimodally (bottom) trained infants in chapter II (left; 32 electrodes) and adults in the current study (right; 64 electrodes). Voltages time-averaged between 167 and 217 ms after stimulus onset for adults, between 100 and 500 ms for infants. Blue is negative, red positive, white zero.

Table III.2. Mean MMR amplitudes (in μV) for the adults in the current study and the infants in chapter II. With within-group standard deviations (SD) and 95% confidence intervals, calculated per Distribution Type and Standard Vowel.[a]

Age Group	Distribution Type	Standard Vowel	N	Mean	SD	Confidence Interval	
						Lower limit	Higher limit
Adult	**Unimodal**	**[ɛ]**	10	–1.12	0.99	–1.82	–0.41
		[æ]	9	–1.05	1.65	–2.31	+0.22
	Bimodal	**[ɛ]**	11	–0.35	0.86	–0.93	+0.23
		[æ]	9	–1.21	1.32	–2.23	–0.19
Infant	**Unimodal**	**[ɛ]**	6	–0.59	0.86	–1.71	+0.52
		[æ]	5	+1.21	1.23	–0.71	+3.14
	Bimodal	**[ɛ]**	5	+2.26	0.83	+0.97	+3.55
		[æ]	6	+0.48	0.80	–0.55	+1.50

a) For the infants the alpha level for the confidence intervals is 2.5% instead of 5%, because the infant study included an additional group of sleeping infants. For details see chapter II.

3.2. No significant effect of distributional vowel training in Dutch adults

Recall (section 2.1) that in order to test whether there was a difference between the unimodally and bimodally trained participants, while controlling for differences in the presented standard, we performed an ANOVA with the MMR amplitude at Fz as the dependent variable, and with Distribution Type (unimodal vs. bimodal) and Standard Vowel ([æ] vs. [ɛ]) as between-subject factors. The main and interaction effects were not significant (for Distribution Type: mean difference bimodal – unimodal = +0.30 μV, 95% CI = –0.50 ~ +1.10 μV, $F < 1$, $p = 0.45$; for Standard Vowel: mean difference [æ] – [ɛ] = –0.40 μV, 95% CI = –1.19 ~ +0.40 μV, $F[1, 35] = 1.02$, $p = 0.32$; for the interaction: $F[1, 35] = 1.41$, $p = 0.24$). Because the

effects involving Standard Vowel were not significant, the amplitude data do not show proof of any perceptual asymmetry (section 2.1). The insignificance of all effects involving Distribution Type implies that the amplitude data do not provide sound evidence that bimodally trained Dutch adult learners have a different amplitude (mean = –0.78 μV, 95% CI = –1.34 ~ –0.23 μV) and thus benefit differently from distributional training than unimodally trained learners (mean = –1.08 μV, 95% CI = –1.65 ~ –0.51 μV). For comparison, the corresponding ANOVA for the infants in chapter II, which also included Time Bin and Electrode as within-subject factors (see Method 6), had yielded a significant effect of Distribution Type (mean difference bimodal – unimodal = +1.06 μV, 95% CI = +0.08 ~ +2.04 μV, F [1, 18] = 7.03, p = 0.016), with a larger positive MMR, and thus a larger effect of distributional training, for the bimodally trained infants (mean = +1.37 μV, 95% CI = +0.68 ~ +2.05 μV) than for the unimodally trained infants (mean = +0.31 μV, 95% CI = –0.38 ~ +1.00 μV).

3.3. Smaller effectiveness of distributional training in adults than in infants

From the statistical significance of the distributional effect in infants (chapter II) and the statistical non-significance of the effect in adults (the present paper) we cannot yet conclude that the effect is greater in infants than in adults. A valid test requires a direct comparison of the two age groups. The difference in MMR amplitude between the Bimodal and Unimodal groups (i.e., Bimodal MMR – Unimodal MMR) for the adults was +0.30 μV (= – 0.78 μV – – 1.08 μV; i.e., in the unexpected direction, though non-significant), whereas that for the infants (chapter II) was +1.06 μV (= +1.37 μV – + 0.31 μV). This age difference does not appear to be due to adults having a smaller MMR amplitude in general than infants, because the literature review in the Method section (section 2.7) suggested that this amplitude is probably *greater* in adults than in infants. The age difference could therefore be due to a truly smaller effect of distributional training in adults than in infants. To verify this, the current section presents a numerical comparison of the infant and adult MMR amplitudes. As determined by the literature review in the

Method section (section 2.7), the comparison requires a normalization of the MMR amplitudes, which should include a correction for the opposite polarity of adult and infant MMRs and a scaling of the size of the MMR. To implement the normalization (or something equivalent to normalization), we multiplied each adult's MMR amplitude by –1 to correct for the negative polarity, and we multiplied each infant's MMR amplitude by a scaling factor to correct for the smaller size. Before applying the scaling factors estimated from the literature, which were 1.18 and 1.41 (section 2.7), we present the results for a more conservative scaling factor of 1.00 (i.e. no scaling), which is smaller than the estimates; this scaling turns the mean MMR for adults into –0.30 μV, and that for the infants into +1.06 μV, giving a difference of 1.36 μV.

3.3.1. Scaling factor of 1

Using a conservative scaling factor of 1, we performed an ANOVA with the normalized MMR amplitude as the dependent variable, and Age Group (infant vs. adult), Distribution Type (unimodal vs. bimodal) and Standard Vowel ([æ] vs. [ɛ]) as between-subject factors (given that in chapter II a strong interaction was observed between Distribution Type and Standard Vowel, Standard Vowel was included to be able to extract possible interactions with this variable). The ANOVA yielded the following normalized MMR amplitudes per Age Group and Distribution Type (as visible in Figure III.4): infant unimodal 0.31 μV (CI = –0.38 ~ +1.00 μV), infant bimodal 1.37 μV (CI = +0.68 ~ +2.05 μV), adult unimodal 1.08 μV (CI = +0.56 ~ +1.60 μV) and adult bimodal 0.78 μV (CI = +0.27 ~ +1.29 μV).

Crucially, the interaction between Age Group and Distribution Type was significant ($F[1,53] = 5.05$, $p = 0.029$). Thus, the effect of distributional training differed between infants and adults (see below). Further, the interaction between Distribution Type and Standard Vowel was significant ($F[1,53] = 4.85$, $p = 0.032$), as well as the triple interaction between Age Group, Distribution Type and Standard Vowel ($F[1,53] = 13.99$, $p = 0.0005$). The other interaction effect

(between Age Group and Standard Vowel) and the main effects were not significant (all *p*-values > 0.21).

As the number of participants was not the same in all groups, it is relevant to note that the crucial interaction between Age Group and Distribution Type did not depend much on the way the terms for the ANOVA were entered in the linear model. With "Type-III sums of squares", the *p*-value for each main or interaction effect is calculated from a comparison between the full model (i.e. the model with all main and interaction terms) and the full model from which only this one term was dropped. This led to the above-mentioned *p*-value of 0.029 for the interaction between Age Group and Distribution Type. With "Type-I sums of squares", the terms are entered into the linear model one by one and the *p*-value for each term depends on when the term is added. Under the constraint that the three two-way interaction terms are added after the three main terms and before the three-way interaction term, the *p*-value for the interaction between Age Group and Distribution Type depended only slightly on the order in which the two-way interactions entered into the model: it was 0.027 if this term was entered first, 0.024 if it was entered after Distribution Type × Standard Vowel but before Standard Vowel × Age Group; 0.025 if it was entered after Standard Vowel × Age Group but before Distribution Type × Standard Vowel; and 0.023 if it was entered last. By contrast, the interaction between Distribution Type and Standard Vowel was not robust to such variation. With Type-III sums of squares, the *p*-value of the interaction was as shown above (i.e., $p = 0.032$), while with Type-I sums of squares the effect was non-significant, irrespective of the chosen order of factors (i.e., the *p*-value ranged from 0.23 to 0.27). This difference in significance is due to the strong effect of the three-way interaction term: only if this triple term is present and has taken away much of the variance does the interaction between Distribution Type and Standard Vowel provide a significant improvement to the model. The robustness of the interaction of Age Group and Distribution Type, together with the lack of robustness of the interaction of Distribution Type and Standard Vowel, means that the former effect has been shown more credibly than the latter.

The observed interaction between Age Group and Distribution Type is pictured in Figure III.4. The figure suggests that the difference in the normalized MMR amplitude between unimodally and bimodally trained participants was larger (i.e., more positive after normalization) for the infants than for the adults. When controlling for a possible effect of Standard Vowel, this difference is significant for the infants (mean difference normalized bimodal – unimodal = +1.06 μV, 95% CI = +0.09 ~ +2.03 μV), thus indicating an effect of distributional training, and not significant for the adults (mean difference normalized bimodal – unimodal = –0.30 μV, 95% CI = –1.03 ~ +0.43 μV). In view of the significance of the interaction between Age Group and Distribution Type, it is now possible to interpret the significant effect of distributional training for the infants as indeed being larger (i.e., +1.06 – –0.30 μV = +1.36 μV, 95% CI = +0.15 ~ +2.57 μV) than the non-significant effect for the adults (if that effect exists at all).

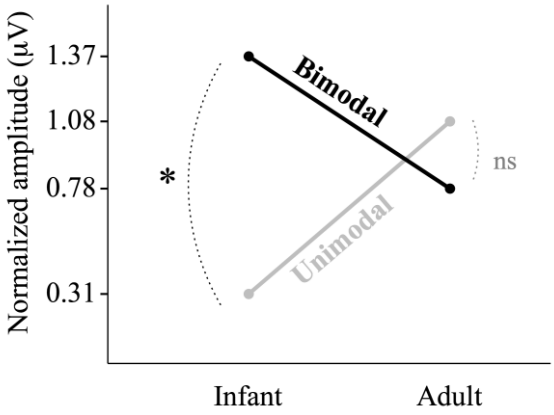


Figure III.4. The interaction between Age Group and Distribution Type. Age group: infant, left vs. adult, right. Distribution Type: unimodal, grey vs. bimodal, black.

3.3.2. Other scaling factors

The statistical significance of the result depended on the size of the scaling factor by which the infant MMR amplitude was multiplied. With the conservative value of 1.00 used above, the *p*-value for the interaction between Age Group and Distribution Type was 0.029 (Type-III sums of squares). With the scaling factors estimated above (section 2.7), namely 1.18 and 1.41, which express the idea that adult MMRs are bigger than infant MMRs, the *p*-value would be lowered to 0.018 and 0.010, respectively. With a scaling factor of 0.8172, which expresses the opposite assumption from that derived from the literature, namely that infants have a somewhat *larger* MMR amplitude than adults, the *p*-value would become exactly 0.05. We can conclude that for a large range of plausible scaling factors, the effect of distributional training is reliably smaller for adults than for infants.

4. Discussion

The current study provides the first evidence in a direct comparison that distributional training of speech sounds is less effective in adulthood, when new languages must be mastered, than in the first months of life, when infants start acquiring native speech sounds. Specifically, an earlier study (chapter II) showed that Dutch 2-to-3-month-old infants who are exposed to a bimodal distribution encompassing the Southern British English vowel contrast /æ/~/ɛ/, have a larger MMR amplitude, and thus supposedly discriminate the two test vowels [æ] and [ɛ] better, than infants exposed to a unimodal distribution. The current study demonstrates that this bimodal advantage is smaller (if at all present) in Dutch adults than in Dutch infants.

The presence of a bimodal advantage in Dutch adults is uncertain, because the difference in test vowel perception between bimodally and unimodally trained adults was not significant. It may be hypothesized that this non-significance was due to a ceiling effect (i.e., top discrimination) in both groups. After all, in the Netherlands English is a compulsory subject of study in middle school and high school, and it is also a language that can be listened to easily on television and

other media. However, such a ceiling effect is unlikely. The MMR amplitudes in both groups were rather small (with 95% confidence intervals close to zero), suggesting relatively poor discrimination (cf., the amplitudes in adults listed in Table III.1). Moreover, it has been shown that *despite* their experience with English, Dutch adults have trouble distinguishing the English vowels that were used in the current study (Schouten, 1975; Weber and Cutler, 2004; Broersma, 2005; Escudero et al., 2008). Similar results have also been obtained with other languages: for instance, adult native speakers of Spanish have considerable trouble in discriminating tokens of Dutch /ɑ/- and /a/, irrespective of the length of exposure to the Dutch language (Escudero and Wanrooij, 2010).

Notwithstanding our efforts to make a sound comparison of the effect of distributional training in infants and adults, it is clear that future research is needed to replicate our results and to confirm the feasibility of our approach. For confirming this feasibility, it will be particularly important to ascertain that infant MMRs truly reflect behavioural discrimination just as adult MMRs do (section 4.1). Relatedly, future research should aim to get a much more detailed understanding of the neural processes underlying infant and adult MMRs, so that differences between them can be explained better (section 4.2 presents a tentative rough explanation for the current results).

4.1. Measuring learning in adults and infants

The comparison of the effect of distributional training in adults versus infants was based on the MMR amplitudes. Our approach featured a minimization of methodological differences between testing infants and testing adults, and a normalization of the MMR amplitudes prior to statistical analysis in order to filter out possible residual differences between adult and infant MMRs. We presented a range of feasible normalization factors to account for the scarcity of information available for estimating such a factor in the literature, and to accommodate different possibilities of calculating such a factor.

Still, an important concern in our approach remains, which, notably, also applies to other outcome measures (such as looking times) in other paradigms. This concern is that the MMR may not reflect the same processes in adults as in infants. In particular, it is important to ascertain in future research whether the infant MMR indeed reflects behavioural discrimination. This has been assumed widely on the basis of evidence in adults (for a review see Näätänen et al., 2007), but has never been verified experimentally. In this context it is noteworthy that a discrepancy between behavioural and neurophysiological measures also exists in the literature on auditory thresholds. These thresholds appear to be much higher in infants than in adults in the behavioural literature (Werner and Gillenwater, 1990), but less so in research where auditory brainstem responses have been measured (Werner, 1996). It has been suggested that this discrepancy occurs due to the co-existence of a mature auditory system and an immature system necessary for making efficient use of this auditory system; the discrepancy can then arise when behavioural measures tap the immature system, while neurophysiological measures tap the mature system (Werner, 1996; Kushnerenko, 2003). To examine whether the infant MMR truly reflects behavioural discrimination, it seems therefore important to relate behavioural measures (such as high-amplitude sucking measures for the youngest infants, and eye-tracking measures for older infants) with MMR recordings.

4.2. Top-down influence on bottom-up learning

It is not certain whether the observed smaller effect of distributional training in adults than in infants is due to a weakened distributional learning mechanism, which is generally considered to represent a purely stimulus-driven, and thus *bottom-up* learning mechanism (Lacerda, 1995; Guenther and Gjaja, 1996), or rather to strengthened *top-down* processing, or perhaps to both of these factors. Top-down processing refers to the modulation of stimulus-driven neural activity in lower-level areas (e.g., the primary auditory cortex) by higher-level linguistic representations (e.g., phonological word forms). In 2-to-3-month olds such top-

down influence is lacking, because they do not have such higher-level representations yet (Jusczyk and Aslin, 1995; Hallé and De Boysson-Bardies, 1996; Jusczyk and Hohne, 1997; Fikkert, 2010; Bergelson and Swingley, 2012).

The first scenario (i.e., a weakened bottom-up learning mechanism) matches the decline of neural plasticity in the course of childhood, which has been related to an increase in the difficulty of "learning" with age (Kral et al., 2001), and which has been included in successful computer simulations of distributional learning (Lacerda, 1995; Guenther and Gjaja, 1996). The second scenario (i.e., strengthened top-down processing) is in accordance with the observation that distributional learning of human speech sounds can be measured in adult rats (Pons, 2006), thus suggesting that it is a low-level mechanism that remains in place after neural plasticity has reduced to adult levels. In this scenario, distributional learning can be observed in the rats, because, similarly to the 2-to-3-month olds, they do not have linguistic representations that could modulate lower-level neural activity.

A top-down influence of higher- on lower-level representations may already emerge after 4 to 5 months of life, as implied by research on the histological structure and development of the human auditory cortex (Moore and Guan, 2001; Moore, 2002; Moore and Linthicum, 2007). This research shows that the six cortical layers that children and adults have, are not present from birth but develop in the first year of life and become visible in post-mortem tissue around 4 to 5 months. Crucially, the division into multiple layers seems to be a prerequisite for top-down influence from higher- to lower-level cortical areas (Kral and Eggermont, 2007). A look at the functional organization of the layers may clarify this. Roughly, layer IV receives input from the thalamus and projects primarily to layers II and III ("supragranular layers"), which in turn target other parts of the cortex; layers V and VI ("infragranular layers") receive input from the supragranular layers and project to the thalamus and other subcortical structures (Bastos et al., 2012). This functional division suggests that in order to make top-down influence from higher- to lower-level representations possible, the infant cortex must first develop supragranular layers, so that incoming signals can reach

higher-level areas, where higher-level representations can be formed, and it must develop infragranular layers that receive top-down influence from these higher-level representations. At 4 to 5 months, rudimentary layering becomes visible in the tissue (Moore, 2002). Although it is possible that some top-down influence from higher-level to lower-level cortical areas occurs before this time via layer I, which is the only layer that is clearly visible in post-mortem tissue at birth (Moore and Guan, 2001; Moore, 2002; Moore and Linthicum, 2007), the infrastructure for canonical top-down cortical influence thus emerges just before infants begin to perceive speech sounds in a language-specific way, which is from 6 months of life (e.g., Werker and Tees, 1984; Kuhl et al., 1992; review in chapter II). This opens up the possibility that this language-specific speech perception relies on top-down influence of higher-level speech sound representations. At the same time, neural plasticity is still high at 6 months (e.g., Huttenlocher and Dabholkar, 1997), so that the possibility remains that the onset of language-specific speech perception (also) relies on bottom-up learning.

If in adults the distributional learning mechanism tends to be "suppressed" by top-down influence of higher-level native linguistic representations, the previous significant effects of adult distributional training might have been obtained because the experimental setting (entailing the absence of a natural language context) reduced the influence of these representations on perception. Alternatively, the way the training stimuli were presented may have attracted participants' attention to the differences between the speech sounds in the tested contrast. If this is true, the observed effects of distributional training would be due to "attention", which can be related to top-down processes in the brain (e.g., Posner and Petersen, 1990; Roelfsema, 2011) rather than to distributional training, which is a strictly bottom-up mechanism.

In this respect it is noteworthy that for the adult Spanish learners of the Dutch vowel contrast /ɑ/~/aː/ in Escudero et al. (2011) and in chapters V and VI, *enhanced* bimodal training in particular seemed effective. Here the acoustic difference between the minimum and the maximum value along the presented continuum of the training distribution was made larger. From previous research in

the second-language literature where other training paradigms than distributional training were used, it is known that widening the acoustic distance between presented stimuli in the training phase can draw participants' attention to the differences between these stimuli and improve subsequent discrimination and categorization performance (Jamieson and Morosan, 1986; Iverson et al., 2005; Kondaurova and Francis, 2010). Thus, it is possible that the previous observations of "distributional learning" in adults were related to attention instead.

All in all, distributional learning as a mechanism for learning speech sounds seems to be weaker later in life than in infancy. The reduced prominence in adulthood may be due to fainter bottom-up learning as well as to the presence (versus the virtual absence in newborns) of higher-level linguistic representations and of a cortical infrastructure that enables top-down influence of these representations on bottom-up learning.

Appendix to Chapter III

Further exploring the ERP method for adult distributional training

Karin Wanrooij, Titia L. Van Zuijen, and Paul Boersma.

The data in this appendix were presented in a poster presentation: "MMN declines after distributional vowel training", the Sixth Conference on Mismatch Negativity (MMN) and its Clinical and Scientific Application, New York, May 1-4, 2012.

1. Purpose of the appendix and summary of conclusions

Chapter III mentions that adult participants performed a pre-test, a training and a post-test. However, the pre-test data were not relevant for the study in chapter III, and were thus not reported (see section 2.1 in chapter III). This appendix shows the results of the pre-test, and relates them to those in the post-test. This report is added to this thesis, because the experiment in chapter III is the first to examine distributional learning with event-related potentials (ERPs) rather than with a behavioural method. Accordingly, it is possible that the new method somehow prevented us from obtaining a significant effect of distributional training (see chapter III). The purpose of the appendix is to explore this option, in addition to presenting the pre-test data.

The conclusions that can be drawn on the basis of the analyses in this appendix are as follows. First, when including the pre-test data in the analysis of the effect of distributional training, the results yield the same conclusion as that in chapter III: they show a non-significant effect of distributional training (Appendix, section 2). Second, there are no indications that the observed smaller MMR amplitude in the post-test than in the pre-test is an artefact of the ERP-method (Appendix, section 3).

2. Including the pre-test data in the analysis

The MMR amplitudes of the pre-test were calculated in the same way as those of the post-test (see section 2.6 in chapter III). The MMR latencies in the pre-test were 204 ms for Unimodal [æ], 177 ms for Unimodal [ɛ], 191 ms for Bimodal [æ] and 188 ms for Bimodal [ɛ]. Table III.3 lists the mean MMR amplitudes in the pre-test for each of these four groups. The corresponding post-test data are reported in chapter III (Table III.2 in section 3.1.3 of chapter III).

In order to test whether there is a difference between the unimodally and bimodally trained participants, while controlling for differences in the presented standard, an analysis of variance (ANOVA) can be performed with the MMR amplitude as the dependent variable, Distribution Type (uni- vs. bimodal) and

Table III.3. Mean pre-test MMR amplitudes (in (μV) at Fz, averaged across participants per Distribution Type presented in the training (uni- vs. bimodal) and per Standard Vowel used in the test ([æ] vs. [ɛ]). With within-group standard deviations (SD) and 95% confidence intervals (CI).

Distribution Type	Standard Vowel	N	Mean	SD	CI
Unimodal	**[ɛ]**	10	–1.62	1.18	–2.47 ~ –0.78
	[æ]	9	–2.06	1.50	–3.21 ~ –0.90
Bimodal	**[ɛ]**	11	–1.56	1.23	–2.38 ~ –0.73
	[æ]	9	–1.35	1.38	–2.41 ~ –0.30

Standard Vowel ([æ] vs. [ɛ]) as between-subject factors, and Test (pre- vs. post-test) as within-subject factor. Thus, this analysis is equivalent to that in chapter III, except for the inclusion of the pre-test data: the effect of distributional training is now reported on the basis of the change in discrimination performance (post-test versus pre-test), rather than on the basis of the post-test data alone.

This ANOVA shows a main effect of Test (mean difference = +0.72 μV, 95% confidence interval, henceforth CI, = +0.31 ~ +1.13μV, $F[1,35] = 12.63$, $p = 0.001$): participants had a smaller, i.e., a less negative MMR after (–0.93μV, CI = –1.33 ~ –0.54 μV) than before distributional training (–1.65μV, CI = –2.08 ~ –1.22μV), which implies a decline in participants' ability to discriminate the two test vowels after training (Aaltonen et al., 1997; Näätänen et al., 1997). Note that both the pre-test and the post-test MMR amplitudes have confidence intervals below zero, indicating that overall participants discriminated the test vowels before and after training.

Other main and interaction effects in the ANOVA are not significant. Crucially, the interaction between Distribution Type and Test is not significant ($p = 0.83$). Thus, the amplitude data do not provide evidence for distributional learning of English /æ/~/ɛ/ in Dutch adult learners. Because the main effect of Standard Vowel ($p = 0.48$) is not significant, the amplitude data do not show proof of a

perceptual asymmetry making [æ] and [ɛ] easier to discriminate when either one or the other vowel is the standard in the oddball paradigm (see section 2.1 in chapter III). In sum, these results based on both pre- and post-test data yield the same conclusions (i.e., no clear evidence for distributional learning and perceptual asymmetry) as the results based on only the post-test data presented in chapter III. In addition, they show that participants' neural discrimination of the test vowels declined in the post-test as compared to the pre-test. It was important to explore whether this was an artefact of the method (see the next section).

3. Exploring the decline in MMR amplitude

The overall smaller MMR (and thus supposedly worse discrimination of the test vowels) after than before training may be expected for unimodally trained listeners (for whom the unimodal training may reinforce their native Dutch experience that the two test stimuli are exemplars of the same vowel), but is contrary to expectation for all participants combined. Because the current study is the first to test distributional learning in adults with ERPs rather than behavioural measures, it was necessary to explore the possibility that the overall decline in MMR amplitude was an artefact of the ERP-method, i.e., that it occurred during the pre- and post-tests (where the ERPs were measured) rather than during the training. Previous research has shown that the MMR may decline in the course of the experiment (McGee et al., 2001). Therefore, we examined the development of MMR amplitudes in each test (i.e., pre- and post-test) over time. To this end, the correlation between time and the MMR amplitude was calculated in three steps. First, for each participant two tables were created, one for the pre- and one for the post-test, with MMR amplitudes in the course of each test. Each MMR amplitude was calculated as the difference between a non-rejected deviant response and three preceding non-rejected standard responses. (Only standards that came after the preceding deviant were included). Second, the correlation between time and MMR amplitude was calculated for each participant and each test. This resulted for each test in 39 correlations for 39 participants. Finally, for each test it was examined if

the mean correlation (i.e., across participants) differed from zero in a one-sample *t*-test.

The correlation between time and MMR amplitude did not differ from zero significantly neither in the pre-test nor in the post-test (pre-test: $r = 0.006$, CI = – 0.023 ~ + 0.035, $t = 0.43$, one-tailed $p = 0.34$; post-test: $r = 0.017$, CI = – 0.012 ~ + 0.047, $t = 1.19$, one-tailed $p = 0.12$). Thus, the MMR amplitude did not change significantly during either test. This is also visible in Figure III.5, which shows the correlations (each dot is a correlation for a single participant) between the MMR amplitude and test time in the pre-test (left) and the post-test (right) for unimodally (grey) and bimodally (black) trained participants. The lines reflect the mean correlations across participants, which are virtually zero.

In sum, the MMR amplitude is not significantly changing over time in the pre- and post-tests. Thus, there is no clear evidence that the smaller MMR amplitude in the post-test than in the pre-test is an artefact of the tests, i.e., of the method in which we measure ERPs rather than behaviour.

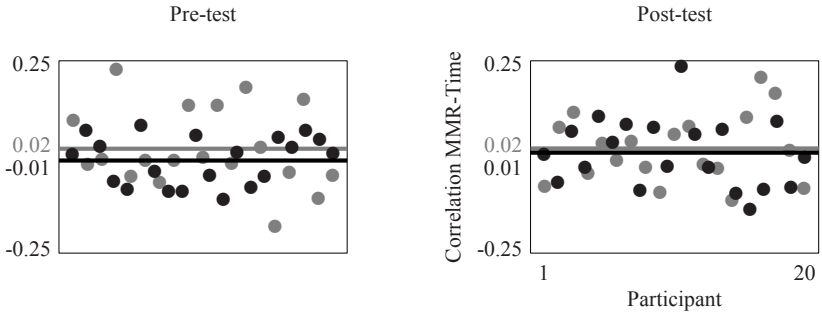


Figure III.5. The MMR amplitude is not significantly changing over time, in the pre-test (left) and post-test (right), for unimodally (grey) and bimodally (black) trained participants. Each dot represents the correlation between the MMR amplitude and test time for a single participant. Mean correlations (lines) in each group (unimodal and bimodal) and test are virtually zero.

Chapter IV

Is distributional vowel training effective for Dutch adults? A behavioural control study

Karin Wanrooij, Johanna de Vos and Paul Boersma

(to be submitted)

Abstract

Distributional vowel training for adults has been reported as "effective" for Spanish and Bulgarian learners of Dutch vowels. Chapter III examined whether such training can also be useful for Dutch adult learners of the English vowel contrast /æ/~/ɛ/. A significant effect of the training was not found. To exclude the possibility that the absence of significance was related to the new method (using the measurement of event-related potentials, or ERPs, instead of behaviour), the current study was conducted. Specifically, it was tested again whether distributional training of English /æ/~/ɛ/ is effective for Dutch adult learners. However, this time behaviour was measured to assess the effect of distributional training, in a design identical to that used previously for the Spanish learners of Dutch. The results failed to provide clear support for distributional learning, thus "replicating" the outcomes of the ERP study. The hypothesis is put forward that a relatively large number of native vowels (as in Dutch versus Bulgarian and Spanish) could reduce the effectiveness of distributional vowel training for adults.

1. Introduction

Several previous studies report an effect of distributional vowel training in adults: one study observes such an effect in Bulgarian adult learners of the Dutch vowel contrasts /ɑ/~/aː/ and /ɪ/~/i/ (Gulian et al., 2007), and three studies observe it in Spanish adult learners of the Dutch vowel contrast /ɑ/~/aː/ (Escudero et al., 2011; chapters V and VI). Chapter III examined whether distributional training can also be effective for *Dutch* adult learners of the Southern British English (SBE) vowel contrast /æ/~/ɛ/. For this, a novel method was used, namely the measurement of the mismatch negativity (MMN), a brain response that can be computed from event-related potentials (ERPs). This neurophysiological method, which had yielded a significant effect of distributional training in Dutch 2-to-3-month old infants (chapter II), differed from the behavioural methods employed in the just-mentioned earlier experiments. Chapter III obtained a non-significant effect of distributional training in Dutch adults.

Even though this null result cannot be interpreted as the absence of distributional learning, it was important to investigate the possibility that the new method caused the insignificance of the effect. For this purpose, we conducted the present experiment. Specifically, we examined Dutch adults' capacity for distributional learning of the English contrast /æ/~/ɛ/, just as in chapter III, but in contrast to chapter III we used the same behavioural method as in the previous studies on distributional vowel learning in Spanish adults (Escudero et al., 2011; chapters V and VI; section 2 in the current chapter). If we found an effect of distributional training, we would attribute the lack of clear positive evidence for it in chapter III to the ERP method. If, just as in chapter III, an effect of distributional training did not surface, the ERP method would not necessarily be inadequate for measuring distributional learning in adults.

2. Method

The method was identical to that used in previous research on Spanish adults' distributional learning of the Dutch contrast /ɑ/~/aː/ (Escudero et al., 2011; chapters V and VI), except for the use of a different contrast (SBE /æ/~/ɛ/) and of participants sampled from a different population (Dutch listeners), and except for the inclusion of two rather than three groups of participants (section 2.1).

2.1. Design

The design consisted of a pre-test, a training phase and a post-test. During the training phase (sections 2.3 and 2.4.1), participants listened to either an enhanced bimodal distribution of /æ/~/ɛ/ (the "Bimodal group"), or to instrumental classical music (the "Music group"). The stimuli in an enhanced bimodal distribution have acoustic values (here: the values for the first and second formants) that are wider apart (i.e., have a larger range and a larger difference between the two means) than those in a non-enhanced bimodal distribution. In this way the difference between the two vowel categories in the distribution becomes more pronounced, and presumably easier to perceive (Kuhl et al., 1997; Liu et al., 2003). In contrast to Escudero et al. (2011) and chapter V, a group presented with a non-enhanced bimodal distribution was not included, because in these two studies clear evidence for distributional learning in such a group could not be found, and thus only exposure to an enhanced bimodal distribution was significantly more effective in improving participants' classification accuracy than exposure to music.[1] Training Type (enhanced bimodal, henceforth simply "bimodal", vs. music) was included as the between-subject factor in the statistical analysis.

During the pre- and post-tests (sections 2.3 and 2.4.2), classification accuracy (in percent correct) of multiple tokens of [æ] and [ɛ] was assessed in a forced-choice XAB-task (section 2.3). Classification rather than discrimination (as

1 Notice that in chapter III the range of F1 and F2 values was similarly "enhanced". Specifically, it encompassed 4.1 ERB for F1 and 2.7 ERB for F2, which is even wider than the 2.9 ERB for F1 and F2 in Escudero et al., 2011, and chapter V.

measured in chapter III) was promoted by the use of many different natural tokens (the Xs; section 2.4.2) and the use of a relatively long inter-stimulus interval between the three stimuli in each trial (i.e., 1.2 seconds; Van Hessen and Schouten, 1999; Werker and Logan, 1985). Test (pre-test vs. post-test) was included as a within-subject factor in the statistical analysis.

To sum up, the analysis of the effect of distributional training on adults' classification performance had the percentage of correctly classified vowels as the dependent variable, Training Type (bimodal vs. music) as a between-subject factor and Test as a within-subject factor.

2.2. Participants

The criteria for participation were the same as in chapter III: participants were native speakers of Dutch who were raised monolingually, had not lived abroad during childhood, and had never passed more than four weeks in countries where English is the national language. The 100 participants were assigned semi-randomly (i.e., sex was controlled for, as in chapter III: 18 men and 32 women in each group) to either the Bimodal group (mean age = 22.3 years; age range = 18–30 years) or the Music group (mean age = 22.3 years; age range = 18–28 years).

2.3. Procedure

In each *test* (i.e., the pre- and post-test), participants heard 80 trials, each containing three stimuli, i.e., an X, A and B stimulus. Participants had to classify each X stimulus as either A or B. Specifically, they were asked to indicate whether the first vowel (X) was more similar to the second vowel (A) or to the third vowel (B), by clicking on "1" (for A) or "2" (for B) on a computer screen after hearing the three vowels. The next trial would only begin after clicking on a response option. The possibility to take a short break was available every 20 trials. The presentation of the A and B stimuli was counterbalanced across trials and trial order was randomized per participant.

Before the *training*, participants in the Bimodal group were instructed to listen to the vowels carefully, because they would perform a second task similar to the first one after the training. Participants in the Music group were told that they would listen to classical music and could relax, after which they would perform another task similar to the first one.

Total experimental time was about 16 minutes (i.e., precisely 1.9 minutes for the training and circa 7 minutes for each test).

2.4. Stimuli

2.4.1. Training

Just as in Escudero et al. (2011) and in chapter V, the training distribution was discontinuous, i.e., it consisted of eight acoustically different vowels that were each repeated a certain number of times in order to create the desired distribution. (For a discussion of continuous distributions as used in chapter III and discontinuous distributions as used in the current experiment, see chapter VI, in which outcomes were not influenced significantly by the choice of one or the other kind of distribution). The vowels were created with the Klatt synthesizer in Praat (Boersma and Weenink, version 2012). The manipulated acoustic properties were the first and second formants (F1 and F2). Table IV.1 lists the F1 and F2 values (in ERB, Equivalent Rectangular Bandwidth), and the frequency of presentation for each of the eight stimuli.

Table IV.1. The F1 and F2 values (in ERB) and the frequency of presentation of the eight training stimuli.

Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Frequency	8	32	16	8	8	16	32	8
F1	9.93	10.37	10.81	11.25	11.69	12.13	12.57	13.01
F2	20.74	20.44	20.13	19.83	19.53	19.23	18.92	18.62

The *bimodality* of the distribution is visible in the higher frequency of presentation of stimuli 2 and 7 (i.e., each 32 times) than of intermediate stimuli along the F1 and F2 ranges (i.e., 16 times for stimuli 3 and 6, and 8 times for stimuli 4 and 5). The respective frequencies of stimuli 1 through 8 add up to a total of 128 presentations.

The *enhancement* of the distribution is reflected in the F1 and F2 values. These enhanced values were calculated in the same way as in Escudero et al. (2011) and chapter V, apart from adaptations to the different vowel contrast. First, the mean F1 and F2 values of /ɛ/ and /æ/, and the standard deviation for the F1 and F2 values were determined on the basis of values reported in Hawkins and Midgley (2005; These values were also used in chapter III. For details see chapter II). Then the edges of the F1 range were calculated by subtracting the standard deviation of F1 from the mean F1 value of /ɛ/ (stimulus 1) and adding it to the mean F1 value of /æ/ (stimulus 8). Similarly, the edges of the F2 range were computed by adding the standard deviation of F2 to the mean F2 value of /ɛ/ (stimulus 1) and subtracting it from the mean F2 value of /æ/ (stimulus 8). The F1 and F2 values of the intermediate stimuli along the F1 and F2 ranges (i.e., stimuli 2 through 7) were calculated by linear extrapolation, where each step between consecutive stimuli was roughly equal on the psychoacoustic ERB scale. The resulting step sizes (i.e., 0.44 ERB for F1 and 0.30 ERB for F2) were comparable to the step sizes in Escudero et al. (2011) and chapter V (i.e., 0.4 ERB for F1 and F2).

Further, each stimulus was filtered with eight additional formants (i.e., the third through tenth formant), which had the same values as in chapter III (F3=2400 Hz, F4=3400 Hz, F5=4050 Hz, F6 through F10: previous formant plus 1000 Hz). Each stimulus had a duration of 140 ms and a fundamental frequency (F0) that fell from 150 Hz to 100 Hz. The inter-stimulus interval was 750 ms. Total training time was thus less than 2 minutes (=128 stimuli *[140 + 750] ms).

2.4.2. Test

The 80 X stimuli (40 for /æ/ and 40 for /ɛ/) in the pre- and post-tests were unique natural tokens of English /æ/ and /ɛ/ produced by six female and five male native speakers of Standard SBE. Two productions of /æ/ and /ɛ/ each were provided by Daniel Williams (Williams and Escudero, 2014). Another two productions of /æ/ and /ɛ/ each were extracted from a subset of stimuli reported in Escudero et al. (2012).

Most of the tokens were extracted from a /h-V-d/ context (*head / had*) or a /f-V-f/ context (*fef / faf*). To add additional variation, some vowels were extracted from other contexts, namely /s-V-s/, /b-V-s/, /h-V-s/, /m-V-s/ and /t-V-s/. Table IV.2 lists the average F0, F1, F2 and duration of /æ/ and /ɛ/ for the female and male speakers separately.

Table IV.2. Mean F1, F2, F0 (in ERB) and duration (in ms) of the X stimuli in the XAB-test. Standard deviations between tokens are given between parentheses.

	F1		**F2**		**F0**		**Duration**	
	Females	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**	**Males**
/æ/	14.96	13.26	19.18	18.20	5.34	3.41	123.88	113.57
	(0.84)	(0.45)	(0.69)	(0.73)	(0.41)	(0.39)	(26.50)	(23.83)
/ɛ/	12.38	11.35	20.97	19.35	5.51	3.43	118.20	97.40
	(1.01)	(0.57)	(0.83)	(0.69)	(0.49)	(0.39)	(23.00)	(26.69)

In contrast to the X stimuli, the response options A and B were synthetic tokens created in Praat (Boersma and Weenink, 2012). For [ɛ], F1 and F2 were 10.95 ERB and 20.04 ERB respectively. For [æ], F1 and F2 were 11.99 ERB and 19.32 ERB respectively. Just as the training stimuli, both response options had eight additional formants with the same values as those for the training stimuli (section 2.4.1), a duration of 140 ms and an F0 that fell from 150 Hz to 100 Hz.

3. Results

For each participant the percentage of correctly classified vowels was computed. Table IV.3 shows the mean pre-test and post-test accuracy percentages, and the difference scores (= the post-test – pre-test accuracy percentage) for the Bimodal and Music groups separately.

Table IV.3. Pre- and post-test accuracy percentages, and the difference (= post-test – pre-test accuracy percentage), for the Bimodal and Music groups. Standard deviations between participants in each group are given between parentheses.

Group	Pre-test	Post-test	Difference
Bimodal	64.98 (12.03)	68.55 (14.29)	3.58 (7.51)
Music	64.33 (9.97)	71.50 (12.53)	7.18 (7.68)

A repeated-measures ANOVA with the accuracy percentage as the dependent variable, Training Type (bimodal vs. music) as between-subject factor and Test (pre- vs. post-test) as within-subject factor showed a significant main effect of Test (mean difference = +5.38%, CI = +3.87 ~ +6.88%, $F[1,98] = 50.03$, $p < 0.001$): the accuracy percentage was higher after (70.03%, CI = +67.36 ~ +72.69%) than before (64.65%, CI = +62.46 ~ +66.84) the training phase. The main effect of Training Type was not significant ($p = 0.62$). Thus, the Bimodal group did not score significantly higher or lower than the Music group across the two tests. Crucially, the interaction between Training Type and Test was significant ($F[1,98] = 5.61$, $p = 0.02$). This indicates that the two groups did not improve equally after as compared to before the training phase. Table IV.3 illustrates, however, that the Bimodal group did not improve more than the Music group, as was the expectation, but rather the other way around.[2] Since the Music

2 The fact that the control group performed better after the training phase than the experimental group is not without precedent in adult distributional learning experiments: Hayes-Harb (2007) also obtained a better discrimination (as expressed in an accuracy percentage) after the training phase for a group that received no training (32.6%) than for a group that received a bimodal training (22.7%).

group did not receive distributional training, this larger improvement cannot be attributed to distributional learning.

4. Discussion

4.1. No clear evidence for distributional vowel learning in Dutch adults

In the current study, we did not find a straightforward effect of distributional training in Dutch adult learners of the SBE vowel contrast /æ/~/ɛ/ when repeating a behavioural experiment that had shown such an effect in Spanish adult learners of Dutch /ɑ/~/aː/ (Escudero et al., 2011; chapters V and VI). The present result "replicates" the outcome reported in chapter III, where similarly a non-significant effect of distributional training in Dutch adult learners of /æ/~/ɛ/ was obtained with a very different method, namely the measurement of the MMN. Thus, the present outcome prevents us from rejecting this MMN method as unsuitable for measuring distributional training effects in adults.

4.2. A possible influence of the native vowel space structure

The fact that we did not obtain clear evidence for distributional vowel learning in Dutch adults does not necessarily mean that distributional training is not effective for them. However, the pattern that arises from the current experiment in combination with chapter III and all previous publications on adult distributional vowel training (Gulian et al., 2007; Escudero et al., 2011; chapters V and VI), hints at such ineffectiveness. This pattern consists of a significant effect of distributional training in Bulgarian adults once (Gulian et al., 2007) and in Spanish adults three times (Escudero et al., 2011; chapters V and VI), and the absence of a straightforward effect in Dutch adults two times (chapter III and the current study). These combined results suggest the possibility that the effectiveness of distributional training is related to the structure of the native vowel space. (Here the vowel space is a two-dimensional space defined by F1 and F2, which were also the manipulated properties in the training distributions in all studies on adult

distributional vowel learning, i.e. in Gulian et al., 2007, Escudero et al., 2011, chapters III, V and VI, and the current study). Specifically, in a roughly equally large perceptual vowel space (Meunier et al., 2003), Dutch has many more vowels (15; Adank et al., 2004) than Spanish (5; Bradlow, 1995) or Bulgarian (6; Klagstad, 1958). Consequently, a training distribution with a specific range of F1 and F2 values[3] is likely to overlap with more native vowels in Dutch than in Spanish and Bulgarian. Also, a possible overlap could only occur at one side of the training distribution for the Spanish and Bulgarian listeners, and at both sides of the distribution for the Dutch listeners. Specifically, the Spanish listeners were exposed to distributions encompassing the Dutch contrast /ɑ/~/aː/ (Escudero et al., 2011; chapters V and VI). The vowels in this contrast are perceived by these listeners as the Spanish /a/, and this vowel lies at a corner of the Spanish vowel space (i.e., Spanish does not have vowels with a higher mean F1 value). The Bulgarian listeners were presented with distributions encompassing the Dutch vowels contrasts /ɑ/~/aː/ and /ɪ/~/i/ (Gulian et al., 2007). The vowels in these contrasts are perceived by these listeners as Bulgarian /a/ and /i/ respectively, and these vowels lie at corners of the Bulgarian vowel space (i.e., Bulgarian does not have vowels with a higher mean F1 than /a/ and with a higher mean F2 than /i/). The Dutch adults listened to distributions encompassing the SBE contrast /æ/~/ɛ/ (chapter III and the current study). The vowels in this contrast do not lie in a corner of the Dutch vowel space (i.e., Dutch has front vowels with a lower mean F1 and higher mean F2 than /ɛ/, and the vowel /aː/ with a higher mean F1 and lower mean F2 than /æ/). Thus, whereas Spanish and Bulgarian listeners had to create a new non-native boundary within a native category, Dutch listeners probably had to shift the native boundary between Dutch /ɛ/ and /aː/. It is conceivable that a distributional training of only a few minutes is more effective for achieving the former than the latter goal.

3 The ranges of F1 and F2 values presented to the Dutch listeners in the current study (namely 3.1 ERB for F1 and 2.1 ERB for F2; see section 2.4.1) were similar to those presented to the Spanish listeners in Escudero et al. (2011), and in chapters V and VI. In these studies the F1 range spanned 2.9 ERB for F1 and F2.

At first sight, the hypothesis that distributional vowel training is less effective for native speakers with relatively many native vowels than for those with fewer native vowels is contrary to the observation by Iverson and Evans (2009): in this study identification accuracy of English vowels improved more and faster for German listeners (18 vowels) than for Spanish listeners. However, it is well-known that differences in the method can exert pivotal influence on the outcomes (e.g., Strange, 1992; Díaz et al., 2012), and such differences could very well account for the seemingly contradicting conclusions. For instance, the training in Iverson and Evans's study extended over 5 sessions within two weeks (as opposed to one distributional training session lasting only a few minutes), participants listened to words (as opposed to bare vowels) and they received feedback (as opposed to only exposure).

Importantly, the training in Iverson and Evans's study and distributional training may also target different mechanisms. Iverson and Evans attribute the improved identification of English vowels by German and Spanish listeners to a facilitated application of (first- and second-language) category knowledge to the presented stimuli, and not to changes in the category representations (described as changes in cue weightings) themselves. In contrast, in an earlier study on *distributional* training (chapter V) it was observed that listeners who improved *had* changed their cue weightings. Unfortunately, it was not tested if these listeners retained these cue weightings over a longer time span. (In Iverson and Evans's study the improved performance was retained in tests several months later). Further, the application of existing linguistic category knowledge to the auditory processing of presented stimuli (i.e., the proposed mechanism in Iverson and Evans's study) implies a top-down neural mechanism, where higher-level linguistic representations in the brain affect lower-level processing. In contrast, distributional learning can be viewed as an exclusively bottom-up, stimulus-driven mechanism (Lacerda, 1995; Guenther and Gjaja, 1996). A final indication that the training in Iverson and Evans's study and the training in distributional learning studies target different mechanisms is that in Iverson and Evans's study feedback seemed to contribute to listeners' improvement, whereas feedback has been shown to hamper

distributional learning (Gulian et al., 2007). In sum, the results reported in Iverson and Evans do not invalidate the hypothesis that the effectiveness of adult *distributional* vowel training is smaller for native speakers of languages with relatively many native vowels, as compared to native speakers of languages with relatively few vowels.

5. Conclusion

The current study replicates the absence of a clear effect of distributional training in Dutch adult learners of the SBE vowel contrast /æ/~/ɛ/ obtained earlier by measuring brain responses, with a behavioural method. In earlier research, the same behavioural method yielded a significant effect of distributional training in Spanish adult learners of Dutch /ɑ/~/aː/ three times. This pattern suggests a difference in the usefulness of distributional vowel training between native speakers of languages with relatively many vowels (such as Dutch) and native speakers of languages with relatively few vowels (such as Spanish). Future research should examine this hypothesis.

Chapter V

What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning

Karin Wanrooij, Paola Escudero, and Maartje E.J. Raijmakers

Journal of Phonetics 2013, 41(5), 307-319, doi: 10.1016/j.wocn.2013.03.005

Abstract

This study first confirms the previous finding that Spanish learners improve their perception of a difficult Dutch vowel contrast through listening to a frequency distribution of the vowels involved in the contrast, a technique also known as *distributional training*. Secondly, it is demonstrated that learners' initial use of acoustic cues influences their performance after distributional training. To that end, types of unique *listening strategies*, i.e., specific ways of using acoustic cues in vowel perception, are identified using *latent class regression models*. The results before training show a split between "low performers", who did not use the two most important cues to the Dutch vowel contrast, namely the first and second vowel formants, and "high performers", who did. Distributional training diversified the strategies and influenced the two types of listeners differently. Crucially, not only did it bootstrap the use of cues present in the training stimuli but also the use of an untrained cue, namely vowel duration. We discuss the implications of our findings for the general field of distributional learning, and compare our listening strategies to the developmental stages that have been proposed for the acquisition of second-language vowels in Spanish learners.

1. Introduction

Learning speech sounds on the basis of frequency distributions is commonly known as "distributional learning" (Gulian et al., 2007; Maye and Gerken, 2000, 2001; Maye et al., 2008). Distributional learning is considered to be the main mechanism that underlies the acquisition of speech sounds in the first year of life, when infants' sensitivity to native speech sound contrasts (which occur frequently in the infant's environment) increases (e.g., Cheour et al., 1998b), while that to non-native speech sound contrasts (which occur infrequently) declines (e.g., Werker and Tees, 1984/2002). Since infants' vocabularies are non-existent or small in the first months of life, another way of learning speech sounds, namely from noticing the difference in meaning between words whose forms differ in only one speech sound, cannot play a dominant role yet (Maye et al., 2002; Stager and Werker, 1997). The probable existence of distributional learning as a real mechanism for learning speech sounds has been supported by computer simulations (Guenther and Gjaja, 1996; Lacerda, 1995) and also by observations in the lab, not only for infants (Cristià et al., 2011; Maye et al., 2002; Maye et al., 2008; Yoshida et al., 2010), but also for adults (Escudero et al., 2011; Gulian et al., 2007; Hayes-Harb, 2007; Maye and Gerken, 2000, 2001).

In distributional learning experiments in the lab, listeners hear a randomly presented series of stimuli that vary in steps along a continuous dimension. Crucially, each stimulus is presented with a certain frequency, such that some stimuli appear more often than others. In this way listeners hear a *distribution* of speech sounds. Two groups of listeners usually participate; one presented with a bimodal and another with a unimodal distribution of speech sounds (e.g., Gulian et al., 2007; Hayes-Harb, 2007; Maye and Gerken, 2000, 2001; Maye et al., 2008). In the former distribution, stimuli with properties near the two endpoints of the acoustic continuum are presented most often, while in the latter, stimuli with properties near the middle of the acoustic continuum are most frequent. After the training phase, both groups of listeners are tested on their ability to discriminate the same two stimuli, which had occurred equally often in both trained distributions. If there is an effect of distributional learning, better discrimination is expected after

exposure to bimodal than unimodal distributions. This is because exposure to a bimodal distribution induces the perception of the two test stimuli as exemplars of two different speech sound categories, while listening to a unimodal distribution leads to hearing the same two test stimuli as exemplars of a single speech sound category (e.g., Gulian et al., 2007; Hayes-Harb, 2007; Maye and Gerken, 2000, 2001; Maye et al., 2008).

Although the original distributional learning studies (Maye and Gerken, 2000, 2001; Maye et al., 2002) aimed at demonstrating that this mechanism underlies the learning of phonetic categories, recent studies have exploited the technique to train difficult non-native speech sound contrasts. Gulian et al. (2007) exposed native Bulgarian speakers to bimodal distributions of the Dutch vowel contrasts /ɑ/~/aː/ and /ɪ/~/i/, which these listeners tend to perceive as the single Bulgarian vowels /a/ and /i/ respectively. After a training phase of only 5 minutes per vowel contrast, listeners exposed to a bimodal distribution classified the vowels in each contrast more accurately than those exposed to a unimodal distribution. More recently, Escudero et al. (2011) presented Spanish-speaking learners of Dutch with bimodal distributions of Dutch /ɑ/~/aː/. In natural speech, these Dutch vowels differ both in their spectral (/aː/ has higher first and second formants) and durational (/aː/ is longer) properties (Adank et al., 2004; Pols et al., 1973). When classifying the vowels, Spanish learners of Dutch tend to rely on the durational differences, while Dutch natives use spectral differences primarily (Escudero et al., 2009; Giezen et al., 2010). To direct Spanish listeners' attention to the dimension that is most important to native Dutch listeners, Escudero et al.'s (2011) training vowels differed in spectral properties only. Further, rather than comparing the effect of bimodal and unimodal training, the authors presented listeners with either a natural bimodal (hence "bimodal") or an enhanced bimodal (hence "enhanced") distribution. In the former distribution, the endpoint stimuli had average values for the first and second formants (Pols et al., 1973; section 2.2.2 of the present manuscript), while the stimuli in the latter had an enlarged spectral difference, i.e., the endpoint tokens had exaggerated properties similar to those of infant-directed (Burnham et al., 2002; Kuhl et al., 1997; Sundberg, 2001; Sundberg and Lacerda,

1999) and foreigner-directed speech (Uther et al., 2007). In this way, the acoustic difference between training stimuli in the enhanced distribution was more pronounced and presumably easier to perceive than the difference between training stimuli in the bimodal distribution (section 2.2.2 and Table V.3). As expected from previous studies that suggest facilitation of speech discrimination with enhanced differences between stimuli (e.g., Kuhl et al., 1997; Liu et al., 2003), the results showed that vowel classification accuracy (as measured in pre- and post-tests; section 2.2.1) increased after enhanced training, and that this improvement was larger than in the control condition. (Improvement after bimodal training was not larger than in the control condition). The authors concluded that difficult non-native contrasts can be trained effectively with a distributional learning paradigm, which requires only a few minutes of stimuli exposure and no feedback.

In the present study we first aimed to show again Escudero et al.'s (2011) distributional training results in adult second-language (L2) learners (section 2.2). To this end, we exposed two new groups of Spanish learners of Dutch to the same bimodal and enhanced distributions of the Dutch vowel contrast /ɑ/~/aː/. Their classification performance of multiple natural realizations of the two vowels was evaluated in pre- and post-tests, which were identical to those used in Escudero et al. The second and primary aim of the present study was to probe the causes of the increase in vowel classification accuracy after enhanced training, found in Escudero et al. (2011). Specifically, we examined whether distributional training could promote the use of the main acoustic cues for distinguishing the Dutch vowels, i.e., their first and second formants, and whether an enhanced distribution is more effective in this respect than a bimodal distribution, for which the difference in formant values between the training vowels is smaller. To investigate listeners' use of acoustic cues, we employed a statistical technique called *latent class regression analysis* (Huang and Bandeen-Roche, 2004; sections 1.2 and 2.3 of the present manuscript). With this technique one can identify classes of listeners, with each class representing a subgroup with a unique *listening strategy*, i.e., a specific way of using acoustic cues. This approach thus allowed us to examine the

relationship between initial listening strategies, improvement after training, and exposure to bimodal versus enhanced distributions.

1.1. Theoretical background and definition of listening strategies

Recall that, as mentioned in section 1, we use the term *listening strategy* to refer to a specific use of acoustic cues in the perception of speech sound contrasts (also known as *acoustic cue-weighting*). Accordingly, we do *not* address general learning strategies (as in e.g., Oxford et al., 1988), or individual differences in L2 speech sound perception that may result from a variety of other factors such as the length of residence in an L2 country (e.g., Flege et al., 1997) or the type of task presented to the listeners (Díaz et al., 2012).

Extensive research has demonstrated cross-linguistic differences in acoustic cue-weighting (e.g., Bohn and Flege, 1990; Escudero and Boersma, 2004; Escudero et al., 2009; Iverson et al., 2005; Iverson et al., 2003; Morrison, 2008, 2009). These studies show that when discriminating speech sounds, native and non-native listeners may favour different acoustic cues. For instance, the well-known observation that Japanese adults have trouble perceiving English /r/ and /l/ as two different speech sounds (e.g., Goto, 1971; Iverson et al., 2003; Miyawaki et al., 1975; Yamada, 1995) has been attributed to the Japanese focus on the irrelevant second formant rather than the relevant third formant, which is used by English natives (Iverson et al., 2003). Similarly and as mentioned above, Dutch natives favour spectral cues when distinguishing between Dutch /ɑ/ and /aː/, while Spanish learners of Dutch tend to resort to duration (Escudero et al., 2009).

