TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

PLASTICITY IN SPEECH PERCEPTION

Effects of Learning, Age and Bilingualism

Henna Tamminen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. B OSA – TOM. 583 | HUMANIORA | TURKU 2022

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

PLASTICITY IN SPEECH PERCEPTION

Effects of Learning, Age and Bilingualism

Henna Tamminen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. B OSA – TOM. 583 | HUMANIORA | TURKU 2022

University of Turku

Faculty of Technology
Department of Computing
Phonetics
Doctoral Programme in Technology (DPT)

Supervised by

Professor, Maija S. Peltola
University of Turku

Reviewed by

Associate Professor, Henrike K. Blumenfeld
San Diego State University

Associate Professor, Piia Astikainen
University of Jyväskylä

Opponent

Adjunct Professor, Stefan Werner
University of Eastern Finland

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

ISBN 978-951-29-8925-6 (PRINT)
ISBN 978-951-29-8926-3 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)
Painosalama, Turku, Finland 2022

UNIVERSITY OF TURKU
Faculty of Technology
Department of Computing
Phonetics
HENNA TAMMINEN: Plasticity in Speech Perception – Effects of Learning, Age and Bilingualism
Doctoral Dissertation, 163 pp.
Doctoral Programme in Technology (DPT)
August 2022

ABSTRACT

Brain plasticity enables us to learn to perceive languages. Theories and models of non-native language learning predict learning difficulties in phonological areas where the native and non-native phonological systems overlap and where a category boundary of one language may lie somewhere within a category of another language. Phonologically, these areas are interesting within the two native languages of bilinguals from birth also.

In this thesis, both psychophysiological and behavioural measurements were used to explore the neural plasticity in relation to learning, age, and bilingualism. The research questions were, whether phonological processing is different between different kinds of bilinguals and whether it is different between bilinguals from birth and monolinguals. A further research question was, whether different background factors, such as manner and age of learning, affect memory trace formation. The bilinguals from birth, simultaneous bilinguals, seem to have a shared phonological system for the two languages, whereas later bilinguals, sequential bilinguals, seem to have separate phonological systems for the languages. The shared system in the simultaneous bilinguals leads to slower processing compared to sequential bilinguals and monolinguals. The languages seem to be active all the time in the shared phonological system irrespective of the language context. Whereas the sequential bilinguals, with the separate systems, can even ignore their native language in a second language context. The studies on the manner and age of learning revealed that both classroom and laboratory training lead to similar functioning of memory traces. The laboratory training effects also seem to be permanent. However, the age of learning affects memory trace formation: While the young adults show memory trace formation during laboratory training, the elderly do not show the same effect.

In conclusion, different types of bilinguals process speech differently and simultaneous bilinguals' speech processing is different from that of monolinguals'. Training in classroom and in laboratory lead to similar functioning of memory traces. Age, however, affects plasticity and laboratory training does not lead to similar training effects in the elderly compared to young adults.

KEYWORDS: speech perception, phonological processing, plasticity, bilingualism, training, aging, mismatch negativity (MMN)

TURUN YLIOPISTO
Teknillinen tiedekunta
Tietotekniikan laitos
Fonetiikka
HENNA TAMMINEN: Plastisiteetti puheen havaitsemisessa – Oppiminen, ikä ja kaksikielisyys
Väitöskirja, 163 s.
Teknologian tohtoriohjelma (DPT)
Elokuu 2022

TIIVISTELMÄ

Kielen omaksuminen tai oppiminen on mahdollista muovautuvien aivojemme vuoksi. Vieraan kielen oppimisteoriat ennustavat oppimisvaikeuksia sellaisilla fonologisilla alueilla, joilla äidinkielen ja vieraan kielen äännejärjestelmät ovat limittäiset eli alueilla, joilla toisen kielen äännekategorian raja sijoittuu toisen kielen äännekategorian sisään. Tällaiset fonologiset seikat ovat mielenkiintoisia myös syntymästään asti kaksikielisten kohdalla.

Tämän tutkimuksen tarkoituksena oli tutkia neuraalista muovautuvuutta vieraan kielen havaitsemisessa oppimisen, iän ja kaksikielisyyden suhteen käyttäen sekä psykofysiologisia että behavioraalisia menetelmiä. Tutkimusten oli toisaalta tarkoitus selvittää, eroaako puheen neuraalinen prosessointi samanaikaisilla ja peräkkäisillä kaksikielisillä sekä eroaako puheen havaitseminen samanaikaisilla kaksikielisillä ja yksikielisillä. Toisaalta selvitettiin, muokkaantuuko yksikielisten puheen prosessointi samankaltaiseksi kuin peräkkäisillä kaksikielisillä kuuntele ja toista harjoittelun myötä ja ovatko harjoittelun vaikutukset pysyviä. Lisäksi selvitettiin eri taustatekijöiden – oppimistavan ja -iän – vaikutuksia muistijälkien syntyyn. Tutkimusten mukaan samanaikaisilla kaksikielisillä näyttää olevan yksi yhtenäinen fonologinen järjestelmä, kun taas peräkkäisillä kaksikielisillä näyttää olevan kaksi erillistä järjestelmää. Yhtenäisessä järjestelmässä molemmat kielet ovat koko ajan aktiivisia kielikontekstista riippumatta, mikä hidastaa prosessointia suhteessa erilliseen järjestelmään. Peräkkäiset kaksikieliset taas voivat toisen kielensä kontekstissa jättää äidinkielensä täysin huomiotta. Oppimistapaan ja -ikään liittyvät tutkimukset osoittivat luokkahuone- ja laboratorioharjoittelun johtavan samanlaiseen muistijälkien toimimiseen. Laboratorioharjoittelun vaikutukset näyttivät myös olevan pysyviä. Oppimisikä vaikuttaa kuitenkin muistijälkien syntyyn eri tavoin: nuorille aikuisille syntyy muistijäljet laboratorioharjoittelun vaikutuksesta, mutta vastaavaa ei nähdä iäkkäämmillä, jo eläkkeellä olevilla henkilöillä.

Näyttää siis siltä, että eri tyyppiset kaksikieliset prosessoivat puhetta eri tavoin ja että samanaikaisten kaksikielisten puheen havaitsemisen prosessointi eroaa yksikielisistä. Kuuntele ja toista harjoittelulla voidaan saavuttaa puheen neuraalisen prosessoinnin taso, joka on samankaltainen kuin peräkkäisillä kaksikielisillä. Ikä kuitenkin vaikuttaa muovautuvuuteen eikä sama harjoittelu toimi iäkkäämmillä henkilöillä samalla tavalla kuin nuorilla aikuisilla.