In addition to reporting group differences, previous research reveals substantial individual differences in the use of acoustic cues (e.g., Chandrasekaran et al., 2010; Escudero and Boersma, 2004; Escudero et al., 2009; Morrison, 2008, 2009). For instance, Escudero et al. (2009) report that over a third of their Spanish learners of Dutch relied more on spectral cues than on duration when categorizing Dutch /ɑ/ and /aː/. Accordingly, it is likely that, in the current study, not all Spanish learners of Dutch will solely focus on duration before training.

Individual differences in cue-weighting are not commonly addressed in theories and models on L2 speech perception, which tend to focus on general group differences. That is, well-known theoretical accounts of non-native speech perception explain the *general* difficulty that Spanish listeners have with discriminating and classifying certain L2 vowels. For instance, for Dutch /ɑ/ and /aː/, both Flege's Speech Learning Model (SLM; Flege, 1995, 2002, 2003; Flege and MacKay, 2004) and Best's Perceptual Assimilation Model (PAM; Best, 1994) posit that the difficulty arises from the similarity of both Dutch vowels to a single Spanish vowel, namely /a/. Mayr and Escudero (2010) present an extensive review of these and other explanations for listeners' difficulties in perceiving non-native speech sounds.

In the current study, where we expect to find differences in the perceptual patterns of Spanish learners of Dutch vowels, we will compare our results to Escudero's Second Language Linguistic Perception (L2LP) model (Escudero, 2005; see also Escudero, 2000), which in contrast to the models mentioned above addresses the possibility that L2 speech sound perception may develop in steps, and that adult listeners may differ in both their perception of L2 speech sounds (see the individual differences in cue use mentioned above) and the way in which this perception develops. Escudero (2000, 2005) explicitly posits successive developmental stages with differential cue weightings for Spanish listeners who learn the English vowels /i/ (as in "beat") and /ɪ/ (as in "bit"). Specifically, Escudero proposes the following stages: (0) no distinction between the two vowels, (1) use of duration to distinguish them, (2) a main reliance on duration with a subtle use of spectral cues, and (3) a main focus on spectral cues with an additional use of duration, which is in accordance with native speaker performance. Morrison (2008) suggests an extra stage between 0 and 1. In this stage ½, listeners use spectral cues to classify the vowels as "good" or "bad" examples of Spanish /i/, while they also start using durational differences, which are not distinctive in Spanish. Given Spanish learners' difficulty to perceive spectral differences between both English /i/ and /ɪ/ and Dutch /ɑ/ and /aː/, and the tendency, in both cases, to resort to the use of duration (for English /i/~/ɪ/: Escudero, 2000, 2005;

Morrison, 2008; for Dutch /ɑ/~/aː/: Escudero et al., 2009), we expect to find listening strategies that are roughly similar to the ones suggested by Escudero (2000, 2005) and Morrison (2008).

1.2. Latent class modelling

The statistical technique that we use to identify types of listening strategies is based on latent class regression analysis (Huang and Bandeen-Roche, 2004). It is an increasingly popular method for identifying groups of participants with similar latent (i.e., non-overt) individual characteristics in a statistically reliable way. For instance, the technique has been used to study children's reasoning strategies (Bouwmeester et al., 2004) and Japanese women's gender-role attitudes (Yamaguchi, 2000). Also, it has been used to identify groups among psychotic patients (Schmitz et al., 2007) and to pinpoint the sources of knowledge in Artificial Grammar Learning (Visser et al., 2009). In this paper, we introduce the technique to the field of speech perception and its development.

The proposed analysis detects groups of listeners with the same listening strategy within the experimental groups. Thus, it is more fine-grained than a standard group analysis, in which individual deviations from group patterns are not accounted for. At the same time, it goes beyond describing the strategy for each listener separately by highlighting similarities between individuals with the same listening strategy.

Previous research has investigated individual strategies and the clustering of individuals separately. For instance, Chandrasekaran et al. (2010), who examined the effect of native American English speakers' cue weighting of pitch height and direction on their ability to learn Mandarin lexical tone, divided listeners in "good" and "poor" learners on the basis of performance scores, before analysing group differences in the use of specific cues. Escudero and Boersma (2004) first examined listening strategies per individual and then listed the number of listeners who utilized each type of strategy. Unlike these previous proposals, we follow Morrison (2007, 2008, 2009) in that we cluster listeners' strategies with a

statistical technique, but we do so in a more far-reaching way. This is because Morrison's hierarchical cluster analysis still requires the researcher to choose the number of groups. In contrast, the latent class regression analysis groups participants *simultaneously* with the extraction of the strategies. The strategies are represented by a latent variable in the model and are not defined *a priori*. We use a common model selection technique (section 2.3) to determine the optimal, most parsimonious number of strategies within the group, which makes the method statistically more robust.

Crucially, the applied method for strategy detection does not use performance scores to assign a participant to a class with a certain listening strategy. Rather, it extracts listening strategies through determining the degree to which acoustic cues predict an individual's vowel classifications, regardless of the correctness or incorrectness of the responses. Thus, an acoustic vowel dimension that is a statistically significant predictor of a participant's classifications is considered a cue that he or she used, and consequently a significant part of that listener's strategy. Because the outcome variable, i.e., an individual's vowel classification (section 2.3), is categorical, we applied logistic regression models, which have many advantages compared to ANOVA techniques (Jaeger, 2008). Since the proposed analysis relies on cues rather than accuracy, it is specifically suited for our purpose of determining what is learned in the distributional learning process.

2. Method

2.1. Participants

The present study included 150 adult native speakers of Spanish (M = 36.8 years, Range = 19–60 years; 123 female and 27 male), who were living in the Netherlands at the time of testing, and had arrived in the Netherlands after the age of 15 years. They were divided into three groups of 50 each: the Enhanced, Bimodal, and Music groups. All these participants completed a pre-test, a training phase and a post-test. Only the training phase differed per group. The Bimodal and

Enhanced groups listened to vowel distributions (section 2.2.2), while the Music group (or control group) was exposed to classical music.

These Spanish-speaking participants had enrolled in a longitudinal project on the perception of Dutch vowels, which included a larger participant pool (n=500) and was led by the second author. They had all taken part in the first session of the longitudinal project six months earlier. During this first session, participants in the Music group had performed the same pre- and post-tests and had listened to the same classical music as in the present study, while participants in the Bimodal and Enhanced groups had only performed the pre-test, and had not received any training.[1] In the first session, assignment to groups had been random. In the present study, which reports results of the second session, participants were assigned to the Bimodal and Enhanced groups while considering their first-session pre-test scores, which were matched with those of the Music group. Other than that, assignment to the two training groups was random.

Table V.1 lists each group's age at the time of testing (AaT), age of arrival (AoA), length of residence (LoR) in the Netherlands, and Dutch proficiency score, i.e., the level of general comprehension of Dutch as measured by the language comprehension component of the Dialang test (www.dialang.org; Alderson and Huhta, 2005). The groups were not significantly different in any of these measures (LoR: $F[2,149] = 0.52$, $p = 0.60$; AaT: $F[2,149] = 1.6$, $p = 0.20$; AoA: $F[2,149] = 1.2$, $p = 0.29$ and Dutch proficiency $F[2,148] = 0.34$, $p = 0.71$). Additionally, the median and range for AaT was comparable across groups: Enhanced: 36 (range: 21–56), Bimodal: 34 (22–55), and Music: 37 (19–60). All participants reported normal hearing.

1 Results of the Music group's first session (i.e., of the pre- and post-test) are reported in Escudero et al. (2011). Results of the Bimodal and Enhanced group's first session (i.e., of the pre-test only) are reported in Escudero and Wanrooij (2010).

Table V.1. Mean age at testing (AaT), age of arrival (AoA) and length of residence (LoR) in the Netherlands (in years), and Dutch proficiency score (see text), per Spanish group. Standard deviations are given between parentheses.

Group	AaT	AoA	LoR	Dutch Proficiency
Enhanced	37.3 (8.0)	31.9 (6.9)	5.4 (5.0)	3.9 (2.2)
Bimodal	35.0 (8.7)	29.9 (7.0)	5.2 (5.4)	4.2 (2.2)
Music	38.0 (9.0)	31.7 (7.2)	6.3 (6.8)	4.0 (2.1)

Further, just as in the larger longitudinal project where out of 500 registrations only 50 were from men, the number of female participants in the present study (38, 41 and 44 in the Enhanced, Bimodal and Music groups respectively) was larger than that of male participants (12, 9 and 6 respectively). In section 3, we examine whether our results are representative of both men and women.

Unlike Escudero et al. (2011), we also included an age-matched group of 25 adult native speakers of Dutch (M = 32 years, Range = 18–60 years; 21 female). These Dutch natives performed the same test as the Spanish listeners but only once, and they received no training. We will compare the Dutch results for this single test to both the pre- and post-tests that Spanish listeners performed in order to assess these listeners' L2 development after training.

2.2. Stimuli and procedure

2.2.1. Test

Spanish and Dutch listeners performed a forced-choice classification task in an XAB format, designed to assess classification performance of Dutch /ɑ/ and /aː/. To promote classification rather than discrimination, the inter-stimulus-interval (ISI) between the three stimuli in each trial (i.e., X, A and B) was chosen to be relatively long (1.2 s) (Van Hesse and Schouten, 1999; Werker and Logan, 1985),

and the X stimuli were chosen to be natural tokens containing much variability, as explained below.

Prior to performing the XAB-task, participants had a practice session of five trials where it was ascertained that they heard the stimuli well and that they understood the task. None of the listeners demonstrated hearing problems or failed to correctly identify the vowels in this practice session. As mentioned above (section 2.1), only the Spanish listeners performed the XAB task a second time after training, i.e., they had a pre- and a post-test. The test procedure, which was the same as in Escudero et al. (2011), was as follows. In each trial, listeners heard a natural token of /ɑ/ or /aː/ (the X stimulus), followed by two synthetic response options (A and B). There were 20 unique X stimuli for each vowel, which were a subset of the vowels reported in Adank et al.'s (2004) corpus and which were produced by 10 male and 10 female speakers of Standard Northern Dutch in an /s–V–s/ context. The average fundamental frequency (F0), first formant (F1), second formant (F2) and duration of the X stimuli are listed in Table V.2, for females and males separately.

Table V.2. Average F1, F2, F0 (in Hz) and duration (in milliseconds) of the X stimuli in the XAB-test. Standard deviations are given between parentheses.

Vowel	F1		F2		F0		Duration	
	Females	Males	Females	Males	Females	Males	Females	Males
/ɑ/	719	584	1239	1156	223	154	93	94
	(100)	(99)	(168)	(127)	(50)	(24)	(13)	(24)
/aː/	923	652	1552	1424	183	132	216	204
	(75)	(144)	(107)	(98)	(36)	(18)	(43)	(14)

Unlike in Escudero et al. (2011), where each X stimulus was presented once and the response options were randomly ordered, we included two repetitions of each X stimulus by counterbalancing the response options. Thus, our XAB task included 80 trials (=20 unique X stimuli×2 vowels×2 repetitions). The two

response options A and B were synthetic stimuli (created using the Praat program of Boersma and Weenink, 2011), because the acoustic properties had to be compatible with those of the training stimuli (section 2.2.2). They were based on typical tokens of /ɑ/ and /aː/ (Pols et al., 1973), with F1-values of 687 and 770 Hertz (Hz) and F2-values of 1104 and 1303 Hz respectively, which five Dutch natives had judged as better exemplars of the Dutch vowels than tokens generated using Adank et al.'s (2004) values (Escudero and Wanrooij, 2010). For both response options, the duration was 140 ms and F0 fell from 150 to 100 Hz, which represents a male voice (e.g., Hollien et al., 1971).

The task was self-paced: listeners were told that the next trial would only appear after their response. They were encouraged to respond as quickly as possible and were asked to guess if uncertain. Also, they were told that they could take a short break (available every 20 trials) if needed. Spanish and Dutch listeners took approximately 7 minutes to complete the task.

2.2.2. Training

Only the Spanish listeners were presented with the training phase. The training stimuli and procedure, which were the same as in Escudero et al. (2011), were as follows. The stimuli during the training phase differed across Spanish groups: Bimodal and Enhanced listeners heard, respectively, bimodal and enhanced training distributions of the Dutch vowel contrast /ɑ/~/aː/, while the Music group listened to instrumental classical music. The goal of the bimodal and enhanced training was to expose participants to the spectral difference between Dutch /ɑ/ and /aː/. Because Spanish listeners tend to classify /ɑ/ and /aː/ on the basis of their duration while ignoring their spectral differences (section 1), the training stimuli differed from one another only in the spectral values for F1, F2 and F3 (the third formant) and not in duration. Table V.3 lists the F1 and F2 values for each of the eight stimuli in the bimodal and enhanced training distributions separately, which were synthesized in the computer program Praat (Boersma and Weenink, 2011).

Table V.3. F1 and F2 values (in Hz) and frequency of presentation for each stimulus in the enhanced and bimodal training distributions (Escudero et al., 2011).

Token number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Frequency	8	32	16	8	8	16	32	8
Enhanced								
F1	600	637	675	714	755	797	840	885
F2	1000	1055	1112	1171	1233	1296	1362	1430
Bimodal								
F1	700	713	726	740	753	767	781	795
F2	1115	1144	1174	1204	1235	1266	1298	1330

The endpoint values (i.e., stimulus 1 and 8 in the table) of the bimodal distribution were similar to the average production values of Dutch /ɑ/ (stimulus number 1) and /aː/ (stimulus number 8), as measured by Pols et al. (1973). The endpoint values of the enhanced distribution were calculated as the average production of /ɑ/ minus one standard deviation (stimulus 1) and the average production of /aː/ plus one standard deviation (stimulus 8). The standard deviations were based on Pols et al. (1973). In each distribution, the steps between consecutive values were approximately equal on the psychoacoustic ERB scale (Bimodal: 0.1 ERB for F1, 0.2 ERB for F2; Enhanced: 0.4 ERB for F1 and F2). F3 was calculated for each stimulus as the stimulus' F2 plus 1000 Hz. All training stimuli had an F0 that fell from 150 to 100 Hz, and a duration of 140 ms. The table also shows the frequency of presentation for each training stimulus. There were 128 stimuli in total, which were presented with an ISI of 750 ms, for a total training duration of less than 2 minutes. The Music group listened to classical music for the same time.

Before the training phase, all participants were told that they would perform another test afterwards. Listeners in the Enhanced and Bimodal groups

were instructed to listen to the training vowels carefully, while listeners in the Music group were asked to relax while listening to the classical music.

2.3. Statistical analysis

A traditional comparison of mean accuracy across groups served to demonstrate the same distributional training results as in Escudero et al. (2011) and thus to demonstrate the validity of our data for the subsequent analysis of listening strategies, i.e., specific uses of acoustic cues in perception, in each group. To identify listening strategies, we used *latent class regression* (LCR) analysis (Huang and Bandeen-Roche, 2004), as mentioned in section 1. LCR analysis explains correlations between responses to different items by introducing a *latent* variable. This variable is nominal, which indicates the existence of a number of different types (*classes*) of behaviour rather than a dimension on which people vary continuously. Furthermore, a finite number of types of behaviour, each with a unique set of *regression* coefficients (and intercepts), is assumed.

We identified the five most important acoustic components for the classification of the natural vowel productions (i.e., the X stimuli) that were presented in the XAB task: duration, F1, F2, F3, and F0. Correct classification needed to be based primarily on F1, F2, duration or a combination of these cues (section 1), and secondarily on higher formants such as F3, which adds subtle information but cannot be used as a single cue to distinguish the two vowels. Further, F0 could not be used to classify the vowels correctly, because it is not a cue for vowel identity.

When participants took only duration into account when classifying the vowels, their listening strategy was confined to the use of this cue. We described such a listening strategy with a binomial regression model, i.e., with a binomially distributed dependent variable and multiple predictors. The dependent variable was the number of times a participant chose the category /aː/ for each specific X stimulus. Since every specific X stimulus was presented twice, the number of times a participant opted for response /aː/ when presented with a token of /ɑ/ or /aː/ was

0, 1 or 2. Note that we thus modelled the categorization of stimuli and not the accuracy of the categorization (section 1.2). The predictors were the five acoustic components of the vowels mentioned above.[2]

In a standard regression analysis, the same regression coefficients apply to each participant. In LCR analysis, the same regression coefficients apply only to members of the same latent group. It is important to note that group membership is not a manifest variable (i.e., an observable variable) but is assigned only after fitting the LCR model to the data. The specified LCR model had the following form:

$$L(y_i) = \mu_c + \beta_{Dc}D + \beta_{F0c}F0 + \beta_{F1c}F1 + \beta_{F2c}F2 + \beta_{F3c}F3$$
$$c = 1...\mathrm{N}_c, i = 1...n \quad (1)$$

Here y_i is the number of times (0, 1, or 2) that a specific X stimulus was classified as /aː/ and L is the standard link function[3] for a binomial regression model (Jaeger, 2008; McCullagh and Nelder, 1989), i.e., the logit function $\log p/(1-p)$, where p is the mean of the binomial distribution; μ_c is the intercept of latent class c; parameters β_{Dc}, β_{F0c}, β_{F1c}, β_{F2c}, and β_{F3c} are the regression coefficients for latent class c; N_c is the number of latent classes; and n is the number of participants. The value of the intercept is a measure of the bias in responding /ɑ/ or /aː/. Because the absolute value is not easy to interpret, we will calculate the bias for each latent class after fitting the model. The regression coefficients indicate how much the logit of the probability of answering /aː/ changed with a one-unit change in the predictor. Note that the regression parameters are not normalized, so that the absolute values are still interpretable given the different ranges for each predictor.

2 We used logarithmic scales for the five acoustic cues to account for the fact that the human ear is better at discriminating small differences in shorter durations and lower frequencies than in longer durations and higher frequencies (e.g., Allan and Gibbon, 1991; Kewley-Port and Watson, 1994; Stevens et al., 1937).

3 The link function provides the relationship between the linear predictor and the mean of the distribution.

Exploratory LCR models with an increasing number of latent classes were fitted to the Spanish groups' pre-test and post-test classification data and to the Dutch natives' classification data in their single test. To establish the optimal number of latent classes in each condition, we used the Bayesian Information Criterion (BIC; Schwarz, 1978).[4] The BIC is commonly used to compare non-nested competing models, in this case models with an increasing number of latent classes (see Lin and Dayton, 1997, for details on the specific uses of BIC in latent class models). The BIC provides a trade-off between goodness of fit (the log likelihood) and the number of parameters in the model. For each added latent class, seven extra parameters are estimated, namely, the intercept and regression coefficients of that class (in our case five regression coefficients for the five predictors), and the proportion of participants that it contains. Lower values for BIC denote better models in which goodness of fit and parsimony are balanced. After fitting the model to the data, each individual participant was assigned to a class. To this end, the posterior probabilities of participants' responses were calculated given each latent class of the model. Subsequently, each participant was assigned to the latent class with the largest likelihood for that participant's data. For fitting models to the data, we used the statistical R-package of FlexMix (Leisch, 2004; see also Grün and Leisch, 2007, for an example of fitting mixtures of logistic regressions in R).

3. Results

Table V.4 shows the group results for the Dutch and Spanish listeners, which are given in accuracy percentages, i.e., the percentage of time listeners correctly classified the 80 test stimuli. The Dutch accuracy was substantially higher than that in all Spanish groups for both the pre- and the post-tests, which confirms previous Dutch results on the same task (Escudero and Wanrooij, 2010), and thus ascertains that the stimuli and the response options were good examples of the Dutch vowels

4 The BIC is defined as minus 2 times the log likelihood of the model, plus the number of parameters times ln(N), with N being the number of participants.

/ɑ/ and /aː/. The Dutch accuracy also shows that the task was relatively difficult, since Dutch listeners did not score at ceiling.

Table V.4. Mean Spanish (pre- and post-test) and Dutch (single-test) accuracy percentages. Standard deviations are given between parentheses.

Test	Enhanced	Bimodal	Music	All Spanish	Dutch
Pre-test	60.4 (11.7)	60.4 (12.2)	61.7 (11.1)	60.8 (11.6)	83.1 (9.6)
Post-test	67.1 (13.5)	64.2 (14.5)	63.7 (13.3)	65.0 (13.7)	–

To investigate if our results for the Spanish participants were similar to those of Escudero et al. (2011), we ran a mixed design analysis with Test as a within-subjects factor (pre-test vs. post-test accuracy) and Group as a between-subjects factor (Bimodal, Enhanced and Music). The results revealed no main effect of Group (F [2,147] = 0.20, p = 0.82), which supports the homogeneity of the groups, and a main effect of Test (F [1,147] = 29.70, $p < 0.001$), which indicates that the improvement between pre- and post-test shown in Table V.4 is statistically significant. Further, the analysis yielded a significant Test×Group interaction (F [2,147] = 3.12, p = 0.047), which indicates that some group(s) improved more than others.

Post hoc t-tests on difference scores (i.e., post- minus pre-test accuracy percentages, as shown in Table V.4) using Tukey's HSD revealed that the Enhanced group improved more than the Music group (difference = 4.63%, with a 95% Confidence Interval, CI = +0.21 ~ +9.04%, p = 0.038), and that the differences in improvement between the Bimodal and Enhanced groups, and between the Bimodal and Music groups were not significant (ps > 0.05). These results are the same as those reported in Escudero et al. (2011). Further, to test whether each group improved significantly in the post-test as compared to the pre-test, the difference score of each group was compared to 0 (which represents no improvement) in a one-sample t-test. Again in accordance with Escudero et al., a significant improvement was found for the Enhanced group (6.63% with CI =

+4.05 ~ +9.20%, t[49] = 5.17, p < 0.001), and not for the Music group (2.00% with CI = –0.50 ~ +4.50%, t[49] = 1.61, p = 0.12). Unlike in Escudero et al., there was also a significant improvement for the Bimodal group (3.83% with CI = +0.97 ~ +6.68%, t[49] = 2.69, p = 0.010).

We also examined whether pre-test accuracy and difference scores (n=150) were significantly correlated with Spanish listeners' LoR, AaT, AoA and Dutch proficiency (section 2.1) using non-parametric correlations (Spearman's ρ). There was a significant correlation between *pre-test accuracy* and both AaT (ρ = –0.19, p = 0.023) and AoA (ρ = –0.23, p = 0.005), indicating that the younger participants were when they performed the task and the younger they were when they arrived in the Netherlands, the higher their accuracy at pre-test. There was no significant correlation between pre-test accuracy and LoR or Dutch proficiency (both ps ≥ 0.71).

Further, there was no significant correlation between *difference scores* and AaT, AoA or Dutch proficiency (all ps ≥ 0.13). Difference scores were significantly correlated with LoR (ρ = 0.17, p = 0.033).

3.1. Listening strategies before distributional training

Table V.5 summarizes the optimal latent class regression models for Spanish learners' pre-test and Dutch natives' single test. It contains the identified classes per group, and the cues that each class used, i.e., their listening strategy. In the regression model the cues are the predictors (section 2.3). None of the Spanish and Dutch classes exhibited a response bias to /ɑ/ or /aː/ (one-sample ts < 2.2, ps > 0.05).[5]

5 As mentioned in Section 2.3, the number of /aː/ responses for any specific stimulus /ɑ/ or /aː/ could be 0, 1 or 2. For the response bias analysis, we thus used the null hypothesis that the average number of /aː/ responses in each class was 1.

Table V.5. Spanish (pre-test) and Dutch (single test) classes, including number of participants per class (N), their mean accuracy, statistically significant predictors (Cues), estimated regression coefficients (Betas) and *p*-values. D=duration.

Group	Class	N	Accuracy (SD)	Cues	Beta (SE)	*p*-value
Spanish	1	33	53.2 (6.2)	D	0.71 (0.33)	0.032
Enhanced				F1	1.36 (0.55)	0.013
	2	17	74.4 (5.4)	D	4.16 (0.58)	< 0.0001
				F1	4.00 (0.85)	< 0.0001
				F2	5.58 (1.54)	0.00028
Spanish	1	39	55.1 (7.1)	D	0.91 (0.30)	0.0028
Bimodal				F3	2.22 (0.94)	0.019
	2	11	79.1 (6.9)	D	4.39 (0.69)	< 0.0001
				F1	4.09 (1.11)	0.00023
				F2	10.08 (2.12)	< 0.0001
Spanish	1	38	56.6 (6.6)	D	1.25 (0.31)	< 0.0001
Music				F1	1.93 (0.51)	0.00015
	2	12	78.0 (4.6)	D	4.96 (0.69)	< 0.0001
				F1	7.14 (1.13)	< 0.0001
				F2	4.81 (1.90)	0.012
Dutch	1	13	75.3 (6.1)	D	5.21 (0.66)	< 0.0001
				F1	4.14 (0.96)	< 0.0001
				F3	6.04 (1.90)	0.0015
	2	12	91.5 (3.3)	D	8.58 (0.87)	< 0.0001
				F0	–5.28 (1.66)	0.0015
				F1	6.15 (1.63)	0.00016
				F2	14.80 (2.95)	< 0.0001

It can be observed that each Spanish group had two latent classes: one with the majority of participants with low mean accuracy (hence "low performers"), and the other with the minority of participants with high accuracy (hence "high performers"). These two pre-test classes per group confirm the equality of the groups at pre-test and are also visible in Figure V.1, left column, which shows the number of participants (*y*-axis) for each accuracy percentage (*x*-axis). The figure clearly shows that most, if not all, low performers (black bars) indeed had lower accuracy than high performers (white bars).

There was a strong correlation between the accuracy percentage obtained in each Spanish class and the number of cues used: Spearman's $\rho = 0.88$, p(one-tailed)[6] = 0.011. Thus, not surprisingly, Spanish learners of the Dutch contrast /ɑ/~/aː/ tend to score higher when they use more cues. Low performers used two cues, namely duration and either F1 (in Enhanced and Music) or F3 (in Bimodal), while high performers used three, namely duration and a combination of F1 and F2. Overall, all six Spanish classes used duration, five classes used F1, three used F2, one used F3, and none used F0, which suggests that Spanish listeners tend to favour certain cues above others. Interestingly, high performers not only used more cues than low performers, but also tended to use cues more intensely, as reflected by their betas (i.e., the regression coefficients in the model; section 2.3). For example, Table V.5 shows that low performers had duration betas of 0.91, 0.71 and 1.25, while high performers had duration betas of over 4.

Because our participant group contained a larger number of females than males, we examined whether the division into low and high performers in the pre-test was representative of both women and men. For this, we counted the number of low and high performers who were female (94 low and 29 high performers) versus male (16 low and 11 high performers). A chi-square test showed no significant difference in listening strategies between the sexes ($\chi[1] = 3.34$, $p = 0.068$).

6 The significance test is one-tailed because we expect a positive correlation between the number of predictors and vowel classification accuracy.

Dutch natives also had two different listening strategies: half of them focused on three cues (duration, F1 and F3) and had moderate accuracy (M = 75.3%), while the other half used four cues (duration, F0, F1 and F2) and had very high accuracy (M = 91.5%). A comparison of the Spanish and Dutch performance shown in Table V.5 suggests that Spanish high performers approximated the Dutch natives who performed moderately well.

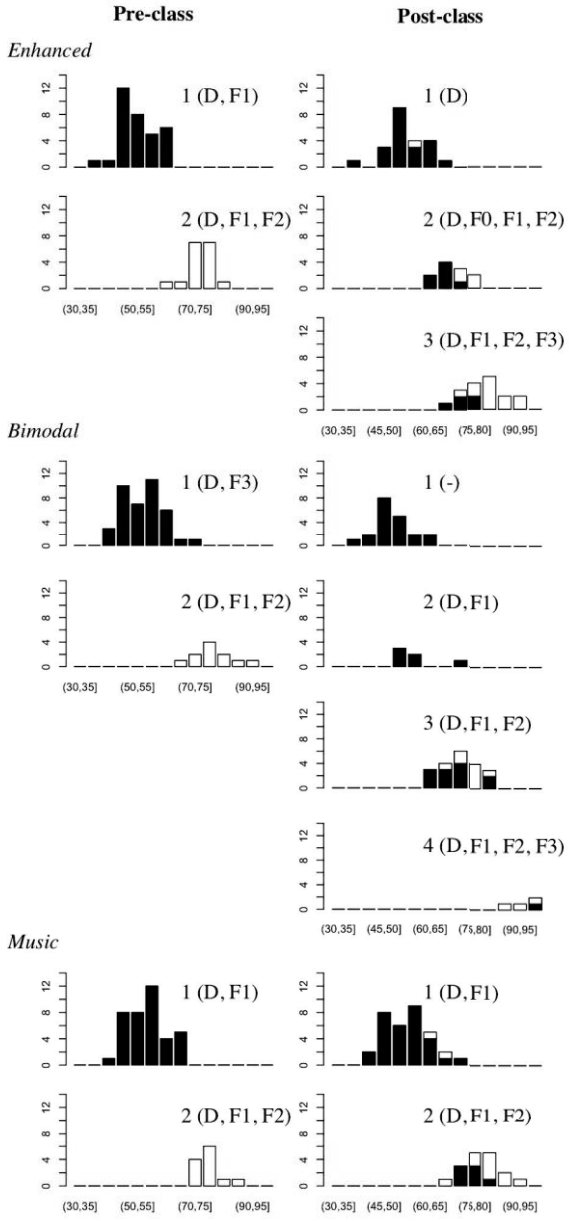
3.2. Listening strategies after distributional training

The Spanish post-test classes are shown in Table V.6, where it can be observed that the post-test yielded three, four and two classes in the Enhanced, Bimodal and Music groups respectively.

Similarly to the pre-test, significant cues for classes with 60% or lower accuracy did not include a combination of F1 and F2 and the maximum number of cues was two, while learners in classes with 70% or higher accuracy used at least three cues including duration, F1 and F2. Classes with 80% or higher accuracy also included F3. Again, a strong correlation was found between accuracy and the number of cues identified for a class: Spearman's $\rho = 0.89$, p(one-tailed)[7] = 0.001, which indicates that when Spanish learners focus on more cues, accuracy of classification of /ɑ/ and /aː/ increases. Duration was also the most consistently used

Figure V.1 (opposite page). Histograms showing the number of Spanish learners (y-axis) for each accuracy percentage (x-axis) per pre-test class (left) and post-test class (right) in each group (Enhanced, Bimodal, and Music). For each class the listening strategy (one or more of the acoustic cues duration, F1, F2, F3 and F0) is given. In both the pre-test and the post-test column black bars represent pre-test low performers and white bars pre-test high performers.

7 See the previous note.


Pre-class
Post-class
Enhanced
1 (D, F1)
1 (D)
2 (D, F1, F2)
2 (D, F0, F1, F2)
3 (D, F1, F2, F3)
(30,35]
(50,55]
(70,75]
(90,95]
(45,50]
(60,65]
(75,80]
Bimodal
1 (D, F3)
1 (-)
2 (D, F1, F2)
2 (D, F1)
3 (D, F1, F2)
4 (D, F1, F2, F3)
Music
1 (D, F1)
1 (D, F1)
2 (D, F1, F2)
2 (D, F1, F2)

cue (8 out of 9 classes), followed by F1 (8 classes), F2 (5 classes), F3 (2 classes) and F0 (1 class). Also as in the pre-test, learners with higher accuracy appeared to use cues more intensely, i.e., they had higher betas, than those with lower accuracy. For instance, duration betas ranged between 0.88 and 2.83 for classes with accuracy below 60%, while they were between 4.68 and 8.72 for classes with higher accuracy.

When comparing the Spanish post-test classes in Table V.6 to those of the Dutch single test in Table V.5, we observe that more than 20 percent (11 out of 50) of the learners in the Enhanced group ended up using the same cues (duration, F0, F1, and F2) as half of the Dutch natives (12 out of 25), but the Dutch had a higher accuracy (70.1% versus 91.5%). This difference may be due to a more efficient use of duration and F2 in the Dutch natives, as reflected by their higher betas. Remarkably, one class of four Bimodal listeners obtained similar accuracy (93.4%) as the best performing Dutch class, despite the fact that they used a different strategy than the Dutch.

Finally, one-sample *t*-tests for each post-class ($\alpha = 0.0056$, 05/9 tests) showed that a bias toward the /aː/ response developed in Bimodal class 2 ($M = 1.43$, CI = +1.22 ~ +1.64, $t[5] = 5.35$, $p = 0.0031$) and Enhanced class 2 ($M = 1.34$, CI = +1.24 ~ +1.44, $t[10] = 7.70$, $p < 0.001$).

3.3. Improvement with training

A comparison of Table V.5 (pre-test) and Table V.6 (post-test) shows that after training an increase in number of classes is only observed for the Enhanced (from 2 to 3) and Bimodal (from 2 to 4) groups. Also, while the Music group has the *same* listening strategies in both tests, listening strategies typically changed after distributional training. These observations suggest that distributional training, and not listening to music, diversified listening strategies. Furthermore only after distributional training, Spanish listeners came closer to the Dutch listening strategies and accuracy (section 3.2).

Table V.6. Spanish post-test classes, with the same variables as in Table V.5.

Group	Class	N	Accuracy (SD)	Cues	Beta (SE)		*p*-value
Spanish	1	22	54.6 (6.6)	D	0.88 (0.40)		0.030
Enhanced	2	11	70.1 (6.0)	D	4.68 (0.71)	<	0.0001
				F0	3.23 (1.33)		0.015
				F1	7.08 (1.34)	<	0.0001
				F2	5.84 (2.16)		0.0067
	3	17	81.2 (6.8)	D	6.23 (0.60)	<	0.0001
				F1	4.31 (0.96)	<	0.0001
				F2	5.01 (1.69)		0.0031
				F3	–5.59 (2.12)		0.0085
Spanish	1	20	51.4 (6.2)	–	–		–
Bimodal	2	6	57.7 (7.3)	D	2.83 (0.85)		0.00091
				F1	3.04 (1.44)		0.035
	3	20	73.1 (6.1)	D	4.03 (0.50)	<	0.0001
				F1	5.53 (0.80)	<	0.0001
				F2	3.64 (1.50)		0.015
	4	4	93.4 (4.1)	D	8.72 (1.64)	<	0.0001
				F1	13.66 (3.95)		0.00054
				F2	41.73 (8.9)	<	0.0001
				F3	–22.40 (7.64)		0.0034
Spanish	1	33	55.5 (7.1)	D	1.15 (0.33)		0.00051
Music				F1	1.29 (0.54)		0.017
	2	17	79.6 (6.0)	D	5.50 (0.57)	<	0.0001
				F1	5.54 (0.92)	<	0.0001
				F2	6.28 (1.65)		0.00014

Figure V.1 illustrates how pre-test performance relates to post-test class membership, as follows. In both the pre-test column (Figure V.1, left) and the post-test column (Figure V.1, right) black bars represent *pre-test* low performers and white bars *pre-test* high performers. Post-test classes are numbered from worst- (1) to best-performing (2 and above). It can be observed that pre-test low and high performers tended to move to the worst and best performing post-test classes respectively, as shown by the higher number of black and white bars in the right column for low and high post-test accuracy respectively.

Specifically, in the *Enhanced group*, out of the 33 pre-test low performers (who used duration and F1 in the pre-test) 21 listeners (64%) moved to the worst-performing post-test class 1 (duration only), 7 (21%) to post-test class 2 (duration, F0, F1 and F2) and 5 (15%) to post-test class 3 (duration, F1, F2 and F3). Out of the 17 Enhanced pre-test high performers (who used duration, F1 and F2 in the pre-test), 1 (6%) moved to post-test class 1 (duration only), 4 (24%) to post-test class 2 (duration, F0, F1 and F2) and 12 (71%) to post-test class 3 (duration, F1, F2 and F3). In the *Bimodal group*, out of the 39 pre-test low performers (who used duration and F3 in the pre-test) 20 (51%) moved to post-test class 1 (no cues), 6 (15%) to post-test class 2 (duration, F1), 12 (31%) to post-test class 3 (duration, F1 and F2) and 1 (3%) to post-test class 4 (duration, F1, F2 and F3). Out of the 11 Bimodal high performers (who used duration, F1 and F2 in the pre-test) 8 (73%) retained the same strategy in post-test class 3, while 3 (27%) moved to post-test class 4 (duration, F1, F2 and F3). In the *Music group*, out of the 38 pre-test low performers (who used duration and F1 in the pre-test) 31 (82%) retained the same strategy in post-test class 1 and 7 (18%) moved to post-test class 2 (duration, F1 and F2), while out of the 12 pre-test high performers (who used duration, F1 and F2 in the pre-test), 2 (17%) moved to post-test class 1 (duration and F1) and 10 (83%) retained the same strategy in post-test class 2.

Figure V.1 also illustrates that if listeners used new cues after training, these cues were always F1 and/or F2 for pre-test low performers, while pre-test high performers, who continued using F1 and F2, also used F3. Some Enhanced listeners also started to use F0. To quantify new cue use more precisely, we

counted the number of pre-test low and high performers who started to use new relevant cues (i.e., the primary cues F1 and F2, and the secondary subtle cue F3)[8] versus those who did not, which are listed in Table V.7. It can be inferred from the table that 36.4% (12 listeners) of the pre-test low-performers who were trained in the Enhanced condition began using F1 and/or F2 in the post-test, as compared to 48.7% (19 listeners) in the Bimodal group and only 18.4% (7 listeners) in the Music group. Further, 70.6% (12 listeners) of the pre-test high performers in the Enhanced group started using F3 after training, versus only 27.3% (3 participants) in the Bimodal group. In the Music group none of the pre-test high performers started using new cues.

Table V.7. Number of low and high performers in the pre-test, who started to use F1, F2 and/or F3 after training (new users) versus those who did not (others).

	Low performers			**High performers**		
	Enhanced	**Bimodal**	**Music**	**Enhanced**	**Bimodal**	**Music**
New-cue users	12	19	7	12	3	0
Others	21	20	31	5	8	12

Two chi-square tests, for pre-test low and high performers separately, showed significant group (Enhanced, Bimodal and Music) differences (low: $\chi[2] = 7.88$, $p = 0.019$, high: $\chi[2] = 15.63$, $p < 0.001$). For *pre-test low performers*, *post hoc* chi-square tests showed that more Bimodal than Music listeners started using F1 and/or F2 ($\chi[1] = 7.90$, $p = 0.005$), and that there was no significant difference between the Bimodal and Enhanced groups in this respect ($\chi[1] = 1.11$, $p = 0.29$). Thus, for pre-test low performers, enhanced training did not significantly improve the use of F1 and/or F2 more than bimodal training. For *pre-test high performers*, *post hoc* chi-square tests demonstrated that more Enhanced than Bimodal listeners

8 Including the irrelevant cue F0 in the analysis strengthens the significance values reported and does not change the main findings.

started using F3 after training (χ[1] = 5.04, p = 0.025). Thus, for pre-test high performers enhanced training was more effective for learning to use F3 than bimodal training. Further, for the Enhanced group, relatively more pre-test high than low performers started using new cues after training (χ[1] = 5.27, p = 0.022), indicating that the enhanced training was more effective for pre-test high than low performers. In the Bimodal group a comparison between pre-test low and high performers was not significant (χ[1] = 1.60, p = 0.21).

Interestingly, Figure V.1 also shows that listeners started to use new cues in a certain order, viz., duration, F1, F2 and F3. That is, listeners who started to use F1 after training, always continued to use duration, those who started using F2 also started or continued to use duration and F1, and those who started to use F3, also started or continued to use duration, F1 and F2.

Recall that duration was the only cue used by all pre-test classes (section 3.1). Table V.8 shows how many pre-test low and high performers in each group (Enhanced, Bimodal and Music) increased their use of duration after training versus those who did not. An increase was reflected in a higher beta for duration in the post- as compared to the pre-test. For *pre-test low performers*, a chi-square test showed that the groups differed in this respect (χ[2] = 45.04, p < 0.001). In *post hoc* chi-square tests, the number of low performers in the pre-test, who increased their use of duration after training was larger in the Enhanced than Music and Bimodal groups (both χ[1] > 20.71, ps < 0.001), and in the Bimodal than Music groups (χ[1] = 7.90, p = 0.005). For *pre-test high performers post hoc* Fisher Exact tests showed that fewer Bimodal than Enhanced (p < 0.001) and Music (p = 0.012) listeners increased their use of duration. In sum, across low and high performers listeners increased their use of duration after enhanced training in particular. Notice that the numbers in Table V.7 and Table V.8 are similar. In fact, all listeners who started using new cues also increased their use of duration.

Table V.8. Number of low and high performers in the pre-test, who increased their use of duration after training versus those who did not.

	Low performers			High performers		
	Enhanced	**Bimodal**	**Music**	**Enhanced**	**Bimodal**	**Music**
Increased use	33	19	7	16	3	10
Others	0	20	31	1	8	2

As in section 3.1, we examined possible sex differences in our results. Specifically, we examined whether men and women differ in their ability to use new cues after training. For this, we counted the number of new-cue users versus other participants, who were female (37 new-cue users and 86 others) and male (16 new-cue users and 11 others). A chi-square test showed a significant difference between men and women ($\chi[1] = 8.25$, $p = 0.005$). Additionally, we examined the sex distribution of new-cue users versus others in *post hoc* chi-square tests for pre-test low and high performers separately. For *pre-test low performers*, there was no significant difference in the ability to use new cues after training between men (7 new-cue users, 9 others) and women (31 new-cue users, 63 others; $\chi[1] = 0.70$, $p = 0.402$). For *pre-test high performers*, the *post hoc* test showed that men (9 new-cue users, 2 others) were more likely to use new cues after training than women (6 new-cue users, 23 others; Fisher Exact Test: $p = 0.001$).

4. Discussion

The present study confirmed Escudero et al.'s results (2011) in two ways. First, our new group of Spanish learners that was exposed to an enhanced distribution of the Dutch vowel contrast /ɑ/~/aː/ (the Enhanced group) classified the Dutch vowels significantly better after than before training, and the control group exposed to classical music (the Music group) did not. Second, this improvement for the Enhanced group was greater than that for the Music group. Unlike Escudero et al.,

Spanish learners who were exposed to a bimodal distribution of the contrast (the Bimodal group) also improved significantly in the post- as compared to the pre-test. Our findings confirm that distributional vowel training, with enhanced distributions in particular, leads to improvement in the classification of difficult L2 contrasts. This result allowed us to pursue our main objective of identifying listeners' strategies and examining the effect of bimodal versus enhanced training on the different strategy types, which will be discussed below.

We found a negative correlation between Spanish listeners' age at testing and pre-test accuracy and also between age of arrival and pre-test accuracy. This is in line with earlier observations for the influence of age of L2 learning on speech perception (e.g., Flege et al., 1999) and on production (see Piske et al., 2001, for a review). Further, neither higher general comprehension of Dutch nor longer exposure to Dutch as reflected in the length of residence in the Netherlands were significantly related to higher pre-test perception accuracy. Although a number of previous studies have shown an effect of these two factors on L2 sound perception (e.g., Escudero et al., 2009; Flege et al., 1997), others have failed to find these effects (e.g., Cebrian, 2006; Escudero and Wanrooij, 2010). For the second factor (amount of exposure) this discrepancy in outcomes is probably due to the unreliability of length of residence as a measure of the amount of exposure to the target language (e.g., Moyer, 2009; Piske et al., 2001). It is a poor measure when, for instance, learners have little contact with native speakers or when the quality of the new language input is bad (e.g., Moyer, 2009). Nevertheless, if length of residence in the current study reflected the participants' amount of exposure to Dutch, the observed significant relation between length of residence in the Netherlands and improvement after training could be interpreted as a sign that our distributional training facilitated perceptual learning that had started outside the lab.

The latent class analysis of listening strategies indicated a split in initial listening strategies between listeners who did not focus on the critical combination of F1 and F2, and listeners who did. As expected, the former ("pre-test low performers") had relatively low and the latter ("pre-test high performers")

relatively high accuracy. After the training phase, listeners in the control group did not change strategies, while listeners in the Bimodal and Enhanced groups diversified their strategies. Improvers among the pre-test low performers started to use F1 and/or F2, while pre-test high performers refined their strategies mainly by adding the subtle secondary cue F3. Further, the outcomes revealed no significant difference between bimodal and enhanced training in learning to use F1 and/or F2 for pre-test low performers, while pre-test high performers profited more from enhanced than bimodal training for learning to include F3 in their listening strategies. This shows the importance of looking beyond group results, which can be considerably affected by group composition. The results for pre-test high performers extend previous research, which shows that enhanced differences in critical acoustic cues can facilitate learning by directing listeners' attention to these cues (e.g., Iverson et al., 2005; Jamieson and Morosan, 1986; Kondaurova and Francis, 2010). Because the usefulness of enhanced training was particularly evident in pre-test high performers' new use of F3 in the post-test, it seems that for listeners who are already attentive to the critical cues, enhancement may facilitate attention to additional, more subtle cues.

Further, our participant groups had mainly female participants. Although in our lab we had not observed sex differences in vowel perception earlier (e.g., sex differences in the data of Escudero and Chladkova, 2010, could not be found), Obleser and colleagues (2001) report a larger left-hemispheric activity for women than for men when listening to vowels. Even though this observation does not necessarily mean that men and women use different acoustic cues when listening to vowels, we explored whether women and men showed different listening strategies and learning behaviour. We did not find sex differences in pre-test listening strategies, and in the ability to use new cues after training for pre-test low performers. However, we found that among pre-test high performers (who were already using F1 and F2 in the pre-test) men were more likely to start using F3 after training than women. The precise meaning of this observed sex difference is not clear and should be examined in future research.

Listeners who used new cues after training simultaneously increased their use of duration (section 3.3). This may be a sign of *cue integration*, i.e., the use of both duration and formant frequencies for vowel perception, as predicted in the L1 distributional learning model of Boersma and colleagues (2003), which was more explicitly formulated and extended in Escudero (2005). The model predicts that, in building a phonological contrast, learners initially use a single cue (e.g., relating a certain duration to a phonological category "short") and then start to integrate additional cues (e.g., also relating an F1 with a certain frequency value to a phonological category "short") on the basis of their correlational distributions. Listeners in the current study may have been in the process of relating a relatively low F1 and/or F2 that they heard during training to the short duration (for /ɑ/, or the high F1 and F2 to the long duration for /aː/) that they were already able to use before the training. Longitudinal studies are needed to confirm this cue integration pattern, but if it indeed takes place in development, it is remarkable that it can surface after only 2 minutes of training.

Some listeners in the Enhanced group started to use F0 after training, which may be related to their response bias to /aː/ (section 3.2). This is because the average F0 of the natural test stimuli, both for the male and female voices, was somewhat lower for /aː/ than for /ɑ/ (Table V.2, section 2.2.1), and thus more similar to the male voice of the response options. Given that F0 is not relevant for determining vowel identity and that the response options did not differ in this cue, this new strategy was likely to have hampered listeners' performance. Indeed, the average accuracy for the Enhanced pre-test high performers decreased when they started to use F0 (compare Table V.6 Enhanced post-test class 2 and Table V.5 Enhanced pre-test class 2), while the average higher accuracy for the Enhanced pre-test low performers who started to use this cue could be based entirely on their new use of F2 and their increased use of duration (Table V.6 Enhanced post-test class 2 versus Table V.5 Enhanced pre-test class 1).

Further, listeners tended to adopt cues in the order duration, F1, F2 and F3. That is, classes that started to use F1 always continued using duration, classes that started to use F2 also started to use or continued using duration and F1, and classes

that started to use F3 also started to use or continued using duration, F1 and F2. In other words, although the analysis of listening strategies after training could have identified several other logically possible strategies (such as F2 alone or F1, F2 and F3), it yielded only four strategies, namely (1) duration, (2) duration and F1, (3) duration, F1 and F2, and (4) duration, F1, F2 and F3. With respect to vowel formants, the observed order seems to reflect a ranking from most to least salient, since lower formants have higher amplitudes in the acoustic signal than higher formants (Klatt, 1980) and differences between two vowels in lower formant frequencies are somewhat easier to discriminate than those in higher formant frequencies (Kewley-Port and Watson, 1994). Possibly because of this perceptual difference listeners started using F1 before F2, despite a larger difference in F2 than in F1 between [ɑ] and [aː] (e.g., the difference between the average natural [ɑ] and [aː] stimuli in the test was 1.76 ERB in F2 and 1.47 ERB in F1). The perceptual difference between F1 and F2 may be related to the larger number of distinctions between vowels in the F1 dimension (three levels in Spanish) than in the F2 dimension (two levels in Spanish) that is observed in the vowel inventories of the world's languages (Ladefoged and Maddieson, 2007).

As for duration, it is not certain whether it is intrinsically more salient as a cue than formants. Spanish listeners in almost all pre- and post-strategy types used duration, despite the fact that they do not use it to distinguish Spanish vowels. This finding is in line with these listeners' attested tendency to resort to duration in order to compensate for their failure to use differences in formant frequencies between non-native vowels (e.g., Escudero and Boersma, 2004; Escudero et al., 2009), and shows that this cue must be fairly accessible. Since duration is also used consistently in non-native speech perception by speakers of other languages than Spanish without native durational differences (e.g., Iverson and Evans, 2007), it has been suggested that the cue is relatively easy to parse for humans in general (Bohn, 1995). Alternatively, the accessibility of duration can stem from the absence of a phonemic contrast along this acoustic dimension in the listeners' L1, as suggested by Escudero and Boersma (2004). Specifically, Escudero and Boersma propose that, when presented with a distribution of speech sounds

differing in duration, speakers of languages without phonemic contrasts along this dimension can form durational categories "from scratch", without interference from existing L1 categories.

Nevertheless, if saliency is indeed the driving force underlying the order in which listeners started to use new cues, this order suggests that in a two-minute distributional training not only the frequency of presentation *across* stimuli affects perception, but also the relative saliency of the acoustic components *within* the presented stimuli. With exposure to a language where the distributional properties of an acoustic cue do not contain linguistically relevant information, it seems that listeners can learn to ignore such a cue, even if it is acoustically salient or accessible. For instance, Spanish listeners without L2 experience do not use duration to distinguish native vowels (e.g., Morrison, 2008). Future research is needed to unravel the precise dynamics between saliency and frequency in distributional learning over a longer time span.

Regarding the nature of development in distributional learning, we expected to find roughly the same developmental stages as posited by Escudero (2000, 2005) and Morrison (2008), as discussed in the Introduction. Although we can only ascertain the existence of these stages with longitudinal data, they can indeed be related to the identified listening strategies. Low performers in the pre-test can be interpreted to be in stage 0, because they could not distinguish /ɑ/ and /aː/ and used duration only slightly. The majority of pre-test low performers started to use duration more intensely after distributional training, which signals a transition to stage 1. Most of them simultaneously started to use F1 (and F2), which corresponds to a transition to stage 2. Moreover, pre-test high performers, who used duration, F1 and F2 in the pre-test, could have started in stage 2 or 3, where listeners attend primarily to F1 and F2, as native listeners do. Indeed, the accuracy of the best-performing Spanish classes (in pre- and post-test) was similar to that of native speakers. This is in line with previous research by Díaz et al. (2012), which showed that in categorization tasks L2 listeners' performance may well reach native-speaker levels. Spanish learners came closer to native speakers'

listening strategies after exposure to distributional training as opposed to classical music.

Interestingly, our approach of focusing on the *content* of what was learned rather than on *attained accuracy* also made it possible to detect progress that was not associated with high performance scores. For instance, the majority of the Enhanced low performers in the pre-test, who turned to duration exclusively after the training and who continued performing badly in the post-test (21 listeners, see section 3.3), could still have progressed from stage 0 to stage 1 because duration, which was introduced in stage 1, was irrelevant for distinguishing the response options. Also, the bias toward /aː/ in the post-test of some Bimodal pre-test low performers who continued to perform rather poorly, could reflect Morrison's developmental stage, where listeners classify vowels as good or bad examples of Spanish /a/. It is conceivable that the Spanish learners in this group only labelled the tokens that were acoustically furthest away from the Spanish vowel /a/ as Dutch /ɑ/.

Importantly, Escudero (2000, 2005) and Morrison (2008) did not view development as necessarily discrete jumps from one stage to another, while we implicitly assumed such categorical transitions because we modelled the listening strategies as distinct types. The current data show that cues can be adopted one by one, as reflected in the strategy types duration, duration–F1, duration–F1–F2 and duration–F1–F2–F3, and that the use of cues can be intensified (or weakened), as reflected by the beta coefficients. However, the clear increase in accuracy when using more cues (i.e., when comparing classes in Figure V.1, accuracy seems to increase dramatically when a cue is added) suggests that the actual transition between stages is categorical rather than gradual. Longitudinal studies are needed to ascertain the developmental stages shown in the present data and their categorical nature.

In sum, we have demonstrated that distributional vowel training can help learners to improve their classification of difficult non-native contrasts. We show that the changes in perceptual cue use after training are related to participants' listening strategies before training. Latent class regression analysis is a way to

identify such strategies. The strategies identified here can be related to previously reported developmental stages for Spanish learners of English and Dutch vowels, which suggests that our method can shed light on the development of second language speech perception.

Chapter VI

Distributional training of speech sounds can be done with continuous distributions

Karin Wanrooij and Paul Boersma

The Journal of the Acoustical Society of America 2013, 133(5), EL398–EL404
doi: 10.1121/1.4798618

Abstract

In previous research on distributional training of non-native speech sounds, distributions were always discontinuous: typically, each of only eight different stimuli was repeated multiple times. The current study examines distributional training with continuous distributions, in which all presented tokens are acoustically different. Adult Spanish learners of Dutch were trained on either a discontinuous or a continuous bimodal distribution of the Dutch vowel contrast /ɑ/~/aː/. Both groups improved their perception of the contrast; this shows that continuous training works equally well as discontinuous training. Using the more natural continuous distributions is therefore recommended for future distributional learning experiments.

1. Introduction

Earlier research has shown that adult learners can improve their discrimination or classification of a non-native speech sound contrast simply by listening for a few minutes to a bimodal distribution representing this contrast (Maye and Gerken, 2000, 2001; Hayes-Harb, 2007; Gulian et al., 2007; Escudero et al., 2011; chapter V). This phenomenon is called "distributional learning." The stimuli differ from one another in steps along an acoustic continuum. For a bimodal distribution, two stimuli with acoustic properties near the end points of the continuum (e.g., the two stimuli with F1 values of 11.9 and 14.0 ERB in Figure VI.1, top) are presented more often than the other stimuli (as represented by the varying line lengths in the figure). Through the differences between the stimuli in their frequency of presentation, listeners supposedly start to treat these two most frequently presented stimuli (and their acoustic neighbours, which are presented slightly less often) as exemplars of two different speech sounds.

1.1. Discontinuous and continuous distributions

In all previous studies on distributional learning, bimodal distributions were based on stimuli with 8 or 10 different values for voice onset time (e.g., Maye and Gerken, 2000, 2001; Hayes-Harb, 2007; Maye et al., 2002, 2008; Yoshida et al., 2010), vowel formants (e.g., Gulian et al., 2007; Escudero et al., 2011; chapter V), or fricative frequencies and formant transitions (Cristià et al., 2011), and these stimuli were repeated in certain proportions. In Figure VI.1 (top), for instance, the eight stimuli (the thin vertical lines) are spaced at equal distances along the F1 continuum, and some stimuli are presented more often than others (the height of the vertical lines), while acoustic values in between those of the eight stimuli are never presented. We therefore label such distributions "discontinuous."

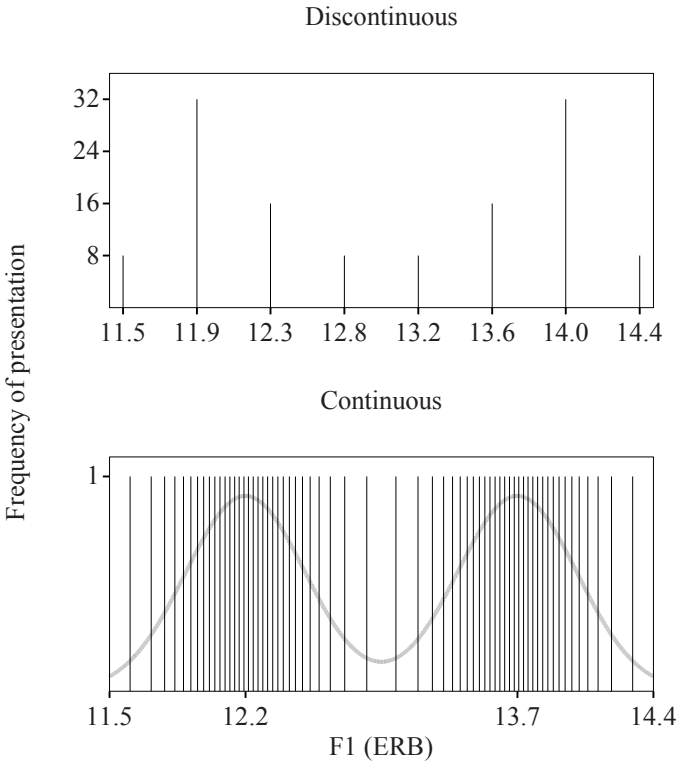


Figure VI.1. A discontinuous (top) and a continuous (bottom) stimulus distribution. Each vertical line represents a stimulus with a specific F1 value. The height of each vertical line shows how often the stimulus is presented to the listener. The grey curve in the bottom picture is the underlying probability density function (see section 2.2).

In a natural environment, however, acoustic values are never repeated exactly. Rather, naturally occurring speech tokens can have any value (between certain bounds) along the relevant acoustic dimension. When applying this idea to a bimodal stimulus distribution for distributional training, we obtain Figure VI.1 (bottom), where the stimuli (the thin vertical lines) are spaced more densely around

12.2 and 13.7 ERB and more sparsely elsewhere, and each stimulus is presented only once. We therefore label such distributions "continuous."

In the current study, we aimed to examine whether previous observations obtained with discontinuous distributions might have been artefacts of the unnatural sampling method. After all, it is known that input variability can influence category formation and discrimination (Lively et al., 1993; Rogers and Davis, 2009), so that one could hypothesize that the observed changes in participants' behaviour after training were due to the artificially sparse (eightfold) sampling of the acoustic space. To find out whether the effects reported in the distributional learning literature have not been methodological artefacts, one would have to test whether adult listeners also improve classification performance through listening to a more ecologically valid continuous distribution, with more variation in acoustic values, and without stimulus repetition. This is done in the present article, which compares three groups of participants: one group was presented with a discontinuous training (hence, the Discontinuous group), another group with a continuous training (hence, the Continuous group), and the third group was a control group that listened to classical music (the Music group). As explained in section 2.1, the Discontinuous and Music groups were taken from chapter V.

1.2. A vowel contrast and its appropriate participant group

For the acoustic continuum we chose the Dutch vowel contrast /ɑ/~/aː/. For this contrast, appropriate listeners are native speakers of Spanish. This group is known to have difficulty classifying the two Dutch vowels when the durational difference (/aː/ is longer; Adank et al., 2004) is eliminated, so that only the spectral difference (/aː/ has higher first and second formants; Pols et al., 1973; Adank et al., 2004) can be used to classify the vowels correctly (Escudero and Wanrooij, 2010; Escudero et al., 2011; chapter V). To train the Spanish listeners on this spectral difference only, the manipulated acoustic dimensions in both distributions (i.e., discontinuous and continuous) were the first and second formant values (F1 and F2), and the duration

of the training vowels was kept constant (see also Escudero et al., 2011; chapter V). Note that Figure VI.1 shows the discontinuous and continuous distributions of the F1 values only; because F2 values varied linearly with those of F1, the pictures for the discontinuous and continuous F2 distributions look identical.

2. Method

The method was identical to that of Escudero et al. (2011) and chapter V. Participants performed a pre-test (section 2.3), a training phase (section 2.2) and a post-test (section 2.3).

2.1. Participants

Participants were adult native speakers of Spanish who were learning Dutch. Only the Continuous group was new, and it consisted of 50 participants. The Discontinuous group was taken from an earlier study as follows. To ensure a high-level benchmark for the Continuous group, we chose for our Discontinuous group the group that had shown the most improvement of all four groups that received discontinuous distributional training in two recent studies in our lab (Escudero et al., 2011; chapter V). These two studies used identical pre- and post-tests and an identical procedure as those used for the Continuous group in the present study, and in both studies the results were the same. Specifically, in both studies three groups of Spanish listeners participated, one presented with a discontinuous bimodal distribution representing the Dutch contrast /ɑ/~/aː/ (the Bimodal group), one exposed to a discontinuous *enhanced* bimodal distribution of the same contrast (the Enhanced group), and one, the control condition, presented with classical music (the Music group). In the enhanced bimodal distribution, the perceptual distance between the end point acoustic values of the training stimuli was larger than that in the non-enhanced bimodal distribution. Accordingly, the difference between the two speech sounds was "exaggerated" and thus presumably easier to perceive (Kuhl et al., 1997; Liu et al., 2003).

With new Spanish participants in the Bimodal and Enhanced groups (50 in each group), chapter V replicated the results obtained for the participants (53 in each group) in Escudero et al. (2011), i.e., that (1) the Enhanced group improved significantly in accuracy of classification of Dutch /ɑ/ and /aː/, (2) the Music group did not show significant progress, and (3) the Enhanced group improved significantly more than the Music group. Table VI.1 shows the difference scores (i.e., post-test minus pre-test classification accuracy in percentages) for each group (i.e., Enhanced, Bimodal, Music) in both studies.

Table VI.1. Difference score (= post-test minus pre-test accuracy percentage) for groups of Spanish listeners presented with enhanced, bimodal and musical training phases in two previous studies. 95% confidence intervals are given between parentheses.

Previous study	Enhanced	Bimodal	Music
Escudero et al.	6.04	0.80	–0.15
(2011)	(+2.76 ~ +9.31)	(–2.22 ~ +3.83)	(–3.50 ~ +3.21)
Chapter V	6.63	3.83	2.00
	(+4.05 ~ +9.20)	(+0.97 ~ +6.68)	(–0.50 ~ +4.50)

It can be observed that the Enhanced group in chapter V had the highest absolute improvement after training of all four groups, with a 95% confidence interval (CI) that appeared narrower and further from zero than in Escudero et al. (2011). Therefore we used this group as a stringent standard against which to compare our Continuous group. In addition, we compared the results of the Discontinuous and Continuous groups to the Music group's results as obtained in chapter V. In Escudero et al. (2011) and in chapter V, a music condition was preferred over a unimodal control condition for ethical reasons, because all participants were learners of Dutch and previous research had shown that a unimodal distribution may reduce discrimination performance (Maye et al., 2002; Appendix to chapter III). Table VI.2 lists the mean age, age range, and length of

residence in the Netherlands as a measure of previous exposure to Dutch, for the Discontinuous (12 male, 38 female), Continuous (15 male, 35 female), and Music (6 male, 44 female) groups separately.

Table VI.2. Participants' age (standard deviation between parentheses), age range, and length of residence (in years) in the Netherlands.

Group	Mean age	Age range	Length of residence
Music*	38.0 (9.0)	19.0–60.0	6.3 (6.8)
Discontinuous*	37.3 (8.0)	21.0–56.0	5.4 (5.0)
Continuous	33.2 (9.8)	21.6–63.2	3.1 (4.9)

*The Discontinuous and Music groups were taken from chapter V.

2.2. Training: stimuli and procedure

The stimuli in the continuous and discontinuous training distributions were made with the Klatt synthesizer in the computer program PRAAT (Boersma and Weenink, 2011). Each stimulus in both distributions had a fundamental frequency (F0) contour that fell from 150 to 100 Hz. Also the stimulus duration was 140 milliseconds (ms), and the inter-stimulus interval was 750 ms. Total training time was nearly 2 minutes.

The stimuli in the *discontinuous* distribution are described in detail in Escudero et al. (2011) and in chapter V. The F1 values (range: 11.5–14.4 ERB) and F2 values (range: 15.3–18.2 ERB) varied in eight steps of approximately 0.4 ERB apart. Stimuli 1 through 4 with the lower F1 and F2 values can be thought of as representing the Dutch vowel /ɑ/, and stimuli 5 through 8 with the higher F1 and F2 values can be thought of as representing the Dutch vowel /aː/. Stimuli 1, 4, 5, and 8 in the tails (see Figure VI.1, top) were each presented eight times, stimuli 2 and 7 at the peaks each occurred 32 times, and stimuli 3 and 6 were each presented 16 times. Thus the total number of presentations was 128.

To make a *continuous* distribution that would correspond as closely as possible to the discontinuous one, we first had to match the shapes of the distributions. For this, we approximated the ratio of the least to most frequent stimuli; i.e., this ratio is 1 to 4 in earlier studies with discontinuous distributions and is approximately 1 to 4 in the current continuous distribution (see Figure VI.1). Further, we created the underlying continuous distribution as the sum of two Gaussian curves the means of which were positioned at 25% and 75% of the F1 range (and consequently also of the F2 range), and the standard deviations of which were set to 11% of the total F1 (or F2) range. This distribution is the probability density function shown in Figure VI.1 (bottom).

The next step was the determination of the F1 and F2 values for each stimulus. We created the same total number of stimuli (128) as for the discontinuous distribution. This time none of the stimuli was repeated, so that each stimulus had a unique combination of F1 and F2 values. As the procedure for the calculation of the F2 values is the same as that for the F1 values, we restrict the description to the F1 values, as follows.

After determining the precise shape of the underlying continuous distribution (the grey curve in Figure VI.1, bottom), the F1 values of the 128 stimuli (the thin vertical lines in Figure VI.1, bottom; for the purpose of clarity only 64 stimuli are shown) were calculated in the following way. First, the area under the curve was normalized, i.e., it was set to 128, the number of stimuli. Then the distribution was sampled evenly, i.e., the F1 values were chosen in such a way that the area between consecutive F1 values under the curve was always 1. Thus there were 127 unit areas between the 128 F1 samples. The additional leftmost area (running from the left edge of the F1 continuum to the first F1 sample) and rightmost area (running from the last F1 sample to the right edge of the F1 continuum) were 0.5 each.

The task of the participants in the training phase was merely to listen. Participants in the Discontinuous and Continuous groups were instructed to listen to the vowels carefully because they would perform a post-test afterward.

Participants in the Music group were asked to relax while listening to the music and were informed that they would perform a post-test afterward.

2.3. Pre- and post-tests: stimuli and procedure

The pre- and post-tests, which were equal to those used in Escudero et al. (2011) and in chapter V, were identical classification tasks, which were the same for all participants. Listeners heard an X-stimulus and two subsequent response options A and B. They were forced to choose which option was from the same vowel category as X.

The X-stimuli were chosen to be natural vowels to promote classification rather than discrimination; they were a subset of the vowels reported in the corpus by Adank et al. (2004), which were produced by male and female speakers of standard Northern Dutch. The response options A (F1 = 12.5 ERB, F2 = 16.1 ERB) and B (F1 = 13.3 ERB, F2 = 17.4 ERB) were chosen to be synthetic; they were created with the computer program PRAAT (Boersma and Weenink, 2011) and had an equal duration of 140 ms to prevent participants from resorting to durational differences between /ɑ/ and /aː/ (recall section 1.2).

In each test, participants were asked to classify 80 X-stimuli. Listeners were told that the next trial would only appear after their response, but they were encouraged to answer as quickly as possible and to guess if they were unsure. To test hearing and understanding of the test, the participants performed a practice test before the pre-test and before the post-test.

3. Results

Table VI.3 gives the pre- and post-test percentages correct (i.e., the percentage of correct classifications of the 80 test stimuli) and the difference (i.e., the post- minus pre-test percentage correct) for the Music, Discontinuous, and Continuous groups. An ANOVA on pre-test accuracy did not display a significant difference between

the three groups ($F[2,147] = 0.40$, $p = 0.67$). This supports the equality of the groups before training.

Table VI.3. Pre- and post-test percentages correct, and difference (= post- minus pre-test percentage correct) per group. Standard deviations between participants in each group are given between parentheses.

Group	Pre	Post	Difference
Music*	61.73 (11.12)	63.73 (13.31)	2.00 (8.81)
Discontinuous*	60.43 (11.71)	67.05 (13.48)	6.63 (9.06)
Continuous	62.40 (10.74)	72.08 (13.12)	9.68 (10.13)

*Discontinuous and Music groups from chapter V.

The difference between pre- and post-test accuracy is a measure of improvement after training. For the Continuous group, this difference was 9.68% (95% CI = +6.80% ~ +12.55%), which was significantly different from zero (one-sample $t[49] = 6.75$, $p < 0.001$). As reported in chapter V, the difference score also differed from zero significantly for the Discontinuous group (one-sample $t[49] = 5.17$, $p < 0.001$), and it did not for the Music group (one-sample $t[49] = 1.61$, $p = 0.12$) (95% CIs: see Table VI.1). This confirmed that both the Discontinuous and the Continuous groups improved their accuracy percentages robustly after training. An ANOVA with difference scores as the dependent variable revealed a significant difference between groups ($F[2,147] = 8.54$, $p < 0.001$). *Post hoc t*-tests on the difference scores using Tukey's HSD for multiple-comparison corrections showed a significant difference between the Music and Discontinuous groups of +4.63% (CI = +0.20% ~ +9.05%, $p = 0.04$) and between the Music and Continuous groups of +7.68% (CI = +3.25% ~ +12.10%, $p < 0.001$), and no significant difference between the Discontinuous and Continuous groups (difference = +3.05%, CI = −1.38% ~ +7.48%, $p = 0.24$). Thus participants who received distributional training improved more than participants who listened to music instead, although we cannot say with confidence that the

progress of the Continuous group (9.68%) was larger than that of the Discontinuous group (6.63%).

4. Conclusion

We showed that listeners' performance in classifying a non-native phoneme contrast can be improved not only by training them with a discontinuous distribution but also by training them with a continuous distribution. We can therefore erase the fear that earlier results demonstrating an effect of training with discontinuous distributions (e.g., Maye and Gerken, 2000, 2001; Maye et al., 2002, 2008; Hayes-Harb, 2007; Gulian et al., 2007; Yoshida et al., 2010; Escudero et al., 2011; Cristià et al., 2011; chapter V) could have been artefacts of the discontinuous sampling method; after all, these results have now been replicated with the arguably more natural continuous distributions, so it has become more likely that the observed perceptual improvements are a realistic result of bimodal training. However, as both types of sampling have now been shown to exhibit distributional learning effects and continuous distributions can be considered more ecologically valid than discontinuous distributions, we recommend for future distributional learning experiments not to artificially reduce the variation in the stimuli to 8 or 10 auditory values but to solely employ continuous distributions.

Chapter VII

Observed effects of "distributional learning" may not relate to the number of peaks. A test of "dispersion" as a confound.

Karin Wanrooij, Paul Boersma, and Titia Benders

(under review)

Abstract

Distributional learning of speech sounds is learning from simply being exposed to frequency distributions of speech sounds in one's surroundings. In laboratory settings, the mechanism has been reported to be discernible already after a few minutes of exposure, in both infants and adults. These "effects of distributional training" have traditionally been attributed to the difference in the *number of peaks* between the experimental distribution (two peaks) and the control distribution (one or zero peaks). However, none of the earlier studies fully excluded a possibly confounding effect of the *dispersion* in the distributions. Additionally, some studies with a non-speech control condition did not control for a possible difference between *processing speech and non-speech*. The current study presents an experiment that corrects both imperfections. Spanish listeners were exposed to either a bimodal distribution encompassing the Dutch contrast /ɑ/~/a/ or a unimodal distribution with the same dispersion. Before and after training, their accuracy of categorization of [ɑ]- and [a]-tokens was measured. A traditionally calculated *p*-value showed no significant difference in categorization improvement between bimodally and unimodally trained participants. Because of this null result, a Bayesian method was used to assess the odds in favour of the null hypothesis. Four different Bayes factors, each calculated on a different belief in the truth value of previously found effect sizes, indicated the absence of a difference between bimodally and unimodally trained participants. The implication is that "effects of distributional training" observed in the lab are not induced by the number of peaks in the distributions.

1. Introduction

1.1. Distributional learning

The term "distributional learning" refers to learning from simply being exposed to frequency distributions of stimuli in one's surroundings (Lacerda, 1995; Guenther and Gjaja, 1996). Distributional learning is considered one of the mechanisms with which infants start learning the speech sounds of their native language (e.g., Maye et al., 2002). There is also evidence of this mechanism in adults who try to master difficult non-native speech sound contrasts (e.g., Maye and Gerken, 2000).

Distributional learning of speech sounds can be explained as follows. When one acoustic property (e.g., the first formant, F1) is measured across many tokens of a certain speech sound category (e.g., a certain vowel), most values are likely to be observed close to the mean of that category. This is illustrated in Figure VII.1. The x-axes represent an F1 continuum, for which the F1 values are expressed in ERB (Equivalent Rectangular Bandwidth); each vertical line marks the F1 value hypothetically measured in a token of the Spanish vowel /a/ (Figure VII.1, top), and in a token of the Dutch vowels /ɑ/ or /a/ (Figure VII.1, bottom). It is apparent that the F1 values tend to cluster around certain values, which are the means of the categories. Accordingly, the probability density functions (the grey curves in Figure VII.1) of the F1 values have peaks here. Conversely, the number of peaks observed in a probability density function is indicative of the number of speech sound categories along the corresponding acoustic continuum. Frequency distributions such as the schematic one in Figure VII.1 have been observed for several speech sound categories (e.g., Lisker and Abramson, 1964; Newman et al., 2001; Lotto et al., 2004).

Distributional learning implies that exposure to such speech sound distributions induces listeners to perceive tokens with acoustic values that occur within one peak as exemplars of the same speech sound category. The idea is that exposure to the Dutch language, and thereby to the F1 distribution at the bottom of Figure VII.1, prepares Dutch listeners for perceiving vowel tokens with F1 values of around 12.2 ERB as belonging to one speech sound category (namely /ɑ/), and

vowel tokens with F1 values of around 13.6 ERB as belonging to another speech sound category (namely /a/), while exposure to the Spanish language, and thereby to the F1 distribution at the top of Figure VII.1, prompts Spanish listeners to perceive these same vowel tokens as exemplars of one single speech sound category (namely Spanish /a/).

The just-described distributional-learning mechanism has been tested empirically in the lab, where perceptual tuning to the number of peaks in the input distribution has been reported to occur already after a few minutes of exposure, for both infants and adults (for infants: Maye et al., 2002; Maye et al., 2008; Yoshida et al., 2010; Capel et al., 2011; chapter II; for adults: Maye and Gerken, 2000, 2001; Hayes-Harb, 2007; Gulian et al., 2007; Escudero et al., 2011; chapters V and VI; Escudero and Williams, 2014). In a typical distributional-learning experiment, two groups of participants (e.g., native speakers of Spanish) are exposed to speech sound distributions encompassing a not yet acquired speech sound contrast (e.g., the Dutch vowel contrast /ɑ/~/a/): one group is presented with a *unimodal* training distribution (i.e., with *one peak*, as in an F1 distribution of the Spanish vowel /a/) and another group with a *bimodal* training distribution (i.e., with *two peaks*, as in an F1 distribution of the Dutch vowel contrast /ɑ/~/a/). Such training distributions have been "discontinuous" or "continuous" (chapter VI). Discontinuous distributions contain only a limited number of acoustically different stimuli, which are each repeated a certain number of times according to the respective distribution. Examples of discontinuous distributions are shown in Figure VII.3 (section 1.2.2). Continuous distributions consist of a large number of acoustically different stimuli, each of which is presented only once. The acoustic values are chosen to be such that they match the intended probability density function. Examples of continuous distributions are shown in Figure VII.4 (section 2.2.1). After exposure to the speech sound distribution, participants are tested on their discrimination or categorization of representative tokens of the contrast involved (e.g., [ɑ]- and [a]-tokens). If the distributional-learning mechanism is effective, it is expected that bimodally trained participants will discriminate or categorize these test stimuli better than unimodally trained participants. This difference between the

groups is expected because only the bimodally trained participants have been exposed to a distribution that suggests the existence of a contrast between the two categories.

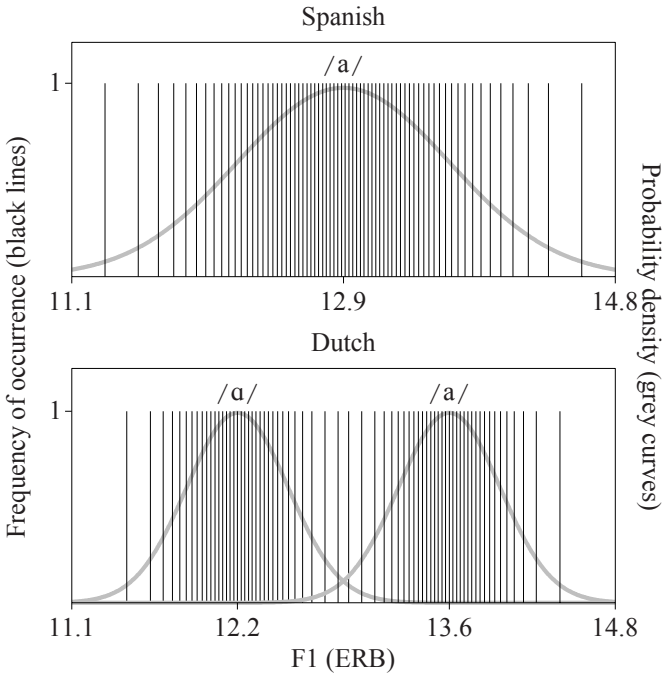


Figure VII.1. Distributions of first formant (F1) values (in ERB), representative of the Spanish vowel /a/ (top) and the Dutch vowel contrast /ɑ/~/a/ (bottom). Each solid vertical line represents a hypothetically measured vowel token with a specific F1 value. The grey curves are the underlying probability density functions.

1.2. Problems in previous research on distributional learning

Studies on distributional learning (previous section) have focused on the *number of peaks* as the relevant factor that shapes the distributional learning process. Unfortunately, it is not certain that the reported effects of distributional learning in

these studies were truly due to perceptual changes induced by the number of peaks in the distributions. The chosen methodologies leave open the possibility that other factors caused these reported effects. Specifically, none of the earlier studies fully equated the training distributions on the amount of *dispersion,* as expressed in for instance the range and the standard deviation of the acoustic values (section 1.2.2). The lack of control for dispersion may be an important omission in the light of indications that the dispersion of acoustic values in the training stimuli can affect speech sound acquisition (section 1.2.1). Evidence even exists that measures of dispersion (such as the range and the standard deviation) in a training distribution may exert more influence on perception than measures of central tendency (such as the mean; Holt and Lotto, 2006: 3066). A second possible confounding effect in some studies with a non-speech control group, is the effect of *processing speech versus non-speech* (section 1.2.3). The two potential confounds are discussed in turn.

1.2.1. The role of dispersion in speech sound learning

Indications that the dispersion of the acoustic values in speech sound distributions can influence adults' speech sound learning can be found in studies reporting that training with "enhancement" leads to changes in adults' perception (e.g., Jamieson and Morosan, 1986). Enhancement refers to the widening of the acoustic distance between speech sound categories, thereby affecting the dispersion in the presented stimulus distributions. The precise effect of enhancement on the dispersion depends on the way in which it is implemented in the training paradigm. In distributional training experiments, it has been implemented by giving enhanced bimodal distributions a larger acoustic difference between the means (i.e., the two peaks in the distribution[1], each of which represents a speech sound category), a wider range, and a larger standard deviation than non-enhanced bimodal

1 The true bimodal means are somewhat closer together than the two peaks.

distributions (Escudero et al., 2011; chapter V).[2] These three factors are of course strongly interdependent. Figure VII.2 demonstrates the difference between the non-enhanced (top) and enhanced (bottom) distributions.

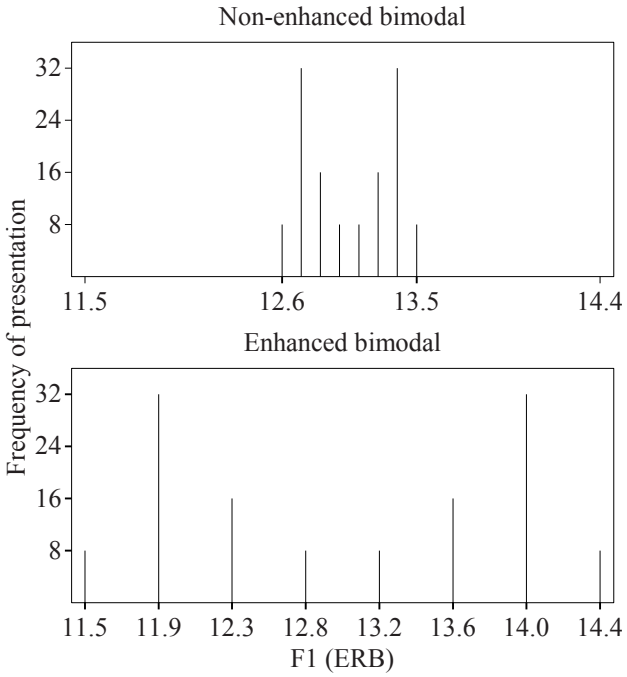


Figure VII.2. Non-enhanced (top) and enhanced (bottom) bimodal distributions of F1 values in the Dutch vowel contrast /ɑ/~/a/, as used in Escudero et al., 2011, and chapter V.

In other training experiments, where participants typically receive feedback during categorization training, enhancement has been implemented by "perceptual fading"

2 Specifically, the values in Escudero et al. (2011) and chapter V were as follows. In the non-enhanced bimodal distribution, the distance between the peaks was 0.67 ERB, the range was 12.60 to 13.54 ERB, and the standard deviation of the pooled distribution was 0.31 ERB. In the enhanced bimodal distribution, the distance between the peaks was 2.02 ERB, the range was 11.52 to 14.35 ERB, and the standard deviation was 0.93 ERB.

(Jamieson and Morosan, 1986), a technique originally applied to visual discrimination learning in birds (Terrace, 1963). With this technique, participants are first presented with exemplars of each speech sound category whose acoustic properties are "enhanced", thus presumably making it easier to hear a difference between the categories. If the participant categorizes the exemplars well, the acoustic difference between the categories is reduced in small steps. As the actually presented distributions depend on participants' performance and thus vary per participant, studies using this technique do not always specify the distribution in terms of means and measures of dispersion. Nevertheless, the initial enhancement is likely to widen the dispersion of the presented distributions in comparison to distributions without such enhancement.

Although direct comparisons between the effects of enhanced and non-enhanced training tend to yield non-significant results (e.g., Iverson et al., 2005; Escudero et al., 2011), enhanced training (both enhanced distributional training and training with perceptual fading) generally leads to improved categorization or discrimination of the trained speech sound categories after as compared to before training (Jamieson and Morosan, 1986; Iverson et al., 2005; Kondaurova and Francis, 2010) and in addition sometimes also as compared to a control group that received no training with speech sound stimuli (McCandliss et al., 2002; Escudero et al., 2011; chapters V and VI). These improvements leave open the possibility that enhancement of the speech sounds presented during training (likely affecting the range and the standard deviation of a speech sound distribution) indeed affects speech sound learning in adults.

The observed benefit of enhancement in distributional training studies could be due to better distributional learning (Escudero et al., 2011; chapter V). However, the assumed benefit of enhancement in perceptual fading studies is usually not attributed to better distributional learning but to a facilitation of "attentional learning", i.e., learning through focusing one's "attention" on the relevant differences between speech sound categories (e.g., Jamieson and Morosan, 1986; Francis and Nusbaum, 2002; Iverson et al., 2005; Kondaurova and Francis, 2010). Such attentional learning is also raised as an additional explanation (apart

from better distributional learning) for improved categorization after training in distributional training studies. Perceptual fading studies that focus on attentional learning generally leave the concept of attention undefined, but it looks as if attention in these studies is mediated by existing knowledge (about, for instance, native speech sound categories; Logan et al., 1991: 882) or knowledge obtained during the experiment in the form of feedback (e.g., McCandliss et al., 2002). Such attention can be related to top-down processes in the brain (Posner, 1990; Roelfsema, 2011). Attentional learning thus seems to contrast with distributional learning, which is viewed as a purely stimulus-driven, bottom-up process (Lacerda, 1995; Guenther and Gjaja, 1996).

At the same time, our understanding of attentional learning and distributional learning (assuming that they exist) is poor, and it is difficult to establish that they are truly separate processes. For instance, *both* predict that the learning of a speech sound contrast should improve from enhancement if enhancement is implemented by only pulling the means of the two categories wider apart without changing each peak's standard deviation. Such an enhancement method could draw participants' attention to the differences between the categories (thus advancing attentional learning) *and* would reduce the overlap between the two peaks (thus promoting distributional learning)[3]. Accordingly, improvement of discrimination or categorization performance after such enhanced distributional training could be accounted for by both distributional learning and attentional learning. Experiments designed to demonstrate the existence of the distributional learning mechanism must exclude the possibility that the results can be explained through attentional learning, and must thus use the same dispersion in the experimental (two peaks) and the control (one or zero peaks) distributions.

In sum, even though it is still unclear precisely what role measures of dispersion in distributions play in adults' speech sound learning, there are several indications that such measures do play a role. Accordingly, it is important to

3 Note that enhancement of the contrast reduces the overlap between the categories if the standard deviations of each peak remain the same. The overlap is not necessarily reduced if the standard deviation of each peak is increased as well (as it is in Figure VII.2).

exclude a possibly confounding influence of dispersion in distributional training experiments. An equal dispersion in the distributions to be compared would also reduce the possibility that differences in attentional learning between training conditions could account for the results, rather than differences in distributional learning.

1.2.2. No adequate control for dispersion across distributional learning studies

None of the previous studies on distributional learning, neither those with infants nor those with adults (mentioned in section 1.1), fully excluded dispersion as a possible factor that can account for the observed differences between the bimodal training groups and the control groups. Three possible measures of dispersion are the range, the standard deviation, and the "edge strength". These are discussed here in turn.

The first measure of dispersion is the range. Typical bimodal and unimodal distributions such as those in Maye et al. (2008) have the same range within a study: the minimum and maximum presented values are the same in the one as in the other distribution (see Figure VII.3). Range was not excluded as a possibly confounding effect in four studies on distributional learning that used a music control group instead of a unimodal control group (Escudero et al., 2011; chapters V and VI; Escudero and Williams, 2014). These four studies investigated the effect of distributional training on Spanish listeners' categorization of vowel tokens representing the Dutch vowel contrast /ɑ/~/a/. In all four studies, listeners to an enhanced bimodal distribution improved significantly more in categorization accuracy than listeners to music.[4] This result could be due to distributional learning, and thus to the presence of two peaks in the enhanced bimodal distribution. However, the use of a music control group instead of a unimodal

4 In Escudero and Williams (2014), who investigated longer-term effects of distributional training (i.e., after 6 and 12 months rather than only after a few minutes), a significant difference between listeners to an enhanced bimodal distribution and listeners to music, was only found in a subset of the tests.

control group leaves open the possibility that the reported effect is related to the wide range of presented acoustic values in the enhanced bimodal distribution.

The second measure of dispersion, the standard deviation, is larger for the bimodal distribution than for the unimodal distribution across studies with a unimodal control group. For instance, if we take typical unimodal and bimodal distributions with stimulus frequencies as in Maye et al. (2008) and if we take a hypothetical acoustic continuum in which each step along the continuum has an identical psychoacoustic distance of 1 (see Figure VII.3), the standard deviation of the unimodal distribution is 1.7 and that of the bimodal distribution is 2.3.[5] In studies with a music control group, the standard deviation of the (enhanced) bimodal distribution cannot be compared to that of the music condition, so that here too (i.e., just as in the studies with a unimodal control group) the possibility remains open that the reported effects of distributional training are related to the large standard deviation in the bimodal distribution rather than to the presence of two peaks.

Our third measure of dispersion is the "edge strength". This term refers to the density of stimuli in the leftmost and rightmost tails of the distribution (the "edges"). It is conceivable that a large edge strength can draw participants' attention to the relevant differences between stimuli, just as a wide range and standard deviation may do (see section 1.2.1). Specifically, the more stimuli are sampled at the edges rather than in the middle of the distribution, the more the listeners' attention can be drawn towards the end points of the continuum, rather than towards the middle. In view of the above, the reported effect of distributional training in the studies with a music control group may have been due to the large

5 Notice that the standard deviations of the *distributions* are compared, not those of the *individual peaks*. (In Figure VII.3, the standard deviations of the individual peaks would be 0.8 for each peak in the bimodal distribution and 1.7 for the unimodal peak). A smaller standard deviation of each bimodal peak than of the unimodal peak is not problematic in a distributional-learning experiment, because it supports the experimental design. Specifically, in the bimodal distribution both the presence of two peaks and the smaller standard deviation of each peak than in the unimodal distribution promote the distributional learning of two separate categories, while conversely in the unimodal distribution both the presence of a single peak and the larger standard deviation of this peak than in the bimodal distribution promote distributional learning of a single category (Guenther and Gjaja, 1996).

edge strength in the enhanced bimodal distribution rather than to the presence of two peaks. Many studies with a *unimodal* control group and an eight-step discontinuous distribution ensured that the stimuli with minimum and maximum values were equally frequent in the unimodal and the bimodal training (e.g., Maye et al., 2008; see Figure VII.3: stimuli number 1 and 8 were each presented eight times in both distributions). Thus, when computed with edges at 1/8 of the range, the bimodal and unimodal distributions in these studies have equal edge strengths. However, when computed with edges at a larger portion (e.g., 1/6) of the range, the bimodal distributions have a greater edge strength. This illustrates that the edge strength depends on the chosen width of the edges. Since it is not known how wide edges must be to avoid a confounding influence of attention to the edges, it remains a possibility that the reported effect of distributional training in the studies with a unimodal control group (just as in the studies with a music control group) was based on a larger edge strength in the bimodal group than in the control group.

In sum, previous research on distributional learning has not fully excluded a possible learning effect based on measures of dispersion, such as the range (in some studies), the standard deviation (in all studies), and the edge strength (depending on the choice of the edges in some or all studies).

1.2.3. No adequate control for processing speech versus non-speech

A significant difference in categorization improvement after distributional training between a group exposed to an enhanced bimodal distribution and a group exposed to music (Escudero et al., 2011; chapters V and VI; as discussed in section 1.2.2 of the current chapter) could not only be attributed to a difference in the number of peaks or to a difference in the dispersion of the acoustic values between the two conditions (as explained in section 1.2.2), but also more generally to a difference between *processing speech* as during the enhanced bimodal training and *processing non-speech* as during the musical training phase. Differences in processing speech versus non-speech are well-documented and include indications that speech is processed along different routes in the brain than non-speech (e.g.,

Dehaene-Lambertz et al., 2005). Such differences are not related to distributional learning, which is supposedly not based on different processing routes during the bimodal training than the control training, but rather, as supported by computer simulations, on a different tuning of neurons in low-level cortical areas such as the primary auditory cortex (Guenther and Gjaja, 1996).

In sum, the previously reported effects of distributional training in studies with only a non-speech control group, could be related to a difference between processing speech and processing non-speech rather than to a difference in the number of peaks in the distribution.

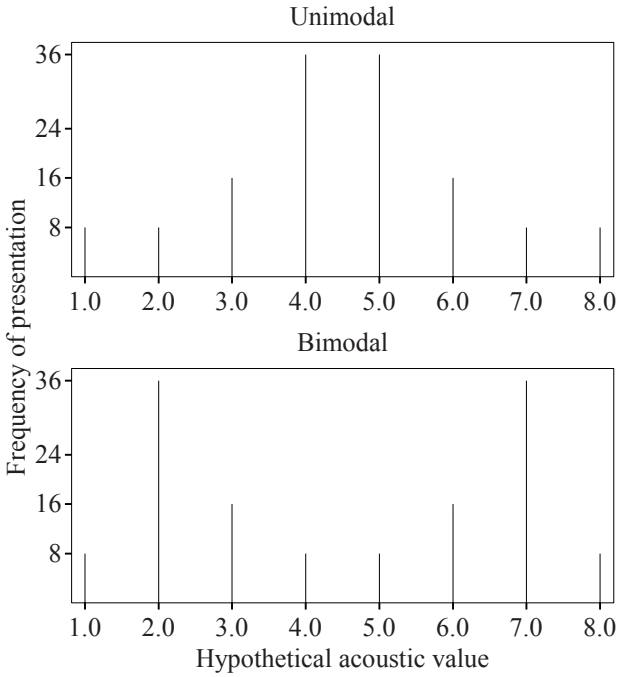


Figure VII.3. Unimodal (top) and bimodal (bottom) training distributions of a hypothetical acoustic value (with an equal psychoacoustic distance of 1 between subsequent values along the continuum), with the frequencies of presentation as used in Maye et al. (2008: figure on page 125).

1.3. Solving the problems: an equally wide unimodal control distribution

The present study followed four previous distributional training studies (Escudero et al., 2011; chapters V and VI; Escudero and Williams, 2014) in the choice of the population and of the vowel continuum appropriate for these listeners: native speakers of Spanish were exposed to distributions along the spectral contrast between the Dutch vowels /ɑ/ and /a/. /a/ has a higher F1 and a higher second formant, F2 (Pols et al., 1973; Adank et al., 2004). This spectral contrast is difficult to learn to perceive for Spanish listeners (Escudero et al., 2009; Escudero and Wanrooij, 2010), but it is the main cue for most native speakers of Dutch (Escudero et al., 2009; Van Heuven et al., 1986). Also in line with the four previous studies, participants were tested on their categorization accuracy of naturally produced [ɑ]s and [a]s before and after training.

In order to determine whether the *number of peaks* (factor 1) in a speech sound distribution tunes participants' perception, and is thus the factor behind the results in distributional-learning experiments, it was necessary to exclude *dispersion* (factor 2) and *processing differences between speech and non-speech* (factor 3) as possible confounds. This can be done by using an experimental distribution and a control distribution that only differ in the number of peaks (factor 1 still present), and which thus have an equal dispersion (factor 2 excluded) and are both speech sound distributions (factor 3 excluded).

The experimental distribution in the current study was based on the "enhanced" bimodal distribution used in Escudero et al. (2011) and chapter V for the same continuum and population, because these studies found a significantly better improvement in vowel categorization after exposure to this distribution than after exposure to music. The control distribution in the present study was a unimodal distribution of speech sounds with the same dispersion (as defined by the range, standard deviation and edge strength; section 1.2.2) as this bimodal distribution. We will henceforth refer to the participants listening to the bimodal distribution as the *Bimodal* group, and to the participants presented with the unimodal distribution as the *Unimodal* group.

By using bimodal and unimodal distributions with an equal dispersion, we rule out the possibility that differences in improvement of categorization between the Bimodal and Unimodal groups can be due to differences in dispersion (factor 2). By using only speech sound distributions, we preclude that dissimilar processing of speech versus non-speech (factor 3) plays a role in any differences found between the two groups. Thus, if we find that the Bimodal group improves significantly more than the Unimodal group, we can confidently attribute this difference to an effect of the number of peaks (factor 1). There will be no straightforward explanation if the reverse result occurs, i.e., if the Unimodal group improves more than the Bimodal group.

If no significant difference (in terms of *p*-values) between the two groups emerges, we are confronted with a *null result* that does not allow us to conclude whether the number of peaks plays a role or not. This problem will be addressed by the computation of Bayes factors (e.g., Kass and Raftery, 1995; Rouder et al., 2009), which allow us to quantify the relative credibilities of the alternative hypothesis (e.g., that the Bimodal group will improve by a certain amount more than the Unimodal group) *and* the null hypothesis (that there will not be a difference in improvement between the two groups).

2. Method

Unless stated otherwise, the method was identical to that used in Escudero et al. (2011) and in chapters V and VI. Spanish adult learners of Dutch (section 2.1) went through a training phase (section 2.2.1), and before and after this training they performed a test that assessed their categorization of several Dutch [ɑ]- and [a]-tokens (section 2.2.2). A comparison of post-test to pre-test accuracy scores determined participants' improvement in categorization performance.

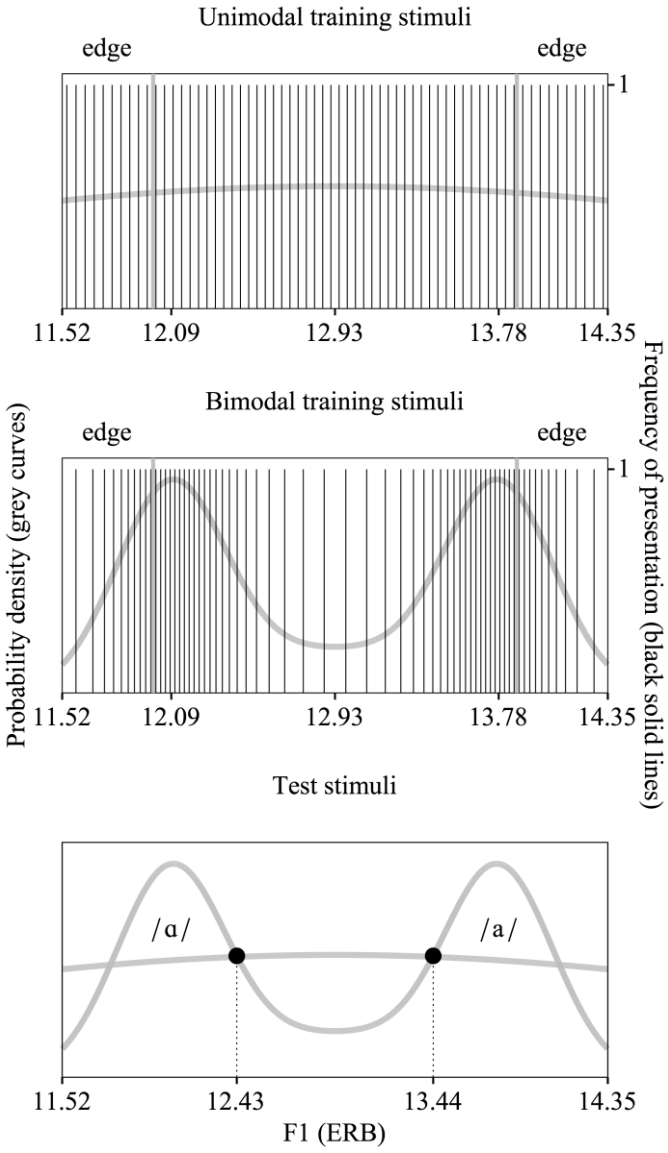
2.1. Participants

The participants were adult native speakers of Spanish. They were semi-randomly assigned to either the Unimodal group or to the Bimodal group (section 1.3), each eventually containing 60 participants. Assignment to the groups was not completely random, because we balanced the groups in terms of age, sex and length of residence in the Netherlands, in this order of importance. Table VII.1 presents the mean age, age range and mean length of residence, in the Unimodal (32 men, 28 women) and Bimodal (26 men, 34 women) groups.

Table VII.1. Participants' age, age range, and length of residence (in years) in the Netherlands, for the Unimodal and Bimodal groups. For age and length of residence, standard deviations within each group are given between parentheses.

Group	Mean age	Age range	Mean length of residence
Unimodal	30.2 (7.3)	20.0 – 56.3	1.2 (1.4)
Bimodal	31.0 (8.0)	18.7 – 52.6	1.4 (2.0)

Figure VII.4 (opposite page). The unimodal (top) and bimodal (middle) training distributions of F1 values used in the present experiment, with an equal range and a nearly equal standard deviation and edge strength (explanation: see text). The unimodal distribution represents the Spanish vowel /a/ and the bimodal distribution is representative of the Dutch vowel contrast /ɑ/~/a/. Each vertical line shows the F1 value of a single stimulus. (For the purpose of clarity only 64 values are shown, rather than the 256 values used). The F1 values of the test stimuli lie at the intersections of the two distributions (bottom).


Unimodal training stimuli
edge
edge
1
11.52
12.09
12.93
13.78
14.35
Bimodal training stimuli
edge
edge
1
11.52
12.09
12.93
13.78
14.35
Probability density (grey curves)
Frequency of presentation (black solid lines)
Test stimuli
/ɑ/
/a/
11.52
12.43
13.44
14.35
F1 (ERB)

2.2. Stimuli and procedure

2.2.1. Training

Figure VII.4 shows the unimodal (top) and bimodal (middle) training distributions used in the current experiment. The unimodal distribution is representative of the Spanish vowel /a/ and the bimodal distribution is representative of the Dutch vowel contrast /ɑ/~ /a/. As is apparent in Figure VII.4, we created continuous (section 1.1) distributions, just as in chapter VI and in contrast to Escudero et al. (2011) and chapter V. The training stimuli were made with the Klatt synthesizer in the program Praat (Boersma and Weenink, 2013) in line with the procedure described in chapter VI. The manipulated acoustic dimensions were F1 and F2. Only the F1 continuum is shown in Figure VII.4.

Just as in chapter VI, the bimodal distribution was created on the basis of two Gaussian curves. The means and standard deviations were slightly adapted from the previously used values (see below) to accommodate the requirement that both distributions should have the same dispersion (section 1.3). The unimodal distribution was created on the basis of a single Gaussian curve.

We defined the dispersion of the distributions with the three variables that were also mentioned in the Introduction (section 1.2.2): the range, the standard deviation and the edge strength. The *range* of both distributions was set to run from 11.52 to 14.35 ERB for F1 (as is visible in Figure VII.4) and from 15.29 to 18.15 ERB for F2. The term "range" below applies to both F1 values and F2 values. We positioned the means of the underlying bimodal Gaussians at 20% and 80% of the range, and set the standard deviation of these underlying Gaussians at 10% of the range. In addition, we skewed the two peaks in the distribution slightly outwards.[6] The mean of the underlying unimodal Gaussian was placed at 50% of the range and had a standard deviation of 100% of the range. With these settings, the *standard deviations* of the bimodal and unimodal training distributions were

6 The formula used for the skewed bimodal distribution is: exp (-0.5 * ((x – μ1) / σ) ^ 2) + exp (-0.5 * ((x – μ2) / σ) ^ 2) + 0.2 * exp (-0.5 * ((x - 0.50) / σSkew) ^ 2), where μ1 and μ2 are 20% and 80% of the range respectively, σ is 10% of the range, and σSkew is set at 15% of the range. (The first two elements are the sum of the two Gaussian curves, the last element adds the skew).

similar, namely 29.3% and 28.4% of the range respectively.[7] The two edges for determining the *edge strength* were each placed at 1/6 of the range of the distribution (see Figure VII.4). With the settings for the range and the standard deviations as outlined above (this section), the edge strength was 0.954 for the unimodal distribution and 0.933 for the bimodal distribution. These numbers are based on a normalized distribution, i.e., a distribution with a range from 0 to 1 and a mean probability density of 1. Table VII.2 summarizes the ranges of F1 and F2 values, the standard deviations and edge strengths of the unimodal and bimodal distributions.

Table VII.2. Three measures for the dispersion of the unimodal and bimodal distributions: the range of F1 and F2 values, the standard deviation (SD) and the edge strength.

Distribution	Range F1 (ERB)	Range F2 (ERB)	SD (% of range)	Edge strength
Unimodal	11.52 to 14.35	15.29 to 18.15	28.4	0.954
Bimodal	11.52 to 14.35	15.29 to 18.15	29.3	0.933

It was not simple to obtain a unimodal and bimodal distribution that were as equal as possible in all three measures of dispersion. The chosen *range* was identical to the range of the enhanced bimodal distributions in Escudero et al. (2011) and chapters V and VI. Widening the F1 and F2 range would lead to including vowels extending into the /ɔ/- region, so that the bimodal distribution would be more representative of the /ɔ/~/a/ contrast than the /ɑ/~/a/ contrast. Shrinking the F1 and F2 range would make the test stimuli too similar. (In order to

7 Notice that the standard deviations of the Gaussians defining the shape of the distributions (e.g., 100% of the range for the unimodal distribution) are not identical to the standard deviations of the peaks in the distributions used in the experiment (e.g., 28.4% of the range for the unimodal distribution), which are not truly Gaussian. This is because the tails of the unimodal and bimodal distributions are cut off at the maximum and minimum acoustic values of F1 and F2, and because the bimodal distribution is a *sum* of two Gaussians.

ensure the discriminability of the test stimuli, we required them to be at least 1 ERB apart in F1 and F2. As will be explained in below (section 2.2.2), the acoustic values of the test stimuli were based on the intersections of the training distributions. Shrinking the range would shorten the acoustic distance between the intersections too much).

The *standard deviations* of the unimodal and bimodal distributions could only be made similar by adapting the distribution in chapter VI. That distribution had been created on the basis of the sum of two Gaussians with means at 25% and 75% of the range, and each with a standard deviation of 11% of the range. The standard deviation of the resulting distribution was 26.8% of the range. In order to make the standard deviation of the unimodal distribution similar to this percentage, while at the same time ensuring that (1) the range would remain as determined, (2) the acoustic distance between the test stimuli [ɑ] and [a] would not become too small (as just explained), and (3) the edge strength in 1/6 of the edges remained similar in both distributions, the enhanced bimodal distribution of chapter VI had to be adapted by changing the means and standard deviation of the Gaussians, and introducing some skewness (as specified above).

If distributional learning would occur, a small effect size (i.e., of the difference in categorization improvement between unimodally and bimodally trained participants) could be expected. This is because Escudero et al. (2011) and chapters V and VI found 95% confidence intervals close to zero when they quantified the difference in improvement in the categorization of Dutch [ɑ]- and [a]-tokens between Spanish listeners exposed to an enhanced bimodal distribution of Dutch /ɑ/ ~/a/ and Spanish listeners in the control condition. To increase the chance of detecting such a small effect, we used twice as many stimuli in the training distributions as in these previous studies, namely 256 in each distribution. (For the purpose of clarity, only 64 stimulus values are shown in each distribution in Figure VII.4).

The 256 experimental training stimuli were supplemented by 128 fillers, of which 64 were tokens of Dutch [i] and 64 were tokens of Dutch [u]. The F1 values of these fillers were sampled randomly from Gaussian distributions (one for each

vowel), with a mean set at 50% of the range and a standard deviation of 30% of the range. The F1 range was 5.81 to 6.93 ERB for both vowels. The F2 values were generated in the same way. The F2 range was 22.10 to 23.46 ERB for [i] and 10.84 to 12.20 ERB for [u]. Just as the stimuli in the training distributions, the fillers were created with the Klatt synthesizer in Praat (Boersma and Weenink, 2013).

Each stimulus presented during the training phase (i.e., each experimental stimulus and each filler) had a fundamental frequency (F0) contour that declined from 150 to 100 Hz and a duration of 140 milliseconds (ms). The durational difference between /ɑ/ and /a/ (/a/ is longer; Adank et al., 2004) did not appear in the training distributions, so that participants could only hear the spectral difference, which is difficult to perceive for these Spanish listeners (Escudero et al., 2009; Escudero and Wanrooij, 2010; section 1.3).

The order of presentation of the 384 stimuli (= 256 experimental stimuli + 128 fillers) was randomized for each participant individually. The stimuli were presented with an offset-to-onset inter-stimulus interval (ISI) of 750 ms. The total duration of the training was 5.7 minutes. Participants were asked to listen to the training vowels carefully, because they would perform a post-test afterward.

2.2.2. Pre- and post-tests

The pre- and post-tests were identical XAB categorization tasks, which were the same as in Escudero et al. (2011) and chapters V and VI except for the two response options A and B (see below). Each of the 80 trials presented participants with a natural token (the X-stimulus) of [ɑ] or [a], followed by two synthetic response options (the A- and B-stimuli), which were [ɑ] followed by [a] or reverse. There were 40 unique X-stimuli, which were a subset of the corpus reported by Adank et al. (2004). Twenty stimuli were [ɑ] and 20 were [a]. Ten stimuli of each vowel were produced by men and 10 by women. Each X-stimulus appeared twice in each test, once with the response options in the order [ɑ] – [a] and once with the response options in the reverse order.

The response options A and B were created with the Klatt synthesizer in Praat (Boersma and Weenink, 2013). In order to ensure that the F1 and F2 values of these response options were trained equally intensively in the unimodal and bimodal distributions, we calculated the intersections of the two distributions (the circles in Figure VII.4, bottom). These values differed slightly from the ones used in Escudero et al. (2011) and chapters V and VI, namely for [ɑ] F1=12.44 ERB, F2=16.21 ERB, and for [a] F1=13.43 ERB, F2=17.23 ERB.[8] Each response option had the same F0 contour (i.e., declining from 150 to 100 Hz) and duration (140 ms) as the training stimuli. The duration was the same for both options in order to isolate participants' learning of the spectral contrast (section 2.2.1).

Before the pre-test and the post-test, participants performed a practice test with [i] and [y] stimuli to make sure that they understood the test, and that they did not have problems hearing the vowels.[9]

3. Analyses and results

3.1. Descriptives

Table VII.3 lists the pre-test and post-test accuracy percentages, and the difference (i.e., the post-test minus the pre-test accuracy percentage), for the Unimodal and Bimodal groups separately. This difference is a measure of improvement after training, and thus reflects the *improvement score*.

8 The F1 and F2 values of the two response options in the test in Escudero et al. (2011) and in chapters V and VI were for [ɑ]: F1 = 12.5 ERB, F2 = 16.1 ERB and for [a] F1 = 13.3 ERB, F2 = 17.4 ERB.

9 In the region of Dutch /i/ and /y/ in the F1-F2 vowel space, Spanish has the vowel /i/ only. However, Spanish listeners tend to hear a rather clear difference between tokens of Dutch /i/ and /y/, possibly because the rounding of /y/ makes them perceive tokens of /y/ as close to Spanish /u/ (Escudero and Wanrooij, 2010). Listeners in the current experiment, as in Escudero et al. (2011) and in chapters V and VI, did not show any difficulties with the practice test.

Table VII.3. Pre- and post-test accuracy percentages, and improvement score (= post- minus pre-test accuracy percentage) per group. Standard deviations between participants in each group are given between parentheses.

Group	Pre	Post	Difference
Unimodal	60.35 (10.28)	66.33 (12.07)	5.98 (8.32)
Bimodal	59.98 (10.03)	65.25 (13.57)	5.27 (9.62)

3.2. Significance tests

The first set of analyses is based on common (frequentist) significance testing. This was done to assess the outcomes in the context of the previous results on distributional learning in Spanish adults presented with distributions of Dutch /ɑ/~/a/ (Escudero et al., 2011; chapters V and VI), which were all based on such tests.

In line with Escudero et al. (2011) and with chapters V and VI, we performed a one-sample *t*-test for each group (i.e., one for Unimodal and one for Bimodal), that compared the group's improvement score against zero. The results show a significant difference from zero, and thus better categorization accuracy after than before training, for both groups (Unimodal: 95% confidence interval [henceforth CI] = +3.83 ~ +8.13%, $t[59] = 5.56$, $p < 0.0001$; Bimodal: CI = +2.79 ~ +7.76%, $t[59] = 4.25$, $p < 0.0001$). Accordingly, both unimodal and bimodal training yield improved categorization performance for Spanish learners of Dutch /ɑ/~/a/.

An independent-samples (Unimodal vs. Bimodal) *t*-test, with the improvement score as the dependent variable, did *not* show a significant difference between the Unimodal and Bimodal groups (mean difference in improvement score, i.e., Bimodal – Unimodal score = –0.71%, CI = –3.96 ~ +2.54%, $t[118] = -0.43$, $p = 0.67$). This result does not enable us to say with confidence that Spanish learners' perception of Dutch /ɑ/~/a/ is affected by the number of peaks in a training distribution.