ASIASANAT: puheen havaitseminen, fonologinen prosessointi, muovautuvuus, kaksikielisyys, harjoittelu, ikääntyminen, mismatch negativity (MMN)

Acknowledgements

For too long my answer was “next year” when someone asked me when I would defend my thesis. Well, next year turned out to be year 2022. During the long thesis journey, I have become grateful for several people and institutions.

The journey begun when I had the opportunity to work as a research assistant in a project funded by the Academy of Finland (grant 206352). Later, with the Emil Aaltonen Foundation funding, I was able to concentrate on my thesis research. I am highly grateful for these funders. My research has also been supported by Turku University Foundation, Doctoral Program MATTI, and Doctoral Program Utuling. And finally, I would like thank the University of Turku Graduate School and Doctoral Program in Technology (DBT).

I address my greatest and sincerest thanks to my supervisor, Professor Maija S. Peltola. I am deeply grateful for your ideas, guidance, and patience during the preparation process of my thesis. We have had endless discussions about past, present, and future research topics and that has made me the researcher I am today. You have navigated with me through dense fog and clearer skies. Thank you Maija.

I also wish to thank Associate Professor Henrike K. Blumenfeld and Associate Professor Piia Astikainen for agreeing to act as the pre-examiners of my thesis. Thank you for your valuable comments. Adjunct Professor Stefan Werner kindly accepted the invitation to be my opponent on the most exciting day of my thesis journey, thank you. I also want to thank my research director Professor Tapio Salakoski.

The research and the articles of my thesis would not have been finished without all the people who volunteered as participants. Thank you all. I would also like to thank Heidi Toivonen, Jaana Luotonen, Heli Kurttila, Katri Jähi, and Eetu Laine for their help in the lab during testing. Professor Teija Kujala and Emeritus Professor Risto Näätänen, the co-authors and research partners, I am deeply grateful for you both for the time you have spent reading and commenting the article manuscript versions and for sharing your vast expertise with me.

Past and current colleagues at the Phonetics and Learning, Age & Bilingualism –laboratory, or the former Department of Phonetics, thank you for the scientific discussions, the endless argumentations on not so scientific topics and all the laughs over the years. Maija S. Peltola, Kimmo Peltola, Antti Saloranta, Katja Haapanen, Tomi Rautaoja, and Jemina Kilpeläinen being the most recent colleagues, thank you. I would also like to thank all the people, but especially Mia Ek, Teemu Laine and Heikki Hämäläinen, at the former Centre for Cognitive Neuroscience during the years I spent there. Jyrki Tuomainen, in between the Department of Phonetics and CCN, I appreciate your help at the early stages of this project, before it became my thesis project. I am also very grateful for all the students on my phonetics courses over the years for your enthusiasm on my research, you have been teaching me as well. Heidi, Maija, and others, thank you for your company and all the fun in the conferences where we have been presenting the research of this thesis.

And then there are the union guys from TYT. What a terrific bunch of people you are. Thank you all for counterbalancing the academic world and work as you represent many different fields of science. And most of all, thank you for all the get-togethers, (bad) jokes and fun that steered my mind off this thesis.

Friends and family, thank you for being there. Jenni, thank you for your help with the finishing touches of the thesis.

Pessi, how can I ever enough express my gratitude for the years together and for everything you have done for me and on behalf of me. Thank you. And finally, I want to thank Martti, our four-legged stress reliever.

Lieto, June 2022
Henna Tamminen

Table of Contents

Tables

Figures

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Peltola, M. S., Tamminen, H., Toivonen, H., Kujala, T. & Näätänen, R. Different kinds of bilinguals – Different kinds of brains: The neural organisation of two languages in one brain. *Brain & Language*, 2012; 121(3): 261–266.

II Tamminen, H., Peltola, M. S., Toivonen, H., Kujala, T. & Näätänen, R. Phonological processing differences in bilinguals and monolinguals. *International Journal of Psychophysiology*, 2013; 87(1): 8–12.

III Tamminen, H., Peltola, M. S., Kujala, T. & Näätänen, R. Phonetic training and non-native speech perception – New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures. *International Journal of Psychophysiology*, 2015; 97(1): 23–29.

IV Tamminen, H., Kujala, T. & Peltola, M. S. Training non-native speech sounds results in permanent plastic changes – Hard-wiring new memory traces takes time. Submitted.

V Tamminen, H., Kujala, T., Näätänen, R. & Peltola, M. S. Aging and non-native speech perception: A phonetic training study. *Neuroscience Letters*, 2021; 740: 135430.

1 Introduction

In modern society, monolingualism is more and more scarce while bilingualism, may the definition be loose or strict, is increasingly common. The learning or acquisition of a foreign or second language is possible because of our plastic brain. A second native language may be acquired simultaneously with the first one, later in childhood, or even later. A foreign language may be learned in a natural environment through immigration, for instance, or in a classroom. Learning a foreign language comprises of, among other things, the learning of foreign language speech perception patterns. Learning of these patterns has been shown to be more effective, i.e., leading to native like perception, through natural environment than in a classroom.

Phonologically the interesting learning situations are those where theories and models of foreign language learning predict difficulties. When the speech sound categories overlap in the mother tongue and the foreign language, difficulties of some level are usually anticipated. Phonological overlapping is obviously inevitable in bilinguals' two native languages as well, also providing interesting research questions.

This thesis explores plasticity as shown by neural processing or behaviour from the viewpoint of learning, age and bilingualism in speech perception. Broadly speaking, the research questions are, whether speech is perceived differently by different kinds of bilinguals and how different background factors, such as manner and age of learning, affect the formation of memory traces. Hence, speech perception, native language (NL) acquisition, foreign language learning, bilingualism, language learning by training, and the effects of age on learning will be discussed in the theoretical part of the thesis. First, speech perception and some of the speech perception theories, as a ground for this thesis, will be contemplated. The discussion then moves forward to discuss the acquisition of the mother tongue and the learning of foreign language, highlighting some of the models of native and second language learning. Next, different aspects of bilingualism, including terminological issues, will be examined. Then the focus will be on language learning by different training methods, by reviewing speech perception and production related training studies. Finally, the effects of age on foreign language learning are in focus.

Before the empirical part, the research questions, aims and hypotheses are discussed. The empirical part summarises the methodologies used in the studies. Speech perception was measured with traditional behavioural listening tasks as well as with psychophysiological measures at neural, pre-attentive level. Then the results are summarised first focusing on the experiments on different kinds of bilinguals' behavioural and neural speech perception, different types of bilinguals being those from birth and those who learned the second language (L2) later in life in classroom environment. Then the training experiments, which explore the effects of training on speech perception in young and older adults, will be introduced.