3.3. Bayes factors

From having found a *p*-value above 0.05 we cannot draw any conclusions about whether the null hypothesis is true or false. Because we wanted to be able to quantify evidence in favour of both the alternative *and* the null hypothesis, we computed Bayes factors (henceforth "BFs") (e.g., Kass and Raftery, 1995; Rouder et al., 2009). A BF denotes the likelihood ratio of the data occurring under the null hypothesis (H_0) versus the data occurring under the alternative hypothesis (H_1):

$$\mathrm{BF}_{01} = \frac{p(data|H_0)}{p(data|H_1)}$$

The "01" in this equation refers to H_0 and H_1 respectively. Thus, if $\mathrm{BF}_{01} = 10$, the observed data are 10 times more likely to occur if H_0 is true than if H_1 is true; if BF01 = 0.1, the observed data are 10 times more likely to occur if H_1 is true than if H_0 is true. If we assume that H_0 and H_1 are equally likely *a priori* (as is common and as we do henceforth), the Bayes factor BF_{01} can be said to quantify the evidence in support of H_0 over H_1. Thus, if $\mathrm{BF}_{01} = 10$, H_0 is 10 times more likely to be true than H_1 (i.e., the odds are 10 to 1 in favour of H_0); if $\mathrm{BF}_{01} = 0.1$, H_1 is 10 times more likely to be true than H_0; (i.e., the odds are 10 to 1 in favour of H_1). Whether a clear choice between the two hypotheses is possible, depends on the height of the Bayes factor. If $\mathrm{BF}_{01} > 20$, there is said to be strong support for H_0, and if $\mathrm{BF}_{01} < 1/20$, there is said to be strong support for H_1; if, however, BF_{01} lies between 3 and 20, the data are said to moderately favour H_0, and if BF_{01} lies between 1 and 3, the data are said to only trivially favour H_0 (Kass and Raftery, 1995).

In the current paper, the null and alternative hypotheses are defined in terms of the effect size of the difference in the improvement score (= the post-test minus the pre-test accuracy percentage) between the Unimodal and Bimodal groups, i.e., in terms of how much the two groups differ in their improvement of categorization accuracy after as compared to before training. An observed effect

size *d* can be calculated as the number of standard deviations difference between two improvement scores:

d = (improvement score of group 1 – improvement score of group 2) / standard deviation

where the standard deviation is the standard deviation across the two groups. In our case group 1 is the Bimodal group and group 2 the Unimodal group.

The null hypothesis (Figure VII.5, top) is always the same, namely that there is no difference in the improvement score between the Unimodal and Bimodal groups, and that accordingly the effect size *d* is exactly zero:

H_0: $d = 0$

The value of the BF depends on the definition of the alternative hypothesis. To accommodate different *a priori* beliefs about the effect size, we computed the BF in four different ways, i.e., with four different alternative hypotheses, which are increasingly less specific about the expected value of the effect size. The first and second alternative hypotheses (H_1 and H_2) include information about the effect size obtained from Escudero et al. (2011) and chapters V and VI; the third and fourth alternative hypotheses (H_3 and H_4) do not. Table VII.4 provides an overview of the

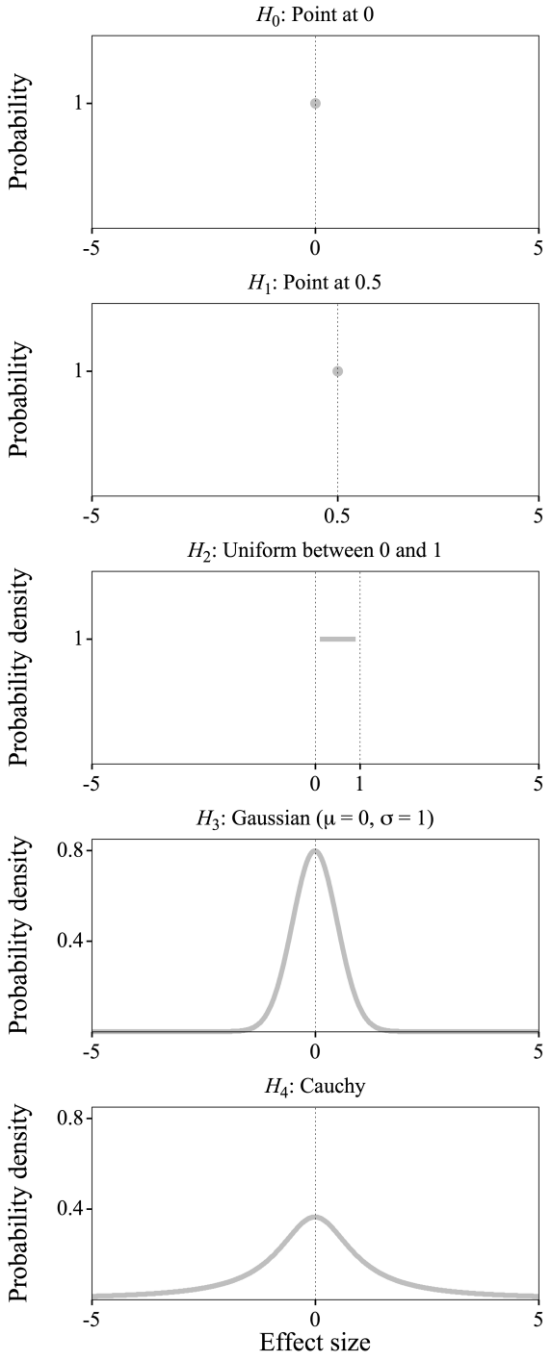
four alternative hypotheses and the resultant BFs, which we will now discuss in detail.[10]

Table VII.4. The four alternative hypotheses (H) and the resulting Bayes factors (BF).

H		BF
H_1:	*d* = + 0.50	BF_{01} = 137.86
H_2:	*d* is a random value drawn from a uniform distribution between 0 and 1.	BF_{02} = 5.97
H_3:	*d* is a random value drawn from a Gaussian distribution with mean 0 and standard deviation 1.	BF_{03} = 5.32
H_4:	*d* is a random value drawn from a Cauchy distribution	BF_{04} = 4.73

Figure VII.5 (opposite page). Null hypothesis (H_0) and four alternative hypotheses (H_1 through H_4) about the effect size: a point distribution at 0 (H_0), a point distribution at 0.5 (H_1), a uniform distribution between 0 and 1 (H_2), a Gaussian distribution with mean = 0 and sigma = 1 (H_3) and a Cauchy distribution (H_4). Explanation: see text.

10 The four Bayes factors can be computed in R (REF) with the equation **dt** (*t*, *df*) / (**mean** (*weight* * **dt** (*t*, *df*, **ncp** = *d* * **sqrt**(*n*))) / **mean** (*weight*)). In this equation, **dt** is the R function that computes the *t* probability density, and **ncp** is the non-centrality parameter of this density; *t* is the between-groups *t* value of our experiment, i.e. -0.43; *df* is the number of degrees of freedom for a *t* test, i.e. 60+60-2 = 118; *n* is half the geometric mean of the two group sizes (Rouder et al. 2009, p.234), i.e. 60*60/(60+60) = 30; *d* is the hypothesized range of possible effect sizes, and *weight* is the shape of the distribution for all these *d* values. For H_1, *d* is 0.5 and *weight* is 1. For H_2, *d* is (-0.5+1:1e5)/1e5 and *weight* is 1. For H_3, *d* is ((-10e5*width*+0.5):(10e5*width*-0.5))/1e5 and weight is exp(-0.5*(*d*/*width*)^2), where *width* is 1. For H_4, *d* is ((-1000*1e4*width*+0.5):(1000*1e4*width*-0.5))/1e4 and *weight* is 1/(1+(*d*/*width*)^2)), where *width* is sqrt(2)/2 (our equations for H_3 and H_4 are formulated in such a way that they will also work for other values of *width*). At the time of writing the computations for H_3 and H_4 are also available on Rouder's website (http://pcl.missouri.edu/bayesfactor).


H_0: Point at 0
Probability
1
-5
0
5
H_1: Point at 0.5
Probability
1
-5
0.5
5
H_2: Uniform between 0 and 1
Probability density
1
-5
0
1
5
H_3: Gaussian (μ = 0, σ = 1)
Probability density
0.8
0.4
-5
0
5
H_4: Cauchy
Probability density
0.8
0.4
-5
0
5
Effect size

Alternative hypothesis 1 (Figure VII.5, second from top) stipulates that the effect size *d* is a specific value:

H_1: $d = +0.50$

This value of +0.50 is based on effect sizes derived from the improvement scores observed in Escudero et al. (2011) and chapters V and VI, as follows. In Escudero et al. (2011) and chapter V, one group of listeners was exposed to a non-enhanced bimodal distribution (the Bimodal group), a second group to an enhanced bimodal distribution (the Enhanced group), and a third group to music (the Music group). In chapter VI, improvement in categorization was compared between a Music group and two Enhanced groups, one presented with a discontinuous distribution and the other to a continuous distribution. As mentioned in the Introduction (section 1.2.2), in all three studies the improvement score was significantly larger for the Enhanced group than for the Music group. In Escudero et al. (2011) and chapter V, the improvement score for the Bimodal group was not significantly different from that of the Music group and also not from that of the Enhanced group. For the current analysis, we considered the improvement scores of the previous Enhanced groups as proxies for the expected improvement score of our Bimodal group (which was also exposed to an enhanced bimodal distribution, just as the Enhanced groups in the previous studies; section 1.3). Because it was not clear whether our Unimodal group would behave more similarly to the previous Music groups or to the previous Bimodal groups, we considered the improvement scores of the previous Music and Bimodal groups as proxies for the expected improvement score of our Unimodal group. When calculating the effect sizes observed in the three studies, we used the above-mentioned formula for the effect size *d*, and took a previous Enhanced group as group 1, and either a previous Bimodal group or a previous Music group as group 2. The improvement scores for the Enhanced, Bimodal and Music groups were 6.04% (CI = +2.76 ~ +9.31%), 0.80% (CI = –2.22 ~ +3.83%) and –0.15% (CI = –3.50 ~ +3.21%) respectively in Escudero et al. (2011), and 6.63% (CI = +4.05 ~ +9.20%), 3.83% (CI = +0.97 ~ 6.68%) and 2.00% (CI = –0.50 ~ +4.50%)

respectively in chapter V. The improvement scores for the Enhanced and Music groups in chapter VI were 9.68% (CI = +6.80% ~ +12.55) and 2.00% (CI = –0.50 ~ +4.50) respectively.[11] The standard deviation across groups was 11.93% in Escudero et al. (2011), 9.46% in chapter V and 9.81% in chapter VI. Table VII.5 shows the resulting effect sizes *d*.

Table VII.5. Effect size *d* in previous studies (see text).

Previous study	Enhanced–Bimodal	Enhanced–Music
Escudero et al. (2011)	+0.44	+0.51
Chapter V	+0.30	+0.49
Chapter VI		+0.78

The average of the five listed effect sizes is +0.50, the value in hypothesis 1. Notice that this value is explicitly positive, i.e., it reflects the belief that our Bimodal group will have a *higher* improvement score, and thus improve *more* after distributional training than the Unimodal group. The BF calculated on the basis of the null hypothesis versus this first alternative hypothesis expresses strong support for the null:

$$BF_{01} = 137.86$$

Specifically, BF_{01} indicates that the observed data are 137.86 times more likely to have occurred under H_0 (that *d* is exactly 0), than under H_1 (that *d* is exactly 0.5).

In alternative hypotheses 2 through 4, the effect size is no longer defined as a specific value, but as a probability density function (Figure VII.5, as explained below): *d* is expected not to be one specific value, but a random value drawn from

11 The Enhanced group referred to here is the group presented with a continuous enhanced distribution in chapter VI (the Continuous Enhanced group). In chapter VI , the group presented with a discontinuous enhanced distribution (the Discontinuous Enhanced group) and the Music group were taken from chapter V.

a distribution whose form defines the likelihood of that value. In alternative hypothesis 2, the effect size is any value between 0 and 1 with equal probability (Figure VII.5, middle):

H_2: *d* is a random value drawn from a uniform distribution between 0 and 1.

The hypothesis still includes the information mentioned in Table VII.5 about previously obtained effect sizes (i.e., all effect sizes in Table VII.5 fall within the range of the distribution), but it is vaguer about the precise value of the expected effect size than hypothesis 1. Since *d* is defined as 0 or positive, hypothesis 2 expresses the belief that the Bimodal group will improve *at least as much* as the Unimodal group. The BF calculated on the basis of the null hypothesis versus this second alternative hypothesis also expresses support for the null:

$BF_{02} = 5.97$

That is, BF_{02} implies that the observed data are 5.97 times more likely to have occurred under H_0 (that *d* is exactly 0) than under H_2 (that *d* is somewhere between 0 and 1).

Hypotheses 1 and 2 show that previous observations can be incorporated in the alternative hypothesis to different extents, depending on the researcher's belief in the truth value of these observations. Previous observations can also be deemed inappropriate for incorporation in the alternative hypothesis, for example if concerns (such as mentioned in the Introduction, section 1.2) about the earlier observations create uncertainty about the applicability of the information to the experiment to be performed. In this case, the alternative hypothesis should reflect the assumption that we do not have a clear expectation about the effect size. This is done in alternative hypotheses 3 and 4. In alternative hypothesis 3, the effect size is any value around 0, with values closer to the mean being more likely than values

further away from the mean as defined by a Gaussian distribution (Figure VII.5, fourth from top):

> H_3: *d* is a random value drawn from a Gaussian distribution with a mean of 0 and a standard deviation of 1.

Since *d* can be positive, zero or negative, the belief that the Bimodal group will improve at least as much as the Unimodal group, which was inherent in alternative hypotheses 1 and 2, is now dropped. The BF calculated on the basis of the null hypothesis versus the third alternative hypothesis still expresses support for the null:

$$BF_{03} = 5.32$$

In other words, BF_{03} indicates that the observed data are 5.32 times more likely to have occurred under H_0 (that *d* is exactly 0) than under H_3, (that *d* is a value around zero, whose probability is defined by a Gaussian distribution).

It is possible to be even less specific about the expected value of the effect size than in alternative hypothesis 3, by loosening the belief that the effect size is more likely to occur close to zero. This is done with a Cauchy distribution (for an explanation, see Rouder et al., 2009), as used in alternative hypothesis 4 (Figure VII.5, bottom):

> H_4: *d* is a random value drawn from a Cauchy distribution, with a width of (√2)/2.[12]

Notice in Figure VII.5 that the tails of the Cauchy distribution are much heavier than those of the Gaussian distribution, thus reflecting a much smaller confidence

12 The tails of a Cauchy distribution are so heavy that the mean and the standard deviation do not exist (Rouder et al., 2009: p.231). The equation used for the Cauchy distribution is: ((-1000*1e4**width*+0.5):(1000*1e4**width*-0.5))/1e4, where *width* is sqrt(2)/2 (see also note 10).

that the effect size should be relatively close to zero. Again, the BF calculated on the basis of the null hypothesis versus the fourth alternative hypothesis expresses support for the null:

$$BF_{04} = 4.73$$

Thus, BF_{04} indicates that the observed data are 4.73 times more likely to have occurred under H_0 (that *d* is exactly 0) than under H_4 (that *d* is a value around zero, whose probability is defined by a Cauchy distribution, i.e., with more uncertainty as to the effect size than expressed in the Gaussian distribution used for H_3).

In sum, four different calculations of the Bayes factor, which differ in the extent to which they incorporate *a priori* beliefs about the expected effect size, unanimously support the null hypothesis that there is no difference between bimodally and unimodally trained Spanish participants in improvement of categorization of Dutch [ɑ]- and [a]-tokens. If we follow the interpretation of Bayes factors by Kass and Raftery (1995; section 3.3), the support for the null hypothesis ranges from moderate support (hypotheses 2 through 4, which represent less strong *a priori* beliefs about the effect size than hypothesis 1) to strong support (hypothesis 1, which incorporates the most explicit *a priori* beliefs).

4. Discussion

In the present study we trained Spanish adult participants on a bimodal or a unimodal distribution encompassing the Dutch vowel contrast /ɑ/~/a/, and then tested their improvement in categorization of Dutch [ɑ]- and [a]-tokens after training. For the first time in the research on distributional learning of speech sounds, the bimodal and unimodal distributions had nearly identical dispersions, as defined by the range, standard deviation and edge strength. The results show that Spanish adult participants improve their categorization of Dutch [ɑ]- and [a]-tokens irrespective of the training distribution, and that categorization accuracy does not improve significantly more after exposure to one distribution than after exposure to

the other distribution. Additionally, four different Bayes factors (ranging from incorporating *a priori* beliefs about the expected effect size as much as possible to not incorporating previous knowledge at all) provided unanimous evidence for the null hypothesis that there is no difference between bimodally and unimodally trained Spanish listeners in categorization improvement. In other words, the number of peaks in the distribution does not play a role in the observed improved categorization.

The number of peaks must now also be dismissed as the factor that explains the earlier results on Spanish listeners' larger improved categorization of Dutch [ɑ]- and [a]-tokens after enhanced bimodal training than after listening to music (Escudero et al., 2011; chapters V and VI; Escudero and Williams, 2014). Future research should determine which factor(s) do account for these results. At least two factors, which were also mentioned in the Introduction, appear to be viable candidates: "processing speech versus non-speech" (since the earlier studies compared learning from exposure to a speech distribution to learning from exposure to non-speech) and the "wide dispersion" of the enhanced bimodal distributions (since the earlier studies compared learning from exposure to an enhanced bimodal distribution to learning from exposure to music, which has no relevant dispersion).

The conclusion that the number of peaks in the distributions cannot explain the observed perceptual learning in Spanish adults may very well extend to *all* previous results on distributional learning in infants and adults. Although other studies included a control group exposed to a unimodal speech distribution (so that "processing speech versus non-speech" cannot be a factor accounting for the reported effects), none of the studies controlled for dispersion as was done in the current study. Results from other paradigms than distributional training suggest that enhancement of training stimuli (i.e., a wide dispersion in the training distributions) can advance the learning of speech sound categories through drawing participants' attention to the relevant differences between the categories (e.g., Jamieson and Morosan, 1986; Iverson et al., 2005; Kondaurova and Francis, 2010). In view of this potential influence of dispersion on attentional learning, dispersion

is a high-ranking potential confound whose role should be separated from that of the number of peaks before we can conclude that distributional learning based on the number of peaks is a mechanism that tunes speech perception.

Chapter VIII

Neural correlates of distributional speech sound learning: a literature review

Karin Wanrooij

(to be submitted)

Abstract

Distributional speech sound learning is learning speech sound categories from plain exposure to speech, i.e., without feedback or instruction. In linguistic theory and related computer simulations, the mechanism is viewed as a low-level, bottom-up process, possibly related to neuronal tuning in the primary auditory cortex (A1). However, neuroscientific evidence has been presented scarcely. This article reviews possible neural correlates of distributional learning in infancy and adulthood, obtained with various research techniques, which target different levels of analysis (i.e., the population of neurons, the neuron, and the synapse), and which are applied *in vivo* and *in vitro*, in animals and humans.

The resultant picture is that in infancy distributional learning can indeed be viewed as bottom-up induced changes in the firing properties of neurons in A1, and possibly at other low levels of auditory cortical processing. Natural sound distributions in infants' environment contribute to the formation of balanced auditory parameter maps, which are necessary for normal perception, and to the creation of basic categorical representations, in A1. Vowel distributions in particular may help shaping the parameter map that represents spectral properties of sound (the "tonotopic map"). What precisely triggers the onset of language-specific speech sound perception in the second half of the first year of life remains unsolved, but is probably best viewed as a mix of experiential and maturational factors.

In adulthood, distributional learning requires "attention" that triggers neuromodulatory influence from subcortical structures such as the nucleus basalis. Similar subcortical influence is also present in infancy, when "attention" is not required to induce it. In contrast to subcortical influence, top-down influence coming from higher-level (linguistic) representations in the neocortex only becomes prominent in the course of childhood.

1. Introduction

Learning from mere exposure to the frequency distributions of ambient stimuli, without receiving any feedback, is called "distributional learning". It is considered an important learning mechanism in infancy, for instance for learning the speech sounds of the native language (Lacerda, 1995; Guenther and Gjaja, 1996). In the lab, the mechanism of distributional learning can appear already after only a few minutes of exposure to speech sound distributions, as demonstrated in several distributional training experiments with not only infant learners of native speech sounds, but also adult learners of non-native speech sounds (for infants: Maye et al., 2002, 2008; Yoshida et al., 2010; Capel et al., 2011; chapter II; for adults: Maye and Gerken, 2000, 2001; Gulian et al., 2007; Hayes-Harb, 2007; Escudero et al., 2011; chapters V and VI; Escudero and Williams, 2014).

The concept of distributional learning is discussed in more detail below (section 1.1). Linguistic theories portray the mechanism as a low-level, bottom-up process (section 1.2). However, concrete neuroscientific evidence is poorly presented in the literature (section 1.3). The aim of this review is to fill this gap (section 1.4).

1.1. The concept of distributional learning

The concept of distributional learning is illustrated schematically in Figure VIII.1. Speech sounds are characterized by certain acoustic properties. The two most important acoustic properties that characterize vowels are the first and second formant frequencies (F1 and F2) (Peterson and Barney, 1952). The values of these formants vary for vowels in different languages, and also for vowels pronounced by different speakers of a language and for different pronunciations of each speaker. When measuring for instance the F1 values in multiple pronunciations of a certain vowel (say the Dutch vowel /ɛ/, as in the Dutch word *pet*, "cap"), it is likely that the distribution of the values is similar to that shown in Figure VIII.1 (top): the distribution has a gradually increasing number of values near the mean (top left; each vertical line represents a unique F1 value) and thus a larger underlying

frequency density near the mean than further away from the mean (top right; grey curve). The frequency density curve is likely to approximate a Gaussian distribution (Lisker and Abramson, 1964; Newman et al., 2001; Lotto et al., 2004). In other languages than Dutch, other Gaussian-like frequency distributions may appear. For instance, the F1 values measured in multiple pronunciations of English /ɛ/and /æ/, as in the English words *pet* and *pat* respectively, will together cover a similar range in the F1 continuum as the F1 values of Dutch /ɛ/ (Figure VIII.1, bottom left). Accordingly, English has two Gaussian-like distributions along this range, representing two vowels (Figure VIII.1, bottom right). Not surprisingly, vowels with F1 values as shown in Figure VIII.1 will be *perceived* as Dutch /ɛ/ by Dutch listeners, and as either English /ɛ/ (for the lower values along the continuum) or /æ/ (for the higher values) by English listeners. The concept of distributional learning is the idea that such language-specific speech sound perception results from merely experiencing the frequency distributions of the language to be learned, i.e., without the influence of prior knowledge and without any feedback or instruction.

1.2. Distributional learning in linguistic theory

Linguistic theory generally formulates distributional learning of speech sounds as a low-level (section 1.2.1), bottom-up (section 1.2.2) process.

1.2.1. A low-level process

The mechanism of distributional learning is "low-level", because it is usually described as implicating only the lower levels of auditory processing. For instance, in the computer simulation of distributional learning by Guenther and colleagues (Guenther and Gjaja, 1996; Guenther and Bohland, 2002), distributional learning involves the thalamus, which is the last site that the auditory signal passes before

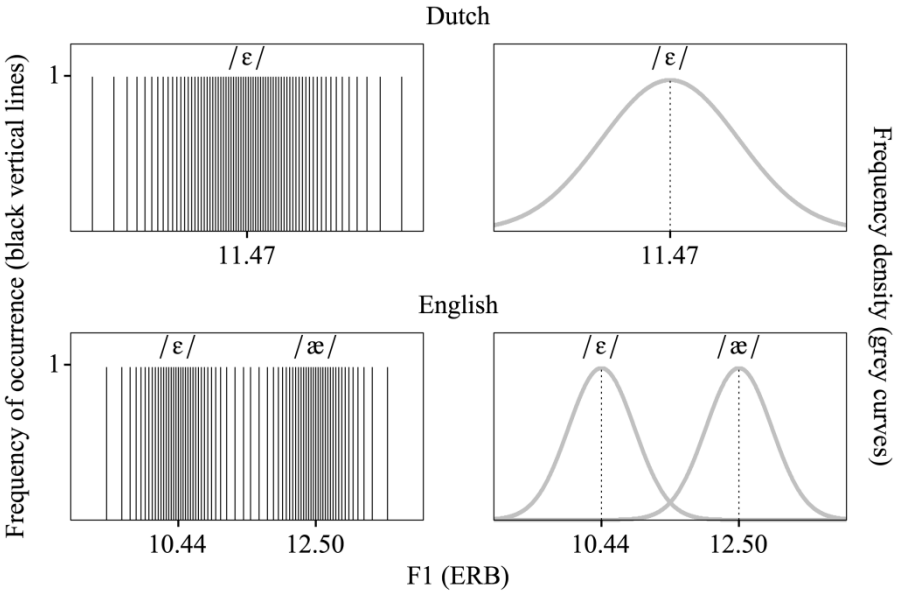


Figure VIII.1: Frequency distributions of first-formant (F1) values of the Dutch vowel /ɛ/ (top) and the English vowel contrast /ɛ/~ /æ/ (bottom). Each token in the environment (left: each vertical line) is unique and thus occurs only once (left: the height of each vertical line = 1). The frequency density is highest around the mean F1 values of each vowel category (right: the grey curves have peaks here).

reaching the cortex, and the primary auditory cortex, which represents the lowest level of auditory processing in the cortex. In the BiPhon model (the Model of Bidirectional Phonology and Phonetics; Boersma, 1998-2014, as in e.g., Boersma, 2011), distributional learning pertains to only the two lowest representations that are considered to be involved in comprehension and in perceptual learning, as shown in Figure VIII.2, namely the Auditory Form and the Surface Form. The Auditory Form is a continuous phonetic representation of acoustic information such as pitches and formants; the Surface Form is a discrete sublexical phonological representation such as a phoneme. The two representations are black in the figure, as opposed to the higher-level grey representations. In adult auditory

comprehension, the acoustic waveform is mapped onto a phonetic Auditory Form (e.g., an F1value of 11.47 ERB in Figure VIII.1) and a phonemic Surface Form (e.g., the Dutch vowel category /ɛ/). Surface Forms are then mapped onto other representations, including an Underlying Form in the lexicon (e.g., the Underlying Form |pɛt| in the Dutch word *pet*) and a representation pertaining to meaning (in this case the meaning of the Dutch word *pet*, "cap").

In *distributional learning*, the activation of Auditory Forms (e.g., several F1 values around 11.47 ERB in Figure VIII.1) leads to perceptual "warping", the outcome of which is a rudimentary Surface Form (e.g., the Dutch vowel category /ɛ/). Warping refers to a distortion of the "perceptual space" in such a way that some speech sounds come to be perceived as closer and thus as more similar to one another, than other speech sounds, even if the acoustic distance between the speech sounds in both situations is the same. For instance, on the /ɛ/~ /æ/ continuum in Figure VIII.1, which is a one-dimensional perceptual space, native speakers of English are expected to find it easier to hear a difference between vowel tokens with F1 values of 11 ERB and 12 ERB, respectively, than between vowel tokens with F1 values of 10 ERB and 11 ERB, respectively, even though the acoustic difference is 1 ERB in both cases. Notice that the former difference straddles a category boundary, while the latter difference does not. Thus, warping distorts perception in such a way that the sensitivity to perceive differences between two speech sounds increases near category boundaries (11.47 ERB in Figure VIII.1) and shrinks near category means (10.44 ERB and 12.50 ERB in Figure VIII.1). In the BiPhon model, mere exposure to speech sound distributions induces such perceptual warping. As just mentioned, this warping does not involve higher-level representations than Auditory and Surface Forms.

Other linguistic models than the BiPhon model also confine the scope of distributional speech sound learning to lower-level representations, but the details on how distributional learning affects these representations differ. For instance, perceptual warping under the influence of exposure to speech is also incorporated in the Native Language Magnet theory (NLM; Kuhl, 1994; Kuhl et al., 2008), but here the storage of "prototypes" is required for warping to occur. Specifically, it is

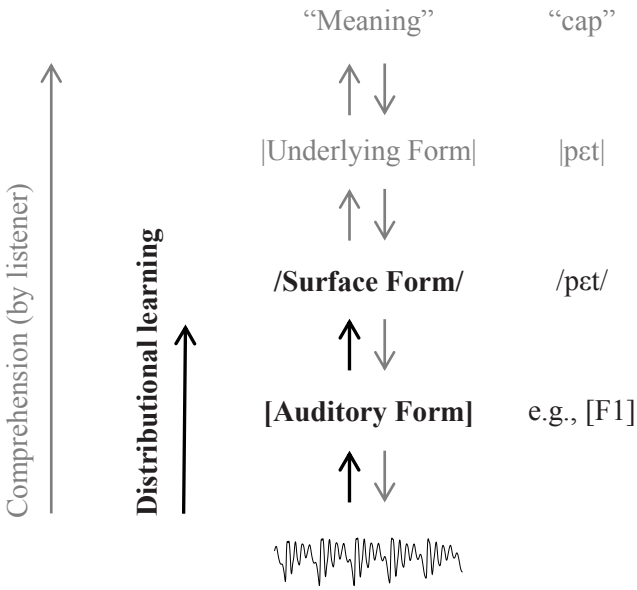


Figure VIII.2: Distributional learning as a low-level, bottom-up process, in an example linguistic model, the BiPhon model (Boersma, 1998-2014, as in e.g., Boersma, 2011).

claimed that "infants store sensory information" (Kuhl et al., 2008: 985), and that at a certain point, "the representations most often activated (*prototypes*) begin to function as perceptual magnets for other members of the category" (p. 982), i.e., the stored prototypes begin to warp the perceptual space. In the BiPhon model, such storage of prototypes is not necessary. In this respect, the model resembles the computer simulations by Lacerda (1995) and Guenther and Gjaja (1996; replicated by Boersma in Wanrooij, 2009).

The BiPhon model does also not need the storage of concrete instantiations or "exemplars" of speech sound categories, for warping to occur. In this regard, the model resembles the computer simulation by Guenther and Gjaja (1996), but differs from the proposal by Lacerda (1995) and also from models that embrace the exemplar theory (e.g., Pierrehumbert, 2003; Werker and Curtin, 2005). The difference between the exemplar approach on the one hand and the warping

approach as proposed in the BiPhon model and in Guenther and Gjaja's simulation on the other hand, can be illustrated by means of Figure VIII.1: in the exemplar approach, distributional learning corresponds to storing each concrete, experienced token separately (the vertical lines in Figure VIII.1, left), and in the warping approach, it corresponds to deriving a more abstract, overall representation (the grey curve in Figure VIII.1, right) without such storage. Setting aside these differences between models (storage of some kind versus no storage), models are quite unanimous in confining the effect of distributional speech sound learning to lower-level representations (such as the Auditory and Surface Forms in the BiPhon model in Figure VIII.2).

Linguistic models also differ in "how low" precisely the representation is that is assumed to arise from distributional learning. The BiPhon model focuses on the *language-specificness* of the representation, and views it as a rudimentary *phonemic* representation (Surface Form in Figure VIII.2), one level higher than a non-language-specific *phonetic* representation (Auditory Form in Figure VIII.2). A similar focus on the language-specificness of the representation resulting from distributional learning is found in the NLM, even though in this model the representations are called "phonetic" rather than "phonemic".

Other opinions emphasize the *rudimentariness* of the representation resulting from distributional learning. According to Pierrehumbert (2003), for example, natural distributions contain so much variation and tend to overlap to such an extent, that they cannot possibly induce phonemic representations; at best, they can induce more basic (i.e., lower) representations of "positional variants of phonemes" (p.129), since the overlap between distributions of phonemes is reduced greatly if each distribution only reflects tokens of a phoneme as pronounced in a specific "prosodic position and segmental context" (p. 129). Thus, Pierrehumbert proposes that the speech sound representations induced by distributional learning are confined to *context-specific* representations. In the BiPhon model such context-specific representations could be located at a level lower than the phonemic Surface Forms (but still higher than the phonetic Auditory Forms).

The NLM, which as just mentioned concentrates on the language-specificness of the representations resulting from distributional learning, mitigates the problem of overlaps between natural distributions by stressing that infants are exposed to distributions of speech sounds pronounced in infant-directed speech (Kuhl et al., 2008). Such speech contains exaggerated pronunciations, in particular of corner vowels, which cause the means of the distributions to be pulled apart and concomitantly the overlap between distributions to be reduced[1] (e.g., Kuhl et al., 1997; Burnham et al., 2002; Uther et al., 2007; Cristia and Seidl, 2013; McMurray et al., 2013; Englund, 2005).

Yet another model, the PRIMIR model ("a developmental framework for Processing Rich Information from Multidimensional Interactive Representations"; Werker and Curtin, 2005: 197), posits that distributional learning produces "phonetic categories" at the so-called "General Perceptual plane". This level of representation is a more basic (i.e., lower) level than the "Phonemic plane", where, according to this model, phonemic representations arise. In sum, despite differences in the formulation of the level of representation evolving from distributional learning, a consensus exists that distributional learning only leads to a low-level representation, while the creation of full-fledged, adult-like phonemic representations requires the influence of higher-level representations, in particular lexical ones (e.g., Boersma et al., 2003; Pierrehumbert, 2003; Werker and Curtin, 2005).

1.2.2. A bottom-up process

Linguistic theory does not only view distributional learning as a low-level process (section 1.2.1), but also as a strictly "bottom-up" process. The bottom-up nature is illustrated in Figure VIII.2 by the black, *upward* arrows from the acoustic

1 The overlap between two distributions is reduced when pulling the means apart, provided the dispersion of the two distributions remains the same. The dispersion in infant-directed speech is larger than in adult-directed speech. The enhancement of the acoustic distance between the means of two distributions in infant-directed speech thus contributes to reducing the overlap that would result without enhancement.

waveform to the Auditory Form to the Surface Form, as opposed to the grey, downward arrows. This is in accordance with the definition of distributional learning as learning from mere exposure to external stimuli (the acoustic waveforms in Figure VIII.2), without the need for prior knowledge (the knowledge embodied in the grey higher-level linguistic representations in Figure VIII.2) and without receiving feedback or instruction (which can also be viewed as triggering top-down influence of higher-level representations). Computer models show that it is indeed possible to model distributional learning as a strictly bottom-up process (Lacerda, 1995; Guenther and Gjaja, 1996).

Linguistic theories often attribute adults' difficulties in learning certain non-native speech sound contrasts to the influence of native speech sound categories and other phonological knowledge of the native language (Polivanov, 1931/ translation 1974; see also several models used to describe second-language speech sound perception and learning, such as the Perceptual Assimilation Model, PAM, Best, 1994, and the Speech Learning Model, SLM, Flege, 1995). In the case of distributional learning, this hampering influence of the native language can be formalized as a top-down influence on the bottom-up mechanism of distributional learning, at least in models that allow for top-down processing of language (e.g., BiPhon; TRACE, McClelland and Elman, 1986). In the BiPhon model, this influence comes from existing Surface Forms and Underlying Forms (Figure VIII.2). Since infants most probably do not come into the world with speech sound representations[2], distributional learning in infants can hardly be affected by such top-down influence. Therefore, it can be expected that distributional learning is a less hampered mechanism in infants than in adults. Recently, we have collected evidence that the capacity for distributional learning in adults is indeed smaller than that in infants (chapter III).

2 The possibility that speech sound representations are innate (Chomsky and Halle, 1968; Eimas et al., 1971; Eimas, 1975) has largely been abandoned (Guenther and Gjaja, 1996; Boersma, 1998; Kuhl et al., 2008; see also Karmiloff-Smith, 2006).

1.3. Limited formulation of neural correlates

Linguistic theory has presented limited explicit neuroscientific evidence for distributional learning as a low-level, bottom-up process. Most theories that are used to account for first- and second-language speech sound learning, lack neuroscientific explanations (e.g., Pierrehumbert, 2003; PRIMIR, Werker and Curtin, 2005; PAM, Best, 1994; SLM, Flege, 1995; the Second Language Linguistic Perception model, L2LP, Escudero, 2005). Only the NLM relates distributional learning explicitly to "neural commitment", without, however, offering a specification of the neural mechanisms (Kuhl et al., 2008).

The computer model of distributional learning developed by Guenther and colleagues (already mentioned in section 1.2.1) is related to neuroscientific data more explicitly (Guenther and Gjaja, 1996; see also subsequent computational and experimental refinements in e.g., Guenther and Bohland, 2002). The model relates distributional speech sound learning explicitly to changes in the firing properties of neurons in the auditory cortex, on the basis of similar changes observed in animal brains. Also, learning is modelled as changes in the strength of synapses between cells in the thalamus and cells in the primary auditory cortex. These changes in synaptic strength reflect Hebb's idea that the synapse between two neurons strengthens when their activity is paired (Hebb, 1949). Such Hebbian learning indeed occurs as a real process in the cortex (e.g., Wang et al., 1996). Other researchers have also proposed accounts of speech sound learning based on Hebbian learning (e.g., McCandliss et al., 2002). Apart from these specifications of neural mechanisms in terms of Hebbian learning, the literature lacks a more elaborate review of neuroscientific evidence for distributional learning.

1.4. Aim and approach

The current article is an attempt to fill the gap mentioned in the previous section: it aims to explore whether it is possible to find concrete neuroscientific evidence that supports linguistic theoretical hypotheses (section 1.2) and experimental observations (studies mentioned in section 1) pertaining to distributional speech

sound learning. Concomitantly, it also aims to contribute to narrowing the gap between linguistic research, which commonly does not present supporting neuroscientific evidence, and neuroscientific research, which mostly does not incorporate linguistic theory.

In line with the concept of distributional learning in linguistic theory as a low-level, bottom-up mechanism (section 1.2) and as proposed by Guenther and Gjaja (1996; section 1.3), this review of neuroscientific evidence concentrates on *bottom-up induced changes in the firing properties of neurons in the primary auditory cortex (A1)*. A1 represents the lowest level of auditory processing in the cortex (Kaas and Hackett, 2000).

It is true that changes in neural firing properties ("plasticity") can occur at all levels along the auditory pathway, including A1 and including subcortical levels below and cortical levels above A1, but the current review does not address such lower- and higher-level plasticity. Lower-level plasticity remains undiscussed, because linguistic theory envisions distributional learning as ensuing directly from plain exposure to the input (sections 1.1 and 1.2) and plasticity *directly* induced by input occurs primarily in the cortex rather than subcortically (reviews in: Buonomano and Merzenich, 1998; Froemke and Jones, 2011). It is possible that higher-level areas in the cortex than A1 are also affected by distributional input, either indirectly via A1 or directly, since there are direct connections between the thalamus and higher-level auditory areas (Hackett, 2011). Similarly, it is possible that same-level areas are affected, because A1 is only one of the primary auditory regions that receive information on the incoming sound from the thalamus directly (Hackett, 2011). Unfortunately, these same-level and higher-level areas have been studied far less than A1. Accordingly, results observed in these areas are not discussed.

The body of research on experience-induced neural changes in A1 is vast, and comprises different subfields that use different measurement techniques. This paper includes *evidence obtained from non-human animal brains and from human brains*. The non-human animal data give important insights into auditory learning that cannot be obtained as easily from human brains. This is because this evidence

is obtained with invasive measurements, which allow for a spatial and temporal resolution that is superior to that obtained from non-invasive measurement techniques commonly used for human participants. It is obvious that animals do not process and learn language in a human way and that, accordingly, correspondences should be assumed with caution. Even so, speech sound processing and learning are likely to be more similar between non-human animals and humans at lower than higher levels of representation, since the higher the level the more linguistically complex the representation becomes (section 1.2.1). Distributional learning as viewed in linguistic theory pertains to the lowest levels of linguistic representation in the cortex (section 1.2.1), where the chance of finding similarities should thus be highest.

The review pays special attention to *possibly different neural correlates of distributional learning in infants than in adults*. The theoretical prediction is that distributional speech sound learning in adults may be hampered by top-down influence of higher-level linguistic representations (section 1.2). On the other hand, distributional learning in adults can be demonstrated in the lab already after a few minutes of exposure and such a short-term result has been reported more often for adults than for infants (studies mentioned in section 1). This paper aspires to provide a better understanding of possible age differences in distributional learning from a neuroscientific perspective.

In sum, this review presents data from various subfields of neuroscience, that can shed light on the mechanism of distributional speech sound learning in infants and in adults. The data focus on plasticity of neurons in A1, first in babyhood (section 3), and then in adulthood (section 4). Mechanisms that may underlie this plasticity at different levels of neuroscientific analysis are also discussed (section 5). In order to understand the description of the neuroscientific correlates of distributional learning, it may be helpful to get an idea of the anatomical organization of the adult A1 (section 2).

2. Anatomical organization of the adult A1

A1 in adult mammals is anatomically organized into auditory parameter maps, of which the most investigated one is the tonotopic map representing the frequency parameter (e.g., Merzenich and Brugge, 1973; Schreiner, 1998). The organization of the tonotopic map can be explained as follows. Each neuron in A1 is sensitive to certain frequency values of sound.[3] The range of frequencies that affect the neuron's firing rate is its receptive field (RF). A typical neuron in A1 also has a preferred frequency to which it is particularly sensitive. This preferred frequency can be defined in different ways. It is often expressed as the neuron's "characteristic frequency" (CF), which is the frequency that makes the neuron fire at its threshold intensity (this is the lowest intensity level that excites the neuron, i.e., makes it fire significantly above its spontaneous firing rate); another way of expressing the preferred frequency is the neuron's "best frequency" (BF), which is the neuron's highest response at a certain intensity level (Eggermont, 2008).[4] The neurons in A1 are positioned together in such a way that when schematically picturing their CFs (or BFs) in a flattened cortex, these CFs run from low to high frequencies along one axis (the frequency gradient) and stay the same along the other axis (the isofrequency bands). Along the isofrequency axis, neurons vary systematically in firing properties in response to other stimulus parameters, such as intensities (Schreiner, 1998) and the direction of frequency-modulated sweeps (Zhang et al., 2003). Typical firing patterns for still other parameters vary in a patchy manner across the same area. A1 thus reflects several auditory parameter maps, which are partially overlapping anatomically (Schreiner, 1998; Recanzone et al., 1999).

3 "Frequency" refers to the acoustic property of sound that can be expressed in hertz (Hz), not to the frequency of occurrence (as in section 1.1).

4 The characteristic frequency (CF) and the best frequency (BF) do not have to be identical (Eggermont, 2008). This shows that the preferred frequency depends on the intensity level of the sound. When the preferred frequency is measured with fMRI (functional Magnetic Resonance Imaging), the BFs are reported, because the noise generated by the fMRI equipment prevents the measurement of the CFs.

Such maps have been observed in adult animals (e.g., Merzenich and Brugge, 1973; Schreiner, 1998; Recanzone et al., 1999) and there is multiple evidence that they exist in human adults too (Pantev et al., 1989; Langner et al., 1997; Formisano et al., 2003). The human A1 probably consists of at least two areas, which are each tonotopically organized (Merzenich and Brugge, 1973; Formisano et al., 2003).

Acoustic characteristics of speech sounds are traceable in activation patterns in A1, as evidenced in animals. For instance, the F1 and F2 values of vowels elicit activation in the tonotopic map (e.g., Steinschneider et al., 1990; Ohl and Scheich, 1997; Schreiner, 1998).

3. Plasticity in A1 in babyhood: the impact of plain exposure

The functional development of A1 is well-studied in animals, particularly in cat kittens and rat pups (for cat: e.g., Eggermont, 1996; Kral et al., 2002; for rat: e.g., Zhang et al., 2001, 2002). These studies show that the auditory parameter maps that are present in the adult A1 (section 2) are not present at birth yet. They develop after birth, purely under the influence of sound in the environment, i.e., the animal does not have to perform a task and is not rewarded or punished for certain behaviour. The development of the maps reflects changes in neurons' firing properties in response to ambient sounds. These experience-induced changes affect perception (Han et al., 2007) and cannot be undone later in life (e.g., Zhang et al., 2001, 2002), at least not easily and radically (section 4). Because the changes depend on the kind of sound distribution that the animal is exposed to, they should be viewed as the result of distributional learning. Below I first list the kind of sound distributions that baby animals have been exposed to (section 3.1), before addressing the main results and the conclusions that can be drawn from these results (sections 3.2 through 3.5).

3.1. Distributions used in animal experiments

Studies have examined the influence on A1 cortical map development of several acoustic parameters (Eggermont, 1996; Insanally et al., 2009). Most extensively studied is the influence of frequency on tonotopic map formation. I distilled five kinds of sound distributions from these studies, that baby animals have been exposed to, usually for several weeks. These distributions include (1) *flat distributions*, where the baby animal is presented with white noise (Zhang et al., 2002; Chang and Merzenich, 2003; Speechley et al., 2007), (2) *single-point distributions*, where the baby animal is exposed to a single pure tone (Stanton and Harrison, 1996; Zhang et al., 2001; Zhou et al., 2008), and (3) *multiple-point distributions*, where the baby animal hears multiple tones (Nakahara et al., 2004; Köver et al., 2013). In addition, the effect of (4) "*empty distributions*" has been examined in studies with cats that were born deaf (Ponton and Eggermont, 2001; Kral et al., 2001, 2002). Maps after exposure to these abnormal distributions (1 through 4) have been compared to the maps of animals raised normally in a quiet environment, i.e., without exposure to other sounds than those occurring naturally in a litter. These multiple (5) *natural distributions* are supposedly Gaussian-like. Recently, rat pups have also been exposed to more complex natural distributions than those occurring in a normal quiet environment, namely to naturalistic sounds recorded in the jungle, containing vocalizations of several species (Bao et al., 2013). The different kinds of distributions are summarized schematically in Figure VIII.3.

3.2. The importance of natural distributions

The first conclusion that can be drawn from the results of early exposure to the five kinds of distributions (section 3.1) is straightforward: natural input in the form of multiple natural distributions (multiple distributions 5) is necessary for healthy map development. The effect of exposure to distorted or impoverished input (distributions 1 through 4) is severe. For instance, exposure to pulsed noise (flat

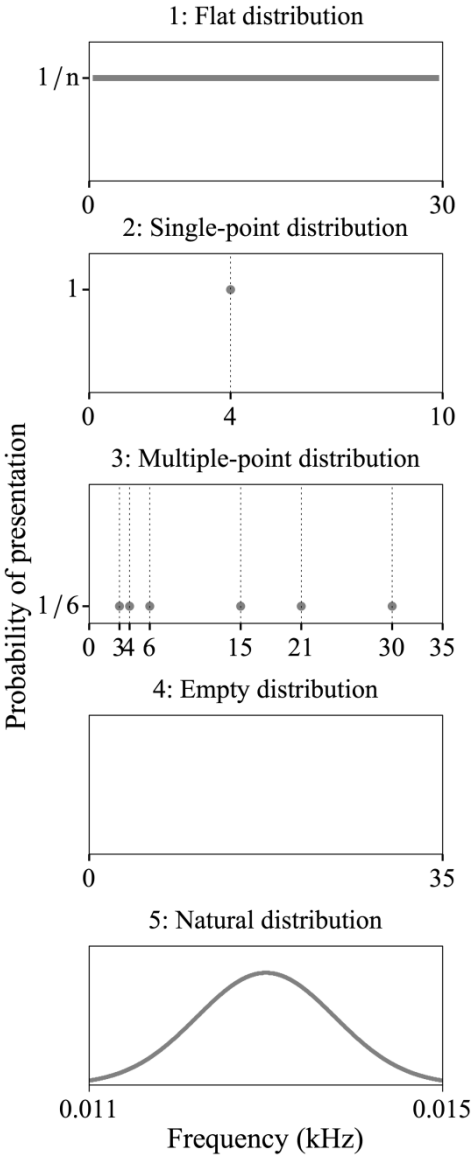


Figure VIII.3: Examples of distributions used in experiments with animals: (1) flat distribution, (2) single-point distribution, (3) multiple-point distribution, (4) empty distribution, (5) natural distribution. Explanation: see section 3.1.

distribution 1) causes the RFs to be incomplete[5] and multiple-peaked (Zhang et al., 2002), and exposure to repeated tones (single-point distribution 2) keeps the RFs broad and thus unselective, and produces an over-representation of the tone, as reflected in a larger area of neurons responsive to the presented tone than in normal development[6] (for cats: Stanton and Harrison, 1996; for rats: Zhang et al., 2001).

Similarly, exposure to a tone sequence (multiple-point distribution 3) alters map formation (as is discussed in section 3.4). These changes seem more or less permanent, since they are still present in adulthood, and since exposing adult animals to the same distributions does not lead to changes in tuning and map organization (e.g., Zhang et al., 2001). Cortical map development fails to proceed in deaf baby animals whose A1 is not stimulated within a certain time frame (empty distribution 4) (Ponton and Eggermont, 2001; Kral et al., 2001). In contrast to the effect of exposure to abnormal distributions (1 through 4), exposure to natural sounds (natural distribution 5) leads to balanced maps, in which CFs are well-ordered tonotopically (e.g., Zhang et al., 2001; Bao et al., 2013). All in all, the A1 firing patterns adapt to the sound distributions in the baby animal's environment.

It is likely that human infants' auditory map development is similarly affected by ambient sounds (Chang and Merzenich, 2003). Especially vowels may contribute to shaping the tonotopic map, because they are prominent in the speech stream and their defining properties, F1 and F2, are coded for in the tonotopic map (section 2).

5 When plotting a neuron's firing behaviour for a range of frequencies at a certain intensity level, normally the tuning curve appears as a continuous line, i.e., the neuron responds to *all* presented frequencies within its RF. However, when exposed to pulsed noise in the critical period after birth, the tuning curve appears as a series of scattered dots, i.e., the neuron does not respond to all frequencies within the RF. As a consequence, the RF is "incomplete".

6 The larger area responsive to the presented tone contrasts with a smaller overall tonotopic area (Zhang et al., 2001; De Villers-Sidani et al., 2007).

3.3. A series of sensitive periods

The baby animal studies on the effect of exposure to sound distributions also show that neurons in A1 tune for different parameters in different sensitive periods, i.e., periods of heightened susceptibility to change (Eggermont, 1996; Insanally et al., 2009). These periods seem to unfold in a cascading fashion: coding of "more basic" parameters precedes that of "more complex" parameters. For instance, neurons tune for frequency before they tune for frequency modulations (Insanally et al., 2009). The cascading pattern of a series of sensitive periods fits a more general pattern of cascading sensitive periods throughout language development, i.e., extending from the development of speech perception in the first year of life to the development in other language domains such as morphosyntax and semantics (Werker and Tees, 2005).

Interestingly, the earlier maturation of frequency tuning than frequency modulation tuning could play a role in the earlier emergence of language-specific perception for vowels than for plosives, in human infants. Speech sound discrimination tasks with infants reveal a transition from universal to language-specific perception in the first year of life: first discrimination performance is the same across infants regardless of the native language (*universal* perception), and then it gets better for native speech sound contrasts (e.g., Cheour et al., 1998b), and worse for non-native irrelevant contrasts (*language-specific* perception) (e.g., Werker and Tees, 1984). The language-specific discrimination appears around 6 months for vowels (e.g., Kuhl et al., 1992) and somewhat later between 8 and 12 months for plosives (e.g., Werker and Tees, 1984; Best et al., 1995). It is conceivable that this order of development is related to the just-mentioned earlier maturation of frequency tuning (necessary for vowel perception) than of frequency modulation tuning (necessary for plosive perception). This would mean that the emergence of language-specific perception of the different kinds of speech sounds is related to the end of the respective sensitive periods.

The notion of a series of sensitive periods has replaced the more traditional notion of a "critical period" with a fixed offset beyond which change is not possible. Instead, both the content and the timing of the presented sounds affect the

offset of the sensitive periods for A1 map development. For instance, the sensitive period for tonotopic map formation in rats appears very short (from postnatal day 11 to 13; De Villers-Sidani et al., 2007) when exposing pups to *single tones*, and longer with exposure to *white noise* (from postnatal day 8 to 28; Zhang et al., 2002). Similarly, the period is shortened when presenting *pulsed* sound (Zhang et al., 2001, 2002) and extended with exposure to *continuous* sound (Chang and Merzenich, 2003; Zhou et al., 2008). This last-mentioned effect is similar to that induced initially in deafness[7] (Ponton and Eggermont, 2001; Kral et al., 2001; Kral, 2013). Notice that the sensitive periods (henceforth the SP, i.e., the overall sensitive period) in rats do not start directly after birth, when rat pups do not hear properly yet, and when the mother rat raises her pups in relative silence. The SP onset only starts at the onset of hearing, which coincides with the onset of a period in which the mother rat increases the number of her vocalizations (De Villers-Sidani et al., 2007).

The onset of hearing and the SP in human infants starts before birth, in the last trimester of gestation (Querleu et al., 1988). As a result, newborns are sensitive to the properties of language that were transmitted relatively well into the mother's womb, such as prosody and rhythm (Nazzi et al., 1998). The offset of the SP depends on several factors (as just mentioned), but is unlikely to extend beyond approximately 3 years of age. This is because congenitally deaf children can develop normal hearing provided that their cortex starts receiving auditory stimulation before this time (Sharma et al., 2002; Kral, 2013). Since the lack of auditory input in deafness (empty distributions 4) *stretches* the duration of the SP (Ponton and Eggermont, 2001; Kral et al., 2001; Kral, 2013; section 3.2), the SP in normal development must end earlier.[8]

7 Sound deprivation and exposure to continuous sound do not have the same effect on the brain (the effect is only similar in extending the critical period). For instance, total sound deprivation lowers the excitation thresholds of neurons in the auditory cortex and reduces their dynamic range across frequencies, while presenting continuous noise raises excitation thresholds and does not reduce neurons' dynamic range (Kral, 2013).

8 This is not to say that auditory development is completed at 3 years of age. Rather, complete maturation of the brain and the maturation of auditory responses continue until adulthood (Moore and Guan, 2001; Ponton et al., 1996, 2000; Werner, 2007).

An interesting possibility to think about the SP in normal development is given by the above-mentioned observation that a mother rat increases the number of her vocalizations during the SP. Specifically, it is tempting to draw a parallel between the mother rat's vocalization behaviour and the use of infant-directed speech by human mothers in the first year of their infants' lives (section 1.2.1). Even though, of course, infant-directed speech is far more complex than rat vocalizations, causing much more variability in its use across mothers and cultures (Benders, 2013), it is striking that the time when mothers use infant-directed speech overlaps with the SP. Kuhl and colleagues (2008) view infant-directed speech as an important "agent of change" (2008: 982) in the transition from universal to language-specific speech perception, and posit that exposure to in particular infant-directed speech leads to "neural commitment" (2008: 983). The animal studies propose a specification of such neural commitment as cortical tuning in A1 during the SP, as also explicitly hypothesized by Chang and Merzenich (2003).

3.4. The influence of context

Another relevant observation in baby animal research on the impact of sound exposure, is that the tuning of A1 neurons in early development is not simply affected by spectral content, but rather by changes in spectral content over different time scales. On a small timescale, neurons tune for frequency modulations (Insanally et al., 2009; section 3.3). On a larger timescale, neurons tuning depends on the *order* (Nakahara et al., 2004) and *range* of tones in a presented sequence (Köver et al., 2013), as can be explained as follows.