General discussion and conclusions aggregate the five studies and evaluate learning, age and bilingualism results in the light of the research questions and hypotheses as well as the theoretical background of the thesis and other relevant research on speech perception.

2 Theoretical background: Speech perception and learning

We usually understand the message conveyed when someone speaks to us. When perceiving speech, we deconstruct or convert the speech signal into phonemes, and hence, we must be able to categorise speech sound units. When the language spoken to us is not our mother tongue the decoding patterns are different, specific to the language in question. If we have learned the foreign language, the process is easier but not necessarily native-like.

This chapter will first concentrate on speech perception and some relevant theories of speech perception. Then it will continue with the acquisition and learning of language in the light of earlier research and relevant models of language learning. Then the focus will be on the special case of bilingualism after which language learning via training and language learning in the elderly are discussed.

2.1 Speech perception

It is obviously essential that a person receiving a message spoken by another person is able to hear. Thus, one important part of speech perception, and in the speech chain (Denes & Pinson, 1963), is naturally the neural activity taking place in hearing and perception. Perceiving a spoken message received by the outer and inner ear, a listener needs to apprehend the acoustic signal, and tackle phonology, lexicon, semantics, grammar, coarticulation and so forth. In its entirety, speech perception is such a vast process that it is difficult to represent by a single theory of speech perception. Further, some of the existing speech perception theories are opposed more strongly than others, and some theories have evidence for their claim, other theories provide counter-evidence for the same theorem. Most theories probably contain some plausible elements, and yet, none of them describes speech perception exhaustively. As Hawkins (1999a, p. 226) reminds us, "a theory need not be wrong because it is hard to verify empirically". The adjacent sound segments affect the acoustics of each other so that there can be excessive variation in the acoustic properties of a given phoneme depending on the context. Nevertheless, despite the lack of invariance in speech, people usually understand the message received. The lack of invariance in speech undeniably hampers the formation of speech perception models and theories. While there have been attempts to discover acoustic-phonetic invariance[1] (e.g., Blumstein & Stevens, 1979, 1980; Kewley-Port, 1982, 1983; Mack & Blumstein, 1983), still a comprehensive series of invariants has not been discovered (e.g., Nygaard & Pisoni, 1995). Speech perception theories may be classified according to the role of the listener, namely to theories in which the listeners play an *active* role and theories in which they are *passive*[2] (Crystal, 1997, p. 148). Theories of active listeners (e.g., the Motor Theory) are those where listeners play an active role and use their knowledge about the articulatory movements in identifying the linguistic input and acoustic signal. In the passive scenario, the listener automatically decodes the speech matching it to abstract speech patterns (e.g., feature detection, template matching). One general classification of speech perception theories divides them into two categories according to the primary object of perception; namely auditory theories, where the target of perception is the acoustic signal itself, and motor theories, where the object is the articulatory gesture. Some auditory and motor theories are described briefly in the following with some aspects of speech perception.

Among the **AUDITORY THEORIES** of speech perception, the **Auditory Enhancement Theory** and the **Quantal Theory of Speech** are presumably the most well-known ones. The **Native Language Magnet (NLM)** model is not an auditory theory proper, but has a strong auditory bias, and is, hence, described with the auditory theories. The acoustic speech signal and auditory processes are the bases of these theories. What is perceived is the sound, or rather, the distinctive features. It is the acoustic signal itself which is perceived, and acoustic invariance is not emphasised. These theories do not, however, rule out motoric systems. They also consider speech perception not to be special or innate. (e.g., Hawkins, 1999b, pp. 240–241; M. S. Peltola, 2003, pp. 7–8) The **Auditory Enhancement Theory** (Diehl et al., 1990; e.g., Diehl & Kluender, 1989) argues that the articulatory patterns are determined by perceptual needs (i.e., perception drives production) and the phonological distinctions of a language are perceptually enhanced due to specific

[1] I.e. the invariance in acoustic speech signal which coincides with a phonological feature even though there is variance in the signal; and further, perception is sensitive to it.

[2] Passive processing refers to automatic processing where “inputs map directly to outputs with no hypothesis testing or information-contingent operations” (Heald & Nusbaum, 2014, p. 1). Whereas active processing refers to such processing where adjusting according to new information or uncertainty is possible; in the presence of new information, hypotheses are generated and tested accordingly (ibid.).

speech articulation control[3]. All the information that the listener needs is provided by the acoustic signal, and the acoustic properties are combined together from nearby syllables enhancing the perception and identification of distinctive features. (Hawkins, 1999b, pp. 241, 246–247) According to the theory, there is covariation in both speech production and the resulting acoustic signal[4] which is beneficial for the perceiver (Diehl & Kluender, 1989; Fowler, 2003). The **Quantal Theory of Speech** (Stevens, 1972, 1989) (is based on acoustic theory and is a theory on both speech production and perception) states that a specific vocal tract shape and the acoustic output do not have a linear relationship as a small change in the vocal tract may auditorily result in a greater scale change, and a large change in the vocal tract may result in no change in the auditory perception. According to the theory, sound category formation is supported by the nonlinear relationship between vocal tract shape and acoustic output. A category lies at the stable auditory or acoustic region. The articulation within a category may, thus, be quite flexible while the received sound remains unchanged. The auditory system is very sensitive to the invariant properties or patterns of each phonological distinctive feature. The theory is not considered to be flawless as there are many opposing views[5]. (Hawkins, 1999a, pp. 215–217, 231)

The **Native Language Magnet model** suggests that the perceptual distance between the best exemplar, the prototype[6], of a speech sound category and the near members of the same category is shrunken, while the perceptual distance is stretched near poor exemplars (e.g., Iverson & Kuhl, 1996; Kuhl, 1991; Kuhl et al., 1992; Kuhl & Iverson, 1995). Iverson and Kuhl (1996) emphasise that the perceptual magnet effect and the peaking discrimination sensitivity at a category boundary, categorical perception, are different processes. The non-prototype members of a category are perceived as more similar to the prototype than to other non-prototypes, regardless

[3] This enhancement creates redundant acoustic features which work together in a way that it is possible to identify a particular value of a distinctive feature, since the signal now carries more of that property, or one acoustic property is enhanced so strongly that it changes the perception of another property (Hawkins, 1999b, p. 241).

[4] E.g., there is covariation in vowel rounding and tongue backness as both lower the F2 and the result of that covariation is back vowels which are easily distinguishable from the (unrounded) front vowels. Hence, independent gestures working together result in acoustic signals maximally distinguishing phonemes (Diehl et al., 2004).