The influence of the *order* of presentation is apparent from the effect of exposure to a repeated sequence of a limited number of tones (multiple-point distribution 3), which yields "sequence-specific" neuronal responses (Nakahara et al., 2004: 7170). Specifically, Nakahara and colleagues (2004) exposed rat pups to a train of relatively low-frequency tones (30-ms tones of 2.8 KHz at 0 ms, 5.6 kHz at 150 ms, and 4 kHz at 300 ms) followed by a train of relatively high-frequency

tones (30-ms tones of 15 kHz at 500 ms, 21 kHz at 650 ms, and 30 kHz at 800 ms) during the SP. Note that the multiple-point distribution in Figure VIII.3 reflects these frequencies, but cannot show the order of presentation. The exposure yielded a different A1 map and different neuronal responses as compared to those in normally developed rats. The most important result in the context of the current section appeared when comparing, after the exposure time, the responses to the presented sequence (i.e., the sequence presented during the exposure time) with the responses to their reversed version: the forward sequence elicited stronger and more reliable responses to each of the tones than the reversed sequence, and this order-selectivity was significantly larger than in control rats. In particular, the response to the third tone in the low-frequency train (i.e., the 4 kHz tone) was mostly absent in the reversed presentation, while being strong in the forward presentation. The zone in A1 representing this frequency was also reduced as compared to zones representing the other tones. In sum, the outcomes of Nakahara and colleagues demonstrate that neurons develop order-dependent firing properties.

Tuning is also affected by the *range* of tones in a sequence. In a study by Köver and co-researchers (2013), three experimental groups of rat pups were exposed to sequences of six tone pips drawn from a logarithmically uniform distribution (from 4 to 32 kHz; multiple-point distribution 3). A control group was reared normally. The only difference between the three experimental groups was the range of tones *within each sequence*: for the first group the frequency of each tone within a sequence was the same (the single-frequency group), for the second group it was drawn from the entire distribution (the full-range group), and for the third group it was drawn from either the lower or the higher half of the distribution (the half-range group). Importantly, all groups were exposed to frequencies across the whole distribution; the range was constrained only within each sequence. Consequently, there were no significant differences between the groups in the distribution of the CFs, hence in the structure of the tonotopic map. However, bandwidths (here: the width of the RFs measured at threshold intensity) were significantly affected, with narrower bandwidths for the single-frequency group, broader bandwidths for the full-range group, and shifted bandwidths for the half-

range group, as compared to the control group. In particular these shifted bandwidths are important in the context of the current section: they were accompanied by behavioural signs of categorical perception, as is explained in more detail below (section 3.5). Thus, the study by Köver and co-researchers (2013) shows that A1 neurons develop range-dependent firing properties, which can be related to behavioural perception.

In sum, neurons tune differently when exposed to different reoccurring contexts. This conclusion drawn from animal research supports the view that if distributional learning of speech sounds in human infants reflects a similar neuronal tuning, it will produce *context-dependent* representations of speech sounds, and not phonemes, which abstract away from the context (e.g., Pierrehumbert, 2003; section 1.2.1).

3.5. "Categorical" representations

Above it was shown that A1 firing patterns come to reflect the sound distributions in the baby animal's environment (sections 3.2 and 3.4). When exposing baby animals to more complex distributions than single-point or flat distributions, this reflection of sound distributions is similar to a basic kind of categorization (section 1.1). An example of "categorical" representations in A1 appears in the above-discussed study by Köver and colleagues, who measured both the changes in firing properties of neurons in A1 and behavioural perception (2013; multiple-point distribution 3; section 3.4). The neural measurements showed shifts in the RFs in the half-range group as compared to the control group: neurons had developed a preference for either responding to the lower or to the higher frequencies. In perception, only the half-range group showed better behavioural discrimination of two tones crossing the boundary frequency than of tones with equal acoustic distance (on a logarithmic scale) off the boundary, and this better discrimination at

the boundary frequency was significantly larger than in the control group.[9] Thus, exposure to sound distributions led to a basic form of categorization in A1 neurons' firing properties and in perception. This strengthens the idea that distributional learning can lead to primitive categorical representations already in A1.

Rat pups in another study were presented with natural jungle sounds, containing repetitions of more than 40 different "song motifs", i.e., distinct vocalization patterns of birds, mammals and insects (Bao et al., 2013; natural distribution 5). Here, categorical representations arose in A1 at the population level, i.e., involving multiple neurons, and at the single-neuron level. Specifically, the population responses to the different song motifs were more different and responses to variations of the same motif were more similar in animals exposed to jungle sounds than in control animals. Thus, animals exposed to jungle sounds became more sensitive to relevant differences and less sensitive to irrelevant differences between the song motifs. At the single-neuron level, the response selectivity was higher in jungle-exposed animals than in controls (i.e., RF bandwidths were narrower and responses tended to be in phase with the discerning features of a motif). Concomitantly, while more neurons responded to all presented sounds, generating a larger area of responsive neurons, a smaller number of neurons responded to each specific motif. Overall, exposure to the complex natural distributions led to distinct groups of A1 neurons firing selectively for song motifs. Thus, the heightened response selectivity at both the single-neuron level and at the population level shows again that simple exposure to sound distributions in early development can lead to activation patterns reflecting "categories" already at the lowest level of auditory processing (A1).

Not surprisingly, there is no similarly precise evidence of categorical speech sound representations as a result of distributional learning in human infant brains. However, studies with adult human participants confirm that such representations can be present at low levels of processing, probably including A1.

9 There was a significant difference in cross-boundary discrimination between the four groups (the groups are mentioned in section 3.4); and in post-hoc comparisons between the groups only the half-range group showed better cross-boundary discrimination than the control group.

For instance, Shestakova and colleagues (2004) observed that despite an impressive variability in the natural vowel stimuli that they presented to participants (the Russian vowels /a/, /i/ and /u/ were each pronounced once by 150 different talkers, yielding a total of 450 different tokens), the activated areas in the supratemporal lobe were consistently the same for stimuli representing the same vowel category. In addition, the areas of activation in response to the vowel stimuli were located orthogonally to the tonotopic map, a result that also appears in other studies with human participants (Diesch and Luce, 1997, 2000; Mäkelä et al., 2003; Obleser et al., 2003). Considering the categorical firing patterns that emerge in baby animals after exposure to sound (see above in this section), it can be speculated that the "speech sound categories" identifiable at low levels of cortical processing in human adults are due to distributional learning in infancy.

3.6. Summary and implications for distributional learning

In sum, the sound distributions that baby animals are exposed to early in life (section 3.1) impact the development of the primary auditory cortex (A1) fundamentally (sections 3.2 through 3.5). They determine the RFs and CFs (at the single-neuron level) and the structure of the auditory parameter maps and the population responses (at larger levels). The resulting firing properties affect perception, and remain into adulthood.

Distributional learning of speech sounds in human babies (section 1) could reflect similar changes in firing properties in A1 (and possibly in other low-level cortical areas). Vowels in particular can be expected to have an important early influence on tonotopic map development, since they stand out in the speech stream and since their main features, the formant frequencies F1 and F2, are probably coded for in an early stage of development (section 3.3). Because sound exposure can lead to context-dependent "categorical" representations in baby animals' A1 (sections 3.4 and 3.5), distributional speech sound learning may also lead to a basic kind of context-dependent speech sound "categories" in human infants' A1.

4. Plasticity in A1 in adulthood: the role of "attention"

The previous section described that distributional learning in baby animals, and thus possibly also the observed effects of distributional training in human infants (e.g., Maye et al., 2008; section 1), could reflect neuronal tuning on the basis of plain exposure to speech sounds. The question addressed below is whether the effects of distributional training in *adults* (e.g., Maye and Gerken, 2000; section 1) can be viewed as an identical process.

4.1. Limited change with passive exposure

The dominant view in the neuroscientific literature is that the tuning properties of neurons in the adult auditory system cannot change by plain exposure to sound distributions (reviews in Keuroghlian and Knudsen, 2007; Weinberger, 2007). Recent research has tempered this dominant view somewhat, by demonstrating changes in the tuning properties of A1 neurons in the adult cat, after persistent exposure (usually extending over months) to certain bandlimited sounds of fairly loud intensity levels (flat distribution 1, in section 3; review in Pienkowski and Eggermont, 2011). This shows that plain exposure *can* lead to representational changes in adult A1 neurons. However, the effect of such persistent exposure is similar to that induced by hearing loss: a smaller representation of the presented frequencies and a larger representation of the frequencies bordering on the presented frequencies. This effect differs from the larger representation of presented tones observed in baby animals after exposure to similar sound distributions (section 3).

Most research with adult animals marks the absence of plasticity in the adult A1 with mere exposure to sound (e.g., Recanzone et al., 1993; Zhang et al., 2001, 2002; Bao et al., 2004; Polley et al., 2006; Blake et al., 2006; Rutkowski and Weinberger, 2005; review in Keuroghlian and Knudsen, 2007). For instance, Zhang and colleagues (2001, 2002) exposed not only rat pups to pulsed noise and pulsed tones (flat distribution 1 and single-point distribution 2; see section 3), but also their mothers. In contrast to the profound changes in the pup A1, changes in

the mothers' A1 were not observed. Another example is provided by Polley and colleagues (2006), who trained two groups of adult rats with identical stimuli. One group was trained to attend to frequency cues (the "frequency recognition" group or FR group), the other group to attend to intensity cues (the "loudness recognition" group or LR group). Specifically, both groups were exposed to the same tone pips with various frequencies and intensities, but the FR group was trained to respond to a frequency of 4987 Hz at any intensity, while the LR group was trained to respond to an intensity of 35 dB sound pressure level (SPL) at any frequency. After training, the FR group showed plasticity in the tonotopic map only, and the LR group in the intensity map only. Thus, mere exposure to intensity cues was not sufficient to induce change in the intensity map of adult rats trained on frequency cues, and mere exposure to frequency cues did not suffice to cause change in the tonotopic map of adult rats trained on intensity cues. The conclusion that plasticity induced by passive exposure does not occur in the adult animal A1, signals that if distributional learning indeed reflects such plasticity (section 3.6), it cannot account for the effects of distributional training in human adults (section 1).

4.2. Change with explicit signals of behavioural relevance

The observation that simple exposure to sound does not yield changes in tuning properties of adult A1 neurons does not mean that such changes are impossible. A1 neurons remain plastic throughout life. However, in order to induce plasticity in A1 at an age beyond the SP, a sound must be coupled with an explicit sign of its behavioural relevance, that makes the organism attend to the sound (reviews in Weinberger and Bakin, 1998; Keuroghlian and Knudsen, 2007; Weinberger, 2007).[10] Such an explicit sign can be a reward or a punishment.

Various tasks have been used to make the animal attend to a stimulus. These include associative learning tasks and perceptual learning tasks (review in

10 Note that it is not mere "behavioural relevance" that is required to induce plastic changes, because, of course, natural sound distributions as they occur with plain exposure are behaviourally relevant in themselves: they provide the organism with information about the structure of the environment that it must adapt to.

Pienkowski and Eggermont, 2011). *Associative learning tasks* are classical or operant conditioning tasks. In classical conditioning, an association is formed between two stimuli. For instance, the animal is presented with a tone (stimulus 1, the conditioned stimulus) that is paired with a light shock (stimulus 2, the unconditioned stimulus). The tone thus acquires behavioural relevance, since it predicts the shock, which the animal will try to avoid (Weinberger and Bakin, 1998). In operant conditioning, an association is formed between a behaviour and its consequence. For instance, when the animal is presented with a certain tone (the conditioned stimulus), it can drink (the consequence) if it pushes a lever (the behaviour). The tone thus acquires behavioural relevance, because it indicates the availability of water (Rutkowski and Weinberger, 2005). *Perceptual learning tasks* are discrimination or identification training tasks, in which the animal is rewarded (e.g., with food) if it discriminates a pair of stimuli and "punished" (e.g., with a timeout) if it does not (e.g., Recanzone et al., 1993; Zhou and Merzenich, 2007; Polley et al., 2006; Blake et al., 2006). In this way, the animal's attention is drawn to the stimuli.

Considering the requirement that the animal must pay attention to the stimuli in order to induce plasticity in the adult animal A1, it is interesting that adult human participants in distributional training experiments are commonly asked to attend to the stimuli, and sometimes also to perform a task that is meant to keep their attention to the stimuli (e.g., to check each stimulus heard on a checklist: Maye and Gerken, 2000; Hayes-Harb, 2007). Thus, if distributional learning indeed reflects neuronal tuning, and if attention is needed for adult distributional learning, then the outcomes in distributional training experiments with adults (section 1) may involve "attentional mediation" of the distributional learning mechanism. Section 5.2.2 discusses this "attentional mediation" in more detail.

4.3. Robustness and the ability to adjust

The changes that are observed in the adult A1 after the different types of training (section 4.2) are similar in kind to those observed in the baby A1 after plain

exposure: tuning curves can change and acoustic properties can become represented by more neurons or by less neurons (see also section 4.4). At the same time, the changes seem less profound than those in baby animals (review in Keuroghlian and Knudsen, 2007). While in babyhood the presented sound determines the basic infrastructure of the cortical maps, in adulthood this basic layout is not altered. Changes in responses to a target frequency (i.e., a tone paired with positive or negative reinforcement) tend to occur only if this frequency is within the neuron's original RF, and are often shifts in preferred tuning frequencies (reviews in Keuroghlian and Knudsen, 2007; Pienkowski and Eggermont, 2011).

Perhaps contrary to expectation, such changes in the adult A1 do not necessarily require long training times and are not necessarily fragile: they can be induced within minutes and can persist for hours and even for months (Edeline et al., 1993; Weinberger and Bakin, 1998; Fritz et al., 2003, 2005; Zhou and Merzenich, 2007). Note that the studies on plasticity in the baby animal A1 always used relatively long training times (usually of several weeks; section 3). Therefore, it is not known whether short training times in baby animals also induce changes in neuronal tuning. Nevertheless, the manifestation of plasticity in the *adult* animal A1 already after short-term exposure to sound supports the possibility that a similar plasticity induced by exposure occurs in the A1 of human participants who demonstrate an effect of short-term distributional speech sound training (section 1).

Both the robustness of the basic layout of auditory maps in A1 and the ability for rapid adjustments are highly functional for the adult organism. The robustness of the basic A1 layout signals that the organism is well-adapted to the acoustic environment, while the ability for swift adaptation ensures that the organism is equipped for dealing with changes in this environment. In addition, plasticity in the adult A1 make the organism optimally geared towards dealing with *various* acoustic situations in the environment. This is apparent from observations that when neurons acquire new RFs in a task, they do not lose the old RFs, allowing the organism to switch between RFs depending on the task at hand (Fritz et al., 2005). These observations trigger the speculation that proficient bilinguals may switch RFs in A1 depending on the language that they listen to.

Interestingly, studies with human participants hint at the possibility that first acquired RFs may be lost if they are not re-activated from time to time during childhood. This possibility is based on the hypothesis by Werker and Tees (2005) that knowledge of a first language can be lost, if the language is not re-used periodically during childhood. The hypothesis was inferred from studies on the perception of Korean by two groups of adults: French adults who had been adopted from Korea when they were between 3 and 9 years of age and who had no subsequent experience with Korean (Pallier et al., 2003; Ventureyra et al., 2004), and second-generation Korean-American adults, most of whom were born in the United States and who had had differential experience with Korean throughout childhood (Oh et al., 2003). The Korean-French adoptees could not remember Korean and were fluent in French. In a study using a behavioural discrimination task, they did not discriminate Korean plosive phonemes significantly better than French control participants, irrespective of whether they had returned to Korea for short periods in adulthood (Ventureyra et al., 2004). Event-related fMRI (functional Magnetic Resonance Imaging) results in another study with participants from this population showed similar locations of activation when listening to French sentences for adoptees and French controls (Pallier et al., 2003). The activation patterns differed from those elicited by listening to sentences in other languages, including Korean. Thus, the adoptees seemed to have lost their knowledge of Korean, including possibly speech-related knowledge contained in A1 RFs. In contrast, the second-generation Korean-American participants still scored better in a Korean phoneme perception task than American controls, even if they had had only limited experience with perceiving Korean in childhood (Oh et al., 2003). The crucial difference between the adoptees and the second-generation participants seemed to be the latter group's occasional re-exposure to Korean throughout childhood (Werker and Tees, 2005). It can be speculated that such re-exposure also involves the occasional re-use of speech-related RFs in A1.

4.4. Area expansion and contraction during learning

A specific change that is often observed in the A1 of adult animals after training is a larger area of representation of trained sounds, the size of which tends to be correlated with the degree of improvement in behavioural discrimination (e.g., Recanzone et al., 1993; Rutkowski and Weinberger, 2005; Polley et al., 2006; review in Keuroghlian and Knudsen, 2007). There is, however, increasing evidence that the overrepresentation does not reflect the end of the learning process, but is an initial, transient stage, which is followed by a stage in which the area of representation shrinks again (Reed et al., 2011; review in Pienkowski and Eggermont, 2011). This process of expansion and contraction has been demonstrated convincingly in rats by Reed and colleagues (2011). The rats retained the discrimination performance that they had achieved during expansion of the representational area, after the size of this area had contracted again. Reed and co-researchers propose that the initial expansion of the representational area reflects a heightened dedication of resources to the task stimuli, and the subsequent normalization of the representational area reflects that the processing of these stimuli becomes more efficient. This "Expansion-Renormalization" model (Reed et al., 2011) of auditory learning can explain results of animal studies that seemed somewhat puzzling before, such as the longer rather than the expected shorter latencies that were found in combination with an expanded area of representation after discrimination training in owl monkeys (Recanzone et al., 1993) and the behavioural improvements *without* any changes in tuning properties and map plasticity (Brown et al., 2004).

A similar pattern of learning has also been proposed for speech sound learning in human participants on the basis of several neuroimaging results, which sometimes find increased activation in the auditory cortex, including in the A1, after speech sound training, and sometimes decreased activation (Zhang and Wang,

2007). [11] It has not yet been examined whether distributional speech sound training also provokes initial area expansion and subsequent area reduction.

4.5. Summary and implications for distributional learning

In contrast to the profound effect of plain exposure to sound distributions on the baby animal A1, the effect of plain exposure on the adult animal A1 is limited (section 4.1). However, plasticity in the adult animal A1 can be induced fairly easily if an explicit sign of behavioural relevance draws the organism's attention to the sound (section 4.2). The plasticity does not affect basic map layout (as it does in baby animals), but it allows the animal to adapt quickly to changes in the environment (section 4.3). In the course of learning, areas in A1 can get larger and smaller without an observable effect on behavioural perception. Hence, neuronal learning proceeds in phases that do not map separately onto behavioural perception (section 4.4). The observation that adult animals should pay active attention to the stimuli for plasticity to be induced hints at a possible crucial influence of "attention" in distributional training experiments with human adults, who have commonly been asked to pay attention to the stimuli during training (see section 5.2.2 for more details on this influence).

5. Factors underlying the different plasticity in adulthood than babyhood

Many differences between the infant and adult brain can contribute to the differential susceptibility for A1 map plasticity in babyhood (section 3) versus adulthood (section 4). At least two kinds of interrelated differences can be

11 Increased activation after speech sound training as measured with functional Magnetic Resonance Imaging (fMRI) in human participants can parallel the larger representational area (implying the activation of more neurons) and the larger activation of individual neurons that are observed in animals. It should be noted, however, that the temporal resolution of fMRI measurements is poor, so that it is possible that the fMRI outcomes reflect *longer* activation of the same neurons rather than activation of more neurons or increased activation of the same neurons.

observed, namely differences in the cortical structure (section 5.1) and differences in synaptic plasticity (section 5.2).

5.1. Cortical structure

The structure of the cortex in human newborns is very different from that in older children and adults. As will be clarified below, the initial structure and its development imply an enhanced capacity for distributional learning in infancy versus adulthood, and offer interesting possible explanations for the transition from universal to language-specific speech perception, which occurs in the first year of life (e.g., Werker and Tees, 1984; Kuhl et al., 1992), and which can be hypothesized to be based on the distributional learning mechanism (Maye et al., 2002; chapter II; section 3.3 in the current chapter).

The following description of the structure of the auditory cortex at different ages relies on research carried out by Jean Moore and colleagues (Moore and Guan, 2001; Moore, 2002; Moore and Linthicum, 2007), unless stated otherwise. In this research, two methods have been used to visualize auditory cortical structure in human post-mortem tissue, namely one method that highlights cell bodies (but not their branchings) and one method that exposes myelinated axons (but not their cell bodies and dendrites).[12] The *cell-body material* discloses the development of the cytoarchitecture, and thus the progressive *differentiation* of cortical cells into cell types, which each have different structures and functions and which each constitute a different cortical layer (see below). The development of the cytoarchitecture is largely genetically programmed, and thus largely independent of experience (Moore and Guan, 2001). The *axon material* visualizes the development of the myeloarchitecture, and thus the progressively increasing *efficiency* of the connections between neurons, once these connections have been established (section 5.2). This is because myelin, which is wrapped around the

12 A typical neuron has two types of processes extending from its cell body, namely *dendrites* for receiving information from other cells and *axons* to send information to other cells. The method used to visualize the axons does not highlight the myelin itself, but neurofilament protein, which is a prominent component in myelinated axons.

axon in a final stage of axonal development, increases conduction velocity and thereby reduces energy costs (Purves et al., 2008). Myeloarchitectural development is largely driven by experience. Figure VIII.4 is a schematic presentation of the results obtained by Moore and Guan (2001) with both methods (cell-body material in the top row; axon material in the bottom row) across the ages 0 (i.e., at birth), 4.5 months, 1 year and 12 years. The structure at 12 years is adult-like. Differences in cyto- and myeloarchitecture between adults (section 5.1.1) and infants (section 5.1.2), provide possible perspectives on the emergence of language-specific speech perception in the first year of life (section 5.1.3) and hint at differences in the capacity for distributional learning between adults and infants (section 5.1.4).

5.1.1. Cortical structure in human adults

The adult cortex (comparable to Figure VIII.4, right: at "12 years") has six well-defined layers, which run from the most superficial layer I to the deepest layer VI. Animal studies have begun to shed some light on the connections between neurons in different layers within and across cortical areas. These connections suggest certain information flows. The precise flows are highly complex and still largely unknown, in particular in the human brain (Hackett, 2011). What is important in the context of the current review, is the observation that the layered structure of the cortex must play a crucial role in information transport from lower- to higher-level areas (the bottom-up, or "feedforward" projections) and from higher- to lower-level areas (the top-down, or "feedback" projections). This can be explained by looking at the "canonically supposed flows" of information, as detected across sensory systems in animals (for the visual system: Van Essen and Maunsel, 1983; Felleman and Van Essen, 1991; for the auditory system: Galaburda and Pandya, 1983; Mitani and Shimokouchi, 1985; Rouiller et al., 1991). According to these canonical flows, bottom-up projections tend to arise from *supragranular layers* (I, II, III), and end in layer IV of higher-level areas, while top-down projections mostly arise from *infragranular layers* (V, VI), and avoid targeting layer IV of lower-level areas (their main targets are layers I and VI).

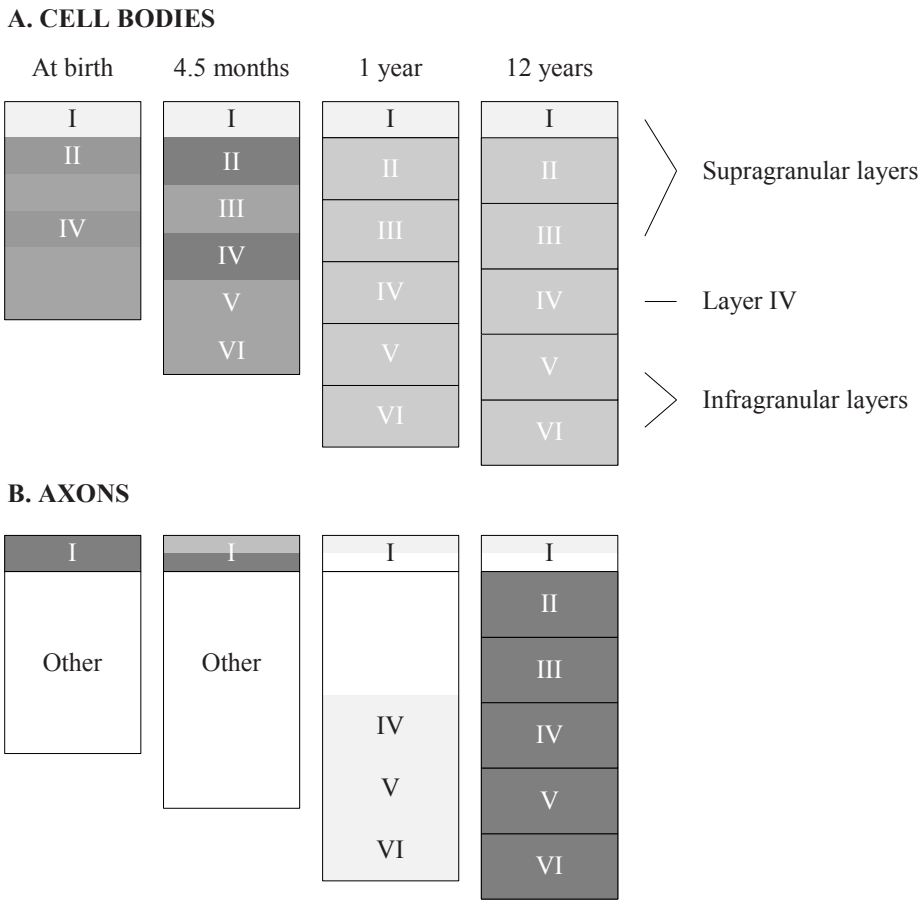


Figure VIII.4: Development (from left to right) of cortical layers in the human auditory cortex from birth to 12 years of life, when cortical structure is adult-like. Figure made on the basis of post-mortem data reported in Moore and Guan, 2001. Top row: cell body development; bottom row: development of myelinated axons. Further explanation: see text.

Note in Figure VIII.4 that the adult layer I contains fewer cell bodies (Figure VIII.4, top right: the shading is lighter) and is relatively *poorly* myelinated (Figure VIII.4, bottom right: the shading is lighter) as compared to the other layers. The reason for this is not clear. The layer contains mainly long axons that extend horizontally across cortical areas, thereby contacting apical dendrites of neurons in deeper layers. Animal research indicates that these axons are the axons of cell bodies in the thalamus (Cetas et al., 1999; Herkenham, 1980) and in adjacent cortical areas (Galaburda and Pandya, 1983). The axons coming from the thalamus originate largely from the *medial* division of the medial geniculate complex (MGC), in contrast to the axons that extend from the thalamus to layer IV of A1, which mainly originate from the *ventral* division of the MGC. Hypotheses as to the function of the information transmitted via layer I include providing "context" for the "content", where information on the content of the stimulus is delivered to layer IV (Llinás et al., 2002: 449), modulation (see sections 5.2.1 and 5.2.2), stimulation of plasticity (see sections 5.2.1 and 5.2.2), and sustenance of feedback interactions through the thalamus (Rubio-Garrido et al., 2009; Crick and Koch, 1998). All in all, the presence of a full six-layer infrastructure (Figure VIII.4, right) and the efficiency of this infrastructure (Figure VIII.4, bottom: the dark shading in layers II through VI denotes a high level of myelin) signal adults' ability to efficiently transport information bottom-up and top-down.

5.1.2. The development of cortical structure in infancy

The neonate cortex still has to acquire the infrastructure for bottom-up and top-down information transport that is visible in the adult cortex (section 5.1.1), and it does so by first emphasizing the development of the bottom-up pathway. This can be understood by looking at the development of cortical structure from birth (when the cortex is thinner than in adults, and largely lacks clearly differentiated and myelinated layers, as visible in Figure VIII.4, left) to the first birthday. This development is outlined below.

In the *cell-body material* around birth (Figure VIII.4, top left), layer I stands out, because the number of cell bodies is low as compared to that in the other layers (the shading is light), and layers II and IV are vaguely visible, because cell bodies in these imminent layers are more closely packed than in the other layers (the shading is darker). The visibility of these two layers may reflect the importance in early life of an infrastructure for bottom-up information transport: as described above (section 5.1.1), bottom-up projections arise from supragranular layers (among which layer II) and target layer IV of higher-level areas.

A rudimentary differentiation into the full six layers is visible around 4.5 months, and differentiation is more or less complete around the first birthday. In tandem with this differentiation, cell bodies also become less closely packed in the course of the first year of life (the shading in the top row of Figure VIII.4 gets lighter). This is probably due to the development of dendrites and axons, which take up increasingly more volume, thus deepening the cortex. Specifically, the complexity of branching rises after birth and peaks between 2 and 4 years of age, after which it declines again (Conel, 1939-1967, partly reproduced in Kral, 2007).

In the *axon material* around birth (Figure VIII.4, bottom left), again layer I stands out: it is the only layer that is heavily myelinated. Part of the axons probably plays a role in the differentiation of the cells (Marin-Padilla and Marin-Padilla, 1982; Moore and Guan, 2001). Axons with this function are concentrated in the lower tier of layer I at 4.5 months (Figure VIII.4, bottom 2nd from left), and have disappeared around the first birthday (3rd from left), when laminar differentiation is complete. Myelinated axons that are concentrated in the upper tier of layer I at 4.5 months also lose prominence in the first year of life. However, they do not disappear completely: some remain throughout life and retain a supporting role in processing (section 5.1.1).

Other layers than layer I do not contain myelinated axons at all in the first few months of life, indicating that information transport to and from higher-level cortical areas is largely inefficient. Between 4.5 and 12 months, myelination begins at axons in the deeper layers (IV to VI). In particular the myelination observed in layer IV hints at incipient efficiency of bottom-up projections (since this layer is

the canonical target of such projections; section 5.1.1). The myelination in layers V and VI could indicate incipient efficiency of top-down projections (since these layers are the canonical source of these projections; section 5.1.1), but could also reflect efficiency of bottom-up projections just as the myelination observed in layer IV (Moore, 2002). In subsequent years, the myelination proceeds to involve all layers, except layer I (where, as just explained, myelinated axons are lost).

The combined cell-body and axon material suggests an earlier maturation of bottom-up projections than top-down projections, which is particularly evident in the relatively early visibility of layer IV in the cell-body material and in the axon material, across auditory cortical areas. Animal studies confirm the earlier maturation of bottom-up projections than top-down projections (Barone et al., 1996; Batardière et al., 2002; see also Kral and Eggermont, 2007; Kral, 2013).

5.1.3. Implications for the onset of language-specific speech perception

As discussed in section 3, neuroscientific evidence supports the possibility that distributional learning reflects neuronal tuning in A1 and that the onset of language-specific speech perception in the second half of the first year of life marks the end of the sensitive period (SP) for neuronal tuning in response to ambient speech sounds. Moore and colleagues (Moore, 2002; Moore and Linthicum, 2007) propose a supplementary account of the transition from universal to language-specific speech perception, which is based on the observation that precisely around the time of the transition the first signs of myelinated axons appear in the deeper layers (IV to VI) of the cortex (Figure VIII.4, bottom: cf., the patterns at 4.5 months and at 1 year of age). The researchers hypothesize that these axons represent thalamocortical projections, and thus the first efficient bottom-up information transport from the thalamus to the auditory cortex. According to this proposal, universal speech perception in the first months of life is based on analyses performed in the brainstem, while the subsequent language-specific speech perception is based on the first analyses performed in A1. Note that in this scenario distributional learning as a bottom-up process is sufficient for producing

language-specific speech perception, and thus for triggering rudimentary categorical perception.

Unfortunately, it is impossible to exclude a role of top-down influences in the onset of language-specific speech perception. First, it is not certain that the myelinated axons that appear in the deeper layers represent thalamocortical projections. This holds in particular for the axons in layers V and VI, which are the canonical source of top-down projections in adults and which are also targets of top-down projections (section 5.1.1). Further, around the time of the transition from universal to language-specific speech perception, there is another development apart from incipient myelination in the deeper layers: neurons differentiate into the six layers that enable the transfer of information to and from higher-level cortical areas (Figure VIII.4, top). This leaves open the option that the onset of language-specific speech perception is based on the budding development of an infrastructure for top-down processing. Such an infrastructure would make it possible that unfolding higher-level representations modulate speech sound processing and learning in A1. Note that the baby animal data discussed in section 3 can also not exclude a possible role of top-down processing in the emergence of categorical firing patterns (section 3.5), even though these patterns emerged with "mere exposure" to sound distributions (hence suggesting purely bottom-up learning). This is because the studies concerned only measured the firing patterns resulting from exposure, without being able to measure whether these patterns arose via bottom-up projections, top-down projections or both. Hence, the precise role of bottom-up and top-down involvement in triggering the onset of language-specific speech perception remains unclear.

5.1.4. Summary and implications for distributional learning

The cyto- and myeloarchitecture of the auditory cortex from the age of 12 shows a well-developed infrastructure for bottom-up and top-down information transport, thus making it possible that higher-level linguistic representations exert a top-down influence on the distributional learning mechanism in older children and adults

(section 5.1.1). In contrast, the architecture in newborns fails such an infrastructure, and its subsequent growth in the first year of life implies a focus on the development of bottom-up before top-down projections, which is supported by research in animals (section 5.1.2).

Of course, it is possible that top-down projections are already effective in newborns via the well-myelinated layer I (section 5.1.2). However, this layer does probably not convey information on the *content* of the auditory stimulus (section 5.1.1; see also Moore, 2002). It is also possible that top-down projections reach the neonate's A1 via other layers than layer I in ways that do not surface with the cell-body method or the axon method (section 5.1). However, if such top-down connections exist in newborns, they do probably not convey information on higher-level *linguistic* representations (since it is unlikely that newborns already have such representations; section 1.2.2). Considering the above, it is likely that the bottom-up mechanism of distributional learning is less hampered by top-down cortical influences in early infancy than later in life. It remains unclear whether the onset of language-specific speech perception in the second half of the first year of life relies purely on bottom-up distributional learning or involves the first top-down influences of higher-level representations (section 5.1.3).

5.2. Functionality: synaptic plasticity

Once the cytoarchitecture is in place (section 5.1.2) and cells have grown axons and dendrites, connections can arise between cells, typically between the axon of one cell and a dendrite of another cell, so that signals can start passing between them, and the architecture can become functional. These structural and functional connections are synapses. Their functional development depends largely on experience: they become stronger or weaker depending on how much and when they are activated. For instance, they can become stronger when activity of the postsynaptic neuron is triggered repeatedly by activity of the presynaptic neuron, and they can become weaker with repeated occurrences of postsynaptic activity that is not related to presynaptic firing (Buonomano and Merzenich, 1998). The

strengthening of the synapse is called *Long-Term Potentiation* (LTP), and the weakening *Long-Term Depression* (LTD). LTP and LTD have been demonstrated in many parts of the brain, including in the animal auditory cortex (Wang et al., 1996) and in the human auditory cortex (Clapp et al., 2005; Zaehle et al., 2007). Hebbian learning, which was predicted to underlie distributional speech sound learning in some accounts (Guenther and Gjaja, 1996; McCandliss et al., 2002; section 1.3), reflects the processes of LTP and LTD. In the neuroscientific literature, synaptic plasticity as reflected in LTP and LTD is generally hypothesized to be the main mechanism underlying the experience-induced neuronal RF tuning and concomitant A1 map plasticity discussed in sections 3 and 4 (Buonomano and Merzenich, 1998). In line with this hypothesis, synaptic plasticity is larger in babyhood (section 5.2.1) than in adulthood (section 5.2.2), thus supporting the larger A1 map plasticity in babyhood (section 3) than in adulthood (section 4).

5.2.1. Synaptic plasticity in babyhood

The degree of synaptic plasticity is higher in babyhood than in adulthood, due to (1) a larger number of synapses and (2) a higher degree of plasticity at each synapse. The *larger number of synapses* is caused by a short period of synaptogenesis in infancy, followed by a prolonged period of synaptic pruning, which reduces the number of synapses by about half (Huttenlocher and Dabholkar, 1997). Roughly, the period of synaptogenesis runs parallel to the period of dendrite and axon development (section 5.1.2). The precise time course of synaptogenesis and pruning, which are both partly genetically programmed and partly influenced by experience (Kral, 2013), depends on the cortical area. In the auditory cortex, synaptic density reaches a maximum value already around 3 months of life (cf., 15 months in the prefrontal cortex); pruning starts after around 3 to 4 years of age and finishes around 12 years (cf., 16 years in the prefrontal cortex) (Huttenlocher and Dabholkar, 1997).

The degree of synaptic plasticity is also larger in early postnatal development than later in life due to a *higher degree of plasticity at each synapse*, which is induced by neuromodulation (Robertson et al., 1991). Neuromodulation is the adjustment of the activity of several neurons by means of neuromodulators, which are dispersed from certain parts of the brain across the cortex. For instance, the nucleus basalis in the basal forebrain is an important source of acetylcholine (Bakin and Weinberger, 1996), of which the neuromodulatory influence is enhanced during sensitive periods (Robertson et al., 1991; Aramakis et al., 2000; Consonni et al., 2009; Picciotto et al., 2012). The higher degree of synaptic plasticity in babyhood than in adulthood probably contributes to the higher map plasticity in the baby A1 than the adult A1 (section 3.2 versus 4.1), and may thus contribute to a higher capacity for distributional speech sound learning in human baby's than in adults (section 5.1; see also chapter III).

There are differences in the pace of maturation between synapses at different levels of processing and between different types of synapses. *Thalamocortical* synapses mature before *corticocortical* synapses (reviews in Froemke and Jones, 2011; Froemke and Martins, 2011), thus supporting a dominant role of bottom-up distributional learning early in life, which was also deduced above from the research on the development of cortical structure (section 5.1.4). Further, *excitatory* synapses mature before *inhibitory* synapses (Dorrn et al., 2010).[13] Both developmental patterns at the synaptic level (i.e., "thalamocortical before corticocortical" and "excitatory before inhibitory") support patterns of change observed at the level of the neurons' RFs in A1 (section 3), as summarized in Table VIII.1. At the level of neurons' RFs (middle row in Table VIII.1), *spectral* tuning precedes tuning for *fast temporal* properties (Insanally et al., 2009; section 3.3). At the synaptic level (top row), the former (spectral tuning) involves excitatory circuits (Kaur et al., 2004; Liu et al., 2007), while the latter (fast temporal tuning) involves the later maturing inhibitory circuits (Zhang et al.,

13 Synapses are excitatory, when firing of the presynaptic cell *increases* the probability that the postsynaptic cell will fire. Conversely, they are inhibitory when presynaptic firing *reduces* the probability of postsynaptic activity.

2003). Additionally, these inhibitory circuits necessary for fast temporal tuning rely on corticocortical connections (Zhang et al., 2003), while the circuits necessary for spectral tuning also rely on the earlier maturing thalamocortical connections (Kaur et al., 2004; Liu et al., 2007). The differential paces of maturation at the synaptic level, which can be related to differential paces of RF tuning at the neuronal level, could contribute to the earlier emergence in behavioural discrimination (bottom row in Table VIII.1) of language-specific *vowel* perception (for which spectral tuning is necessary) than of language-specific *plosive* perception (for which fast temporal tuning is crucial).

Table VIII.1: Potential neural contribution (of maturational processes at the level of the synapse and at the level of the neuron) to the earlier onset of language-specific perception for vowels than for plosives. Explanation: see text.

Level of analysis	Preceding maturation	Subsequent maturation
Synapse	Thalamocortical synapse	Corticocortical synapse
	Excitatory synapse	Inhibitory synapse
	|	|
Neuron	Spectral tuning	Fast temporal tuning
	|	|
Language-specific perception	Vowel perception	Plosive perception

5.2.2. Synaptic plasticity in adulthood

Section 4 discussed that plasticity can only be induced in the adult A1 if an auditory stimulus is coupled with a sign of behavioural relevance, which draws the organism's attention to the stimulus. Interestingly, the same plasticity can be induced if, instead, the auditory stimulus is combined with electrostimulation of certain neuromodulatory nuclei, in particular of the nucleus basalis (NB) in the basal forebrain (Bakin and Weinberger, 1996; Kilgard and Merzenich, 1998; Bao

et al., 2003) and the ventral tegmental area (VTA) in the midbrain (Bao et al., 2001). The involvement of the NB in inducing plasticity in the adult A1 indicates that the role of neuromodulation in boosting synaptic plasticity is not confined to sensitive periods (cf., section 5.2.1), even if the effects of this modulation may differ in adulthood versus babyhood.

The NB and the VTA are part of different modulatory systems: the NB is part of the cholinergic system that uses the neuromodulator acetylcholine (ACh), while the VTA is part of the dopaminergic system that uses the neuromodulator dopamine (DA). The precise functions of the two systems in stimulating synaptic plasticity in adulthood are complex and not fully understood. The complexity is apparent from the multiple connections between brain areas involved in these systems. Both structures receive information from various parts of the brain, including the thalamus and the amygdala (for NB: Morris et al., 1998; for VTA: Phillipson, 1979), and both project information across the cortex, including the A1 (for NB: Kilgard and Merzenich, 1998; for VTA: Bao et al., 2001). There are also projections from the VTA to the NB (Bao et al., 2001). The functions of the two structures are thus at least partly interrelated.

An important function of both the NB and the VTA is thought to be the dispatch of information about the behavioural relevance of stimuli across the cortex, thereby enhancing the organism's responsiveness to behaviourally relevant stimuli and reducing its responsiveness to irrelevant stimuli (for the NB: Bakin and Weinberger, 1996; Kilgard and Merzenich, 1998; Picciotto et al., 2012; for the VTA: Schultz, 1992; Bao et al., 2001). Several specifications of this function have been proposed. For instance, DA is said to draw a participant's attention to a stimulus by enhancing its salience (Schultz, 1992), and to relate a reward to a preceding stimulus (Bao et al., 2001). ACh may help to raise and sustain an organism's attention to incoming stimuli (Himmelheber et al., 2000, 2001; Arnold et al., 2002) Note that these proposed functions of neuromodulation via the NB and VTA tie in with the animal research discussed in section 4, which showed that in order to elicit plasticity in the adult A1, the auditory stimuli must be paired with

explicit signs of *behavioural relevance* that make the animal pay active *attention* to the stimuli.

Another proposed neuromodulatory function of ACh is to stimulate the coding of external stimuli by enhancing the influence of bottom-up processing, while hampering the retrieval of existent representations in memory by diminishing the influence of feedback processing (Hasselmo, 2006). This proposal is endorsed by recent studies with slice preparations of the mouse thalamus and A1, which show that plasticity of thalamocortical synapses is inhibited in adulthood, but can be unmasked (i.e., the synapses can be made plastic again) by a combination of cortical disinhibition (i.e., preventing inhibitory circuits to modulate the excitatory bottom-up processing) and cholinergic neuromodulation from the NB (Blundon et al., 2011; Chun et al., 2013). All in all, neuromodulation plays an important role in revitalizing synaptic plasticity in lower-level cortical areas such as A1.

5.2.3. Summary and implications for distributional learning

Synaptic plasticity (i.e., plasticity at the level of the synapse) is viewed in the neuroscientific literature as a major factor underlying RF plasticity (i.e., plasticity at the level of the neuron). In accordance with this view, synaptic plasticity is higher in babyhood (section 5.2.1) than in adulthood (section 5.2.2), just as RF plasticity was found to be higher in babyhood (section 3) than in adulthood (section 4). Hence, if distributional speech sound learning in humans reflects synaptic and RF plasticity, then the capacity for such learning will be higher in infants than in adults, thus confirming the outcomes of chapter III.

Section 4 showed that although mere exposure to sound stimuli can hardly induce any RF plasticity in A1 in adult animals, such RF plasticity *can* occur in these animals provided that the exposure is coupled with an explicit sign of behavioural relevance (such as a reward or punishment) that draws the animal's attention to the stimuli. Section 5.2 now described in more detail that such attention probably triggers neuromodulatory influence from subcortical structures that can heighten the degree of synaptic plasticity in the adult A1 again. This signals that

observed effects of distributional speech sound training in human adults may rely on neuromodulatory influence triggered by attention to the stimuli in the experiment.

6. Discussion

This paper reviewed possible neural correlates of distributional speech sound learning. The main conclusion is that such learning may reflect *changes in firing properties of neurons in the primary auditory cortex (A1), and possibly at other low levels of cortical auditory processing, under the influence of exposure to speech sound distributions*. This conclusion is in accordance with Guenther and Gjaja's proposal, which they incorporated in their computer simulation of distributional learning (1996) (section 1.3), and also reflects a conjecture put forward in the literature on the development of the rat pup A1 by Chang and Merzenich (2003).

Exposure-induced changes in firing properties have been observed in the A1 of different kinds of animals. They occur predominantly in babyhood (section 6.1), but can also be observed in adulthood, provided bottom-up stimulation of A1 neurons is combined with neuromodulatory influence on these neurons (section 6.2). Such influence comes from subcortical nuclei and should be taken into account when studying distributional learning, in addition to possible top-down influence of (linguistic) representations coming from higher-level cortical areas (section 6.3). Despite the interesting explanations that neuroscientific evidence offers for distributional speech sound learning, it is clear that future research is needed to examine the effects of distributional learning beyond A1 and to unravel the impact of the mechanism from that of other mechanisms on creating categorical perception (section 6.4).

6.1. Distributional learning in infancy

Distributional learning was found to correspond to experience-induced changes in A1 neurons' firing properties. These changes occur predominantly in a certain sensitive period in babyhood when auditory parameter maps are still under construction, and the effects of ambient sounds on neurons' firing properties and concomitantly on the layout of the maps are still profound. There are several reasons to be confident that the neural correlates of distributional speech sound learning in human infants tie in with the experience-induced neural changes observed in baby animals.

First, just as the animal A1 (section 3), the human A1 is not fully developed at birth (Moore and Guan, 2001; section 5.1). The maturation of the auditory cortex, including that of A1, features particularly drastic changes in the first year of life, which is precisely the period that marks an important change in infants' perception, namely the transition from universal to language-specific perception of speech sounds (Werker and Tees, 1984; Kuhl et al., 1992). The unfinished nature of the human auditory system early in life is also visible in the rather large portion of total sleep time spent in so-called non-rapid eye movement (NREM) sleep, a sleep stage that is considered to be indispensable for the development of sensory systems, and of which the prominence declines in the course of the first 8 months of life (Graven and Browne, 2008).

Second, just as mother rats increase the number of their vocalizations in the sensitive period (De Villers-Sidani et al., 2007), human mothers adapt the way in which they address their infants during the just-mentioned period of major auditory sensory development (section 3.3). They do so by using infant-directed speech, in which acoustic characteristics of speech sounds are exaggerated and more varied than in adult-directed speech (Kuhl et al., 1997; section 1.2.1). The mother rats' vocalization behaviour supposedly contributes to balanced auditory map formation. Human mothers' use of more variation in infant-directed speech could serve a similar purpose. The idea that infant-directed speech may contribute to auditory map formation supports the claim by Kuhl and colleagues, that infant-directed

speech is an important "agent of change" in infants' speech sound acquisition, which leads to "neural commitment" (e.g., Kuhl et al., 2008: 982).

Third, even if direct relations between neuronal firing patterns and behavioural perception should be viewed with extreme caution (see also section 6.4.3), a remarkable parallel exists between on the one hand the earlier onset of language-specific speech perception for vowels (around 6 months) than for plosives (between 8 and 12 months) in human infants and on the other hand the earlier maturation of spectral tuning (crucial for vowel perception) than of fast temporal tuning (crucial for plosive perception) as measured in the animal A1 (Insanally et al., 2009; section 5.2.1). The different paces of tuning for spectral and fast temporal properties of sound are supported by certain maturational differences at the level of the synapse (section 5.2.1).

Fourth, there is a match between, on the one hand, the primitive categorical representations of ambient sounds that arise in the baby animal A1, and, on the other hand, the categorical speech sound representations observed at low levels of processing in humans (section 3.5). The observation that exposure to sound distributions can lead to simple categorical representations in the animal A1 thus supports the hypothesis that sound exposure may lead to elementary speech sound representations in humans (section 1.2.1). The context-dependent nature of the representations observed in the animal A1 supports the claim by Pierrehumbert, that representations resulting from distributional speech sound learning are at best "positional variants of phonemes" (Pierrehumbert, 2003: 129; section 1.2.1).

In sum, the four correspondences between observations in baby animal research and observations in research pertaining to humans endorse the proposal that distributional speech sound learning in human infants reflects neuronal tuning in low-level auditory cortical areas, triggered by ambient speech sound distributions.

6.2 Distributional learning in adulthood

Changes in firing properties of A1 neurons are also observed in adult animals, provided the "bottom-up" stimulation of A1 neurons (i.e., induced by external sound distributions) is combined with neuromodulatory influence on these neurons coming from subcortical nuclei such as the nucleus basalis and the ventral tegmental area. Such neuromodulatory influence can be triggered by making the adult animal attend to the presented sounds, and is instrumental in boosting bottom-up plasticity, while obstructing top-down influence of higher-level representations (section 5.2.2).

The often cumbersome acquisition of non-native speech sounds by adults, even when they have been exposed to the speech sounds for years (Cebrian, 2006; Escudero and Wanrooij, 2010) may thus reflect the limited plasticity that is observed in the A1 of adult animals with plain exposure to sound. The effects of distributional speech sound training that are observed in human adults after only a few minutes of exposure time seem to be at odds with such limited plasticity. However, they can be explained by the just-mentioned observation that neuromodulation related to "attention" can regenerate A1 plasticity in adult animals. Distributional training experiments with adult participants have commonly included ways to make participants attend to the stimuli during distributional exposure. Therefore, these experiments may have induced a neuromodulatory and thus facilitating influence on distributional learning.

6.3. Two kinds of neural influence on distributional learning

This review reveals the importance of distinguishing two kinds of neural influence that can affect distributional learning in lower-level areas such as A1, namely *neocortical* influence (i.e., originating from other parts of the neocortex), and *subcortical* influence (i.e., coming from subcortical areas).

Neocortical influence includes influence from higher-level linguistic representations, and is thus the top-down influence that is referred to in linguistic theory (section 1.2). It also includes influence from representations in other

modalities such as vision. Neocortical influence only develops fully in the course of childhood (section 5.1.4).

Subcortical influence refers to influence via older parts of the cortex than the neocortex, such as the amygdala, basal ganglia, and various nuclei. An example of subcortical influence discussed in this paper is the neuromodulation from the nucleus basalis, which stimulates plasticity and which is triggered by "behavioural relevance" of and "attention" to a stimulus (section 5.2.2). The origin of this subcortical influence is not clear. Some studies hypothesize that higher-level cortical representations in for instance the parietal or prefrontal cortex trigger the influence in a top-down way (e.g., Polley et al., 2006; Pienkowski and Eggermont, 2011). However, it is not certain whether such top-down influence is involved (e.g., Kilgard and Merzenich, 1998). Relatedly, the concept of "attention" is often associated with top-down influence, in particular in the case of voluntary attentional control (e.g., Roelfsema, 2011). However, attentional processes are complex and can also be related to bottom-up activity, in particular when attention is triggered by a salient stimulus (Awh et al., 2012). Hence, it is not clear whether the neuromodulatory influence that induces plasticity in the adult A1 should be considered a bottom-up influence, a top-down influence, or both.

Subcortical influence has not received much attention in research on speech processing and learning, even though its scope must be immense across a lifetime. In contrast to the infrastructure for neocortical top-down influence, the infrastructure of subcortical influence is present from birth (for neuromodulatory influence in babyhood: see section 5.2.1. This influence may be projected via layer I, which is well-functioning in neonates: see section 5.1.2).

The importance of distinguishing the two types of influence is also clear from the opposite effects that they can have on the mechanism of distributional learning in human adults who try to master a contrast between non-native speech sounds: while cortical top-down influence from higher-level linguistic representations may hamper distributional learning in A1 (sections 1.2.2 and 5.1.4), subcortical influence may revive plasticity in A1 and thus stimulate distributional learning (section 5.2.2).

6.4. Remaining puzzles

6.4.1. Involvement of areas beyond A1

The main conclusion in the current review relies for a considerable part on research pertaining to the A1 in animals. In view of similarities across animal species of the effects of sound exposure on A1 processing (sections 3 and 4), it is likely that the results also apply to humans. At the same time, there are differences between non-human animals and humans in sound processing and learning. This can be deduced from the many differences between animal species in the number of subareas and their characteristics even at the lowest levels of auditory processing (Hackett, 2011). It is not known to what extent these differences affect the conclusions in this review. It is likely, however, that the direct effect of sound exposure on neurons' tuning properties extends beyond A1 to other low-level areas (as already incorporated in the formulation of the main conclusion; section 6). This is because direct projections from the thalamus to the auditory cortex are not confined to projections to A1, but include parallel projections to other primary auditory subareas and to subareas in the secondary auditory cortex (Hackett, 2011). Even within A1, distributional learning must be a more complex mechanism than reviewed here, in view of the two parallel projections to A1, one to the left and one to the right hemisphere, each of which may have a different focus of analysis (Hickok and Poeppel, 2007). A direct effect of a presented sound distribution on the tuning properties of cortical neurons can thus be expected at several low-level auditory areas. Future research awaits the task of revealing and disentangling such effects.

6.4.2. The creation of categorical representations

The present review shows that even *categorical* representations can arise in A1, when exposing an organism to sound distributions (sections 3.5 and 6.1). Unfortunately, this observation does not solve a persistent problem in linguistic theory (Boersma, 2012), namely what triggers the *creation* of these representations.

Specifically, it is possible that the categorical firing patterns were created in a purely bottom-up way, i.e., via "a sufficient amount" of bottom-up distributional learning. However, it also remains possible that the firing patterns became categorical only once top-down influence from higher-level representations was established. This is because when measuring RFs in living animals (as was done in the studies that report the development of categorical firing patterns; Bao et al., 2013; Köver et al., 2013), possible top-down connections to the cells of which the RFs are measured cannot be severed, and therefore top-down influence cannot be excluded. Just as it is unclear what triggered the emergence of categorical firing patterns in A1, it is unclear what precisely causes the emergence of categorical *perception*. This issue is taken up in the next section.

6.4.3. The relation with perception

Even though correspondences are found between firing patterns in A1 and perception (for baby animals: Han et al., 2007; Köver et al., 2013; for adult animals: see example references in section 4.4), it is clear that a percept is the result of a more complex process than just A1 firing patterns. The involvement of many areas beyond sensory areas is demonstrated nicely in a study by Romo and De Lafuente (2013). It is also clear from the time window needed to yield a percept that a participant can report. This time window is around 150 to 200 ms after stimulus onset, while an acoustic signal already reaches A1 after 9 to 10 ms (Näätänen and Winkler, 1999), after which it disperses across the brain. The reportable percept thus includes contributions of firing patterns across the brain.

Because it is still vague what triggers the creation of categorical firing patterns (the previous section 6.4.2), and because the contribution of such firing patterns to perception is still unclear (this section), it remains unsolved what precisely triggers the onset of categorical perception. In particular, considering the focus of this review on speech sound acquisition, the reviewed neuroscientific data cannot pinpoint what causes the onset of infants' language-specific speech perception (and supposedly the concomitant offset of the sensitive period for

learning a certain speech sound category) at around 6 months for vowels and 8 to 12 months for plosives. Since characteristics of the input can fasten or delay the end of the sensitive period as observed in firing patterns and in perception (section 3.3) and since bottom-up connections become efficient in the infant brain just before the onset of language-specific speech perception (Moore, 2002; section 5.1.3), it is conceivable that the onset is based on long-enough, adequate bottom-up learning alone. On the other hand, the cortical layers that enable top-down influence from higher-level parts of the cortex also evolve just before the onset of language-specific speech perception (Moore and Guan, 2001; section 5.1.3), so that an involvement of top-down influence on this onset cannot be excluded. Unravelling the precise roles of nurture (experience-induced bottom-up learning) and nature (the maturation of cortical structures) in establishing categorical perception thus remains a major topic for future research.