[5] Hawkins (1999a, pp. 226–227), for example, has pointed out that the theory is too narrow as it ignores some factors (consonants modelled only in a few contexts; no attention to prosody driven changes) which might be as important as the ones that are emphasised and also, the easiest explanation is provided to sounds that are rare in the world's languages.

[6] Kuhl's idea of the prototypes and their effects is based on the prototype effects presented by Rosch (1977, 1978).

of the same physical distance/difference, as an effect of the magnet (Grieser & Kuhl, 1989; e.g., Kuhl et al., 1992). The NLMs presupposition is that at birth and innately, babies perceive and are sensitive to all speech sounds, regardless of their mother tongue. Exposure to a specific language, however, alters the infants' phonetic perception by the age of six months (Kuhl et al., 1992) and language specific phoneme representations are formed before the age of 12 months (Cheour et al., 1998). The main modality of the NLM is considered to be neutral, although with a strong auditory bias (Hawkins, 1999b), and hence it is grouped with the auditory theories of speech perception. The native language effects on speech perception during infancy change the modality from auditory to polymodal (e.g., vision and internal feedback) so that perception influences production and vice versa (Hawkins, 1999b, pp. 249, 255). The unifying element of The Quantal Theory of Speech, the Auditory Enhancement Theory, and the Native Language Magnet model is the link between speech perception and non-species-specific auditory properties. Another unifying element is the auditory nature of speech perception, and hence, native language learning is not linguistic in nature, as language is acquired through auditory input. (M. S. Peltola, 2003, p. 8) Another similarity between The Quantal Theory of Speech and the NLM, is the idea of the category lying at a stable auditory or acoustic region in The Quantal Theory and the shrunken perceptual distance between the prototype and the near members. The expanded version of NLM is discussed in detail in chapter 2.2 Acquisition and language learning.

Categorical perception (CP) refers to the ability to divide the perceptual space into divergent categories, e.g., different colours. In speech sound perception the division into different speech categories is accomplished with phonologically relevant features. Speech sound discrimination sensitivity is high and reaction times (RT) short in between category perception, whereas sensitivity decreases and RTs increase in within category discrimination (Liberman et al., 1957). Also, according to CP, the peaks and troughs in the discrimination can be predicted from the identification result. The categorical perception has been shown to have a neural correlate as the mismatch negativity (MMN) response, which elicits to a change in a string of presented stimuli[7], is larger for the between category sound contrast than for the within category contrast (Sharma & Dorman, 1999), indicating a peak in the discrimination sensitivity at native language phoneme boundaries and lower sensitivity within good category exemplars. CP also describes well the nonlinearity between the vocal tract shape and the acoustic output presented by the Quantal Theory (Hawkins, 1999a, p. 216). While in CP the emphasis is on the category boundaries, perceptual magnet effect (PME) (e.g., Iverson & Kuhl, 2000; Kuhl,

7 See chapter 4.3.1 Tasks and measures for a more comprehensive description.

1991, 2000; Kuhl et al., 1992, 2008; Kuhl & Meltzoff, 1996) focuses on the category itself[8]. The prototype of the category, formed according to the most often heard exemplar of the category, determine the hierarchy within the category. The prototype, functioning as a magnet, holds the category together. When all the criteria of CP are not met and the within category discrimination is significantly above chance, however still below the level of discrimination at the boundary, the performance is said to refer to Phoneme Boundary Effect (PBE) (e.g., Strange, 1999). It has been suggested that there are perceptual differences between vowels and consonants (Rosen & Howell, 1987, p. 118), i.e., consonants tend to be perceived categorically (Liberman et al., 1967; Strange, 2007, p. 39), whereas vowels tend to be perceived continuously (Fry et al., 1962), or at least more continuously than consonants (Strange, 2007, p. 39). When, for example, comparing the perception of stops, fricatives, and vowels, it has been argued that stops are most strongly perceived categorically and vowels continuously. Whereas, the perception of fricatives is somewhat more continuous than that of stops. (Kronrod et al., 2012). Most probably, some of the discrepancies between the differences in the research on consonant and vowel perception in relation to CP can be explained by both task-related (e.g., Kronrod et al., 2012) and between-subject (e.g., Repp, 1981) differences. Further, vowels are possibly harder to perceive than consonants (Pereira Reyes, 2014, p. 30) as the acoustical properties of vowels are less stable than those of consonants (Liberman et al., 1967). In addition, the vowel space is probably more or less fully covered in any given language, even in languages with only a few vowel phonemes. Whereas the consonant space functions differently as gradual sliding from a consonant to another does not occur in a similar manner as in vowels, the consonants also have stricter boundaries than vowels.

The two main **MOTOR THEORIES** of speech perception are the **Motor Theory** and **Direct Realism**. According to the later version of the Motor Theory[9], the speakers' abstract intended gestures for phonetic categories are reconstructed by the listener, and there is a special, innate phonetic module in the brain for the automatic deconstruction process (Liberman & Mattingly, 1985, 1989). According to the Motor Theory, coarticulation is part of the execution of movement and, hence, it is not part of the abstract gesture of the "pure" phoneme (e.g., Diehl et al., 2004;

[8] There has been, however, discussion that the CP and PME co-occur (e.g., Kronrod et al., 2012; Tomaschek et al., 2011), but at the same time Kronrod and colleagues (2012) remind that the categorical effects are stronger in consonant perception compared to vowels.

[9] The first editions of the theory postulated that the articulatory movements of the vocal tract or the neural commands to the articulators were the basis for speech perception and that they were reconstructed during the decoding of the auditory patterns (Hawkins, 1999a, p. 200; Liberman et al., 1967; Nygaard & Pisoni, 1995, pp. 84–85).

Hawkins, 1999a, p. 200; Liberman & Mattingly, 1989). **Direct Realism** or the **Direct Realist Theory**, and the speech perception approach[10] of it (Fowler, 1986; Fowler & Rosenblum, 1991), is another motor theory of speech perception. Opposite to Motor Theory, speech was considered to be governed by the same laws as other aspects of perception, and not to be special. According to the Direct Realism objects and events are perceived directly. The theory calls the sensory information the *proximal stimulus*, which is the acoustic signal of speech and also the sensation (seeing and feeling[11]) of the moving vocal tract which actuate the perception. The actual physical environment is the *distal object* or *distal stimulus*, i.e., the gesture. What is experienced, is the perceptual object itself, the motoric gestures, not the representation of it. Invariance is possible both in the acoustic signal and in the object of perception, although, the acoustic invariance is not interesting in the light of the Direct Realism. (Hawkins, 1999b, pp. 233–235) Articulatory phonology (e.g., Browman & Goldstein, 1992), or gestural phonology (e.g., Best, 1995), describes gestures similarly as the Direct Realism, they are the place and degree of constriction produced by the vocal tract articulators. Formant frequencies are the proximal stimuli and the vocal-tract shape is the distal stimulus. (Hawkins, 1999b, p. 235) While the intended gestures are those perceived according to the Motor Theory, the object of perception according to the Direct Realism, are the speaker's actual gestures. No decoding and recoding is needed in Direct Realism, as is the case in the Motor Theory (e.g., Nygaard & Pisoni, 1995, p. 86). Both motor theories, hence, consider speech production mechanism as a crucial aspect of speech and that there is a sturdy link between speech perception and production. Coarticulation is explained in Direct Realism so that the overlapping gestures last longer than the acoustic segment they are associated with. The acoustic information is weak in the beginning and the end of the gesture when occurring simultaneously with another gesture and in the intervening part it is the main information of the gesture in question. The theory considers gestures as units of phonology and production as well, not only of perception, to avoid destruction and reconstruction of units. (Hawkins, 1999b, pp. 236, 238) In summary, the interrelationship between production and perception is emphasised in both motor theories. Peltola (2003, p. 9) recapitulates that, in relation to language acquisition, compared to the Motor Theory,

[10] The speech perception approach of the Direct Realism is based on the general theory of direct perception, or ecological psychology or ecological theory of perception by Gibson (1966, 1979).

[11] Hawkins (1999b, p. 233) sums up two main reasons for the development of the theory; on the one hand there was a need to explain speech perception similarly to other aspects of perception, vision in particular, and on the other hand, some of the premises of the Motor Theory needed to be remodelled.

the Direct Realism is closer to the auditory Quantal Theory, as it suggests that the received information is non-linguistic.

The **Fuzzy Logical Model of Perception** (FLMP) and the **TRACE Model** are considered to be among the **NEUTRAL THEORIES** of speech perception. The neutral theories, instead of describing the modality through which perception takes place, consider how phonetic or phonemic decisions are made. These theories do not assume invariance, rather, they assume potential salience in all information and variance is accepted. The theories are also referred to as *continuous-information models*; information is evaluated continuously to estimate the probability that a particular sound was uttered. (Hawkins, 1999b, p. 266; Massaro, 1994, pp. 245–247, 252–254)

In summary of the speech perception theories and models a few comparisons are discussed. The NLM is compared with the Quantal Theory and the Auditory Enhancement theory by Hawkins (1999b, pp. 254–255) so that there is a connection between NLM and the Quantal Theory in that both have emphasis on natural categories, exposure to a particular language, i.e., linguistic experience with perceived statistical distributions, alters the perceptual space so that new distinct categories can be perceived. The opposite is suggested by the Auditory Enhancement Theory, according to which the distinct categories form when speakers enhance certain types of acoustic properties and when those properties cause additional auditory distortions (due to general psychoacoustic processes; whereas in the NLM, the distortions are due to the linguistic experience). The common factor between NLM and the Auditory Enhancement Theory, however, is that more than one factor can modify the perceptual space. The Motor Theory and the Quantal Theory differ in the basic assumptions of the mechanisms and processes of speech perception, not so much in the outcome itself. The common factors in these theories are that they both try to account for categorical perception, in addition, both theories search for the invariants. However, the invariant is defined differently in these theories as according to the Motor Theory the invariants are the movements of speech production but the Quantal Theory considers them to be in the acoustic signal. (e.g., Hawkins, 1999a, p. 217)

Some twenty years ago Binder (2000) wrote that “our understanding of speech recognition processes has gradually advanced over the 50 years, from a state of almost total ignorance to one of well-informed confusion”. As the neural basis for speech perception and understanding is a complicated network consisting of many brain areas, Binder’s comment seems still valid. The most crucial areas involved in speech processing are located in the auditory cortex in the temporal lobe and its surroundings and in the lower parts of the frontal lobe, especially in the left hemisphere. However, many other areas of the cortex and subcortical structures are also involved, including areas in the right hemisphere. (e.g., Ross, 1984; Scott &

Johnsrude, 2003) The earliest significant theories of the speech areas in the brain where those of Paul Broca and Carl Wernicke. Broca's suggestion was that the back part of the inferior frontal gyrus in the left frontal lobe is connected to speech production. Wernicke, on the other hand, suggested that the back part of the superior temporal gyrus in the left temporal lobe is responsible for speech perception and the storage for word representations. (e.g., Zurif & Swinney, 1994, pp. 1055–1056) Later, as brain research methods have developed, knowledge of the speech areas in the brain have become more accurate. Functional resonance imaging (fMRI) and positron emission tomography (PET) measure changes in the metabolism and blood circulation in the brain, whereas, electroencephalography (EEG) and magnetoencephalography (MEG) measure the electric functioning and magnetic fields of the electric functioning, respectively. On the basis of a vast set of empirical findings, Hickok & Poeppel (2007, 2016) proposed a dual-stream model for speech processing, which was similar to that presented for the visual domain in the 1980s. The two routes in the model are the ventral and dorsal streams, the "sound to meaning" stream and the "sound to action" stream, respectively. The ventral stream is assumed to be bilaterally organised, although there are differences between the hemispheres, whereas the dorsal one is left-hemisphere dominant. Structures in the superior and middle portions of the temporal lobe are involved in the ventral route and structures in the posterior frontal lobe and the posterior planum temporale are involved in the dorsal route (for a thorough discussion, see Hickok & Poeppel, 2007, 2016).

The pre-attentive event-related brain potential (ERP) mismatch negativity (MMN)[12], measured with electroencephalography (EEG), has been greatly used in speech perception research. To mention a few examples, studies have shown the MMN to be sensitive to differences within vowel categories (Sharma & Dorman, 1998). On the other hand, the MMN has also been shown sensitivity towards a between consonant category contrast over a within category contrast (Sharma & Dorman, 1999). Further, new-borns learn to discriminate between speech sounds trained during sleep, as shown by increased MMN amplitudes (Cheour, Martynova, et al., 2002). Also, a speech sound change elicited MMN is equivalent to that of adults in 4 to 7 year-old children showing maturation of speech discrimination (Shafer et al., 2010). In addition, phoneme changes and rule violations are detected implicitly, i.e., explicit awareness and attention is not necessary for neural discrimination (Virtala et al., 2018). Research on speech perception has also shown (native) language processing effects on preconscious as well as on conscious level. Native language effects on pre-attentive speech perception were shown by Näätänen

[12] See chapter 4.3.1 Tasks and measures for a more comprehensive description.

and colleagues (1997)[13] and many other studies since (e.g., Brandmeyer et al., 2012)[14] which have also shown an agreement between behavioural and neurophysiological speech perception. In her division of the speech perception theories (Auditory, Motor, and Neural theories), Peltola (2003, pp. 7–11) discusses that, in the light of the neural theories, the "most fundamental point may be that the perception of speech sounds is – at least in the most natural form, i.e. in the case of mother tongue perception – preconscious." (ibid., pp. 11). In other words, the automatic process does not require any conscious effort in speech sound perception.