Chapter IX

General discussion

1. Introduction

This thesis concentrated on the role of *distributional learning* in the acquisition of *vowel categories*, by *infants* acquiring the vowels of their first language and by *adults* learning the vowels of a new language. The main conclusions are that (1) distributional learning can contribute to the acquisition of native speech sound categories in infancy (section 2.2); (2) the capacity for distributional learning is larger in infants than in adults (section 2.2); and (3) observed effects of distributional training in the lab may not be based on the number of peaks in the training distributions (section 2.4).

These and other conclusions are summarized in Table IX.1. Below, I discuss all conclusions within a discussion of the five research topics and the related research questions, which were presented in the Introduction (section 2). For convenience, I repeat the five topics here: the replicability of distributional training experiments (section 2.1), the possibly changing role of distributional learning with age (section 2.2), potential differences in the effectiveness of distributional training between listener types within conditions (section 2.3), possible effects of manipulations of the training distributions (section 2.4), and neurobiological mechanisms of distributional learning (section 2.5). This chapter ends with some directions for future research (section 3) and concluding remarks (section 4).

Table IX.1. The five research topics, with the conclusions and corresponding chapters (cf. Table I.2 in the Introduction).

Topic	Conclusions	Chapter
1. Replicability of distributional training experiments	1.1. An "effect of distributional training" is replicable in infants and adults. (However, see conclusion 4.3 below).	II, V-VI
	1.2. For adults, the pattern of replications prompts the hypothesis that the chances of finding such an effect may be higher when the learners' native language has a relatively small number of vowels.	III-VI
	1.3. It remains uncertain whether the MMR method can be used to measure effects of distributional training.	II-III
2. The role of distributional learning with age (infants vs. adults)	2.1. Distributional learning can contribute to the emergence of language-specific speech perception in the first year of life, and thus to the acquisition of native-language speech sound categories (cf., conclusion 5.3).	II
	2.2. The capacity for distributional learning is larger in infancy than in adulthood.	III, VIII
3. Possible differences between listener types within conditions	3.1. Some listener types use fewer cues than other listener types when perceiving non-native vowels. Mere exposure to a vowel distribution encourages each listener type to start using increasingly subtle cues. (That is, cues are adopted in an order that probably reflects an order of declining salience).	V

Table IX.1 (continued).

Topic	Conclusions	Chapter
4. Possible effects of manipulations of the distributions	4.1. Exposure to an *enhanced* bimodal vowel distribution improves adult listeners' categorization of representative vowels, but it remains indeterminate whether enhanced distributional training is more effective for learning speech sounds than non-enhanced distributional training.	V
	4.2. Exposure to a bimodal vowel distribution with *high stimulus variability* improves adult listeners' categorization of representative vowels, but it remains unclear whether high-variability training is more effective for learning speech sounds than low-variability training.	VI
	4.3. The number of peaks (i.e., of means) in the distributions may not underlie any of the reported effects of distributional training observed in the lab to date. An important potential confound is a difference in the dispersion of the distributions.	VII
5. Neurobiological mechanisms of distributional learning	5.1. Distributional learning may reflect experience-induced changes in firing properties of neurons at the lowest levels of cortical auditory processing.	VIII (literature review)
	5.2. In adult animals, distributional learning requires "attention" to the stimuli that causes neuromodulatory influence from subcortical structures (such as the nucleus basalis) to revive plasticity. A similar neuromodulatory influence may have occurred in human adult participants in studies reporting a distributional training effect, since these studies commonly required the adult participants to pay close attention to the stimuli.	
	5.3. With exposure to sound distributions, the firing properties of neurons at the lowest levels of cortical auditory processing can come to reflect basic, context-specific categorical representations (cf., conclusion 2.1). However, the precise role of distributional learning in category creation remains unclear.	

2. Conclusions pertaining to the research topics

2.1. Replicability of distributional training experiments

At the beginning of the project in 2009, it was not clear whether an effect of distributional training was replicable, in particular for infants, and whether it was replicable with new contrasts and novel methods. Tables IX.2 (for infants) and IX.3 (for adults) show the experimental studies on distributional learning known at the end of 2014. Six of the 19 studies are presented in this thesis (chapters II through VII). Not all statistical comparisons between experimental groups that were reported in each study are listed in the tables. The focus is on the comparisons measuring an effect of distributional training, i.e., comparisons of a bimodal and a control group[1] (and not on, for instance, a comparison between enhanced and non-enhanced bimodal training). When more than two experimental groups participated, the tables list the *p*-value for the overall comparison between the groups whenever this value was reported and non-significant, thus invalidating a further comparison between the groups.

The first question asked in the Introduction was whether distributional learning can indeed be demonstrated as a mechanism in infants in a distributional training paradigm. Chapter II answers this question in the affirmative, with a novel contrast (the English vowel contrast /ɛ/~/æ/) selected for a sample drawn from a novel, younger population (2-to-3-month old Dutch infants), and with a novel method (electrophysiological measurements).

The overall picture that emerges from the complete list of infant studies known in 2014 (Table IX.2) is that *an effect of distributional training is replicable in infants*, even if not easily: most studies report such an effect on the basis of (nearly) significant results. Experiments with infants in the lab are notoriously difficult, because infants cannot be told to perform a task and the chances of

[1] Table IX.3 contains one comparison between a *unimodal* and a control group (Hayes-Harb, 2007). This comparison can also illustrate an effect of distributional training: if participants can discriminate a contrast before training and if distributional learning occurs during exposure to a unimodal distribution, participants' discrimination performance will *drop* in the post-test as compared to the pre-test. This is because the unimodal distribution trains the participants *not* to perceive a contrast (see chapter I, section 1.3).

dropouts due to fussiness are high. Considering these difficulties, the use of the neurophysiological method for very young infants in chapter II may have facilitated the detection of a distributional training effect: the measurement of the mismatch response, or MMR, does not require the infant to comply with a task, and the age of 2 to 3 months is an age at which infants are relatively quiet as compared to older infants, thus yielding fewer artefacts triggered by crying, fussiness and movement. Notwithstanding the conclusion that a distributional training effect is replicable in infants, it is clear that more replications are welcome for a more conclusive assessment of this replicability: Table IX.2 includes effects that were non-significant (Pons et al., 2006a, 2006b) or ambiguous (Cristià et al., 2011; see note (i) in Table IX.2), and, as mentioned in the Introduction, it is possible that more non-significant effects exist but remain unreported.

Another research question was whether an effect of distributional training can be replicated *in adults* with new speech sound contrasts appropriate for new native-language groups. Following Escudero et al. (2011), chapters V and VI confirm that distributional training in the lab can elicit learning in adults, with a new contrast (Dutch /a/~/ɑ/) appropriate for a new population (native speakers of Spanish). According to all three studies, Spanish listeners to an enhanced bimodal distribution of Dutch /a/~/ɑ/ improve more in their categorization accuracy of several natural [a]- and [ɑ]-tokens than listeners to music.[2]

On the other hand, chapters III and IV failed to find a straightforward effect of distributional training in yet another native-language group, namely native speakers of Dutch who were exposed to distributions encompassing the English contrast /ɛ/~/æ/. A clear effect was found neither with neurophysiological measurements (chapter III), nor with behavioural measurements (chapter IV). These indecisive outcomes for the Dutch adults versus the observed effectiveness of distributional training for the Spanish adults (Escudero et al., 2011; chapters V

2 Escudero and Williams (2014) also tested Spanish learners of Dutch /a/~/ɑ/. Their results are difficult to compare to the results in Escudero et al. (2011), and in chapters V and VI, because their research focussed on the longer-term effects of distributional training (i.e., after 6 and 12 months) than just after a few minutes of training. After 6 months, the participants trained with an enhanced bimodal distribution showed better improvement than participants exposed to music.

and VI) and for Bulgarian adults (Gulian et al., 2007) hint at the possibility that the success of distributional training in adults depends on the trained speech sound contrast in relation to the native-language speech sound inventory. Specifically, distributional training of non-native vowel contrasts might be more successful for native speakers of languages with a relatively small number of native vowels (e.g., Bulgarian, Spanish), who must split a single native-language category, than for native speakers of languages with a relatively *large* number of native vowels (e.g., Dutch), who must *shift* a native category boundary (see chapter IV). Future research should examine this hypothesis.

Taken together (Table IX.3), the studies on distributional vowel training in adults suggest that *an effect of such training can be replicated with new contrasts appropriate for new adult populations. The pattern of replications prompts the hypothesis that the chances of finding such an effect may be higher when the learners' native language has a relatively small number of vowels.*

The thesis also examined whether an effect of distributional training can be obtained with a novel method, namely the measurement of the MMR instead of the behavioural methods that had been used in all other experiments on distributional learning. The infant study in chapter II suggested that this is indeed possible. However, as was touched upon above (this section), the same method did not yield a similar positive effect in adults (chapter III). Since the study in chapter III was the first to use the MMR method to examine adult distributional learning, it was important to explore further whether the new method was possibly inappropriate for examining this in adults. Therefore, an additional behavioural control study was conducted (chapter IV). This control study tested again whether Dutch adults could learn the English vowel contrast /ɛ/~/æ/ via distributional training, just as was done in the neurophysiological study in chapter III. However, this time the method was exactly the same as in the above-mentioned three studies that obtained an effect of distributional training in Spanish adults (chapters V and VI, and Escudero et al., 2011). Just as the neurophysiological study in chapter III, the behavioural study in chapter IV did not yield a straightforward effect of distributional training in the Dutch adults. Hence, it has become less likely that the MMR method caused the

non-significance of the distributional training effect in the adults in chapter III. Still, it is impossible to conclude on the basis of chapters III and IV that the MMR method is suitable for measuring effects of distributional training in adults. In addition, even though the MMR method yielded a significant effect of distributional training in the *infants* (chapter II), it is clear that replications of this result are called for to substantiate the conclusion that the MMR method is suitable for measuring distributional learning in infants (see also section 2.5). All in all, I conclude that *it remains uncertain whether the MMR method can be used to measure effects of distributional training*.

More important than the conclusions mentioned above, however, is the following concern. Even if an effect of distributional training is replicable, this does not necessarily entail that such an effect reflects the mechanism of distributional learning, i.e., learning based on the number of peaks in the input distribution. Chapter VII raises the concern that distributional learning experiments in the lab may *not* tap this mechanism. This point is discussed further in section 2.4.

Table IX.2. Studies on ***infant*** distributional learning known in 2014, with participants' age (in months) and native language (L1), the non-native speech sound contrast in the bimodal training distributions (contrast), the duration of the training (Time, in minutes), the groups that were compared (bi = bimodal, uni = unimodal), the total number of participants in the combined groups mentioned in the groups column included in the analysis (N included) and additional participants tested (N excluded), and the *p*-value of the comparisons.

Study	Age (mths)	L1	Contrast	Time (min.)	Groups	N incl. (N excl.)	*p* value
Consonants							
Maye et al., 2002	6 – 9	English	/d/~/t/[c]	2.3[e]	- bi vs. uni	48 (12)	0.063
Maye et al., 2008	7 – 9	English	/d/~/t/[c]	2.8	- bi vs. uni	97 (56)	0.001
			or		- bi vs.		0.001
			/g/~/k/[c]		non-speech		
Yoshida et al., 2010	10 – 12	English	/d/~/t/[c]	2.3[e]	- bi vs. uni	48 (11)	0.85
	10 – 11	English	/ɖ/~/ɖ̪/ (Hindi)	1.9[f]	- bi vs. flat	48 (14)	0.87
	10 – 11	English	/d/~/t/[c]	4.6[g]	- bi vs. flat	48 (21)	0.036
Capel et al., 2011	11	Dutch	/ɖ/~/ʈ/ (Hindi)	2.7	- bi vs. uni	54 (39)	0.053
Cristià et al., 2011[a]	4 – 6	English	/s̪/~/ɛ/[d] (Polish)	2.6	- bi vs. flat	64 (34)	?[i]
Vowels							
Pons et al., 2006a[b]	6	English	/ɛ/~/ɛː/	?[h]	- bi vs. uni	?[h]	"ns"
Pons et al., 2006b[b]	8	English	/e/~/ɪ/	?[h]	- bi vs. uni	32 (?[h])	"ns"
Chapter II	2 – 3	Dutch	/ɛ/~/æ/ (English)	12.1	- bi vs. uni	35 (1)	0.016

a) Variation in the training stimuli was implemented along two dimensions (representing the fricative and vocalic portions of the training syllables, respectively) so as to yield a *two-dimensional grid* of training stimuli. This is in contrast to the other studies mentioned in the table, where one or more properties of the training stimuli were varied such that a *one-dimensional list* of training stimuli resulted.
b) Unpublished results presented in posters at conferences.
c) The contrast between voiced and voiceless unaspirated plosives (such as /d/ versus /t/ and /g/ versus /k/) is not phonemic in English, even though the orthography suggests that it is. The distinction only appears in allophonic contexts. English has a voicing contrast between "voiceless" unaspirated plosives (such as /d̥/ at the onset of the word "do" and /g̊/ at the onset of "game") and voiceless aspirated plosives (as /tʰ/ and /kʰ/ at the onset of "two" and "came" respectively).
d) The contrast was the contrast between Polish "retroflex and alveolopalatal sibilants" (Cristià et al., 2011: 388). I follow Ladefoged and Maddieson (1996) in the use of the non-IPA symbol /ʂ̺/ for the "retroflex" sibilant rather than the IPA-symbol /ʂ/, because the Polish variant is not truly retroflex.
e) Training duration without fillers was around 1.5 minutes (deduced from the text as follows: 96 training stimuli * (465 ms + 500 ms inter-stimulus interval)).
f) Training duration without fillers was 1.2 minutes.
g) Training duration without fillers was 3.1 minutes.
h) ? = not reported.
i) Results were ambiguous: $p > 0.16$ for the main effect, but $p = 0.007$ for an interaction effect.

Table IX.3. Studies on ***adult*** distributional learning known in 2014. Variables: see Table IX.2.

Study	Age (yrs)	L1	Contrast	Time (min.)	Groups	N incl. (N excl.)	*p* value
Consonants							
Maye & Gerken, 2000	18 – 41	English	/d/~/t/[a]	9[d]	- bi vs. uni	32	< 0.05
Maye & Gerken, 2001	(Students)	English	/d/~/t/[a]	9[d]	- bi vs. uni	32	< 0.01
	(Students)	English	/ɡ/~/k/[a]	9[d]	- bi vs. uni	32	< 0.05
Hayes-Harb, 2007	(Students)	English	/ɡ/~/k/[a]	9[d]	- bi vs. uni	66	0.04
					- bi vs. no training		0.235
					- uni vs. no training[g]		0.007
Peperkamp et al., 2003	(Adults)	French	/ʁ/~/χ/[b]	9	- bi1 vs. bi2 vs. uni[h]	60	> 0.1[i]
Shea et al., 2006	(Students)	English	/dæ/~/dˤɑ/[c]	12[e]	- bi vs. uni	32	< 0.01[j]
	(Adults)	Spanish					
Vowels							
Gulian et al., 2007	16 – 60	Bulgarian	/ɑ/-/a/, /ɪ/-/i/ (Dutch)	5	- bi vs. uni	40	0.029
Escudero et al., 2011	24 – 63	Spanish	/ɑ/-/a/ (Dutch)	1.9	- bi vs. non-speech	159	0.91
					- enhanced bi vs. non-speech		0.020

Table IX.3 (continued).

Study	Age (yrs)	L1	Contrast	Time (min.)	Groups	N incl. (N excl.)	*p* value
Vowels							
Chapter V	19 – 60	Spanish	/ɑ/-/a/ (Dutch)	1.9	- bi vs. non-speech - enhanced bi vs. non-speech	150	0.592 0.038
Chapter VI	19 – 63	Spanish	/ɑ/-/a/ (Dutch)	1.9	- discontinuous bi vs. non-speech - continuous bi vs. non-speech	150	0.04 < 0.001
Escudero & Williams, 2014	24 – 63	Spanish	/ɑ/-/a/ (Dutch)	1.9	- enhanced bi vs. bi vs. non-speech	79	0.055[k]
Chapter III	18 – 30	Dutch	/ɛ/~/æ/ (English)	12.1	- bi vs. uni	39 (5)	0.45
Chapter IV	18 – 30	Dutch	/ɛ/~/æ/ (English)	1.9	- bi vs. non-speech	100	?[l]
Chapter VII	19 – 56	Spanish	/ɑ/-/a/ (Dutch)	5.7[f]	- bi vs. uni	120	0.67

a) See note (c) of Table IX.2.
b) The distinction between voiced and voiceless uvular fricatives is allophonic in French, not phonemic.
c) Participants were exposed to either a unimodal or a bimodal distribution based on either the consonant continuum /dV/~/dˤV/ (vowel kept constant) or the vowel continuum /Cæ/~/Cɑ/ (consonant kept constant). The consonant contrast represents the Arabic contrast between non-emphatic and emphatic (pharyngealized) alveolar plosives, which is accompanied by allophonic variation in the vowel /æ/. After the emphatic plosive, the second vowel formant is lowered, yielding /ɑ/. The vowel contrast is phonemic for English listeners, not for Spanish listeners.
d) Half of the training stimuli consisted of fillers. The precise duration of exposure to the training stimuli (i.e., without the fillers) cannot be calculated from the article.
e) Training duration without fillers was 6 minutes.
f) Training duration without fillers was 3.8 minutes.
g) See note 1 in the main text of this chapter.
h) There were two bimodal groups. In one group each VC-sequence in the training was coupled with a CV-syllable where the C agreed in voicing with the preceding C. In the other group the Cs did not agree in voicing.
i) The *p*-value represents the interaction between the test (post- vs. pre-test) and the distribution (uni- vs. bimodal1 vs. bimodal2).
j) The *p*-value represents the interaction between the test (post- vs. pre-test) and the distribution (uni- vs. bimodal) across language groups.
k) Participants in the three groups did five tests: a pre-test and a post-test in session 1, again a pre-test and a post-test in session 2 six months later, and a single test in session 3, six months after session 2. Several comparisons were made. Only the comparison between the three groups in session 1 is reported in the table.
l) The *p*-value of 0.020 cannot reflect an effect of distributional training, because participants exposed to music improved *more* in classification performance than participants exposed to the enhanced bimodal distribution.

2.2. The role of distributional learning with age

In 2009, distributional training experiments had only been performed with infants from 6 months of age, i.e., from an age when speech sound perception is already turning language-specific (e.g., Kuhl et al., 1992). An ensuing question was whether distributional learning can actually *contribute* to the development of language-specific speech sound perception. Chapter II demonstrated an effect of distributional training in infants aged 2 to 3 months, i.e., well before the onset of language-specific speech perception. We concluded that *distributional learning can contribute to the emergence of language-specific speech perception in the first year of life, and thus to the acquisition of native-language speech sound categories*. The quest for neural correlates of distributional learning in a literature review in chapter VIII supported this conclusion. It showed that when exposing baby animals to sound distributions, categorical representations, as reflected in neurons' firing properties, can come to be observed in the primary auditory cortex.

Another question was whether the capacity for distributional learning is different in adulthood, when speech sounds of new languages must be learned, than in infancy, when the speech sounds of the mother tongue must be mastered. In chapter III, a new method involving the normalization of MMR amplitudes was developed to make a direct comparison possible between infant and adult distributional learning in the lab. The outcomes disclosed that *the capacity for distributional learning is larger in infancy than in adulthood*. The literature review of neural correlates of distributional learning in chapter VIII endorsed this conclusion. It showed that while passive exposure to sound distributions has a large impact on the firing properties of neurons in the baby animal auditory cortex, it does not lead to similar changes in the auditory cortex of adult animals (see also section 2.5).

2.3. Possible differences between listener types within conditions

A further research question, which was inspired by the trend in linguistic research to pay attention to individual differences between participants, was whether

distributional training can affect types of listeners *within* experimental conditions differently. Chapter V introduced "latent class regression analysis" (Huang and Bandeen-Roche, 2004) to the field of speech perception and learning research, to identify such types of listeners within each experimental group (one group exposed to a bimodal distribution, one to an enhanced bimodal distribution and one to music). This analysis is useful for this purpose, because it extracts the types without an a priori definition of these types or of the number of types, i.e., it enables the identification of "latent" (non-overt) types ("classes") in a statistically reliable way. In chapter V, each type was defined as using a different "listening strategy", i.e., a different combination of acoustic cues, in vowel perception. The listening strategies identified in the pre-test were compared to those identified in the post-test after the training phase.

In line with the previous literature (e.g., Escudero and Boersma, 2004; Morrison, 2008, 2009; Chandrasekaran et al., 2010), chapter V confirmed that listeners display different capacities to use the appropriate cues when listening to non-native speech sounds: some listener types used fewer cues than other types. Chapter V then showed that *mere exposure to a vowel distribution (and not exposure to music) encourages each listener type to start using increasingly subtle cues*. That is, *cues are adopted in an order that probably reflects an order of declining salience* (see chapter V, where the identified order was 1. duration – 2. first formant – 3. second formant – 4. third formant).

At this point in time, it is not clear whether (and if so to what extent) these results can be generalized to other native-language groups. As mentioned above (section 2.1), Dutch learners of English /ɛ/~/æ/ did not show an effect of distributional training, neither with neurophysiological measurements nor with behavioural measurements. Hence, it is not clear whether Dutch listeners learn from a distributional vowel training. If we assume that they do not learn from a vowel training *and* if the results in chapter V are generalizable to the Dutch listeners, then these results suggest that exposure cannot not add cues to Dutch listeners' listening strategies, because they are already used to focusing on "subtle" cues (including the second formant, which was a subtle cue for Spanish

participants listening to Dutch /a/~/ɑ/). It is clear that this speculation should be investigated further.

2.4. Possible effects of manipulations of the distributions

The research questions at the beginning of the project focussed on two types of manipulations of the distributions: "enhancement" and "more variability". Enhancement was implemented by pulling apart the means of the two categories in the bimodal distribution, thereby enlarging the range and standard deviation of the distribution (Figure IX.1: middle versus top). More variability was implemented by synthetically adding acoustic variation while avoiding stimulus repetition (Figure IX.1: middle versus bottom). The specific research questions were whether enhancement and more variability can benefit distributional speech sound learning. It was hypothesized that they could, on the basis of previous work on speech sound training, which used other paradigms than distributional training (e.g., Jamieson and Morosan, 1986; for details see chapters V and VI).

In line with Escudero et al. (2011), chapters V and VI indeed showed that *exposure to an enhanced bimodal distribution of vowels improves adult listeners' categorization of representative vowels*, and it does so more than exposure to instrumental music. The replication of this conclusion across three studies adds to its reliability. At the same time, enhanced bimodal training (Figure IX.1, middle) was not significantly more beneficial for adult learners than *non*-enhanced bimodal training (Figure IX.1, top) in the three distributional training studies with both enhanced and non-enhanced bimodal conditions (chapter V; Escudero et al., 2011; Escudero and Williams, 2014). Accordingly, *it remains indeterminate whether enhanced distributional training is more effective for learning speech sounds than non-enhanced distributional training*. Notably, this uncertainty is not confined to *distributional* speech sound training: other paradigms for adult speech sound training also yield non-significant differences between enhanced and non-enhanced training (Iverson et al., 2005; Kondaurova and Francis, 2010).

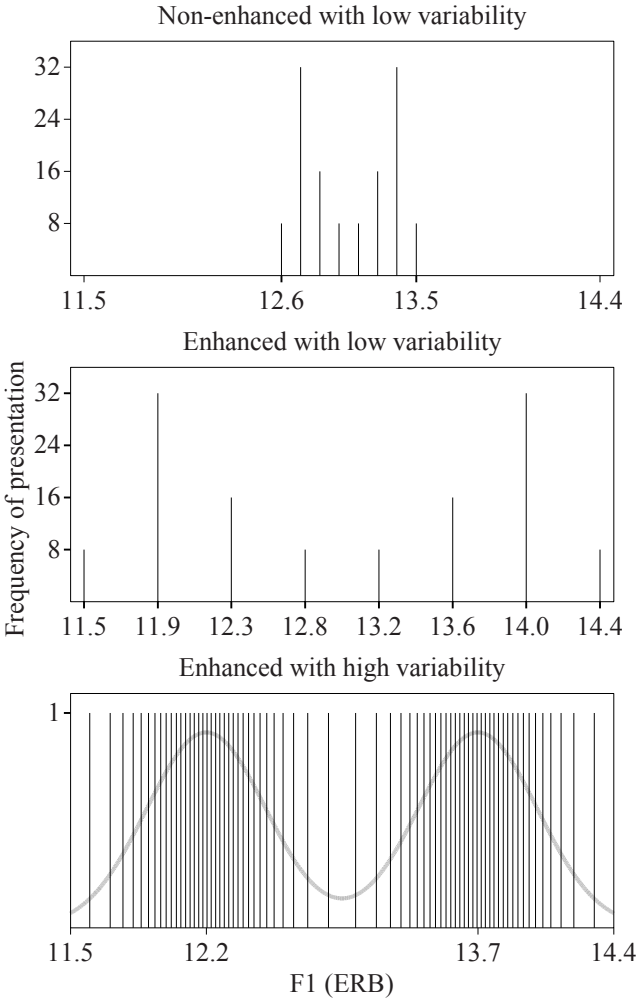


Figure IX.1. Bimodal distributions. Non-enhanced (top) versus enhanced (middle) distributions used in chapter V. Low-variability (middle) versus high-variability (bottom) distributions used in chapter VI. The grey curve (bottom) shows the underlying probability density.

Similarly, Chapter VI showed that *exposure to a bimodal vowel distribution with high stimulus variability improves adult listeners' categorization of representative vowels*, and it does so more than exposure to instrumental music. At the same time, high-variability training (Figure IX.1, bottom), which was used in a distributional training experiment for the first time, did not benefit adult learners significantly more than low-variability training (Figure IX.1, middle). Therefore, *it remains unspecified whether high-variability distributional training is more effective for learning speech sounds than low-variability training*. It should be emphasized that this uncertainty also applies to studies with other kinds of speech sound training than distributional training: when these studies report a beneficial effect of high variability for adult learners, they tend to abstain from a direct statistical comparison between high- and low-variability training (e.g., Logan et al., 1991; Lively et al., 1993; Bradlow et al., 1997). In the rare cases where such a comparison is made, the results are inconclusive, due to possibly confounding factors (high variability is often implemented simultaneously with enhancement in so-called "perceptual fading" paradigms, which are explained in chapter VII; McCandliss et al., 2002; Jamieson and Morosan, 1989) and due to confusing outcomes (e.g., high-variability perceptual fading was beneficial for learners when speech sounds were presented without feedback, but not when presented with feedback: McCandliss et al., 2002; and it was beneficial when speech sounds were synthetic, but not when they were natural: Jamieson and Morosan, 1989).

As mentioned above, Chapters V and VI were intended to examine the effects of enhancement and variability on distributional learning; they were not intended primarily to demonstrate the existence of distributional learning as a mechanism, as had already been done in earlier distributional training studies (Maye and Gerken, 2000, 2001; Gulian et al., 2007; Hayes-Harb, 2007). Hence, the use of a unimodal control group was less important than in these earlier studies, and was in the end avoided altogether in view of the consideration that unimodal training might lead to *impaired* discrimination of the two speech sound categories represented in the bimodal distribution (Maye et al., 2002; Hayes-Harb, 2007).

This was undesirable because most participants were learners of the language that had the presented bimodal contrast (namely they were Spanish learners of Dutch, which has the presented bimodal contrast /a/~/ɑ/). To avoid a potentially harming effect of a unimodal distribution, the control group in chapters V and VI (as well as in Escudero et al., 2011, and in Escudero and Williams, 2014) was exposed to music instead. Be that as it may, the lack of a unimodal control group brought about the unfortunate possibility that the results (i.e., better improvement of categorization after enhanced bimodal training than after exposure to music) were not based on the number of peaks in the distribution, and thus not on distributional learning. Instead, the results could be based on the enhancement of the training distribution (in chapter VI, possibly in combination with the high variability of the training stimuli), or on the processing of speech in the training distribution versus the processing of non-speech in the music condition.

Chapter VII was designed as a control experiment that should assess whether the number of peaks could really underlie the observed distributional training results. For this, we created unimodal and bimodal distributions with equal enhancement and variability. More specifically, the enhancement was matched by making three measures of dispersion as equal as possible in the two distributions: the range, the standard deviation and the edge strength (see chapter VII). The variability was matched by making sure that each of the two distributions contained the same number of acoustically different stimuli, namely 128 (each presented once). By matching the enhancement and variability, these two factors were excluded as possible confounds; by using only vowel distributions, the factor of processing speech or non-speech was excluded as a possible confound. The results showed a non-significant difference between the two conditions in a frequentist significance test, and undisputed evidence for the *absence* of learning based on the number of peaks, in a series of Bayesian tests.

An even more disturbing observation in this control study in chapter VII, was that the absence of distributional learning may not be confined to the distributional training studies with music control groups (chapters V and VI; Escudero et al., 2011; Escudero and Williams, 2014), but may extend to all

distributional training studies to date, including those with infants and those with adults. This is because even though other studies used unimodal control groups and thus controlled for the confound "processing speech versus non-speech", *none* of the previous studies examining distributional learning in a laboratory setting fully excluded a possible confounding influence of enhancement: at least one measure of dispersion was larger in the bimodal distribution than in the control condition across studies. Consequently, *the number of peaks (i.e., of means) in the distributions may not underlie any of the reported effects of distributional training observed in the lab to date, including effects observed in infants and effects observed in adults*. *An important potential confound is a difference in the dispersion of the distributions.* This is not to deny that distributional learning does not exist. In fact, observations of natural speech sound learning (see Introduction section 2.1) and neuroscientific evidence (literature review in chapter VIII) substantiate its existence, at least in infants. The conclusion only shows that the mechanism may not be easy to tap in a few-minute session in the lab. Notice also that although the conclusions drawn in chapters II through VI do not necessarily pertain to an effect of learning based on the number of peaks in the distribution, they still pertain to the effect of exposure to a vowel distribution.

2.5. Neurobiological mechanisms of distributional learning

In linguistic theory, distributional learning is viewed as a low-level, bottom-up mechanism that may produce rudimentary representations of speech sound categories (see section 1.2 in chapter VIII). The final research question in the Introduction was whether it is possible to pinpoint concrete neurobiological processes in the brain that could represent or affect such a low-level, bottom-up mechanism. The literature review of animal and human neuroscientific evidence in chapter VIII illustrated that *distributional learning may reflect experience-induced changes in firing properties of neurons at the lowest levels of cortical auditory processing*.

The review of neuroscientific evidence also suggested *a larger capacity for such experience-induced changes in babyhood than in adulthood*, which endorses the results in chapter III (where the measurement and analysis of MMRs implied a larger capacity for distributional speech sound learning for infants than adults; section 2.2). Factors that could underlie a higher capacity for distributional learning in babyhood than in adulthood were a higher degree of synaptic plasticity, and the apparent lack of an infrastructure for cortical top-down influence from higher-level to lower-level representations (apart from the probable lack of such higher-level representations themselves). Both factors are partly governed by genetically programmed maturation, and partly by experience (see the review in chapter VIII).

Further, the literature review in chapter VIII described that *in adult animals, distributional learning requires "attention" to the stimuli that causes neuromodulatory influence from subcortical structures (such as the nucleus basalis) to revive plasticity in the lower-level auditory cortex* (e.g., Keuroghlian and Knudsen, 2007; see the literature review in chapter VIII). In experiments with adult animals, attention is elicited by making sure that the animal is actively involved in a task. Considering the importance of attention for adult learning, it is of significance that attention commonly features prominently in distributional training experiments with adults. Participants are usually explicitly prompted to pay attention to the training stimuli. In addition, many training distributions that have been used to date had wide dispersions in one or more measures (section 2.4), and such wide dispersions supposedly draw participants' attention to the critical differences between the speech sound categories contained in the distribution (as hypothesized in studies using other training paradigms than distributional training: e.g., Jamieson and Morosan, 1986; Iverson et al., 2005).

Furthermore, in all behavioural distributional training experiments with adults so far, participants not only had to pay attention to the training stimuli, but also to the *test* stimuli (i.e., they received a task in which they had to respond actively to these stimuli). The potential importance of attention to the *test* stimuli for learning is also apparent from evidence that participants improve their

perception *in the course* of the active behavioural test: specifically, the Dutch participants in the behavioural distributional training experiment (chapter IV) improved significantly already *during* the pre-test (De Vos, 2012). Conversely, the adult participants in the *MMR* experiment, who were asked *not* to pay attention to the test stimuli (chapter III), did not significantly improve (or get worse) in their discrimination performance during the test (Appendix to chapter III). If attention paid to the *test* stimuli in the pre-test reinforces adults' learning during the *training*, it is possible that the MMR method, where participants do not have to pay attention to the *test* stimuli, is less suitable for making adults learn during the *training*. This possibility should be inspected in future research.

In sum, attention has been shown to be an important prerequisite for distributional learning in adult animals; it induces a neuromodulatory influence from subcortical structures on low-level auditory cortex, which temporarily revives plasticity here and thereby enables distributional learning again (reviews in Keuroghlian and Knudsen, 2007; and in chapter VIII). *A similar neuromodulatory influence may have occurred in human adult participants in studies reporting a distributional training effect, since these studies commonly required the adult participants to pay close attention to the stimuli.*

Another interesting observation in the literature review in chapter VIII was that *with exposure to sound distributions, the firing properties of neurons at the lowest levels of cortical auditory processing can come to reflect basic, context-specific categorical representations*, at least in babyhood. This observation supports the conclusion of chapter II, that distributional learning can contribute to the acquisition of language-specific speech sound categories (section 2.2). *However, the precise role of distributional learning in* creating *categorical representations could not be specified*, because none of the experiments (neither those with animals or humans reviewed in chapter VIII, nor the experiments described in chapters II through VII) could clarify whether the categorical firing patterns arose purely as a result of "a sufficient amount" of bottom-up learning, or only emerged after the establishment of a top-down influence from higher-level cortical representations (section 6.4.2 in chapter VIII).

The thesis, in combination with earlier research, *did* show some support for the idea that category creation involves not only a sudden event, but also gradual development ("gradual" in the sense of "ongoing", "proceeding in small steps"). The "sudden" emergence of language-specific speech perception in the second half of the first year (e.g., Werker and Tees, 1984; Kuhl et al., 1992) suggests that category creation is a sudden event. Similarly, in chapter V of this thesis, the jumps in accuracy scores when listeners adopted new cues could also be seen as rather abrupt behavioural changes reflecting the emergence of new categorical representations. At the same time, chapter V revealed *gradual* changes in listeners' perception of the non-native test vowels after exposure: cues were added one by one, and the use of duration was intensified, without necessarily being accompanied by jumps in accuracy scores. The idea of gradual category creation, and relatedly the idea of *dynamic* categorical representations (rather than categorical representations that are fixed entities), fit the concept of a category as explained in the Introduction, namely a category with fuzzy boundaries (Rosch, 1973; Rosch and Mervis, 1975; Rosch et al., 1976). Such dynamic categories can be observed at several levels of linguistic analysis. Rosch and colleagues identified categories with fuzzy boundaries at different hierarchical levels within the conceptual level. For speech sound categories, it has been shown that representations of language-specific categories are moulded and improved in the course of childhood (Hazan and Barrett, 2000). In addition, it has been demonstrated that in the course of speech sound training neurophysiological changes precede behavioural changes (Tremblay et al., 1998). In view of the above, it is meaningful that attempts to model learning in a neurobiologically valid manner, incorporate gradual category creation (McClelland and Rumelhart, 1986; see also Boersma, 2012). The idea that *gradual* development of categorical representations can lead to *sudden* behavioural changes may become even more attractive, when considering that the combination of gradual change and sudden outcomes is also omnipresent in natural phenomena far beyond language learning, such as the sudden changes in states of matter with gradual rises in temperature (chemistry), or the sudden movement of a heavy object when gradually increasing

the force directed at it (physics). In summary, this thesis provided some support for the concept of dynamic categorical representations (in chapter V), and showed that distributional learning can contribute to the creation of categorical representations (chapter II and the review in chapter VIII), but could not pinpoint the exact role of distributional learning in the category creation process.

3. Future directions

Many conclusions in this dissertation (discussed in section 2; overview in Table IX.1) are based on results obtained with innovative methods in the field of distributional speech sound learning, i.e., the MMR method to assess an effect of distributional training in chapters II and III, the method to compare infant and adult MMRs in chapter III, the latent class regression analysis to detect what listeners learn from exposure to a vowel distribution in chapter V, and the use of continuous distributions in chapter VI. Hence, replications are indispensable for consolidating the reliability of the results, this time taking into account a possibly confounding influence of the dispersion in the distributions (section 3.1 below), and of the differential processing of speech and non-speech (section 2.4).

These potential confounds are also a complicating factor in a meta-analysis of the results mentioned in Tables IX.2 and IX.3. At first sight, the presence of nearly significant effects and clear null results in these tables may make a meta-analysis an interesting endeavour, in particular because they may partly be due to the smallness of effects of distributional training (see the confidence intervals for the measured effects of distributional training in the chapters of this thesis, which were always close to zero). However, the meaning of such a meta-analysis is greatly reduced if it turns out that the measured effects do not reflect effects of distributional learning at all (section 2.4).

Apart from general replications and a careful examination of the role of dispersion, future research is needed to tackle an issue that could not be resolved in this thesis, namely what *precisely* is the role of distributional learning in the onset of categorical speech sound perception in the second half of the first year of life

(section 3.2). Finally, it is clear that future studies are necessary to investigate speech sound learning beyond the self-imposed boundaries in this thesis (section 3.3).

3.1. The role of dispersion in distributional learning

Chapter VII revealed an important possible confound in all distributional training studies to date, including those in the current thesis, namely the influence on speech sound learning, of different measures of dispersion in the distributions. The distributional training studies that compare the effects of distributions with wide dispersions (the "enhanced" distributions) to the effects of distributions with narrow dispersions (the "non-enhanced" distributions) do not shed light on this issue, because these studies obtained null results (Escudero et al., 2011; chapter V; see section 2.4). Even if the effects had been statistically significant, it would not have been clear whether the effect was caused by the larger range, the larger standard deviation or the wider distance between the two means in the enhanced distribution than in the non-enhanced distribution (or by a combination of two or three of these measures).

Thus, the unravelling of the precise effects of different measures of dispersion of speech sound distributions on speech sound learning is an important topic for future research. For the learning of tones, a step in this direction was already undertaken by Holt and Lotto (2006). This inspiring study shows the importance of the variance of each peak in a bimodal distribution for the learning of tones, and also reveals a complicating factor that must be taken into account when venturing upon an examination of the role of dispersion, namely the extent to which the listener relies on the manipulated cue for categorizing the tone versus his or her reliance on other cues in the tone. In other words, future research should take into account that the role of dispersion depends on the weight of the cue in perception.

3.2. The role of distributional learning in category creation

This thesis did not clarify the precise role (if any) of distributional learning in triggering the onset of categorical perception, particularly the onset of language-specific speech perception in second half of the first year of life (section 2.5). Chapter II showed that distributional learning can *contribute* to the emergence of behavioural language-specific speech perception, i.e., to categorical perception; and the literature review in chapter VIII showed that categorical firing patterns of low-level auditory neurons can be *observed* after mere exposure to sound distributions. Still, it remains unclear to what extent this behavioural and neural categorization is *caused* by distributional learning. Specifically, does the emergence of behavioural and neural categorization rest on "a sufficient amount" of bottom-up distributional learning only, or does it (also) reflect the top-down influence of newly formed higher-level cortical representations? In the former scenario, it is not clear what "a sufficient amount" is. In the latter scenario, it is not clear what triggers the creation of the higher-level representations, and to what extent distributional learning played a role in this (see also the review in chapter VIII). An additional complicating factor is the uncertainty about the precise relation between neural and behavioural categorization, a relation that represents a further imperative topic for future research (see chapter III).

3.3. Research beyond the self-imposed boundaries of this thesis

This dissertation focussed on distributional learning (among many ways of learning) of isolated synthetic vowels (among various naturally pronounced speech sounds in context). This was useful in order to be able to study the mechanism of distributional vowel learning in isolation. Nevertheless, it is clear that future research should reach beyond these self-imposed boundaries, and study the interactions between distributional learning and the many factors that have been shown or hypothesized to affect speech sound learning, such as prosody and stress (Johnson and Jusczyk, 2001; Pierrehumbert, 2003), visual cues (Yeung and Werker, 2009; Ter Schure et al., 2014), word knowledge (Bergelson and Swingley,

2012), and social interaction (Kuhl et al., 2003). Another unsettled issue that should be addressed in the future, is the relation between distributional speech sound learning and statistical learning at other levels of linguistic analysis, such as the sequential learning of patterns in speech (Saffran et al., 1996; see section 1.3 in the Introduction).

This thesis was also confined to distributional learning in infants and adults. In fact, all studies on distributional learning to date have been performed only with infants, at several ages between 2 and 12 months, and with adults. Even though the crucial sensitive period for speech sound learning probably ends around the first birthday in normal development, the auditory cortex retains a higher level of plasticity throughout childhood and adolescence than in adulthood (review in chapter VIII). It is up to future research to discover the possible changes in the capacity for distributional learning between the first and the 18th birthday, as well as the possible changes with growing age in adulthood.

4. Concluding remarks

This thesis studied distributional vowel learning in infants and in adults in a series of behavioural and neurophysiological experiments in the lab, and in a literature review exploring neurobiological evidence for distributional learning. It can be concluded that distributional learning contributes to natural speech sound acquisition and can be traced in neural firing patterns at the lowest levels of cortical auditory processing, at least in babyhood, but is difficult to tap straightforwardly in laboratory settings. Hence, interesting questions remain to be solved, of which the two most pressing ones are the influence on learning of dispersion in speech sound distributions, and the ever-persisting issue of the precise role of statistical learning in the creation of categorical representations.

Summary

Distributional learning of vowel categories in infants and adults

Introduction

In this dissertation I examine how infants learn to perceive the vowels of their mother tongues and how adults learn to perceive the vowels of a new language. I concentrate on one specific learning mechanism, namely learning from simple exposure to the environment, without receiving instruction or feedback. This mechanism is called *distributional learning*. Researchers study distributional learning by exposing participants to speech in the lab. I call this a *distributional training*. Before describing in more detail what I examined, I will explain what is special about perceiving vowels, and what is meant precisely by "distributional learning" and "distributional training".

What is special about perceiving vowels?

One may wonder what is so special about perceiving vowels or other speech sounds. This can be appreciated when considering that each pronunciation of a speech sound is *different* from each other pronunciation of that same speech sound. We can measure these differences in the recorded speech signal. For instance, if 10 native speakers of English were to repeat the word *pet* a 100 times, then any instance of the 1.000 vowel pronunciations would differ from any of the other instances. They would differ in for example their duration, in the frequency values that they are composed of, and in the pitch with which they are pronounced. Still, native speakers of English will *perceive* each of these 1.000 vowels as the *same* kind of vowel. Apparently, our brains ignore *irrelevant* differences between vowel tokens of the same category. At the same time our brains readily detect differences between vowel tokens of different categories. These differences are *relevant*: they signal a change in meaning (between for example *pet* and *pat*).

One might think that this skill of grouping speech sounds into categories is something that we are born with. This is true to some extent: there is a limit to the differences that a human ear can perceive. But the way in which adults perceive speech sounds also depends on the native language, and this is a sign that we *learn*

how to group instances of speech sounds into categories. A well-known example that shows that speech perception is *language-specific* is the trouble that native speakers of Japanese experience in hearing a difference between English "r" as in *rice* and "l" as in *lice*. Because Japanese does not contain different words with "r" and "l", the difference is not relevant in Japanese.

We acquire this language-specific speech perception already before the first birthday. Researchers have determined that at the beginning of life infants perceive speech sounds in a way that is independent of the language that they experience. At that time, Japanese infants hear a difference between "r" and "l", just as English infants do. In the second half of the first year of life, however, English infants become better at hearing the difference, while Japanese infants become worse. In other words, infants' speech perception turns language-specific.

Note that a declining ability to perceive irrelevant differences represents an important *improvement* and not a deterioration. Imagine what would happen if you would continuously hear irrelevant differences between speech sounds. You would constantly experience different words, even though the speaker does not intend different words. Indeed, infants who do not acquire language-specific speech perception in time, have a higher chance of being delayed in further language development.

What is distributional learning?

How do infants actually learn to perceive speech sounds in a language-specific way? After all, they do not get explicit instructions about this from their caregivers. Researchers think that infants acquire language-specific perception by simply being exposed to ambient speech. This way of learning is called *distributional learning*, the topic of this dissertation.

One may wonder why distributional learning is called "distributional". This has to do with the fact that speech sounds occur in *distributions*. What a distribution is can be explained best with an example. Speech sounds have different acoustic properties. An important acoustic property of vowels is the so-called "first

formant" (F1). Now suppose that we measured the F1 value of multiple pronunciations of the Dutch vowel "ɛ"[1] (as in the *Dutch* word *pet*, meaning "cap"), and that we marked each value as a vertical line on an x-axis. This could yield the distribution illustrated in the top picture of Figure A. You can see that most values (most vertical lines) cluster around a mean value "M" and that the fewest values occur far from the mean. (The grey curve also illustrates this: it has a peak where most values occur and it drops where the values are less frequent). Hence, the pronunciations of "ɛ" do not differ from one another in a random way: they constitute a systematic distribution.

That vowel distributions are *language-specific* becomes clear when we compare this distribution of F1 values for the Dutch vowel category "ɛ" to that for the English vowel contrast between "ɛ" as in the *English* word *pet* and "æ" as in *pat*. This distribution is illustrated in the bottom picture of Figure A. Along the same continuum of F1 values, Dutch has one vowel category "ɛ", and English the two categories "ɛ" and "æ".

It will come as no surprise that there is a relation between the way in which speech sounds are *pronounced* and the way in which they are *perceived*. Native speakers of Dutch do not only pronounce Dutch "ɛ" with F1 values that lie around the mean F1 value in Figure A (top picture), they also perceive vowel instances with these F1 values as "ɛ". Native speakers of English perceive these same vowels as either "ɛ" (for the values left in the continuum of Figure A)

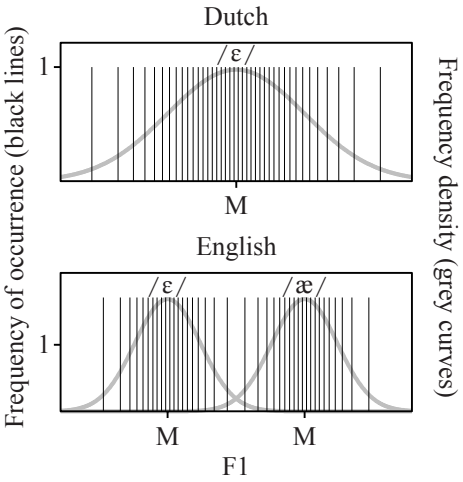


Figure A. Language-specific vowel distributions

[1] For the vowels in this summary I use the symbols of the International Phonetic Alphabet. These symbols give linguists information about the pronunciation.

or “æ” (for the values on the right). The idea of distributional learning is that we acquire such *language-specific speech perception* through exposure to *language-specific distributions*.

What is distributional training?

Researchers study distributional learning by means of a *distributional training*. For this, they create artificial distributions of speech sounds that approximate natural distributions. Table A shows examples of training distributions. You can see that they look like the natural distributions in Figure A.

During the training, participants listen to the speech sounds of a distribution for a few minutes. One group of participants is usually exposed to a *unimodal* distribution (with a single peak, as in het top picture of Figure A), and another group to a *bimodal* distribution (with two peaks, as in the bottom picture). A unimodal distribution normally reflects a native speech sound category, and the bimodal distribution a contrast that has to be learned. For instance, the Dutch adults in chapter III listened to either a unimodal distribution representative of the Dutch vowel “ɛ” or to a bimodal distribution representative of the English vowel contrast between “ɛ” and “æ”.

After the training, researchers assess whether the exposure to the distributions has affected the participants’ perception. They do this by measuring whether the bimodal group of participants has become better in perceiving the bimodal contrast than the unimodal group. If that is the case, they will have found a *distributional training effect*.

What did I examine and what are the results?

In this dissertation, I studied distributional learning on the basis of five main questions:

1. Can we really demonstrate distributional learning in a distributional training experiment?
2. Is the role of distributional learning different in infants (who learn their mother tongues) than in adults (who learn a new language)?
3. Do adults differ in what they learn from a distributional training?
4. Can manipulations of the distributions influence the effectiveness of a distributional training?
5. Does the neuroscientific literature contain evidence for distributional learning?

I examined these questions in a set of experiments (chapters II through VII), and with a literature review of neurobiological processes that possibly reflect distributional learning (chapter VIII). Table A at the end of the summary lists and explains the experiments. The remainder of this summary presents the conclusions for each of the five questions. At the end of the summary, the reader will know more about the role of distributional learning in infants' and adults' vowel acquisition, and about the role of distributional *training* in demonstrating distributional learning.

1. Can we really demonstrate distributional learning in a distributional training experiment?

At the beginning of the project, other researchers had already performed experiments with a distributional training. However, not all of the experiments yielded an effect of distributional training. For infants, for instance, two studies reported an effect, while two other studies did not find an effect. Therefore, it was important to examine whether we could replicate an effect of distributional training. Moreover, the bimodal distributions in the previous studies were confined

to certain contrasts only. Therefore, we also wanted to examine whether an effect was replicable with new contrasts.

The experiment with 2-to-3-month old infants (chapter II) confirms that a distributional training can affect infants' perception. I will explain this study in more detail when addressing the next question in this dissertation.

Among the five experiments with adults, two studies obtained a distributional training effect (chapters V and VI). In both studies, we trained native speakers of Spanish on the contrast between the Dutch vowels "ɑ" and "a" (as in the Dutch words *man*, which means "man" and *maan*, which means "moon", respectively). This contrast is difficult for Spanish listeners, because they perceive instances of both vowels as the *Spanish* vowel category "a". Two other experiments with adults did not yield a clear training effect (Chapters III and IV). In these experiments, we exposed Dutch adults to the contrast between the English vowels "ɛ" as in the English word *pet* and "æ" as in the English word *pat*. This contrast is difficult for Dutch listeners, because they perceive instances of both vowels as the *Dutch* vowel category "ɛ".

The pattern that the training yielded an effect in Spanish listeners two times and not a clear effect in Dutch listeners two times suggests that the number of vowels in the native language may influence the effectiveness of the training. This is because Dutch has many more vowels (15) than Spanish (5), and it might be more difficult to change the refined perceptual abilities that are needed to distinguish many vowel categories than to change a coarser perception. However, I have not studied this speculation further.

What was more important than this speculation, was the result obtained in the fifth and last experiment with adults (chapter VII). This result shows that participants in a distributional training experiment may not learn from the number of peaks in the training distribution at all, and hence not from distributional learning in the way that I described above (see "What is distributional learning?"). I discuss this important result in more detail when addressing question 4.

2. Is the role of distributional learning different in infants (who learn their mother tongues) than in adults (who learn a new language)?

At the beginning of the project, the role of distributional learning in the acquisition of speech sounds was still rather unclear. All infants in the available distributional training studies were 6 months of age or older, and thus had an age at which their speech perception is already turning language-specific (as mentioned above, this happens between 6 and 12 months of life). As a consequence, it was not clear whether distributional learning could also contribute to the *emergence* of this language-specific perception. In order to show this, it was necessary to demonstrate distributional learning at an age *before* the emergence of language-specific speech perception. Therefore, the infants in our experiment were only 2 to 3 months of age (chapter II).

These infants were raised in Dutch homes. We exposed them to either a unimodal distribution of the Dutch vowel "ɛ" or a bimodal distribution of the English vowel contrast between "ɛ" and "æ" (see Table A). After this exposure, the bimodally trained infants were better at discriminating a representative "ɛ" from a representative "æ" than the unimodally trained infants. We concluded that the mechanism of distributional learning is indeed available before infants' perception turns language-specific, and hence that it can contribute to the development of this language-specific perception.

Would distributional learning play an equally large role in adults' acquisition of the speech sounds of a new language? At the beginning of the project, linguists already expected that distributional learning would be more difficult for adults than for infants, because adults' perception is already language-specific. It was also conspicuous that adults can *continue* to experience problems with the discrimination of certain speech sounds of a new language, even when they have been exposed to this language for years (this can happen when they live in a country where the language is spoken). In order to compare the role of distributional learning in adults to that in infants, we repeated the distributional training that we had done with the infants in almost the exact same way, with the adults (chapter III). We obtained a smaller distributional training effect than that

found in the infants. We concluded from this that the capacity to learn vowels through simple exposure to vowel distributions is smaller in adults than in infants. This conclusion is in line with the just mentioned expectations and also with observations reported in the neuroscientific literature (which I discuss when addressing question 5).

3. Do adults differ in what they learn from a distributional training?

At the beginning of the project, it was not clear what participants learn precisely from a distributional training, and whether they can differ in what they learn from this training. We examined this in chapter V.

First, we tested how well native speakers of Spanish can identify examples of the Dutch vowels "ɑ" and "a". Subsequently, we inferred from these test scores what "cues" they had used for this. Cues are acoustic properties of speech sounds that people use unconsciously to perceive these speech sounds. Examples of vowel cues are the duration (D) of the vowel, and frequency components such as the first formant (F1), the second formant (F2) and the third formant (F3). To infer which cues the Spanish listeners had used, we did a so-called "latent class analysis". This is a statistically reliable technique to identify groups ("classes") of participants that use the same cues. These groups are not plainly visible in the data (their presence is non-overt or "latent"). We labelled participants who used the same cues "people with the same listening strategy". Indeed, not all participants had the same strategy: some groups used less cues than other groups.

Subsequently, we examined how these listening strategies that participants had before the training, had changed in a test after the training. We observed that when participants learn from a bimodal training, they add cues to their listening strategies. Notably, we saw only certain combinations of cues in the listening strategies (namely D, D–F1, D–F1–F2 or D–F1–F2–F3). By comparing these new listening strategies with the former strategies, we could infer that listeners learn the new cues *in a specific order* (namely in the order: D, F1, F2 en F3). This order

probably reflects a declining prominence of the cues in the speech signal. The outcomes suggest that people learn vowel categories *in developmental steps*.

In sum, adults can differ in what they learn from a distributional training. They learn to use additional cues that are slightly more subtle than the cues they were already using.

4. Can manipulations of the distributions influence the effectiveness of a distributional training?

At the beginning of the project researchers had not yet studied whether certain manipulations of the training distributions can make participants learn *more* from the training. We investigated this in chapters V and VI. The idea was based on studies showing that mothers change the distributions in their speech when addressing their infants. Two of the unconscious adjustments that they make are as follows.

First, they make the differences between speech sounds larger and hence clearer for the infant. We imitated this enlargement by widening the bimodal training distribution. This is illustrated in the pictures for chapter V in Table A: the upper picture shows a normal bimodal distribution and the middle picture an enlarged ("clearer") distribution.

The second adjustment that mothers make is that they produce more versions of each speech sound than in speech addressed to adults. In this way, they provide the infant with more examples of that speech sound. We imitated this larger variation by including a larger number of different stimuli in the training than the eight stimuli that were common in earlier distributional training studies. This is visible in the pictures for chapter VI in Table A: the upper picture shows a distribution with only eight different stimuli (eight vertical lines), and the middle picture a distribution with many more stimuli (many vertical lines; for the sake of clarity the number of lines is smaller than the real number, which was 128).