To conclude, speech perception theories and models have many obstacles to overcome and not one theory or model is capable to describe the entire process of speech perception or to solve all the related problems. The lack of invariance is not relevant in other theories (Direct Realism, FLMP, TRACE) while others acknowledge it (Motor Theory, Quantal Theory). The object of perception is considered to be the acoustic signal (Auditory Enhancement), or the intended or actual gesture (Motor Theory, Direct Realist, respectively), or the object is not the centre of the theory at all (FLMP, TRACE). The perspective of a theory may be the within (NLM) or across category (CP) perception. Further, speech perception may be approached, both theoretically and experimentally, from the behavioural or the (pre-attentive) neural level, or both.

13 Näätänen and colleagues showed language-specificity in their study testing pre-attentive speech perception in Finns and Estonians. They showed that compared to the standard stimulus /e/, Finns elicited prominent responses to the native vowels /ø/ and /o/. However, they did not elicit prominent response to the Estonian /õ/, even though the acoustic distance from /e/ to Estonian /õ/ was greater than to the Finnish /ø/. The Estonians elicited prominent responses to all three vowels, /ø/, /õ/, and /o/, belonging to their phoneme repertoire.

14 Brandmeyer and colleagues used syllables (/pa/-/ba/) with varying VOTs in their study to test native English and native Dutch (proficient in English) speakers' speech perception. Voiced and unvoiced labial stops are distinguished by aspiration in English whereas prevoicing distinguishes them in Dutch as shown by behavioural and psychophysiological results.

2.2 Acquisition and language learning

"Throughout the lifespan, perceivers continue to refine their perception of speech even for their native language."

Best & Tyler (2007, p. 24)

SPEECH ACQUISITION AND MODELS OF SECOND LANGUAGE SPEECH LEARNING

Speech acquisition starts before birth during the fetal period, and foetuses even recognise their mother's voice from other female voices (Kisilevsky et al., 2003). Further, exposure to non-native-like features of speech during the fetal period results in learning shown post-birth (Partanen et al., 2013). After birth, infants discriminate non-native phonemes (e.g., Best et al., 1988; Werker et al., 1981; Werker & Tees, 1999), however, native language phonemes are preferred over non-native ones (Moon et al., 2013). Native language speech is perceived through language-specific phoneme representations (Näätänen et al., 1997; Sharma & Dorman, 2000) and these representations evolve in early childhood before the age of twelve months (Cheour et al., 1998; Ortiz-Mantilla et al., 2016) when the phonetic perception alters to favouring native language phonetic prototypes (Kuhl et al., 1992; Werker & Tees, 1984).

For example Kuhl (2004; Kuhl et al., 2008) describes a developmental timeline for speech perception and production in typically developing infants during the first year of life. For the first eight months, speech is perceived universally, i.e., infants are able to differentiate all the speech sounds of the world; and, according to her, general auditory processing mechanisms are responsible for that, not a speech-specific mechanism. Language-specific speech perception develops during the last half of the first year of the infants' life, while sensory learning takes place during the whole first year. Infants learn rapidly from language exposure, which is crucial to language learning. They combine pattern detection and computational abilities[15] with social skills and at about six months their perception of the phonetic units of speech has changed to language-specific and they perceive vowels according to their native language. Later (Kuhl et al., 1992), at about eight months, they detect typical

[15] This is called *statistical learning* which refers to the ability to extract statistical regularities from the surrounding world to learn about it. The distributional patterns of sounds help in category formation as the amount of representatives heard is greater at and near the category centre and lesser at the category boundaries. The distributional patterns of syllables, on the other hand, help in learning to segment words, as the probabilities of adjacent syllables are different within and across word boundaries. (e.g., Kuhl, 2004)

stress patterns in words and are able to segment words (Saffran et al., 1996). Nine-month-old infants recognise language-specific sound combinations (Jusczyk et al., 1994) and at the age of eleven months foreign language consonant perception starts to decline (but remains above chance level) at the same time as the native language consonant perception sharpens (Kuhl et al., 2006; Rivera-Gaxiola, Silva-Pereyra, et al., 2005; Tsao et al., 2006). (Kuhl, 2004; Kuhl et al., 2008) The plasticity of the human brain is most evident within the infants as shown by the incredible speed in development during the first year of their life.

When an infant has acquired the ambient or native-language, a neural commitment to the native-language has taken place (Kuhl, 2004; Kuhl et al., 2008). According to the **native language neural commitment** (**NLNC**) (Kuhl, 2004), dedicated neural networks are formed for the acquired native language during the early stages of acquisition. The acquired native language neural network, necessary for native language processing, interfere with foreign language processing later in life, at least the processing of the mismatching patterns. The NLNC also hypothesises that infants whose phonetic perception is more precise advance faster in language acquisition. (e.g., Kuhl, 2000, 2004; Kuhl et al., 2008; Rivera-Gaxiola, Klarman, et al., 2005; Zhang et al., 2005, 2009) Better native speech perception skills at the age of seven months increase language development in infants during the age of 14 to 30 months (Kuhl et al., 2005). Further, better non-native skills in seven-month-old infants lead to slower language development during that age period (Kuhl et al., 2005). These behavioural results have been replicated and extended with psychophysiological measures (MMN) (Kuhl et al., 2008). Curiously, poor L2 (here, the second native language) perception in early, highly skilled bilingual adults (with equal language experience in both languages), correlates with discrimination abilities in both native and non-native phoneme discrimination, compared to good perceivers (Díaz et al., 2008). Nine-month-old infants are able to learn foreign language from natural exposure (native speakers of the target language read children's books and played with toys with the infants), however, exposure through television or audiotape does not lead to learning (Kuhl et al., 2003). Thus, social experience is of importance in infants learning foreign language phonetic aspects.