The outcomes of these two experiments are not straightforward. In chapter V, the Spanish participants that had listened to an enlarged distribution improved

their perception of the Dutch vowels "ɑ" en "a", but it remained undecided whether an enlarged distribution is more effective for learning vowel categories than a non-enlarged distribution. The result in chapter VI was similar: the Spanish participants that had listened to a distribution with many different stimuli improved their perception of the Dutch vowels, but it remained unresolved whether a distribution with much variation is more effective for learning vowel categories than a distribution with limited variation. Hence it is possible that the improvement that the Spanish listeners showed in their perception of the Dutch vowels is not based on the enlargement or the large degree of variation in the training distributions.

The final experiment in this dissertation (chapter VII) features yet another manipulation of the distributions than the manipulations that I just described. The aim of the experiment was to determine whether observed distributional training effects were truly based on the "number of peaks" in the distributions. In all previous studies that report an effect of distributional training (among which the studies in this dissertation), the listeners to a bimodal distribution (with two peaks) had improved more in their perception of the speech sounds in the trained contrast than listeners to a unimodal distribution (with one peak) or than the participants in another control group (such as the listeners to music, for whom the number of peaks is not defined). Researchers had always assumed that the bimodal group had performed better because they had been exposed to a distribution *with two peaks*. However, we detected another possible explanation, namely a *wide dispersion*. The concept of dispersion is more complex than that of "the number of peaks". One of the reasons for this is probably that dispersion can be defined in various ways. (Chapter VII explains three ways: the range, the standard deviation and the edge strength. Interested readers can go to this chapter for more details). Unfortunately, in none of the distributional training studies to date researchers have excluded the possibility that their results may be due to a wide dispersion in the bimodal distribution. Therefore, we tested whether we would also obtain a distributional training effect if we exposed the participants to either a bimodal or a unimodal distribution that had *an equal dispersion* (Table A shows the distributions). Now the distributions *only* differed in the number of peaks. Hence, if we found a

distributional training effect, this effect would truly be based on the number of peaks in the distributions. However, the results prove that in this case a distributional training effect is absent. The implication is that observed distributional training effects are not based on the number of peaks in the training distributions. Future research should now establish whether these effects have to do with the dispersion in those distributions.

5. Does the neuroscientific literature contain evidence for distributional learning?

At the beginning of the project, the linguistic literature about distributional learning had paid hardly any attention to neurobiological processes that could underlie distributional learning. Linguists did have the opinion that distributional learning must be a low-level process in the brain, because it is learning by means of simple exposure to speech and not learning for which people must use certain knowledge or skills (for those we probably need higher-level areas in the brain). In an attempt to narrow the gap between linguistic and neuroscientific knowledge, I examined whether the neuroscientific literature contains evidence for distributional learning at a low level of processing in the brain. This led to the literature review in chapter VIII. I confined the exploration to the literature about the primary auditory cortex (A1). This is the lowest level of auditory processing in the brain.

Studies with adult animals show that each brain cell in the adult A1 is specialized and produces certain characteristic firing patterns in response to sound. Together, the cells constitute a kind of map, on which their specializations are arranged systematically.

Studies with baby animals demonstrate that such maps do not exist in babyhood yet. What is important in the context of the current dissertation is that the brain cells develop their characteristic firing patterns on the basis of sounds in the environment, and thus through *distributional learning*. After a certain sensitive period, however, the influence of ambient sounds on the properties of brain cells in A1 becomes limited. This suggests that the influence of distributional learning has

declined. It should be noted that the way in which an animal perceives sounds is related to the firing properties of the cells in A1. Hence, ambient sounds in the sensitive period determine the animal's *perception* for the rest of its life. Researchers report these observations for different kinds of animals, such as rats, cats and monkeys. Therefore, it is likely that they also apply to human babies. A smaller influence of distributional learning after a sensitive period is in line with the conclusion of the experiment in chapter III that adults are less capable of learning vowels through distributional learning than infants (I discussed this result under question 2).

It is not the case that the properties of brain cells in A1 cannot change at all anymore after the sensitive period. Research with animals also shows that they can change provided that the animals perform a task that attracts their attention to the sounds in the training. Certain nuclei in the brain then send out substances that make the cells in A1 plastic again, so that ambient sounds can affect the firing properties of these brain cells again. Considering this, "attention" might also be important for distributional learning in human adults. More research is needed to confirm this. Also, we do not know yet what "attention" is precisely. Nonetheless, it is conspicuous that when doing a distributional training experiment with adults, researchers commonly ask the participants to pay close attention to the presented speech sounds in the training.

In sum, according to the neuroscientific literature, distributional learning in *infants* may reflect changes in the firing properties of brain cells at a low level of processing in the brain, purely under the influence of ambient speech sounds. Distributional learning in *adults* may require "attention" to these speech sounds in addition.

Conclusion

This dissertation gives us a number of insights about learning vowel categories through simple exposure to ambient speech (hence about *distributional learning*): (1) 2-to-3-month old infants already learn from a distributional training. This means that distributional learning can contribute to the development of language-specific speech perception, which infants start showing from 6 months of life onwards; (2) Adults can also learn from a distributional training. They probably learn to use increasingly subtle cues to identify the vowels of the new language. However, the capacity for distributional learning is smaller than in infants; (3) The observed distributional training effects (both those in this dissertation and those obtained in other studies) may not be due to the number of peaks in the training distributions. Future research should determine whether a prominent other candidate factor, namely the dispersion of the distributions, is responsible for the effects instead.

Table A. Overview of the experiments

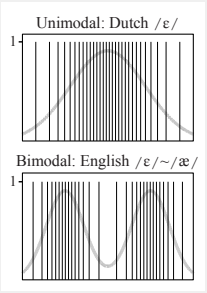
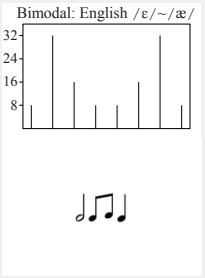
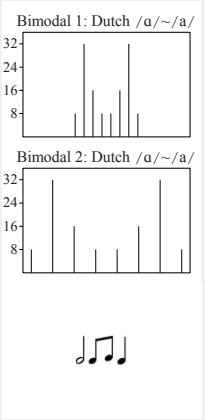
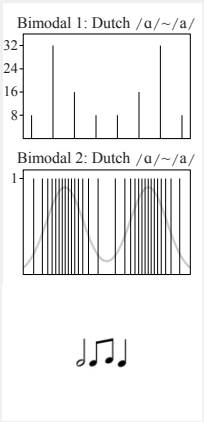
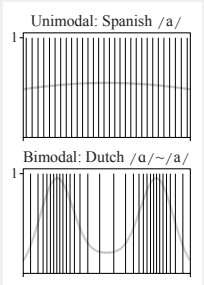
Chapter	Groups	Design	Test method	Training conditions*
Infants				
II	Dutch (2-to-3-mnd)	1. Training 2. Post-test	MMR	
Adults				
III	Dutch	1. Pre-test 2. Training 3. Post-test	MMR	
IV	Dutch	1. Pre-test 2. Training 3. Post-test	Behaviour	
V	Spanish	1. Pre-test 2. Training 3. Post-test	Behaviour	

Table A (continued). Overview of the experiments

Chapter	Groups	Design	Test method	Training conditions*
Adults				
VI	Spanish	1. Pre-test 2. Training 3. Post-test	Behaviour	
VII	Spanish	1. Pre-test 2. Training 3. Post-test	Behaviour	

***: Training conditions**. x-axis: F1 value (see under "what is distributional learning?")
y-axis: how often did the participants hear the stimulus?

Explanation of Table A: How were the experiments performed?

Design

Table A shows that all experiments with adults in this dissertation consisted of a test (the pre-test), a training phase and another test (the post-test). In this way we could examine whether the bimodal group *improved* more than the control group (improvement score = post-test score – pre-test score). The experiment with the infants only had a test after the training. One of the reasons was that otherwise the experiment would become too long for the infants.

Test method

We used different methods to assess whether there was an effect of distributional training. Table A shows that the method was a ***behavioural task*** in chapters V through VII. In the pre- and post-tests, participants were asked to label speech sound tokens: for each presented vowel, they had to indicate on a computer screen whether they perceived an example of the one or the other vowel in the bimodal contrast. Subsequently, we calculated the percentage of correct answers. This was the participant's test score.

Obviously, this task is not suitable for infants. Our infants were also too young for one of the behavioural tasks that researchers had used in other distributional training studies with infants. The chosen alternative was the measurement of brain signals on the basis of which we could calculate the ***mismatch response (MMR)***. The idea behind this method is that if someone perceives a difference between two speech sounds A and B, his or her brain signal in response to A will differ from that to B. This difference is the MMR. The method is suitable for adults *and* infants, because it does not require certain behaviour: the response occurs automatically even when participants do not pay attention to the sounds. A larger MMR indicates a better ability to discriminate the test sounds. We calculated the MMR for each participant. This MMR was the partipant's test score.

Explanation of Table A (continued): How were the experiments performed?

Training conditions

All experiments used at least one ***bimodal*** group (with two peaks). However, not all experiments contained a ***unimodal*** group (with one peak). Table A shows that this was the case in for example chapters V and VI. The participants in these chapters were Spanish speaking learners of Dutch, and we wanted to avoid the risk that a unimodal training would obstruct their efforts to learn Dutch. After all, if the participants learn from the distributions, then we do not only expect that their perception of the test vowels will improve in a bimodal training, but also that their perception may become worse in a unimodal training. Therefore we chose another control group to which we could compare the bimodal group, namely a group of participants who were exposed to classical ***music*** during the training phase.

Also, we created two types of training distributions that served to imitate natural speech sound distributions: discontinuous distributions and continuous distributions. The pictures in Table A show the difference between the two. The ***discontinuous*** distributions consist of eight vertical lines of different lengths (for example in chapters IV and V). This means that the participants were presented with eight different vowel sounds, each of which was repeated a certain number of times during the training. The length of each vertical line demonstrates how many times the sound was repeated. The ***continuous*** distributions consist of many more vertical lines and all these lines are equally long (for example in chapters II and III). This means that there were many acoustically different stimuli, each of which was presented the same number of times during the training, namely once. Continuous distributions look more like natural distributions (as illustrated in Figure A).

Samenvatting

Distributioneel leren van klinkercategorieën bij baby's en volwassenen

Introductie

In dit proefschrift bestudeer ik hoe baby's de klinkers van hun moedertaal leren waarnemen en hoe volwassenen de klinkers van een nieuwe taal leren waarnemen. Ik heb daarbij naar één bepaald leermechanisme gekeken, namelijk leren door simpelweg blootgesteld te zijn aan de omgeving, dus leren zonder dat iemand je instructies of feedback geeft. Dit leermechanisme heet *distributioneel leren*. Onderzoekers bestuderen dit mechanisme door mensen aan spraak bloot te stellen in het lab. Ik noem dit *distributioneel trainen*. Voordat ik vertel wat ik precies heb onderzocht, zal ik uitleggen wat er nu zo bijzonder is aan het waarnemen van klinkers, en wat "distributioneel leren" en "distributioneel trainen" nu precies inhouden.

Wat is er bijzonder aan het waarnemen van klinkers?

Je vraagt je misschien af wat er speciaal is aan het waarnemen van klinkers of andere spraakklanken. Dit is beter te begrijpen als je je realiseert dat elke uitspraak van een bepaalde spraakklank *verschilt* van elke andere uitspraak van diezelfde spraakklank. Dat kunnen we meten in het spraaksignaal. Bijvoorbeeld, als 10 moedertaalsprekers van het Nederlands het woord *maan* 100 keer herhalen, dan is in elk van deze 1.000 woorden de klinker anders. Ze verschillen bijvoorbeeld in hun duur, in de frequenties waaruit ze zijn opgebouwd, en in de toonhoogte waarop ze worden uitgesproken. Desondanks zullen Nederlandse luisteraars deze 1.000 klinkers telkens als *dezelfde* soort klinker waarnemen. Kennelijk negeren onze hersenen *irrelevante* verschillen tussen klinkers van dezelfde categorie. Aan de andere kant signaleren onze hersenen wel verschillen tussen klinkers van verschillende categorieën. Deze verschillen zijn *relevant*, want ze veroorzaken een verandering van betekenis (van bijvoorbeeld *maan* naar *man*).

Nu zou je kunnen denken dat de vaardigheid om spraakklanken te groeperen in categorieën aangeboren is. Tot op zekere hoogte is dit zo: er is een grens aan de verschillen die menselijke oren kunnen horen. Maar de manier waarop volwassenen spraakklanken waarnemen hangt ook af van de moedertaal, en dat is

een aanwijzing dat we *leren* hoe we klinkers moeten groeperen in categorieën. Een bekend voorbeeld dat laat zien dat spraakperceptie *taalspecifiek* is, is de moeite die moedertaalsprekers van het Japans hebben om het verschil te horen tussen de Engelse "r" uit *rice* en de "l" uit *lice*. Omdat het Japans geen verschillende woorden heeft met "r" en "l", is het verschil in het Japans niet relevant.

Deze taalspecifieke spraakperceptie verwerven we al in het eerste levensjaar. Onderzoekers hebben vastgesteld dat baby's aan het begin van hun leven spraakklanken waarnemen op een manier die onafhankelijk is van de taal die ze in hun omgeving horen. Japanse baby's horen dan net als Engelse baby's een verschil tussen "r" en "l". Tussen de 6 en 12 maanden gaan echter Japanse baby's het verschil tussen "r" en "l" steeds minder goed horen, terwijl Engelse baby's daar steeds beter in worden. De spraakperceptie van baby's wordt dan dus taalspecifiek.

Merk op dat een afnemende vaardigheid om irrelevante verschillen te horen een belangrijke *verbetering* is, en geen achteruitgang. Denk je maar eens in hoe het zou zijn als je niet-relevante verschillen tussen klanken zou blijven horen. Dan zou je voortdurend verschillende woorden ervaren, terwijl er geen verschillende woorden bedoeld zijn. Baby's die niet op tijd een taalspecifieke perceptie verwerven, hebben inderdaad een grotere kans om achter te gaan lopen in de verdere taalontwikkeling.

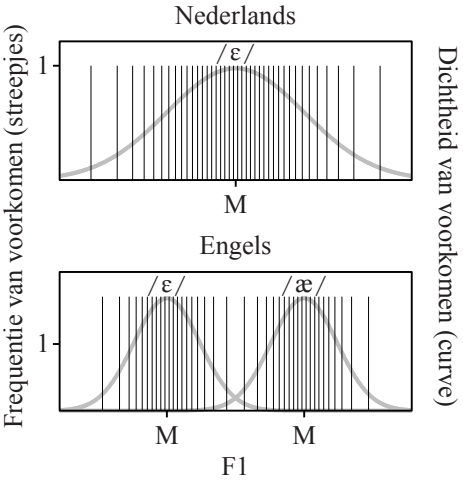
Wat is distributioneel leren?

Hoe leren baby's eigenlijk om spraakklanken op een taalspecifieke manier waar te nemen? Ze krijgen hierover immers geen expliciete uitleg van hun ouders. Onderzoekers denken dat baby's taalspecifieke perceptie leren doordat ze simpelweg blootstaan aan spraak in hun omgeving. Deze manier van leren heet *distributioneel leren*, en daarover gaat dit proefschrift.

Nu vraag je je misschien af waarom distributioneel leren "distributioneel" wordt genoemd. Dit heeft te maken met het feit dat spraakklanken voorkomen als *distributies*. Wat een distributie is, kan ik het beste uitleggen met een voorbeeld. Spraakklanken hebben verschillende akoestische eigenschappen. Een belangrijke

akoestische eigenschap van klinkers is de zogenaamde "eerste formant" (F1). Stel nu dat we de F1-waarde opmeten van een groot aantal uitspraken van de Nederlandse klinker "ɛ"[1] uit het woordje *pet* en dat we die waardes aanstrepen op een x-as. We zouden dan de *distributie* (of "verdeling") van F1-waardes kunnen krijgen die afgebeeld is in het bovenste plaatje van Figuur A. Je ziet dat de meeste waardes (de meeste streepjes) voorkomen rondom de gemiddelde waarde "G" en de minste waardes ver van dit gemiddelde af. (Dit is nog een keer weergegeven door middel van de grijze curve. De curve heeft een piek waar de waardes het meest voorkomen en gaat naar beneden waar de waardes minder voorkomen). De uitspraken van "ɛ" verschillen dus niet willekeurig: ze vormen een systematische distributie.

Dat klinkerdistributies *taalspecifiek* zijn, wordt duidelijk als we deze distributie van F1-waardes voor de Nederlandse klinkercategorie "ɛ" vergelijken met de distributie voor het Engelse klinkercontrast tussen "ɛ" (zoals in het *Engelse* woord *pet*, dat "huisdier" betekent) en "æ" (zoals in het Engelse woord *pat*, "tikken"). Deze distributie staat in het onderste plaatje van Figuur A. Je kunt zien dat langs hetzelfde continuüm van F1-waardes het Nederlands één categorie "ɛ" heeft, en het Engels de twee categorieën "ɛ" en "æ".



Figuur A. Taalspecifieke klinkerdistributies

Het wekt waarschijnlijk geen verbazing dat er een verband is tussen hoe spraakklanken in een taal worden *uitgesproken* en hoe ze in die taal worden *waargenomen*. Nederlanders spreken niet alleen

[1] Om klinkers aan te geven gebruik ik de symbolen uit het Internationale Fonetische Alfabet. Deze geven taalkundigen informatie over de precieze uitspraak.

de “ɛ” uit met F1-waardes die rond de gemiddelde waarde in Figuur A (bovenste plaatje) liggen, ze nemen uitspraken met deze F1 waardes ook waar als “ɛ”. Moedertaalsprekers van het Engels nemen de klanken in dit gebied waar als ofwel “ɛ” (voor F1 waardes links in het plaatje van Figuur A) ofwel “æ” (voor waardes rechts). Het idee van distributioneel leren is dat mensen een *taalspecifieke perceptie* leren door blootstelling aan *taalspecifieke distributies*.

Wat is distributioneel trainen?

Onderzoekers bestuderen distributioneel leren door middel van een *distributionele training*. Ze maken daarvoor kunstmatige distributies van spraakklanken die natuurlijke distributies nabootsen. Voorbeelden van trainingsdistributies staan in Tabel A. Je ziet dat ze lijken op de natuurlijke distributies in Figuur A.

Tijdens de training luisteren deelnemers aan het experiment een paar minuten naar de klanken van een distributie. Eén groep deelnemers krijgt meestal een *unimodale* distributie te horen (met één piek, zoals in het bovenste plaatje van Figuur A), en een andere groep een *bimodale* distributie (met twee pieken, zoals in het onderste plaatje). Een unimodale distributie komt meestal overeen met een klank uit de moedertaal en de bimodale distributie met een klankcontrast dat geleerd moet worden. In het experiment in hoofdstuk III luisterden Nederlandse volwassenen bijvoorbeeld naar ofwel een unimodale distributie representatief voor de Nederlandse klinker “ɛ” ofwel een bimodale distributie representatief voor het Engelse klinkercontrast tussen “ɛ” en “æ”.

Na de training meten onderzoekers of de blootstelling aan de distributies de perceptie van de deelnemers heeft veranderd. Ze doen dat door te kijken of de bimodaal getrainde groep deelnemers beter is geworden in het onderscheiden van het bimodale contrast dan de unimodaal getrainde groep. Als dat zo is, dan hebben ze een *effect van distributioneel trainen* gevonden.

Wat heb ik onderzocht en wat zijn de resultaten?

Ik heb in dit proefschrift distributioneel leren bestudeerd vanuit vijf hoofdvragen:

1. Kan een distributionele training echt distributioneel leren aantonen?
2. Is de rol van distributioneel leren anders in baby's (die hun moedertaal leren) dan in volwassenen (die een nieuwe taal leren)?
3. Verschillen volwassenen in wat ze leren van een distributionele training?
4. Kunnen manipulaties van de distributies de effectiviteit van een distributionele training beïnvloeden?
5. Bevat de neurowetenschappelijke literatuur aanwijzingen voor distributioneel leren?

Ik heb deze vragen onderzocht in een serie experimenten (hoofdstuk II tot en met VII) en met een literatuuroverzicht van neurobiologische processen die mogelijk overeen komen met distributioneel leren (hoofdstuk VIII). Tabel A aan het einde van de samenvatting geeft een overzicht van de experimenten, met daarbij een uitleg over hoe ze zijn uitgevoerd. In de rest van deze samenvatting presenteer ik voor ieder van de net vermelde vragen de conclusies die deze dissertatie heeft opgeleverd. Aan het einde daarvan weet de lezer meer over de betekenis van distributioneel leren voor de klinkerverwerving bij baby's en bij volwassenen, en over de betekenis van een distributionele *training* voor het aantonen van distributioneel leren.

Kan een distributionele training echt distributioneel leren aantonen?

Aan het begin van het project hadden andere onderzoekers al enkele experimenten met een distributionele training gedaan. Deze experimenten lieten echter niet allemaal een effect van distributioneel trainen zien. Voor de baby's waren er bijvoorbeeld twee onderzoeken die zo'n effect rapporteerden, en twee studies die géén effect vonden. Het was daarom belangrijk om te kijken of we een effect van distributioneel trainen konden repliceren. Bovendien waren de bimodale distributies in de eerdere studies beperkt tot bepaalde klankcontrasten. Daarom

wilden we ook kijken of een effect van distributioneel trainen repliceerbaar was met nieuwe contrasten.

Het experiment met 2 tot 3 maanden oude baby's (hoofdstuk II) laat zien dat een distributionele training de perceptie van baby's inderdaad kan beïnvloeden. Ik zal deze baby-studie beter uitleggen bij de behandeling van de volgende hoofdvraag.

Van de vijf experimenten met volwassenen leverden twee studies een effect van distributioneel trainen op (hoofdstuk V en VI). In beide trainden we moedertaalsprekers van het Spaans op het contrast tussen de Nederlandse klinkers "ɑ" zoals in het woord *man* en "a" zoals in het woord *maan*. Dit contrast is moeilijk voor deze luisteraars, omdat ze uitspraken van beide klinkers waarnemen als de *Spaanse* klinkercategorie "a". Twee andere experimenten met volwassenen lieten geen duidelijk effect van de training zien (hoofdstuk III en IV). In deze experimenten lieten we Nederlandse volwassenen luisteren naar het contrast tussen de Engelse klinkers "ɛ" zoals in het Engelse woord *pet* en "æ" zoals in het Engelse woord *pat*. Dit contrast is moeilijk voor Nederlanders, omdat ze uitspraken van beide klinkers waarnemen als de *Nederlandse* klinkercategorie "ɛ".

Het patroon dat de training twee keer een effect opleverde met Spaanse volwassenen en twee keer geen duidelijk effect met Nederlandse volwassenen doet vermoeden dat het aantal klinkers in de moedertaal misschien invloed heeft op de effectiviteit van de training. Het Nederlands heeft namelijk veel meer klinkers (15) dan het Spaans (5) en het zou kunnen dat de verfijnde perceptie die nodig is om veel klinkercategorieën van elkaar te onderscheiden lastiger te veranderen is dan een grovere perceptie. Deze speculatie heb ik echter niet verder onderzocht.

Belangrijker dan de speculatie is het resultaat van het vijfde en laatste experiment met volwassenen (hoofdstuk VII). Dit resultaat laat zien dat deelnemers aan een distributionele training mogelijk helemaal niet leren van het aantal pieken in de trainingsdistributie, en dus niet van distributioneel leren zoals dat hierboven is uitgelegd (onder "Wat is distributioneel leren?"). Ik bespreek dit belangrijke resultaat preciezer bij de behandeling van hoofdvraag 4.

Is de rol van distributioneel leren anders in baby's (die hun moedertaal leren) dan in volwassenen (die een nieuwe taal leren)?

Aan het begin van het project was de rol van distributioneel leren in de spraakklankverwerving nog niet goed onderzocht. De baby's in de bestaande onderzoeken waren allemaal 6 maanden of ouder, en hadden dus een leeftijd waarop ze spraakklanken al op een taalspecifieke manier gaan waarnemen (zoals hierboven vermeld, gebeurt dit tussen 6 en 12 maanden). Het was daardoor onduidelijk of distributioneel leren ook bij kan dragen aan het *laten ontstaan* van die taalspecifieke perceptie. Om dat aan te tonen was het nodig om te laten zien dat baby's distributioneel leren *voordat* hun perceptie taalspecifiek wordt. De baby's in ons experiment waren daarom nog maar 2 tot 3 maanden oud (hoofdstuk II).

Deze baby's groeiden op met Nederlands sprekende ouders. We lieten ze luisteren naar ofwel een unimodale distributie van de Nederlandse klinker "ɛ" ofwel een bimodale distributie van het Engelse klinkercontrast tussen "ɛ" en "æ" (zie Tabel A). Na deze training konden de bimodaal getrainde baby's een representatieve "ɛ" beter van een representatieve "æ" onderscheiden dan de unimodaal getrainde baby's. We concludeerden dat het mechanisme van distributioneel leren inderdaad beschikbaar is voordat de perceptie van baby's taalspecifiek wordt, en dat het dus bij kan dragen aan de totstandkoming van die taalspecifieke perceptie.

Zou distributioneel leren een even grote rol spelen bij het verwerven van klinkers uit een nieuwe taal door volwassenen? Aan het begin van het project hadden taalkundigen al de verwachting dat distributioneel leren moeilijker zou zijn voor volwassenen dan voor baby's, omdat volwassenen al met taalspecifieke oren luisteren. Het was ook opvallend dat volwassenen kunnen problemen *blijven* houden met het onderscheiden van bepaalde spraakklanken uit een nieuwe taal, ook al wonen ze al jaren in het land waar de taal gesproken wordt, en zijn ze dus al jaren blootgesteld aan die taal. Om de rol van distributioneel leren bij volwassenen te kunnen vergelijken met die van baby's, hebben we de distributionele training met de Nederlandse baby's op bijna exact dezelfde manier herhaald met Nederlandse volwassenen (hoofdstuk III). Het trainingseffect was bij de

volwassenen kleiner dan bij de baby's. Daaruit concludeerden we dat het vermogen om klinkers te leren door simpele blootstelling aan klinkerdistributies kleiner is bij volwassenen dan bij baby's. Deze conclusie komt overeen met de hierboven genoemde verwachtingen en ook met observaties uit de neurowetenschappelijke literatuur (waarop ik bij de behandeling van hoofdvraag 5 terugkom).

Verschillen volwassenen in wat ze leren van een distributionele training?

Aan het begin van het project was niet duidelijk wat mensen nu precies leren van een distributionele training, en of mensen kunnen verschillen in wat ze van de training leren. Dit hebben we onderzocht in hoofdstuk V.

Eerst testten we *voor* de training hoe goed moedertaalsprekers van het Spaans voorbeelden van de Nederlandse klinkers "ɑ" en "a" kunnen benoemen. Uit die testscores leidden we af welke "cues" ze daarbij hadden gebruikt. Cues zijn akoestische eigenschappen van spraakklanken, die mensen onbewust gebruiken om die spraakklanken waar te nemen. Voorbeelden van cues voor klinkers zijn de duur (D) van de klinker, en frequentiecomponenten zoals de eerste formant (F1), tweede formant (F2) en derde formant (F3). Om af te leiden welke cues de Spaanse luisteraars hadden gebruikt, deden we een zogenaamde "latente klasse analyse". Met deze analyse kun je op een statistisch betrouwbare manier groepen ("klassen") van mensen die dezelfde cues gebruiken identificeren. Die groepen zijn op het oog niet duidelijk in de data te zien (ze zijn daar verborgen ofwel "latent" in aanwezig). De mensen die dezelfde cues gebruikten noemden we "mensen met eenzelfde luisterstrategie". Inderdaad bleken niet alle deelnemers dezelfde strategie te hebben: sommige groepen gebruikten minder cues dan andere groepen.

Vervolgens hebben we gekeken hoe deze luisterstrategieën van voor de training veranderd waren in de test na de training. We zagen dat als mensen leren van de bimodale training, ze cues toevoegen aan hun eerdere luisterstrategie. Opvallend was dat alleen bepaalde combinaties van cues in de luisterstrategieën voorkwamen, (namelijk D, D–F1, D–F1–F2 of D–F1–F2–F3). Door deze nieuwe luisterstrategieën te vergelijken met de eerdere luisterstrategieën, konden we

afleiden dat luisteraars nieuwe cues er *in een bepaalde volgorde* bijleren (namelijk in de volgorde: D, F1, F2 en F3). Deze volgorde weerspiegelt waarschijnlijk een afnemende prominentie van de cues in het spraaksignaal. De resultaten zijn een aanwijzing dat mensen klinkercategorieën *stapsgewijs* leren.

Mensen kunnen dus inderdaad verschillen in wat ze van een distributionele training leren. Ze leren er cues bij die net iets verfijnder zijn dan de cues die ze al eerder gebruikten.

Kunnen manipulaties van de distributies de effectiviteit van een distributionele training beïnvloeden?

Aan het begin van het project was nog niet onderzocht of we mensen misschien *beter* kunnen laten leren van een distributionele training als we de distributie vervormen. Dit hebben we onderzocht in hoofdstuk V en VI. Het idee was gebaseerd op onderzoeken die laten zien dat moeders de distributies in hun spraak veranderen als ze tegen hun baby's praten. Ze doen dan onbewust onder meer twee dingen.

Ten eerste maken ze de verschillen tussen spraakklanken groter en dus duidelijker voor de baby. We hebben dit nagebootst door de bimodale trainingsdistributie uit elkaar te trekken. Dit is te zien in Tabel A in de plaatjes bij hoofdstuk V: het bovenste plaatje toont een gewone bimodale distributie en het middelste een bimodale distributie die uitgerekt ("verduidelijkt") is.

Ten tweede maken moeders meer verschillende versies van iedere spraakklank dan in hun spraak tegen volwassenen, en geven de baby dus zo meer verschillende voorbeelden van die spraakklank. We hebben dit nagebootst door meer verschillende stimuli in de training op te nemen dan de acht stimuli die in bestaande studies met distributionele trainingen waren gebruikt. Dit is te zien in Tabel A in de plaatjes bij hoofdstuk VI: het bovenste plaatje toont een distributie met maar acht verschillende stimuli (acht streepjes), en het middelste een distributie met heel veel verschillende stimuli (heel veel streepjes). Overigens is

voor de duidelijkheid het aantal streepjes in het plaatje minder dan het echte aantal (dit was 128).

De resultaten van deze twee experimenten zijn niet eenduidig. In hoofdstuk V verbeterden de Spaanstalige deelnemers die naar een uitgerekte distributie hadden geluisterd weliswaar hun perceptie van de Nederlandse klinkers "ɑ" en "a", maar het bleef onduidelijk of zo'n uitgerekte distributie nu effectiever is om klinkers te leren dan een niet-uitgerekte distributie. Hoofdstuk VI had een vergelijkbaar resultaat: de Spaanstalige deelnemers die naar een distributie met veel verschillende stimuli hadden geluisterd verbeterden weliswaar hun perceptie van de Nederlandse klinkers, maar het bleef onduidelijk of zo'n gevarieerde distributie nu effectiever is om klinkers te leren dan een distributie met weinig variatie. Het is dus mogelijk dat de verbeterde perceptie van de Spaanse luisteraars in deze twee experimenten niet gebaseerd is op de uitgerektheid of de gevarieerdheid van de distributies.

Het allerlaatste experiment in dit proefschrift (hoofdstuk VII) bevat nog een andere manipulatie van de distributies dan de manipulaties die ik net besproken heb. Het doel van het experiment was om vast te stellen of de gevonden effecten van een distributionele training wel echt gebaseerd zijn op "het aantal pieken" in de distributies. In alle eerder bestaande studies die een effect van distributioneel trainen rapporteren (waaronder de studies in dit proefschrift) waren steeds de luisteraars naar een bimodale distributie (met twee pieken) meer verbeterd in hun perceptie van de spraakklanken uit het getrainde contrast dan de luisteraars naar een unimodale distributie (met één piek) of dan de deelnemers uit een andere controlegroep (zoals de luisteraars naar muziek voor wie het aantal pieken niet gedefinieerd is). Steeds was aangenomen dat de bimodale groep het beter had gedaan omdat ze waren blootgesteld aan een distributie *met twee pieken*. Wij ontdekten echter nog een andere mogelijke verklaring, namelijk een *wijde spreiding*. Het concept van spreiding is ingewikkelder dan "het aantal pieken". Dit komt onder meer doordat spreiding op verschillende manieren uitgedrukt kan worden. (Hoofdstuk VII legt drie manieren uit: het bereik, de standaarddeviatie en de sterkte van de randen. De lezer kan naar dit hoofdstuk gaan om de details te

lezen). Ongelukkigerwijs hebben onderzoekers in geen van de tot nu toe gedane studies met een distributionele training uitgesloten dat hun resultaten te wijten kunnen zijn aan een wijde spreiding in de bimodale distributie. In hoofdstuk VII testten we daarom of we nog steeds een effect van distributioneel trainen zouden krijgen als we deelnemers zouden blootstellen aan ofwel een bimodale ofwel een unimodale distributie *met een gelijke spreiding* (Tabel A laat deze distributies zien). Nu verschilden de distributies *alleen* nog in het aantal pieken. Als we dus een effect van distributioneel trainen zouden vinden, dan zou dit effect echt gebaseerd zijn op het aantal pieken in de distributies. De resultaten bewijzen echter dat er in dit geval *geen* effect van distributioneel trainen is. De implicatie is dat de eerdere effecten van distributioneel trainen niet gebaseerd zijn op het aantal pieken in de distributies. Toekomstig onderzoek moet nu uitwijzen of die effecten dan misschien te maken hebben met de spreiding van die distributies.

Bevat de neurowetenschappelijke literatuur aanwijzingen voor distributioneel leren?

Aan het begin van het project bevatte de taalwetenschappelijke literatuur over distributioneel leren nauwelijks aandacht voor neurobiologische processen die aan dit leermechanisme ten grondslag kunnen liggen. Taalwetenschappers waren al wel van mening dat distributioneel leren een proces moest zijn op een laag niveau in de hersenen, omdat het leren is door simpele blootstelling aan spraakgeluiden en niet leren waarvoor je bepaalde kennis of vaardigheden nodig hebt (daarvoor zijn waarschijnlijk hogere niveaus in de hersenen nodig). Om mogelijk een brug te slaan tussen taalwetenschappelijke en neurowetenschappelijke kennis, heb ik onderzocht of de neurowetenschappelijke literatuur aanwijzingen bevat voor dit idee van distributioneel leren op een laag niveau in de hersenen. Dit leidde tot het literatuuroverzicht in hoofdstuk VIII. Ik heb me beperkt tot de literatuur over de primaire auditieve cortex (A1). Dit is het laagste niveau van auditieve verwerking in de hersenen.

Onderzoek met volwassen dieren laat zien dat bij deze dieren iedere cel in A1 gespecialiseerd is en met bepaalde karakteristieke vuurpatronen op geluid reageert. Samen vormen de cellen een soort kaart, waarin hun specialisaties gerangschikt zijn.

Onderzoek met baby-dieren laat zien dat in de baby-tijd zo'n kaart nog niet bestaat. Wat belangrijk is in de context van dit proefschrift, is dat hersencellen hun karakteristieke vuurpatronen ontwikkelen op basis van de geluiden waaraan het dier is blootgesteld, dus door middel van *distributioneel leren*. Na een bepaalde sensitieve periode heeft omgevingsgeluid echter nauwelijks meer invloed op de eigenschappen van hersencellen in A1. Dit suggereert dat de invloed van distributioneel leren dan dus is afgenomen. Belangrijk hierbij is dat de manier waarop een dier geluiden waarneemt gerelateerd is aan de vuurpatronen die hersencellen in A1 hebben ontwikkeld. Omgevingsgeluiden in de sensitieve periode bepalen dus hoe een dier geluiden *hoort* voor de rest van zijn leven. Onderzoekers hebben deze dingen vastgesteld voor verschillende soorten dieren, waaronder ratten, katten en apen, en het is dus te verwachten dat ze ook gelden voor mensenbaby's. Een kleinere invloed van distributioneel leren na een sensitieve periode is in overeenstemming met de conclusie van het experiment in hoofdstuk III dat volwassenen minder goed klinkers kunnen leren door middel van distributioneel leren dan baby's (dit resultaat heb ik besproken bij hoofdvraag 2).

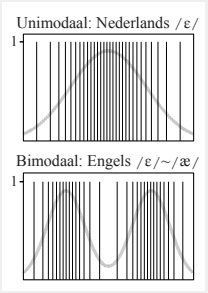
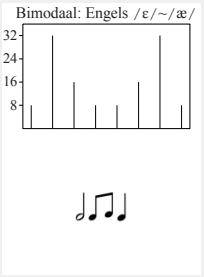
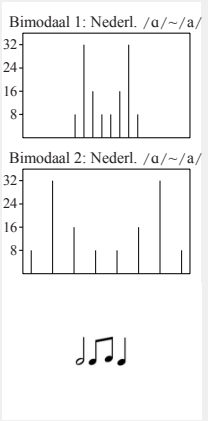
Nu is het niet zo dat de eigenschappen van hersencellen in A1 helemaal niet meer kunnen veranderen na de sensitieve periode. Dieronderzoek laat ook zien dat dit best kan mits de dieren een taak moeten uitvoeren die hun aandacht op de gepresenteerde geluiden vestigt. Bepaalde kernen in de hersenen sturen dan stoffen naar A1 die de cellen weer plastisch maken, en die er dus voor zorgen dat omgevingsgeluiden weer de eigenschappen van de hersencellen kunnen beïnvloeden. "Aandacht" zou dus ook belangrijk kunnen zijn om volwassen mensen distributioneel te laten leren. Hiernaar is meer onderzoek nodig, ook omdat we nog niet precies weten wat "aandacht" nu eigenlijk is. Het is wel opvallend dat in distributionele trainingen met volwassenen onderzoekers doorgaans aan de deelnemers vragen goed op de spraakgeluiden in de training te letten.

Samengevat zou volgens de neurowetenschappelijke literatuur distributioneel leren bij *baby's* dus overeen kunnen komen met veranderingen in de eigenschappen van hersencellen op een laag niveau in de hersenen, puur onder invloed van spraakgeluiden die de baby om zich heen hoort. Bij *volwassenen* zou daarnaast "aandacht" voor dat omgevingsgeluid nodig zijn om van dat omgevingsgeluid te leren.

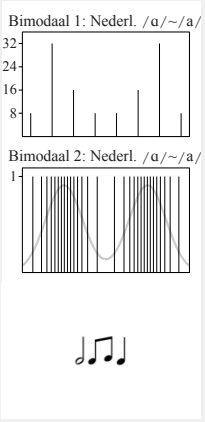
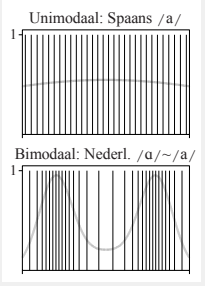
Conclusie

Dit proefschrift vertelt ons een aantal dingen over het leren van klinkercategorieën door simpele blootstelling aan de klinkers die mensen in de spraak om zich heen horen (dus over *distributioneel leren*): (1) Baby's van 2 tot 3 maanden oud leren al van een distributionele training. Dit betekent dat distributioneel leren kan bijdragen aan de ontwikkeling van de taalspecifieke spraakperceptie, die baby's vanaf 6 maanden beginnen te krijgen. (2) Volwassenen kunnen ook van een distributionele training leren. Ze leren dan waarschijnlijk om steeds subtielere cues te gaan gebruiken om de klinkers van de nieuwe taal te identificeren. Het vermogen om distributioneel te leren is bij volwassenen wel kleiner is dan bij baby's. (3) De gevonden leereffecten van een distributionele training (in en buiten dit proefschrift) zijn mogelijk niet gebaseerd op het aantal pieken in de trainingsdistributies, zoals altijd is aangenomen. Toekomstig onderzoek moet vaststellen of een belangrijke tegenkandidaat, namelijk de spreiding in de distributies, dan voor deze effecten verantwoordelijk is.

Tabel A. Overzicht van de experimenten

Hoofdstuk	Groepen	Opzet	Testmethode	Trainingscondities*
Baby's				
II	Nederlands (2 tot 3 mnd)	1. Training 2. Post-test	MMR	
Volwassenen				
III	Nederlands	1. Pre-test 2. Training 3. Post-test	MMR	
IV	Nederlands	1. Pre-test 2. Training 3. Post-test	Gedrag	
V	Spaans	1. Pre-test 2. Training 3. Post-test	Gedrag	

Tabel A (vervolg). Overzicht van de experimenten

Hoofdstuk	Groepen	Opzet	Testmethode	Trainingscondities*
Volwassenen				
VI	Spaans	1. Pre-test 2. Training 3. Post-test	Gedrag	
VII	Spaans	1. Pre-test 2. Training 3. Post-test	Gedrag	

***: Trainingscondities**. x-as: F1-waarde (zie uitleg in "wat is distributioneel leren")
y-as: hoe vaak hoorden de deelnemers de stimulus?

Toelichting bij Tabel A: Hoe zijn de experimenten uitgevoerd?

Opzet

Tabel A laat zien dat in dit proefschrift alle experimenten met volwassenen bestonden uit een test (de pre-test), een training en nog een test (de post-test). We konden met deze opzet bekijken of de bimodale groep meer *verbetert* (post-test score min de pre-test score) dan de controle-groep. Het baby-experiment had alleen een test na de training, onder meer omdat het experiment anders te lang zou worden voor de baby's.

Testmethode

We hebben verschillende methodes gebruikt om te meten of er een effect van distributioneel trainen was. Tabel A geeft aan dat de methode in hoofdstuk IV tot en met VII een ***gedragstaak*** was. De deelnemers moesten in de pre- en post-test spraakklanken benoemen: bij elke gepresenteerde klinker moesten ze op een computerscherm aanklikken of ze een voorbeeld van de ene dan wel de andere klinker uit het bimodale contrast hoorden. Wij berekenden vervolgens hoeveel procent van de antwoorden goed was. Dit was de testscore van de deelnemer.

Zo'n taak is natuurlijk ongeschikt voor baby's. Ook voor de gedragstaken die onderzoekers in eerdere distributionele trainingen met baby's hadden gebruikt, waren onze baby's te jong. Het gekozen alternatief was het meten van hersensignalen op basis waarvan we de ***mismatch response (MMR)*** konden berekenen. Het idee achter deze methode is dat als iemand een verschil waarneemt tussen twee klanken, zijn of haar hersensignaal in reactie op de ene klank zal verschillen van dat op de andere klank. Dit verschil is de MMR. De methode is geschikt voor volwassenen én voor baby's, omdat geen bepaald gedrag nodig is: de response treedt automatisch op ook als mensen helemaal niet op de geluiden letten. Een grotere MMR wijst op een beter vermogen om de geteste klanken uit elkaar te houden. Wij berekenden de MMR voor iedere deelnemer. Dit was zijn of haar de test score.

Toelichting bij Tabel A (vervolg): Hoe zijn de experimenten uitgevoerd?

Trainingscondities

Alle experimenten hadden ten minste één ***bimodale*** groep (met twee pieken). Ze hadden echter niet allemaal een ***unimodale*** groep (met één piek). Tabel A laat zien dat zo'n groep ontbreekt in bijvoorbeeld hoofdstuk V en VI. De deelnemers in deze hoofdstukken waren Spaanstalige leerders van het Nederlands, en we wilden het risico vermijden dat een unimodale training hun leerinspanningen zou dwarsbomen. Immers, als de deelnemers distributioneel leren, dan verwachten we niet alleen dat ze de klanken beter leren onderscheiden tijdens een bimodale training, maar ook dat ze de klanken misschien slechter gaan onderscheiden tijdens een unimodale training. Daarom kozen we voor een andere controlegroep om de bimodale groep mee te vergelijken, namelijk een groep deelnemers die tijdens de trainingsfase werd blootgesteld aan klassieke ***muziek***.

Verder hebben we twee soorten trainingsdistributies gemaakt om natuurlijke spraakklankdistributies na te bootsen: discontinue distributies en continue distributies. De plaatjes in Tabel A laten het verschil zien. De ***discontinue*** distributies bestaan uit acht streepjes van verschillende lengtes (zoals in hoofdstuk IV en V). Dit betekent dat de deelnemers acht verschillende klinkerstimuli hoorden, die ieder een aantal keer herhaald werden tijdens de training. De lengte van ieder streepje geeft aan hoe vaak de stimulus herhaald werd. De ***continue*** distributies bestaan uit veel meer streepjes en deze streepjes hebben allemaal dezelfde lengte (zoals in hoofdstuk II en III). Dit betekent dat er een heleboel verschillende klinkerstimuli waren, die elk even vaak te horen waren tijdens de training, namelijk maar één keer. Continue distributies lijken meer op de echte distributies (zoals afgebeeld in Figuur A).

References

Aaltonen, O., Eerola, O., Hellström, Å., Uusipaikka, E., & Lang, A.H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. The Journal of the Acoustical Society of America, 101 (2): 1090-1105. doi: 10.1121/1.418031.

Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, 24: 197-207. doi: 10.1016/0301-0511(87)90002-0.

Adank, P., Van Hout, R., Smits, R. (2004). An acoustic description of the vowels of Northern and Southern standard Dutch. *The Journal of the Acoustical Society of America*, 116: 1729-1738. doi: 10.1121/1.1779271.

Alderson, J.C., & Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing*, 22, 301–320. doi: 10.1191/0265532205lt310oa.

Allan, L.G., & Gibbon, J. (1991). Human bisection at the geometric mean. *Learning and Motivation*, 22, 39–58. doi: 0.1016/0023-9690(91)90016-2.

Aramakis, V.B., Hsieh, C.Y., Leslie, F.M., & Metherate, R. (2000). A critical period for nicotine-induced disruption of synaptic development in rat auditory cortex. *The Journal of Neuroscience*, 20(16), 6106-6116. PMID: 10934260.

Arnold, H.M., Burk, J.A., Hodgson, E.M., Sarter, M., & Bruno, J.P. (2002). Differential cortical acetylcholine release in rats performing a sustained attention task versus behavioral control tasks that do not explicitly tax attention. *Neuroscience*, 114(2), 451-460. PMID: 12204214.

Aslin, R.N., & Pisoni, D.B. (1980). Some developmental processes in speech perception. In G.H. Yeni-Komshian, J.F. Kavanagh, & C.A. Ferguson (Eds), *Child Phonology, Vol. 2, Perception* (pp 67–96). New York/London: Academic Press.

Bakin, J.S., & Weinberger, N.M. (1996). Induction of physiological memory in the cerebral cortex by stimulation of the nucleus basalis. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 93, 11219-11224. PMID: 8855336.

Bao, S., Chang, E.F., Davis, J.D., Gobeske, K.T., & Merzenich, M.M. (2003). Progressive degradation and subsequent refinement of acoustic representations in the adult auditory cortex. *The Journal of Neuroscience*, 23(34), 10765-10775. PMID: 14645468.

Bao, S., Chan, V.T., & Merzenich, M.M. (2001). Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature*, 412, 79-83. PMID: 11452310.

Bao, S., Chang, E.F., Teng, C-L., Heiser, M.A., & Merzenich, M.M. (2013). Emergent categorical representation of natural complex sounds resulting from the early post-natal sound environment. *Neuroscience*, 248, 30-42. doi: 10.1016/j.neuroscience.2013.05.056.

Bao, S., Chang, E.F., Woods, J., & Merzenich, M.M. (2004). Temporal plasticity in the primary auditory cortex induced by operant perceptual learning. *Nature Neuroscience*, 7(9), 974-981. doi: 10.1038/nn1293.

Barone, P., Dehay, C., Berland, M., & Kennedy, H. (1996). Role of directed growth and target selection in the formation of cortical pathways: prenatal development of the projection of area V2 to area V4 in the monkey. *Journal of Comparative Neurology*, 374 (1), 1-20.

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., & Friston, K.J. (2012). Canonical microcircuits for predictive coding. *Neuron*, 76: 695-711. doi: 10.1016/j.neuron.2012.10.038.

Batardière, A., Barone, P., Knoblauch, K., Giroud, P., Berland, M., Dumas, A-M., & Kennedy, H. (2002). Early specification of the hierarchical organization of visual cortical areas in the macaque monkey. *Cerebral Cortex*, 12 (5), 453-465. doi: 10.1093/cercor/12.5.453.

Benders, T. (2013). *Nature's distributional learning experiment*. Doctoral dissertation. University of Amsterdam.

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the Unites States of America (PNAS)*, 109: 3253-3258. doi: 10.1073/pnas.1113380109.

Best, C.T. (1994). The emergence of native-language phonological influences in infants: a perceptual assimilation model. In J.Goodman, & H. Nusbaum (Eds), *The Development of Speech Perception: the Transition from Speech Sounds to Spoken Words* (pp. 167–224). Cambridge, MA: MIT Press.

Best, C.T., McRoberts, G.W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior & Development*, 18, 339–350. doi: 10.1016/0163-6383(95)90022-5.

Blake, D.T., Heiser, M.A., Caywood, M., & Merzenich, M.M. (2006). Experience-dependent adult cortical plasticity requires cognitive association between sensation and reward. *Neuron*, 52, 371-381. doi: 10.1016/j.neuron.2006.08.009.

Blundon, J.A., Bayazitov, I.T., & Zakharenko, S.S. (2011). Presynaptic gating of postsynaptically expressed plasticity at mature thalamocortical synapses. *The Journal of Neuroscience*, 31(44), 16012-16025. doi: 10.1523/JNEUROSCI.3281-11.2011.

Boersma, P. (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. Doctoral dissertation. University of Amsterdam: LOT Dissertation Series.

Boersma, P. (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. In A. Benz, & J. Mattausch (Eds), *Bidirectional Optimality Theory* (pp. 33-72). Amsterdam: John Benjamins.

Boersma, P. (2012). Modelling phonological category learning. In A.C. Cohn, C. Fougeron, & M.K. Huffman (Eds), *Handbook of Laboratory Phonology* (pp. 207-218). Oxford: Oxford University Press.

Boersma, P., Escudero, P., & Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: an integrated model for the acquisition of language-specific sound categories. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1013–1016. (=*Rutgers Optimality Archive* 585).

Boersma, P., & Weenink, D. (2009–2014). Praat: Doing Phonetics by Computer. Available at: http://www.praat.org (accessed 2009–2014).

Bohn, O.S. (1995). Cross-language speech perception in adults. First language transfer doesn't tell it all. In W. Strange (Ed), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research* (pp. 275-300). Timonium, MD: York Press.

Bohn, O.S., & Flege, J.E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics,* 11, 303–328. doi: 10.1017/S0142716400008912.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, 46, 217–243. doi: 10.1177/00238309030460020801.

Bouwmeester, S., Sijtsma, K., & Vermunt, J.K. (2004). Latent Class Regression Analysis to describe cognitive developmental phenomena: an application to transitive reasoning. *European Journal of Developmental Psychology,* 1, 67-86. doi: 10.1080/17405620344000031.

Bradlow, A.R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America*, 97(3), 1916-1924. doi: 10.1121/1.412064.

Bradlow, A.R., Pisoni, D.P., 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.

Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *The Journal of the Acoustical Society of America*, 117, 3890–3901. doi: 10.1121/1.1906060.

Brown, M., Irvine, D.R.F., & Park, V.N. (2004). Perceptual learning on an auditory frequency discrimination task by cats: association with changes in primary auditory cortex. *Cerebral Cortex*, 14, 952-965. doi: 10.1093/cercor/bhh056.

Buonomano, D.V., & Merzenich, M.M. (1998). Cortical plasticity: from synapses to maps. *Annual Review of Neuroscience,* 21(1), 149-186.

doi: 10.1146/annurev.neuro.21.1.149.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science,* 296(5572), 1435-1435. doi: 10.1126/science.1069587.

Capel, D.J H., De Bree, E.H., De Klerk, M.A., Kerkhoff, A.O., & Wijnen, F.N.K. (2011). Distributional cues affect phonetic discrimination in Dutch infants. In W. Zonneveld, H. Quené, & W. Heeren (Eds), *Sound and Sounds. Studies Presented to M.E.H. (Bert) Schouten on the Occasion of his 65th Birthday* (pp. 33–43). Utrecht: UiL-OTS.

Carral, V., Huotilainen, M., Ruusuvirta, T., Fellman, V., Näätänen, R., & Escera, C. (2005). A kind of auditory "primitive intelligence" already present at birth. *The European Journal of Neuroscience*, 21, 3201–3204. doi: 10.1111/j.1460-9568.2005.04144.x.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics,* 34, 372-387. doi: 10.1016/j.wocn.2005.08.003.

Cetas, J.S., De Venecia, R.K., & McMullen, N.T. (1999). Thalamocortical afferents of Lorente de Nó: medial geniculate axons that project to primary auditory cortex have collateral branches to layer I. *Brain Research*, 830, 203-208. doi: 10.1016/S0006-8993(99)01355-4.

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. doi: 10.1121/1.3445785.

Chang, E.F., & Merzenich, M.M. (2003). Environmental noise retards auditory cortical development. *Science,* 300(5618), 498-502. doi: 10.1126/science.1082163.

Cheng, Y., Wu, H., Tzeng, Y., Yang, M., Zhao, L., & Lee, C. (2013). The development of mismatch responses to Mandarin lexical tones in early infancy. *Developmental Neuropsychology*, 38, 281–300. doi: 10.1080/87565641.2013.799672.

Cheour, M., Alho, K., Čeponiené, R., Reinikainen, K., Sainio, K., Pohjavuori, M., Aaltonen, O., & Näätänen, R. (1998a). Maturation of mismatch negativity in infants. *International Journal of Psychophysiology*, 29, 217–226. doi: 10.1016/S0167-8760(98)00017-8.

Cheour, M., Alho, K., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., Näätänen, R. (1997). The mismatch negativity to changes in speech sounds at the age of three months. *Developmental Neuropsychology*, 13: 167-174. doi: 10.1080/87565649709540676.

Cheour, M., Čeponiené, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998b). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, 1, 351–353. doi: 10.1038/1561.

Cheour, M., Čeponiené, R., Leppänen, P., Alho, K., Kujala, T., Renlund, M., Fellman, V., & Näätänen, R. (2002a). The auditory sensory memory trace decays rapidly in newborns. *Scandinavian Journal of Psychology*, 43, 33–39. doi: 10.1111/1467-9450.00266.

Cheour, M., Martynova, O., Näätänen, R., Erkkola, R., Sillanpää, M., Kero, P., Raz, A., Kaipio, M.-L., Hiltunen, J., Aaltonen, O., Savela, J., & Hämäläinen, H. (2002b). Speech sounds learned by sleeping newborns. *Nature*, 415, 599–600. doi: 10.1038/415599b.

Cheour, M., Leppänen, P., & Kraus, N. (2000). Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. *Clinical Neurophysiology,* 111, 4–16. doi: 10.1016/S1388-2457(99)00191-1.

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., & Näätänen, R. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, 82, 53–58. doi: 10.1016/0378-5955(94)00164-L.

Chomsky, N. & Halle, M. (1968). Sound Pattern of English. Cambridge, MA: MIT Press.

Chun, S., Bayazitov, I.T., Blundon, J.A., & Zakharenko, S.S. (2013). Thalamocortical long-term potentiation becomes gated after the early critical period in the auditory cortex. *The Journal of Neuroscience*, 33(17), 7345-7357. doi: 10.1523/JNEUROSCI.4500-12.2013.

Clapp, W.C., Kirk, I.J., Hamm J.P., Shepherd, D., & Teyler, T.J. (2005). Induction of LTP in the human auditory cortex by sensory stimulation. *European Journal of Neuroscience*, 22, 1135-1140. doi:10.1111/j.1460-9568.2005.04293.x.