Kuhl broadened the NLM model, described in chapter 2.1, later according to the above-mentioned findings. The **NLM-expanded** (**NLM-e**) (Kuhl et al., 2008) includes social and cognitive aspects of language acquisition with five guiding principles. The first concerns the universal to language specific development of phonetic perception which is driven by both the surrounding **language setting** with a distributional set of phonetic units and **infant directed speech** with overemphasised acoustic information (which positively correlate with infants' speech discrimination skills (Liu et al., 2003)). The second principle describes the **NLNC**, the effect of **language exposure** (Kuhl, 2000, 2004), where language input physically changes the neural tissue making the neural networks committed to native language speech patterns. The third principle of the NLM-e consider **social aspects of early phonetic learning** and at the centre of this observation is the finding that social interaction enhances foreign language learning compared to learning through watching a recorded video (Kuhl et al., 2003). The fourth principle states that the **perception-production link** is developmental and builds on perceptual experience (Kuhl & Meltzoff, 1982, 1996). According to the final principle there is a **link between early phonetic perception and later language development**; good native phonetic perception predicts better language development later and good non-native perception predicts slower linguistic development (Kuhl et al., 2005, 2008). The NLM-e is finally described with four phases. In **phase one**, infants discriminate all phonetic units which, according to the model, facilitates development in the next phase, the core of the model. In the **second phase**, environmental input is relevant, and during this phase infants are sensitive to distributional patterns and overemphasised cues of infant directed speech which enhance phonetic learning. During the **third phase** word acquisition starts to develop as a result of improved detection of phonotactic patterns and word-like units, and segregation of phonetic detail. Fairly stable neural representations, which are not disturbed by new utterances, are formed by **phase four**. (Kuhl et al., 2008) Kuhl's models and descriptions of the developmental timeline do not, however, consider the pre-birth speech acquisition at all. According to her models, an infant is a *tabula rasa* that perceives all sounds universally although speech acquisition begins already during the fetal period.

The native language phoneme category hierarchy which is the key in the NLM is crucial also in second language learning. As sounds in the immediate vicinity of the category prototype are hard to discriminate from each other, difficulties in learning to perceive differences between foreign language phonemes are inevitable when the L2 category boundary is located near or at the native prototype.

Best & McRoberts (2003) present that there are two models which do not assume a universal pattern of developmental change as such to be the explanation for speech sound discrimination differences. One assumes a robust/fragile perceptual dimension along which non-native contrasts vary (Burnham, 1986). The fragile contrasts are low in acoustic salience, rare among the world's languages and if these contrasts are not present in the native language, discrimination decline within the first year. The acoustic salience is high in the robust contrasts and they are common in world's languages. Also, even without experience these are well discriminated until early school years. However, as Best & McRoberts (*ibid.*) point out, it is not always straightforward to define the fragility/robustness according to the acoustic salience level, not to mention together with rarity of the contrast[16]. The **Perceptual Assimilation Model** (**PAM**) (e.g., Best, 1994, 1995; Best & Strange, 1992) is another model predicting variations in perception. The basis of PAM is that listeners perceptually assimilate non-native sounds to the most similar native phonemes. The perceptual assimilation of non-native phones to native phonemes happens on the basis of the detection of resemblances in articulators and locations and/or degrees of constriction (Best et al., 2001). PAM describes three of these assimilation types: *Two-Category* (TC) assimilation, *Single-Category* (SC) assimilation, and *Category-Goodness difference* in assimilation (CG). In TC, a pair of non-native phones assimilate to two different categories separately, whereas in SC the phones assimilate to a single native category. They may assimilate either equally well or poorly to a single phoneme. The CG describes a situation where the non-native phones assimilate unequally to one native phoneme, the other is a better representative of that category than the other. PAM also describes two situations where non-native sounds do not assimilate to native phonemes: *Uncategorised-Categorised* (UC), where one phone is assimilated while the other one is not, and *Uncategorised-Uncategorised* (UU), where neither non-native phone is categorised to any native phonemes[17]. Further, PAM describes such a situation where neither non-native phone has any articulatory properties that fit to any native phonemes. These phones would be perceived as non-speech sounds and are described as *Non-Assimilable* (NA). (Best et al., 2001; Best & Tyler, 2007) In principle, it can be said that there are four assimilation types describing perceptual assimilation to the most similar native phones, as in the UC the other part of the contrast assimilates to a similar phone in the native language, and two types which describe no assimilation at all.

The original PAM describes assimilation patterns of a naïve listener of a foreign language but a later extension, **Perceptual Assimilation Model of Second Language Speech Learning** (**PAM-L2**) (Best & Tyler, 2007), describes perceptual assimilation patterns of a second language learner and predicts L2 development during the time of L2 immersion. As in PAM, both phonetic and phonological levels

16 For example, clicks are rare but the psychoacoustic properties of some of them are quite salient (palatal, alveolar, and lateral clicks) while others are less salient (dental clicks) and some are rather weak (bilabial clicks). For more comparisons of contrasts, see Best & McRoberts (2003).

17 Later, the uncategorised phones have been divided into three subgroups, namely focalised, clustered, and dispersed (Faris et al., 2016). Focalised are predominantly similar to a single L1 category, but the categorisation threshold is not achieved. Clustered are similar to more than two L1 categories, and dispersed are not similar to any L1 categories.

of correspondence are engaged in PAM-L2[18]. During the initial stage of the learning process, according to PAM-L2, the L2 phones will be assimilated into or dissimilated from the existing native language categories on phonetic level (Antoniou et al., 2012). Later, when the learner's vocabulary size increases, the perceptual system readjusts to the L2 phonology and new categories corresponding (closely) to L2 categories form, the phones are discriminated according to the L2 phonology (Antoniou et al., 2012; Bundgaard-Nielsen et al., 2011). The model describes four cases which predict how successful L2 perceptual learning may be. In the first case one L2 contrast member is perceptually assimilated to a given native or first language (L1) category and perceptual learning is not needed. Contrasts with other L2 categories would in this case be TC or UC assimilations. This situation is predicted to cause little or no difficulty in learning. In this scenario, L1 and L2 categories are perceived both phonetically and phonologically as equivalent. Or, in the case described in footnote 18, the L2 category is perceived phonetically as deviant, but phonologically as equivalent to the L1 category. In the second case both L2 categories are assimilated to one L1 category, but one better than the other, as CG in PAM. It is predicted that the learner discriminates these phones well, but not as well as in TC, and comes across greater perception difficulties than in the former situation. Further, a new L2 category (phonetic or phonemic) will most probably be formed for the deviant L2 phone, but the better perceived representative will be perceived (phonetically and phonologically) as equal to the L1 category and no new category is formed. The third case is same as PAM's SC and here, the learner has the greatest problems in discrimination at first. A new phonetic category, at least for one of the L2 phones, have to be learned perceptually in order to form a new phonological category or categories. The fourth case describes a situation where the learner does not assimilate the L2 contrastive pair to any one of the L1 categories, but they rather have similarities to several L1 categories, which is the Uncategorised in PAM. The perceptual learning of one or two new L2 phonological categories will be quite easy. For a more thorough speculation of possible learning problems in different stages of L2 learning, see Best & Tyler (2007). The PAM prediction for the discrimination levels for the different assimilation types is gradient as follows: TC>CG>SC (Best et al., 2001). Most of the research on PAM or PAM-L2 have been conducted on consonants, but the principles of the model have been shown to be useful in non-native vowel perception as well (Tyler et al., 2014).