Conel, J.L. (1939 – 1967). *The post-natal development of human cerebral cortex. Vol. I – VIII*. Cambridge MA: Harvard University Press.

Consonni, S., Leone, S., Becchetti, A., & Amadeo, A. (2009). Developmental and neurochemical features of cholinergic neurons in the murine cerebral cortex. *BioMed Central Neuroscience*, 10(18), 1-9. doi:10.1186/1471-2202-10-18.

Crick, F., & Koch, C. (1998). Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. *Nature*, 391, 245-250. doi: 10.1038/34584.

Cristià, A., McGuire, G.L., Seidl, A., & Francis, A.L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics*, 39, 388–402. doi: 10.1016/j.wocn.2011.02.004.

Cristia, A., & Seidl, A. (2013). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, 1-22. (Published in print in 2014: *Journal of Child Language*, 41, 913-934). doi:10.1017/S0305000912000669.

Crowell, D.H., Kapuniai, L.E., Boychuk, R.B., Light, M.J., & Hodgman, J.E. (1982). Daytime sleep stage organization in three-month-old infants. *Electroencephalography and Clinical Neurophysiology,* 53, 36–47. doi: 10.1016/0013-4694(82)90104-3.

De Boer, B., & Kuhl, P.K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online*, 4(4), 129-134. doi: 10.1121/1.1613311.

Dehaene-Lambertz, G. (2000). Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience,* 12, 449–460. doi: 10.1162/089892900562264.

Dehaene-Lambertz, G., & Baillet, S. (1998). A phonological representation in the infant brain. *Neuroreport*, 9, 1885–1888. doi: 10.1097/00001756-199806010-00040.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage,* 24, 21-33. doi: 10.1016/j.neuroimage.2004.09.039.

De Villers-Sidani, E., Chang, E.F., Bao, S., & Merzenich, M.M. (2007). Critical period window for spectral tuning defined in the primary auditory cortex (A1) in the rat. *The Journal of Neuroscience*, 27(1), 180-189. doi: 10.1523/JNEUROSCI.3227-06.2007.

De Vos, J. (2012). *Does enhanced bimodal distributional training improve perception of English /ɛ/ and /æ/ for adult native speakers of Dutch?* Thesis Linguistics, University of Amsterdam. http://www.fon.hum.uva.nl/theses/JohannaDeVosBA2012.pdf.

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: from acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, 22, 680-689. doi: 10.1016/j.lindif.2012.05.005.

Diesch, E., & Luce, T. (1997). Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation. *Psychophysiology,* 34(5), 501-510. doi: 10.1111/j.1469-8986.1997.tb01736.x.

Diesch, E., & Luce, T. (2000). Topographic and temporal indices of vowel spectral envelope extraction in the human auditory cortex. *Journal of Cognitive Neuroscience*, 12(5), 878-893. doi:10.1162/089892900562480.

Dorrn, A.L., Yuan, K., Barker, A.J., Schreiner, C.E., & Froemke, R.C. (2010). Developmental sensory experience balances cortical excitation and inhibition. *Nature*, 465, 932-937, doi:10.1038/nature09119.

Edeline, J-M., Pham, P., & Weinberger, N.M. (1993). Rapid development of learning-induced receptive field plasticity in the auditory cortex. *Behavioral Neuroscience*, 107(4), 539-551. doi: 10.1037/0735-7044.107.4.539.

Eggermont, J.J. (1996). Differential maturation rates for response parameters in cat primary auditory cortex. *Auditory Neuroscience*, 2, 309-327.

Eggermont, J.J. (2008). The role of sound in adult and developmental auditory cortical plasticity. *Ear & Hearing*, 29(6), 819-829. doi: 10.1097/AUD.0b013e3181853030.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, New Series, 171(3968), 303-306. doi: 10.1126/science.171.3968.303.

Ellingson, R.J., & Peters, J.F. (1980). Development of EEG and daytime sleep patterns in normal full-term infants during the first 3 months of life: longitudinal observations. *Electroencephalography and Clinical Neurophysiology,* 49, 112–124. doi: 10.1016/0013-4694(80)90357-0.

Englund, K.T. (2005). Voice onset time in infant directed speech over the first six months. *First Language*, 25(2), 219-234. doi: 10.1177/0142723705050286.

Escudero, P. (2000). *Developmental patterns in the adult L2 acquisition of new contrasts: the acoustic cue weighting in the perception of Scottish tense/lax vowels by Spanish speakers*. Unpublished master's thesis. Scotland, UK: University of Edinburgh.

Escudero, P. (2005). *Linguistic perception and second-language acquisition: explaining the attainment of optimal phonological categorization*. Doctoral dissertation. Utrecht University: LOT Dissertation Series 113.

Escudero, P., Benders, T., & Lipski, S. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: the case of Dutch, German and Spanish listeners. *Journal of Phonetics,* 37, 452-465. doi: 10.1016/j.wocn.2009.07.006.

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America*, 130, EL206–EL212. doi: 10.1121/1.3629144.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition,* 26, 551-585. doi: 10.1017/S0272263104040021.

Escudero, P., & Chladkova, K. (2010). Spanish listeners' perception of American and Southern British English vowels: different initial stages for L2 development. *The Journal of the Acoustical Society of America*, 128: EL254-EL260. doi: 10.1121/1.3488794.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, 36, 345–360. doi: 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, 280-288. doi:10.1016/j.wocn.2011.11.004.

Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech*, 53(3): 343-365. doi: 10.1177/0023830910371447.

Escudero, P. & Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition,* 133, 408-413. doi: 10.1016/j.cognition.2014.07.002.

Felleman, D.J., & Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1, 1-47. doi: 10.1093/cercor/1.1.1.

Fikkert, P. (2010). Developing representations and the emergence of phonology: evidence from perception and production. In C. Fougeron, B. Kühnert, M. d'Imperio, & N. Vallée (Eds), *Laboratory Phonology 10: Variation, Phonetic Detail and Phonological Representation (Phonology & Phonetics 4-4)* (pp. 227–258).

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. 233-277). Timonium, MD: York Press.

Flege, J.E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds), *An Integrated View of*

Language Development: Papers in Honor of Henning Wode (pp. 217-243). Trier, Germany: WVT.

Flege, J.E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer, & N. Schiller (Eds), *Phonetics and Phonology in Language Comprehension and Production (*pp. 319-355). Berlin: Mouton de Gruyter.

Flege, J.E., Bohn, O-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics,* 25, 437-470. doi: 10.1006/jpho.1997.0052.

Flege, J.E., MacKay, I.R.A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America,* 106(5), 2973-2987.

Flege, J.E., & MacKay, I.R.A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition,* 26, 1-34. doi: 10.1017/S0272263104026117.

Formisano, E., Kim, D-S., Di Salle, F., van de Moortele, P-F., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron,* 40(4), 859-869. doi: 10.1016/S0896-6273(03)00669-X.

Francis, A.L., & Nusbaum, H.C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance,* 28(2), 349-366. doi: 10.1037//0096-1523.28.2.349.

Friederici, A.D., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *Neuroreport*, 13, 1251–1254. doi: 10.1097/00001756-200207190-00006.

Friedrich, M., Weber, C., & Friederici, A.D. (2004). Electrophysiological evidence for delayed mismatch response in infants at-risk for specific language impairment. *Psychophysiology*, 41, 772–782. doi: 10.1111/j.1469-8986.2004.00202.x.

Fritz, J.B., Elhilali, M., & Shamma, S.A. (2005). Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hearing Research*, 206: 159-176. doi: 10.1016/j.heares.2005.01.015.

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, 6(11), 1216-1223. doi:10.1038/nn1141.

Froemke, R.C., & Jones, B.J. (2011). Development of auditory cortical synaptic receptive fields. *Neuroscience and Biobehavioral Reviews*, 35, 2105-2113. doi: 10.1016/j.neubiorev.2011.02.006.

Froemke, R.C., & Martins, A.R.O. (2011). Spectrotemporal dynamics of auditory cortical synaptic receptive field plasticity. *Hearing Research*, 279, 149-161. doi: 10.1016/j.heares.2011.03.005.

Galaburda, A.M., & Pandya, D.N. (1983). The intrinsic architectonic and connectional organization of the superior temporal region of the rhesus monkey. *The Journal of Comparative Neurology*, 221, 169-184. doi: 10.1002/cne.902210206.

Giezen, M., Escudero, P., & Baker, A.. (2010). Use of acoustic cues by children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 53(6), 1440-1457. doi: 10.1044/1092-4388(2010/09-0252).

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia,* 9, 317-323.

Graven, S.N., & Browne, J.V. (2008). Sleep and brain development. The critical role of sleep in fetal and early neonatal brain development. *Newborn and Infant Nursing Reviews,* 8(4), 173–179. doi: 10.1053/j.nainr.2008.10.008.

Grün, B., & Leisch, F. (2007). Fitting finite mixtures of generalized linear regressions in R. *Computational Statistics and Data Analysis,* 51, 5247-5252. doi: 10.1016/j.csda.2006.08.014.

Guenther, F.H., & Bohland, J.W. (2002). Learning sound categories: a neural model and supporting experiments. *Acoustical Science and Technology,* 23(4), 213-220. doi: 10.1250/ast.23.213.

Guenther, F.H., & Gjaja, M.N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America*, 100, 1111–1121. doi: 10.1121/1.416296.

Gulian, M., Escudero, P., & Boersma, P. (2007). Supervision hampers distributional learning of vowel contrasts. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp 1893–1896). Saarbrücken: University of Saarbrucken.

Hackett, T.A. (2011). Information flow in the auditory cortical network. *Hearing Research*, 27, 133-146. doi: 10.1016/j.heares.2010.01.011.

Hallé, P.A., & De Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development*, 19: 463-481. doi: 10.1016/S0163-6383(96)90007-7.

Han, Y.K., Köver, H., Insanally, M.N., Semerdjian, J.H., & Bao, S. (2007). Early experience impairs perceptual discrimination. *Nature Neuroscience*, 10(9), 1191-1197. doi:10.1038/nn1941.

Hasselmo, M.E. (2006). The role of acetylcholine in learning and memory. *Current Opinion in Neurobiology*, 16, 710-715. doi: 10.1016/j.conb.2006.09.002.

Hawkins, S., & Midgley, J. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association,* 35, 183–199. doi: 10.1017/S0025100305002124.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research,* 23, 65–94. doi: 10.1177/0267658307071601.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6-12. *Journal of Phonetics*, 28, 377-396. doi:10.1006/jpho.2000.0121.

He, C., Hotson, L., & Trainor, L.J. (2009). Development of infant mismatch responses to auditory pattern changes between 2 and 4 months old. *The European Journal of Neuroscience*, 29, 861–867. doi: 10.1111/j.1460-9568.2009.06625.x.

Hebb, D.O. (1949). *The Organization of Behavior*. New York: Wiley.

Herkenham, M. (1980). Laminar organization of thalamic projections to the rat neocortex. *Science*, 207, 532-535. doi: 10.1126/science.7352263.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience,* 8, 393-402. doi:10.1038/nrn2113.

Hill, J., Dierker, D., Neil, J., Inder, T., Knutsen, A., Harwell, J., Coalson, T., & Van Essen, D. (2010). A surface-based analysis of hemispheric asymmetries and folding of cerebral cortex in term-born human infants. *Journal of Neuroscience*, 30, 2268–2276. doi: 10.1523/JNEUROSCI.4682-09.2010.

Himmelheber, A.M., Sarter, M., & Bruno, J.P. (2000). Increases in cortical acetylcholine release during sustained attention performance in rats. *Cognitive Brain Research*, 9, 313-325. doi: 10.1016/S0926-6410(00)00012-4.

Himmelheber, A.M., Sarter, M., & Bruno, J.P. (2001). The effects of manipulations of attentional demand on cortical acetylcholine release. *Cognitive Brain Research*, 12, 353-370. doi: 10.1016/S0926-6410(01)00064-7.

Hollien, H., Dew, D., & Philips, P. (1971). Phonational frequency ranges of adults. *Journal of Speech and Hearing Research,* 14, 755-760. doi: 10.1044/jshr.1404.755.

Holt, L.L., & Lotto, A.J. (2006). Cue weighting in auditory categorization: implications for first and second language acquisition. *The Journal of the Acoustical Society of America,* 119 (5), 3059-3071. doi: 10.1121/1.2188377.

Huang, G.-H., & Bandeen-Roche, K. (2004). Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika,* 69, 5–32. doi: 10.1007/BF02295837.

Huttenlocher, P.R., & Dabholkar, A.S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology*, 387: 167-178.

Iber, C., Ancoli-Israel, S., Chesson, A.L., & Quan, S.F. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events. Rules, Terminology and Technical Specifications.* Westchester IL: American Academy of Sleep Medicine.

Insanally, M.N., Köver, H., Kim, H., & Bao, S. (2009). Feature-dependent sensitive periods in the development of complex sound representation. *The Journal of Neuroscience*, 29(17), 5456-5462. doi: 10.1523/JNEUROSCI.5311-08.2009.

Iverson, P., & Evans, B.G. (2007). Learning English vowels with different first-language vowel systems: perception of formant targets, formant movement and duration. *The Journal of the Acoustical Society of America,* 122(5), 2842-2854. doi: 10.1121/1.2783198.

Iverson, P., & Evans, 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. doi: 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: 3267-3278. doi: 10.1121/1.2062307.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition,* 87, B47-B57. doi: 10.1016/S0010-0277(02)00198-1.

Jaeger, T.F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59(4), 434-446. doi: 10.1016/j.jml.2007.11.007.

Jamieson, D.G., & Morosan, D.E. (1986). Training non-native speech contrasts in adults: acquisition of the English /ð/ - /θ/ contrast by francophones. *Perception & Psychophysics*, 40 (4): 205–215. doi: 10.3758/BF03211500.

Jamieson, D.G., & Morosan, D.E. (1989). Training new, nonnative speech contrasts: a comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology*, 43(1), 88-96. doi: 10.1037/h0084209.

Johnson, E.K., & Jusczyk, P.W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *Journal of Memory & Language*, 44, 548-567. doi:10.1006/jmla.2000.2755.

Jusczyk, P.W., & Aslin, R.N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29: 1-23. doi: 10.1006/cogp.1995.1010.

Jusczyk, P.W., & Hohne, E.A. (1997). Infants' memory for spoken words. *Science*, 277: 1984-1986. doi: 10.1126/science.277.5334.1984.

Jusczyk, P.W., Luce, P.A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33, 630-645. doi: 10.1006/jmla.1994.1030.

Kaas, J.H., & Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 97(22), 11793-11799. doi: 10.1073/pnas.97.22.11793.

Kahn, A., Dan, B., Grosswasser, J., Franco, P., & Sottiaux, M. (1996). Normal sleep architecture in infants and children. *Journal of Clinical Neurophysiology,* 13, 184–197. doi: 10.1097/00004691-199605000-00002.

Karmiloff-Smith, A. (2006). The tortuous route from genes to behavior: a neuroconstructivist approach. *Cognitive, Affective, & Behavioral Neuroscience*, 6(1), 9-17. doi: 10.3758/CABN.6.1.9.

Kass, R.E., & Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association,* 90(430), 773-795. doi: 10.1080/01621459.1995.10476572.

Kaur, S., Lazar, R., & Metherate, R. (2004). Intracortical pathways determine breadth of subthreshold frequency receptive fields in primary auditory cortex. *Journal of Neurophysiology*, 91, 2551-2567, doi:10.1152/jn.01121.2003.

Keuroghlian, A.S., & Knudsen, E.I. (2007). Adaptive auditory plasticity in developing and adult animals. *Progress in Neurobiology*, 82 (3): 109-121. doi: 10.1016/j.pneurobio.2007.03.005.

Kewley-Port, D., & Watson, C.S. (1994). Formant-frequency discrimination for isolated English vowels. *The Journal of the Acoustical Society of America,* 95(1)*,* 485-496. doi: 10.1121/1.410024.

Kilgard, M.P., & Merzenich, M.M. (1998). Cortical map reorganization enabled by nucleus basalis activity. *Science*, 279, 1714-1718. doi: 10.1126/science.279.5357.1714.

Klagstad, H.L. (1958). The phonemic system of colloquial Standard Bulgarian. *American Association of Teachers of Slavic and East European Languages*, 2(1), 42-54.

Klatt, D.H. (1980). Software for a cascade/parallel formant synthesizer. *The Journal of the Acoustical Society of America,* 67(3), 971-995. doi: 10.1121/1.383940.

Kondaurova, M., & Francis, A. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: comparison of three training methods. *Journal of Phonetics*, 38: 569–587. doi: 10.1016/j.wocn.2010.08.003.

Köver, H., Gill, K., Tseng, Y-T.L., & Bao, S. (2013). Perceptual and neuronal boundary learned from higher-order stimulus probabilities. *The Journal of Neuroscience*, 33(8), 3699-3705. doi: 10.1523/JNEUROSCI.3166-12.2013.

Kral, A. (2007). Unimodal and cross-modal plasticity in the "deaf" auditory cortex. *International Journal of Audiology*, 46, 479-493. doi:10.1080/14992020701383027.

Kral, A. (2013). Auditory critical periods: a review from system's perspective. *Neuroscience*, 247: 117-133. doi: 10.1016/j.neuroscience.2013.05.021.

Kral., A., & Eggermont, J.J. (2007). What's to lose and what's to learn: development under auditory deprivation, cochlear implants and limits of cortical plasticity. *Brain Research Reviews*, 56, 259–269. doi: 10.1016/j.brainresrev.2007.07.021.

Kral, A., Hartmann, R., Tillein, J., Heid, S., & Klinke, R. (2001). Delayed maturation and sensitive periods in the auditory cortex. *Audiology and Neurotology*, 6 (6): 346-362. doi: 10.1159/000046845.

Kral, A., Hartmann, R., Tillein, J., Heid, S., & Klinke, R. (2002). Hearing after congenital deafness: central auditory plasticity and sensory deprivation. *Cerebral Cortex*, 12, 797-807. doi:10.1093/cercor/12.8.797.

Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*. Oxford, UK: Pergamon.

Kraus, N., Burton Koch, D., McGee, T., Nicol, T.G., & Cunningham, J. (1999). Speech-sound discrimination in school-age children: psychophysical and neurophysiologic measures. *Journal of Speech, Language, and Hearing Research*, 42: 1042-1060. doi: 10.1044/jslhr.4205.1042.

Krogh, L., Vlach, H.A., & Johnson, S.P. (2013). Statistical learning across development: flexible yet constrained. *Frontiers in Psychology*, 3, article 598, 1-11. doi: 10.3389/fpsyg.2012.00598.

Kuhl, P.K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 97(22), 11850-11857. doi: 10.1073/pnas.97.22.11850.

Kuhl, P., Andruski, J., Chistovich, I., Chistovich, L., Kozhevnikova, E., Ryskina, V., Stolyarova, E., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science,* 227(5326), 684-686. doi: 10.1126/science.277.5326.684.

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*, 363, 979-1000. doi: 10.1098/rstb.2007.2154.

Kuhl, P.K., Conboy, B.T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: implications for the "critical period". *Language Learning and Development*, 1(3&4), 237-264. doi: 10.1080/15475441.2005.9671948.

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, F13–F21. doi: 10.1111/j.1467-7687.2006.00468.x.

Kuhl, P.K., Tsao, F., & Liu, H. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 100 (15), 9096-9101. doi: 10.1073/pnas.1532872100.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606–608. doi: 10.1126/science.1736364.

Kujala, A., Huotilainen, M., Hotakainen, M., Lennes, M., Parkkonen, L., Fellman, V., & Näätänen, R. (2004). Speech-sound discrimination in neonates as measured with MEG. *Neuroreport*, 15, 2089–2092. doi: 10.1097/00001756-200409150-00018.

Kushnerenko, E. (2003). *Maturation of the cortical auditory event-related brain potentials in infancy.* Doctoral dissertation. University of Helsinki.

Kushnerenko, E., Čeponiené, R., Balan, P., Fellman, V., Huotilainen, M., & Näätänen, R. (2002). Maturation of the auditory event-related potentials during the first year of life. *Neuroreport*, 13 (1): 47-51. doi: 10.1097/00001756-200201210-00014.

Lacerda, F. (1995). The perceptual-magnet effect: an emergent consequence of exemplar-based phonetic memory. In K. Elenius & P. Branderyd (Eds), *Proceedings of the XIIIth International Congress of Phonetic Sciences, Vol. 2* (pp. 140–147). Stockholm: KTH and Stockholm University.

Ladefoged, P., & Maddieson, I. (2007). *The Sounds of the World's Languages.* (10th ed.). Malden, MA/Oxford, UK/Carlton, Australia: Blackwell Publishing.

Lahiri, A., & Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *Journal of Phonetics*, 38, 44–59. doi: 10.1016/j.wocn.2010.01.002.

Lamel, L.F., Kassel, R.H., & Seneff, S. (1986). Speech Database Development: Design and Analysis of the Acoustic-Phonetic Corpus. *Proc. DARPA Speech Recognition Workshop*. Report No. SAIC-86/1546, 100–109.

Lany, J., & Saffran, J.R. (2013). Statistical learning mechanisms in infancy. In J.L.R. Rubenstein, & P. Rakic (Eds.), *Comprehensive developmental*

neuroscience: neural circuit development and function in the brain, volume 3 (pp. 231-248). Amsterdam: Elsevier.

Leisch, F. (2004). FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software*, 11 (8). URL http://www.jstatsoft.org/ v11/i08/.

Lin, T. H., & Dayton, C. M. (1997). Model selection information criteria for nonnested latent class models. *Journal of Educational and Behavioral Statistics*, 22(3), 249–264. doi: 10.3102/10769986022003249.

Lisker, L., & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word,* 20, 384 - 422.

Liu, B-h., Wu, G.K., Arbuckle, R., Tao, H.W., & Zhang, L.I. (2007). Defining cortical frequency tuning with recurrent excitatory circuitry. *Nature Neuroscience*, 10(12), 1594-1600. doi:10.1038/nn2012.

Liu, H-M, Kuhl, P. K., & Tsao, F-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, 6(3), 1-10. doi: 10.1111/1467-7687.00275.

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. *The Journal of the Acoustical Society of America*, 94, 1242–1255.

Llinás, R.R., Leznik, E., & Urbano, F.J. (2002). Temporal binding via cortical coincidence detection of specific and nonspecific thalamocortical inputs: a voltage-dependent dye-imaging study in mouse brain slices. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 99(1), 449-454. doi: 10.1073/pnas.012604899.

Loewy, D., Campbell, K., & Bastien, C. (1996). The mismatch negativity to frequency deviant stimuli during natural sleep. *Electroencephalography and Clinical Neurophysiology,* 98, 493–501. doi: 10.1016/0013-4694(96)95553-4.

Loewy, D., Campbell, K., De Lugt, D., Elton, M., & Kok, A. (2000). The mismatch negativity during natural sleep: intensity deviants. *Clinical Neurophysiology,* 111, 863–872. doi: 10.1016/S1388-2457(00)00256-X.

Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). Training Japanese listeners to identify /r/ and /l/: a first report. *The Journal of the Acoustical Society of America,* 89(2), 874-886. doi: 10.1121/1.1894649.

Lotto, A.J., Sato, M. & Diehl, R.L. (2004). Mapping the task for the second language learner: the case of Japanese acquisition of /r/ and /l/. In J. Slifka, S. Manual, & M. Matthies (Eds), *From Sound to Sense: 50+ Years of Discoveries in Speech Communication* (C181-C186).

Mäkelä, A.M., Alku, P., Mäkinen, V., Valtonen, J., May, P., & Tiitinen, H. (2002). Human cortical dynamics determined by speech fundamental frequency. *NeuroImage,* 17, 1300-1305. doi:10.1006/nimg.2002.1279.

Marcus, G.F., Vijayan, S., Bandi Rao, S., & Vishton, P.M. (1999). Rule learning by seven-month-old infants. *Science*, 283, 77-80. doi: 10.1126/science.283.5398.77.

Marin-Padilla, M., & Marin-Padilla, T.M. (1982). Origin, prenatal development and structural organization of layer I of the human cerebral (motor) cortex. *Anatomy and Embryology*, 164, 161-206. doi: 10.1007/BF00318504.

Martynova, O., Kirjavainen, J., & Cheour, M. (2003). Mismatch negativity and late discriminative negativity in sleeping human newborns. *Neuroscience Letters,* 340, 75–78. doi: 10.1016/S0304-3940(02)01401-5.

Maye, J., & Gerken, LA. (2000). Learning phonemes without minimal pairs. In S.C. Howell, S.A. Fish, & T. Keith-Lucas (Eds), *BUCLD 24 Proceedings* (pp. 522-533). Somerville, MA: Cascadilla Press.

Maye, J., & Gerken, LA. (2001). Learning phonemes: how far can the input take us? In A.H.-J. Do, L. Domínguez, & A. Johansen (Eds), *BUCLD 25 Proceedings* (pp. 480-490). Somerville, MA: Cascadilla Press.

Maye, J., Weiss, D., & Aslin, R. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science,* 11, 122–134. doi: 10.1111/j.1467-7687.2007.00653.x.

Maye, J., Werker, J.F., & Gerken, LA. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101–B111. doi: 10.1016/S0010-0277(01)00157-3.

Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition,* 13(3), 279-297. doi: 10.1017/S1366728909990022.

McCandliss, B., 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. doi: 10.3758/CABN.2.2.89.

McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.

McClelland, J.L., & Rumelhart, D.E. (1986). A distributed model of human learning and memory. In J.L. McClelland, D.E. Rumelhart, & the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models* (pp. 170–215). Cambridge, MA: MIT Press.

McCullagh, P., & J.A. Nelder. (1989). *Generalized Linear Models*. (2nd ed.). Boca Raton, Florida: Chapman & Hall/CRC.

McGee, T.J., King, C., Tremblay, K., Nicol, T.G., Cunningham, J., & Kraus, N. (2001). Long-term habituation of the speech-elicited mismatch negativity. *Psychophysiology*, 38, 653-658. doi: 10.1111/1469-8986.3840653.

McMurray, B., Kovack-Lesh, K.A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: enhancing development or an unintended consequence? *Cognition*, 129, 362-378. doi: 10.1016/j.cognition.2013.07.015.

Merzenich, M.M., & Brugge, J.F. (1973). Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Research,* 50(2), 275-296. doi: 10.1016/0006-8993(73)90731-2.

Meunier, C., Frenk-Mestre, C., Lelekov-Boissard, T., & Le Besnerais, M. (2003). Production and perception of vowels: does the density of the system play a role? In *Proceedings of International Congress of Phonetic Sciences (ICPhS)* (pp. 723-726).

Mitani, A., & Shimokouchi, M. (1985). Neural connections in the primary auditory cortex: an electrophysiological study in the cat. *Journal of Comparative Neurology*, 235(4), 417-429.
(published online in 2004, doi: 10.1002/cne.902350402).

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J., & Fujimura, O. (1975). An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics,* 18(5), 331-340. doi: 10.3758/BF03211209.

Moore, J.K. (2002). Maturation of human auditory cortex: implications for speech perception. *Annals of Otology, Rhinology and Laryngology,* 111, 7–10.

Moore, J.K., & Guan, Y.-L. (2001). Cytoarchitectural and axonal maturation in human auditory cortex. *Journal of the Association for Research in Otolaryngology (JARO)*, 2, 297–311. doi: 10.1007/s101620010052.

Moore, J.K., & Linthicum, F.H. (2007). The human auditory system: a timeline of development. *International Journal of Audiology,* 46, 460–478.
doi: 10.1080/14992020701383019.

Morr, M.L., Shafer, V.L., Kreuzer, J.A., & Kurtzberg, D. (2002). Maturation of mismatch negativity in typically developing infants and preschool children. *Ear and Hearing,* 23, 118–136. doi: 10.1097/00003446-200204000-00005.

Morris, J.S., Friston, K.J., & Dolan, R.J. (1998). Experience-dependent modulation of tonotopic neural responses in human auditory cortex. *Proceedings of the Royal Society of London. Series B*, 265, 649-657. doi: 10.1098/rspb.1998.0343.

Morrison, G.S. (2007). Logistic regression modelling for first- and second-language perception data. In M. J. Solé, P. Prieto, & J. Mascaró (Eds), *Segmental and Prosodic Issues in Romance Phonology* (pp. 219–236). Amsterdam: John Benjamins.

Morrison, G.S. (2008). L1-Spanish speakers' acquisition of the English /i/-/ɪ/ contrast: duration-based perception is not the initial developmental stage. *Language and Speech,* 51, 285-315. doi: 10.1177/0023830908099067.

Morrison, G.S. (2009). L1-Spanish speakers' acquisition of the English /i/-/ɪ/ contrast II: perception of vowel inherent spectral change. *Language and Speech,* 52, 437-462. doi: 10.1177/0023830909336583.

Moyer, A. (2009). Input as critical means to an end: quantity and quality of experience in L2 phonological attainment. In T. Piske, & M. Young-Scholten (Eds), *Input matters in SLA* (pp. 159-174). Bristol (UK) / Buffalo (USA) / Toronto (Canada): Multilingual Matters.

Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Näätänen, R., Gaillard, A.W.K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42, 313–329. doi: 10.1016/0001-6918(78)90006-9.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Livonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385: 432-434. doi: 10.1038/385432a0.

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., & Sams, M. (1989). Do event-related potentials reveal the mechanisms of auditory sensory memory in the human brain? *Neuroscience Letters,* 98, 217–221. doi: 10.1016/0304-3940(89)90513-2.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology,* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin,* 125, 826–859. doi: 10.1037/0033-2909.125.6.826.

Nakahara, H., Zhang, L.I,. & Merzenich, M.M. (2004). Specialization of primary auditory cortex processing by sound exposure in the "critical period".

Proceedings of the National Academy of Sciences of the United States of America (PNAS), 101(18), 7170-7174. doi: 10.1073/pnas.0401196101.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception & Performance*, 24(3), 756-766. doi: 10.1037/0096-1523.24.3.756.

Newman, R.S., Clause, S.A., & Burnham, J.L. (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America,* 109 (3), 1181-1196. doi: 10.1121/1.1348009.

Niedermeyer, E. (2005). Maturation of the EEG: development of waking and sleep patterns. In E. Niedermeyer & F.L. Da Silva (Eds), *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*. 5th Edition (pp. 209–233). Philadelphia: Lippincott Williams & Wilkins.

Obleser, J., Elbert, T., Lahiri, A., & Eulitz, C. (2003). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Cognitive Brain Research,* 15(3), 207-213. PMID: 12527095.

Obleser, J., Eulitz, C., Lahiri, A., & Elbert, T. (2001). Gender differences in functional hemispheric asymmetry during processing of vowels as reflected by the human brain magnetic response. *Neuroscience letters,* 314, 131-134. doi: 10.1016/S0304-3940(01)02298-4.

Oh, J.S., Jun, S-A., Knightly, L.M., & Au, T.K. (2003). Holding on to childhood language memory. *Cognition*, 86, B53-B64. doi: 10.1016/S0010-0277(02)00175-0.

Ohl, F.W., & Scheich, H. (1997). Orderly cortical representation of vowels based on formant interaction. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 94, 9440-9444.

Oxford, R., Nyikos, M., & Ehrman, M. (1988). Vive la Différence? Reflections on sex differences in use of language learning strategies. *Foreign Language Annals,* 21(4), 321-329. doi: 10.1111/j.1944-9720.1988.tb01076.x.

Pakarinen, S., Lovio, R., Huotilainen, M., Alku, P., Näätänen, R., & Kujala, T. (2009). Fast multifeature paradigm for recording several mismatch negativities

(MMNs) to phonetic and acoustic changes in speech sounds. *Biological Psychology*, 82: 219-226. doi: 10.1016/j.biopsycho.2009.07.008.

Pallier, C., Dehaene, S., Poline, J-B., LeBihan, D., Argenti, A-M., Dupoux, E., & Mehler, J. (2003). Brain imaging of language plasticity in adopted adults: can a second language replace the first? *Cerebral Cortex*, 13, 155-161. doi: 10.1016/S1053-8119(01)91925-1.

Pang, E.W., Edmonds, G.E., Desjardins, R., Khan, S.C., Trainor, L.J., & Taylor, M.J. (1998). Mismatch negativity to speech stimuli in 8-month-old infants and adults. *International Journal of Psychophysiology*, 29: 227-236. doi: 10.1016/S0167-8760(98)00018-X.

Pantev, C., Hoke, M., Lehnertz, K., & Lütkenhöner, B. (1989). Neuromagnetic evidence of an amplitopic organization of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology,* 72, 225-231. doi:10.1016/0013-4694(89)90247-2.

Partanen, E., Pakarinen, S., Kujala, T., & Huotilainen, M. (2013). Infants' brain responses for speech sound changes in fast multifeature MMN paradigm. *Clinical Neurophysiology,* 124, 1578–1585. doi: 10.1016/j.clinph.2013.02.014.

Peperkamp, S., Pettinato, M., & Dupoux, E. (2003). Allophonic variation and the acquisition of phoneme categories. In B. Beachley, A. Brown, & F. Conlin (Eds), *Proceedings of the 27th Annual Boston University Conference on Language Development. Volume 2* (pp. 650-661). Sommerville, MA: Cascadilla Press.

Peterson, G.E., & Barney, H.L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America (JASA), 24(2),* 175-184.

Phillipson, O.T. (1979). Afferent projections to the ventral tegmental area of Tsai and interfascicular nucleus: a horseradish peroxidase study in the rat. *The Journal of Comparative Neurology*, 187(1), 117-143. doi: 10.1002/cne.901870108.

Picciotto, M.R., Higley, M.J., & Mineur, Y.S., (2012). Acetylcholine as a neuromodulator: cholinergic signaling shapes nervous system function and behavior. *Neuron*, 76, 116-129. doi: 10.1016/j.neuron.2012.08.036.

Pienkowski, M., & Eggermont, J.J. (2011). Cortical tonotopic plasticity and behavior. *Neuroscience and Biobehavioral Reviews*, 35, 2117-2128. doi: 10.1016/j.neubiorev.2011.02.002.

Pierrehumbert, J.B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46(2-3), 115-154. doi: 10.1177/00238309030460020501.

Pihko, E., Sambeth, A., Leppänen, P., Okada, Y., & Lauronen, L. (2004). Auditory evoked magnetic fields to speech stimuli in newborns – effect of sleep stages. *Neurology and Clinical Neurophysiology,* 6, 1–5.

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, 191-215. doi: 10.006/jpho.2001.0134.

Pisoni, D.B. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *The Journal of the Acoustical Society of America*, 61, 1352–1361. doi: 10.1121/1.381409.

Polivanov, E.D. (1931). La perception des sons d' une langue étrangère. *Travaux du Cercle Linguistique de Prague*, 4, 79-96. English Translation: The subjective nature of the perceptions of language sounds. In: E.D. Polivanov (1974). *Selected works: articles on general linguistics.* The Hague: Mouton, 223-237.

Polka, L., & Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America*, 100, 577–592. doi: 10.1121/1.415884.

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41, 221–231. doi: 10.1016/S0167-6393(02)00105-X.

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: an emergent view of early phonetic development. *Journal of Phonetics*, 39, 467–478. doi: 10.1016/j.wocn.2010.08.007.

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/-/ð/ perception: evidence for a new developmental pattern. *The Journal of the Acoustical Society of America*, 109, 2190–2201. doi: 10.1121/1.1362689.

Polka, L., & Werker, J.F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance,* 20, 421–435. doi: 10.1037/0096-1523.20.2.421.

Polley, D.B., Steinberg, E.E., & Merzenich, M.M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *Journal of Neuroscience,* 26(18), 4970-4982. doi: 10.1523/JNEUROSCI.3771-05.2006.

Pols, L. C. W., Tromp, H. R. C., & Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *The Journal of the Acoustical Society of America,* 53, 1093–1101. doi:10.1121/1.1913429.

Pons, F. (2006). The effects of distributional learning on rats' sensitivity to phonetic information. *Journal of Experimental Psychology: Animal Behavior Processes*, 32: 97-101. doi: 10.1037/0097-7403.32.1.97.

Pons, F., Albareda-Castellot, B., & Sebastián-Gallés, N. (2012). The interplay between input and initial biases: asymmetries in vowel perception during the first year of life. *Child Development,* 1–12. doi: 10.1111/j.1467-8624.2012.01740.x.

Pons, F., Mugitani, R., Amano, S., & Werker, J.F. (2006). Distributional learning in vowel length distinctions by 6-month-old English infants. Poster presented at the International Conference on Infant Studies, June 21, Kyoto, Japan.

Pons, F., Sabourin, L., Cady, J.C., & Werker, J.F. (2006). Distributional learning in vowel distinctions by 8-month-old English infants. Poster presented at the 28th Annual Conference of the Cognitive Science Society, July 29, Vancouver, BC, Canada.

Ponton, C.W., Don, M., Eggermont, J.J., Waring, M.D., & Masuda, A. (1996). Maturation of human cortical auditory function: differences between normal-hearing children and children with cochlear implants. *Ear & Hearing*, 17(5), 430-437. doi: 10.1097/00003446-199610000-00009.

Ponton, C.W., & Eggermont, J.J. (2001). Of kittens and kids: altered cortical maturation following profound deafness and cochlear implant use. *Audiology & Neuro-Otology*, 6, 363-380. doi: 10.1159/000046846.

Ponton, C.W., Eggermont, J.J., Kwong, B., & Don, M. (2000). Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. *Clinical Neurophysiology*, 111, 220-236. doi: 10.1016/S1388-2457(99)00236-9.

Purves, D., Augustine, G.J., Fitzpatrick, D., Hall, W.C., LaMantia, A-S., McNamara, J.O., & White, L.E. (Eds) (2008). *Neuroscience*. Sunderland, Mass.: Sinauer Associates.

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crèpin, G. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 29, 191-212.

Recanzone, G.H., Schreiner, C.E., & Merzenich, M.M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience,* 13(1), 87-103. PMID: 8423485.

Recanzone, G.H., Schreiner, C.E., Sutter, M.L., Beitel, R.E., & Merzenich, M.M. (1999). Functional organization of spectral receptive fields in the primary auditory cortex of the owl monkey. *Journal of Comparative Neurology,* 415(4), 460-481.

Reed, A., Riley, J., Carraway, R., Carrasco, A., Perez, C., Jakkamsetti, V., & Kilgard, M.P. (2011). Cortical map plasticity improves learning but is not necessary for improved performance. *Neuron*, 70, 121-131. doi: 10.1016/j.neuron.2011.02.038.

Robertson, R.T., Mostamand, F., Kageyama, G.H., Gallardo, K.A., Yu, J. (1991). Primary auditory cortex in the rat: transient expression of acetylcholinesterase activity in developing geniculocortical projections. *Developmental Brain Research*, 58, 81-95. doi: 10.1016/0165-3806(91)90240-J.

Rodenbeck, A., Binder, R., Geisler, P., Danker-Hopfe, H., Lund, R., Raschke, F., Weeß, H-G., & Schulz, H. (2007). A review of sleep EEG patterns. Part I: a

compilation of amended rules for their visual recognition according to Rechtschaffen and Kales. *Somnologie*, 10, 159–175. doi: 10.1111/j.1439-054X.2006.00101.x.

Roelfsema, P.R. (2011). Attention – voluntary control of brain cells. *Science*, 332: 1512-1513. doi: 10.1126/science.1208564.

Rogers, J. C., & Davis, M. H. (2009). Categorical perception of speech without stimulus repetition. In *Proceedings of Interspeech 2009* (pp. 376–379). Brighton.

Romo, R., & De Lafuente, V. (2013). Conversion of sensory signals into perceptual decisions. *Progress in Neurobiology*, 103, 41-75. doi: 10.1016/j.pneurobio.2012.03.007.

Rosch, E.H. (1973). Natural categories. *Cognitive Psychology*, 4, 328-350.

Rosch, E., & Mervis, C.B. (1975). Family resemblances : studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review,* 16(2), 225-237. doi: 10.3758/PBR.16.2.225.

Rouiller, E.M., Simm, G.M., Villa, A.E.P., De Ribaupierre, Y., & De Ribaupierre, F. (1991). Auditory corticocortical interconnections in the cat: evidence for parallel and hierarchical arrangement of the auditory areas. *Experimental Brain Research*, 86 (3), 483-505. doi: 10.1007/BF00230523.

Rubio-Garrido, P., Pérez-de-Manzo, F., Porrero, C., Galazo, M.J., & Clascá, F., (2009). Thalamic input to distal apical dendrites in neocortical layer I is massive and highly convergent. *Cerebral Cortex*, 19, 2380-2395. doi:10.1093/cercor/bhn259. doi: 10.1093/cercor/bhn259.

Rutkowski, R.G., & Weinberger, N.M. (2005). Encoding of learned importance of sound by magnitude of representational area in primary auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 102(38), 13664-13669. doi: 10.1073/pnas.0506838102.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, New Series, 274(5294), 1926-1928.

Sambeth, A., Pakarinen, S., Ruohio, K., Fellman, V., Van Zuijen, T.L., & Huotilainen, M. (2009). Change detection in newborns using a multiple deviant paradigm: a study using magnetoencephalography. *Clinical Neurophysiology,* 120, 530–538. doi: 10.1016/j.clinph.2008.12.033.

Schmitz, N., Malla, A., Norman, R., Archie, S., & Zipursky, R. (2007). Inconsistency in the relationship between duration of untreated psychosis (DUP) and negative symptoms: sorting out the problem of heterogeneity. *Schizophrenia Research,* 93, 152-159. doi: 10.1016/j.schres.2007.03.021.

Schouten, M.E.H. (1975). *Native-language interference in the perception of second-language vowels: an investigation of certain aspects of the acquisition of a second language*. Doctoral dissertation, Utrecht University.

Schreiner, C.E. (1998). Spatial distribution of responses to simple and complex sounds in the primary auditory cortex. *Audiology & Neuro-Otology,* 3(2-3), 104-122. doi: 10.1159/000013785.

Schröger, E. (1997). On the detection of auditory deviations: a pre-attentive activation model. *Psychophysiology*, 34, 245–257. doi: 10.1111/j.1469-8986.1997.tb02395.x.

Schultz, W. (1992). Activity of dopamine neurons in the behaving primate. *Seminars in the Neurosciences*, 4, 129-138. doi: 10.1016/1044-5765(92)90011-P.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics,* 6(2), 461-464. doi: 10.2307/2958889.

Shafer, V.L., Yu, Y.H., & Datta, H. (2011). The development of English vowel perception in monolingual and bilingual infants: neurophysiological correlates. *Journal of Phonetics*, 39, 527–545. doi: 10.1016/j.wocn.2010.11.010.

Sharma, A., Dorman, M.F., & Spahr, A.J. (2002). A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation. *Ear & Hearing*, 23(6), 532-539.

doi: 10.1097/01.AUD.0000042223.62381.01.

Shea, C., & Curtin. S. (2006). Learning allophones from the input. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds), *Supplement for the Proceedings of the Boston University Conference on Language Development,* Sommerville, MA: Cascadilla Press.

Shestakova, A., Brattico, E., Soloviev, A., Klucharev, V., & Huotilainen, M. (2004). Orderly cortical representation of vowel categories presented by multiple exemplars. *Cognitive Brain Research,* 21(3), 342-350. doi: 10.1016/j.cogbrainres.2004.06.011.

Smiljanić, R., & Bradlow, A.R. (2009). Speaking and hearing clearly: talker and listener factors in speaking style changes. *Language and Linguistics Compass*, 3(1), 236-264, doi:10.1111/j.1749-818X.2008.00112.x.

Sokolov, E., Spinks, J., Näätänen, R., & Lyytinen, H. (2002). *The Orienting Response in Information Processing*. Mahwah, New Jersey/London: Lawrence Erlbaum Associates.

Speechley, W.J., Hogsden, J.L., & Dringenberg, H.C. (2007). Continuous white noise exposure during and after auditory critical period differentially alters bidirectional thalamocortical plasticity in rat auditory cortex *in vivo*. *European Journal of Neuroscience*, 26, 2576-2584. doi: 10.1111/j.1460-9568.2007.05857.x.

Stager, C.L., & Werker, J.F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature,* 388, 381-382. doi: 10.1038/41102.

Stanton, S.G., & Harrison, R.V. (1996). Abnormal cochleotopic organization in the auditory cortex of cats reared in a frequency augmented environment. *Auditory Neuroscience*, 2, 97-107.

Steinschneider, M., Arezzo, J.C., & Vaughan Jr., H.G. (1990). Tonotopic features of speech-evoked activity in primate auditory cortex. *Brain Research,* 519(1-2), 158-168. doi: 10.1016/0006-8993(90)90074-l.

Stevens, S.S., Volkmann, J., & Newman, E.B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America,* 8, 185-190. doi: 10.1121/1.1915893.

Strange, W. (1992). Learning non-native phoneme contrasts: interactions among subject, stimulus, and task variables. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds), *Speech Perception, Production and Linguistic Structure* (pp. 197-219). Amsterdam/Washington/Oxford: IOS Press.

Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-ð/ in monolingual and bilingual acquisition of English. *Cognition*, 100, 369–388. doi: 10.1016/j.cognition.2005.04.007.

Sundberg, U. (2001). Consonant specification in infant-directed speech. Some preliminary results from a study of Voice Onset Time in speech to one-year-olds. In *Working Papers 49* (pp. 148-151). Department of Linguistics, Stockholm University.

Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica,* 56(3-4), 186-199. doi: 10.1159/000028450.

Swoboda, P.J., Kass, J., Morse, P.A., & Leavitt, L.A. (1978). Memory factors in vowel discrimination of normal and at-risk infants. *Child Development,* 49, 332–339. doi: 10.2307/1128695.

Swoboda, P.J., Morse, P.A., & Leavitt, L.A. (1976). Continuous vowel discrimination in normal and at risk infants. *Child Development,* 47, 459–465. doi: 10.2307/1128802.

Terrace, H.S. (1963). Discrimination learning with and without "errors". *Journal of the Experimental Analysis of Behavior,* 6(1), 1-27.

Ter Schure, S., Mandell, D.J., Escudero, P., Raijmakers, M.E.J., & Johnson, S.P. (2014). Learning stimulus-location associations in 8- and 11-month-old infants: multimodal versus unimodal information. *Infancy*, 19(5), 476-495. doi: 10.1111/infa.12057.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech sound training. *NeuroReport*, t9, 3557-3560. doi: 10.1097/00001756-199811160-00003.

Tsao, F.-M., Liu, H.-M., & Kuhl, P.K. (2006). Perception of native and non-native affricate-fricative contrasts: cross-language tests on adults and infants. *The Journal of the Acoustical Society of America*, 120, 2285–2294. doi: 10.1121/1.2338290.

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: language-specific developmental changes in speech perception abilities. *International Conference on Spoken Language Processing (ICSLP)* (pp. 1695–1698). Yokohama.

Uther, M., Knoll, M.A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication,* 49, 2-7. doi: 10.1016/j.specom.2006.10.003.

Van Essen, D.C., & Maunsell, J.H.R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, 6, 370-375. doi: 10.1016/0166-2236(83)90167-4.

Van Hesse, A.J., & Schouten, M.E.H. (1999). Categorical perception as a function of stimulus quality. *Phonetica,* 56, 56-72. doi: 10.1159/000028441.

Van Heuven, V.J., Van Houten, J.E., & De Vries, J.W. (1986). De perceptie van Nederlandse klinkers door Turken. *Spektator,* 15, 225-238.

Van Leeuwen, T., Been, P., Van Herten, M., Zwarts, F., Maassen, B., & Van der Leij, A. (2008). Two-month-old infants at risk for dyslexia do not discriminate /bAk/ from /dAk/: a brain-mapping study. *Journal of Neurolinguistics,* 21, 333–348. doi: 10.1016/j.jneuroling.2007.07.004.

Van Leussen, J-W., Williams, D., & Escudero, P. (2011). A comparison of Dutch steady-state vowels: contextual effects and a comparison with previous studies. In: W Lee & E. Zee (Eds), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp.1194 – 1197). Hong Kong.

Van Zuijen, T.L., Plakas, A., Maassen, B.A., Maurits, N.M., & Van der Leij, A. (2013). Infant ERPs separate children at risk of dyslexia who become good readers from those who become poor readers. *Developmental Science*, 16(4): 554-563. doi: 10.1111/desc.12049.

Ventureyra, V.A.G., Pallier, C., & Yoo, H-Y. (2004). The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics*, 17, 79-91. doi: http://dx.doi.org/10.1016/S0911-6044(03)00053-8.

Visser, I., Raijmakers, M.E.J., & Pothos, E. M. (2009). Individual strategies in artificial grammar learning. *The American Journal of Psychology,* 122(3)*,* 293-308.

Wang, H., Wang, X., & Scheich, H. (1996). LTP and LTD induced by transcranial magnetic stimulation in auditory cortex. *NeuroReport*, 7, 521-525. doi: 10.1097/00001756-199601310-00035.

Wanrooij, K., & Boersma, P. (2013). Distributional training of speech sounds can be done with continuous distributions. *The Journal of the Acoustical Society of America*, 133, EL398–EL404. doi: 10.1121/1.4798618.

Wanrooij, K., Boersma, P., & Van Zuijen, T.L. (2014a). Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study. *Frontiers in Psychology (Language Sciences)*, 5, article 77, 1-12. doi: 10.3389/fpsyg.2014.00077.

Wanrooij, K., Boersma, P., & Van Zuijen, T.L. (2014b). Distributional vowel training is less effective for adults than for infants. A study using the mismatch response. *PloS One, 9(10)*, 1-13. doi: 10.1371/journal.pone.0109806.

Wanrooij, K., Escudero, P., & Raijmakers, M.E.J. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics*, 41, 307–319. doi: 10.1016/j.wocn.2013.03.005.

Wanrooij, K., Van Zuijen, T., & Boersma, P. (2012). MMN declines after distributional vowel training. Poster presentation at *The 6th Conference on Mismatch Negativity (MMN) and its Clinical and Scientific Application*, May 1–4, New York. http://home.medewerker.uva.nl/k.e.wanrooij/bestanden/MMN2012.pdf (Last viewed 3/11/2013).

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

doi: 10.1016/S0749-596X(03)00105-0.

Weber, C., Hahne, A., Friedrich, M., & Friederici, A.D. (2004). Discrimination of word stress in early infant perception: electrophysiological evidence. *Cognitive Brain Research,* 18, 149–161. doi: 10.1016/j.cogbrainres.2003.10.001.

Weinberger, N.M. (2007). Auditory associative memory and representational plasticity in the primary auditory cortex. *Hearing Research*, 229(1-2): 54-68. doi:10.1016/j.heares.2007.01.004.

Weinberger, N.M., & Bakin, J.S. (1998). Learning-induced physiological memory in adult primary auditory cortex: receptive field plasticity, model, and mechanisms. *Audiology & Neuro-Otology*, 3, 145-167. doi: 10.1159/000013787.

Werker, J.F., & Curtin, S. (2005). PRIMIR: a developmental framework of infant speech processing. *Language Learning and Development*, 1(2), 197-234. doi:10.1080/15475441.2005.9684216.

Werker, J.F., & Logan, J.S. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics,* 37, 35-44. doi: 10.3758/BF03207136.

Werker, J.F., & Tees, R.C. (1984/2002). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development,* 7, 49–63. (Republished in 2002: *Infant Behavior and Development,* 25, 121–133. doi: 10.1016/S0163-6383(84)80022-3).

Werker, J.F., & Tees, R.C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, 46, 233-251. doi: 10.1002/dev.20060.

Werner, L.A. (2007). Issues in human auditory development. *Journal of Communication Disorders*, 40, 275-283. doi: 10.1016/j.jcomdis.2007.03.004.

Werner, L.A., & Gillenwater, J.M. (1990). Pure-tone sensitivity of 2- to 5-week old infants. *Infant Behavior and Development*, 13: 355-375. doi: 10.1016/0163-6383(90)90040-F.

Werner, L.A. (1996). The development of auditory behavior (or what the anatomists and physiologists have to explain). *Ear and Hearing*, 17 (5): 438-446. doi: 0.1097/00003446-199610000-00010.

Williams, D., & Escudero, P. (2014). A cross-dialectal acoustic comparison of Northern and Southern British English vowels. *The Journal of the Acoustical Society of America*, 136(5): 2751-2761. doi: 10.1121/1.4896471.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, H., Iivonen, A., & Näätänen, R. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research*, 7: 357-369. doi: 10.1016/S0926-6410(98)00039-1.

Yamada, R. A. (1995). Age and acquisition of second language speech sounds: perception of American English /r/ and /l/ by native speakers of Japanese. In W. Strange (Ed), *Speech perception and language experience: issues in cross-language research* (pp. 305–320). Baltimore, MD: York Press.

Yamaguchi, K. (2000). Multinomial logit latent-class regression models: an analysis of the predictors of gender-role attitudes among Japanese women. *American Journal of Sociology,* 105(6), 1702-1740. doi: 10.1086/210470.

Yeung, H.H., Chen, K.H., & Werker, J.F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory & Language*, 68, 123–139. doi: 10.1016/j.jml.2012.09.004.

Yeung, H.H., & Werker, J.F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, 113, 234-243. doi: 10.1016/j.cognition.2009.08.010.

Yoshida, K.A., Pons, F., Maye, J., & Werker, J.F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, 15, 420–433. doi: 10.1111/j.1532-7078.2009.00024.x.

Zaehle, T., Clapp, W.C., Hamm, J.P., Meyer, M., & Kirk, I.J. (2007). Induction of LTP-like changes in human auditory cortex by rapid auditory stimulation: an fMRI study. *Restorative Neurology and Neuroscience*, 25, 251-259. PMID: 17943003.

Zhang, L.I., Bao, S., & Merzenich, M.M. (2001). Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nature Neuroscience,* 4(11), 1123-1130. doi:10.1038/nn745.

Zhang, L.I., Bao, S., & Merzenich, M.M. (2002). Disruption of primary auditory cortex by synchronous auditory inputs during a critical period. *Proceedings of the National Academy of Sciences of the United States of America (PNAS),* 99(4), 2309-2314. doi: 10.1073/pnas.261707398.

Zhang, L.I., Tan, A.Y.Y., Schreiner, C.E., & Merzenich, M.M. (2003). Topography and synaptic shaping of direction selectivity in primary auditory cortex. *Nature*, 424, 201-205. doi:10.1038/nature01796.

Zhang, Y., & Wang, Y. (2007). Neural plasticity in speech acquisition and learning. *Bilingualism: Language and Cognition*, 10(2), 147-160. doi:10.1017/S1366728907002908.

Zhou, X., & Merzenich, M.M. (2007). Intensive training in adults refines A1 representations degraded in an early postnatal critical period. *Proceedings of the National Academy of Sciences of the United States of America (PNAS),* 104(40), 15935-15940. doi: 10.1073/pnas.0707348104.

Zhou, X.; Nagarajan, N.; Mossop, B. J., & Merzenich, M.M. (2008). Influences of un-modulated acoustic inputs on functional maturation and critical-period plasticity of the primary auditory cortex. *Neuroscience,* 154(1), 390-396. doi: 10.1016/j.neuroscience.2008.01.026.

Ω

Distributional learning is learning from simple exposure to the environment, without receiving explicit instruction or feedback. This thesis examines to what extent this basic form of learning contributes to learning the vowels of a language, both in infancy, when the mother tongue must be acquired, and in adulthood, when new languages can be learned. The results are based on neurophysiological and behavioural experiments, and on an extensive literature review of possible neural correlates. The main conclusions are that (1) distributional learning can contribute to the acquisition of native vowel categories in infancy, (2) the capacity for distributional learning is larger in infancy than in adulthood, and (3) observed effects of distributional training in the lab may not be based on the number of peaks in the training distributions.