18 Phonetically dissimilar sounds may be considered phonemically equivalent as Best and Tyler (2007) gives an example of an English L2 learner of French perceiving the French voiceless uvular fricative /r/ or [ʁ] and English liquid /r/ or [ɹ] as phonologically equivalent.

Another second language learning model concentrates on similarities between native language phonemes and foreign language speech sounds, namely the **Speech Learning Model (SLM)** (Flege, 1987a, 1988, 1992, 1995a). SLM describes the learnability of second language phonetic segments; not all segments are equally learnable. Further, the model concerns both perception and production and it states that production depends on how well the same phone is perceived[19]. SLM describes the relationship of two sound systems as follows. First, there are phones in the second language which do not match any L1 phonemes. In SLM terms, these sounds are **New**[20]. Learning of the New L2 phones is difficult at first, but L2 experience eventually helps perception. Second, L1 and L2 sound categories may be identical, or highly similar, in which case the L2 category is **Identical** in comparison to the L1 category. Learning Identical categories does obviously not cause problems. Finally, the target language categories which cause most severe learning problems are described as **Similar**, a category that differs only slightly from an L1 category.

The more recent version of the SLM describes the interaction of L1 and L2 phonetic systems as **phonetic category assimilation** and **phonetic category dissimilation** (Flege, 2007). The phonological space of a learner is shared by the L1 and L2 subsystems, and hence, they affect each other. Category formation for L2, whether a category has formed or not, is the core of this view. The situation where the learner has not been able to form a new L2 phonetic category, which would also sound different from the L1 sound, is called the phonetic category assimilation. In this case the L2 sound is perceptually either incorporated or too similar to an L1 sound, or both. Non-nativelike pronunciation of both L1 and L2 is hence possible. The phonetic category dissimilation describes a situation where the learner establishes a new L2 category. In these cases, the learner might produce either an L1 or L2 sound differently enough from the other, even exaggerate to make a difference. The category assimilation equals to both Identical and Similar in the earlier version of the model, and the category dissimilation corresponds to New in the earlier classification.

19 However, not all errors in second language production are due to perception errors according to SLM. Flege (1995a) gives an example of native Spanish learners of English pronouncing ‘school’ as [ɛskul]. The production error is not perceptually motivated but may be caused by the permissible syllable types in the L1.

20 Best & Tyler (2007) point out that, even though this might seem similar to Uncategorised in PAM-L2, there are important differences. In PAM-L2 it is more than the similarity or dissimilarity, it is the dynamics within the interlanguage phonological system. Since there may be several L1 phones that the L2 pair is similar to, the L2 phones may have similarities to different L1 phones or to the same set of L1 phones. In the former case, differences are easily recognised and categories are easily learned perceptually. Whereas in the latter case, discrimination might be difficult.

The state of research for example on Critical Period Hypothesis (CPH), Contrastive Analysis Hypothesis (CAH) and CP as well focus on abstract linguistic items, not phonetic components, and the L1 interference on L2 have affected the early years of the development of the SLM (Flege, 2005, slide 7). The model has a few viewpoints: The phonetic properties of L2 sounds can be accurately perceived by L2 learners if they have had enough proper input. Similarly to L1 development, learning of L2 speech takes time and the characteristics of input have effects on it. And again, similarly to L1 development, perception precedes production (Flege, 1993, 2005, slide 92, 1999; Flege et al., 1999) Flege (2005, slide 93) also states that the SLM proposes that: "The processes and mechanisms that guide successful L1 speech acquisition – including the ability to form new phonetic categories – remain intact and accessible across the life span". And the phonetic elements of L1 and L2 phonetic subsystems are located in a shared phonological space and they affect each other.

The most recent modification of the SLM is the Revised Speech Learning Model (SLM-r) (Flege, 2021; Flege et al., 2021; Flege & Bohn, 2021). The revised model contains some unchanged aspects, some aspects are clarified, and some are new. It explains "how phonetic systems reorganize over the life-span in response to the phonetic input received during naturalistic L2 learning" (Flege & Bohn, 2021, p. 23). One of the unchanged aspects is the assumption that the same mechanisms and processes in L2 learning are used as in L1 learning, regardless of the learners' age. The end result is not, however, the same as in L1 acquisition. One of the clarifications of the revision is that length of residence (LOR), by itself, is not valid when assessing the L2 experience, it is rather LOR and the percentage of L2 use together that need to be taken into account. Hence, LOR is replaced by Full Time Equivalent (FTE) which is a better estimate of L2 input, but not entirely unproblematic as self-reports give estimations, not exact measures of the quantity and quality of L2 input. One of the new aspects is the interest in L2 speech development, instead of the ultimate attainment, as there is no end state for speech development (in L1 or L2). Further, according to the SLM-r, perception and production coevolve (without precedence), as there is evidence for the existence of solid bidirectional interaction between L2 production and perception. Another new aspect is, that instead of focusing on between-group differences, the focus is on how individuals learn L2 sounds, i.e. the attempt is to explain within-subject differences.

Escudero's **Second Language Linguistic Perception** (**L2LP**) model (Colantoni et al., 2015, p. 44; Escudero, 2005, 2009; Escudero & Boersma, 2004)[21] describes, explains, and predicts the sound perception of the second language learners with and

[21] The model basis on the Stochastic Optimality Theory by Boersma (1998) (Escudero, 2009, p. 155), which is out of the scope of this thesis, and hence not introduced here.

at different proficiency levels, and explains individual variation. The L2LP "predicts a developmental sequence for each naïve listener and then tests whether the predicted development is indeed observed in beginner, intermediate, and advanced L2 learners with a shared L1" (Colantoni et al., 2015, p. 44). The L2LP also considers target language (TL) contrasts as PAM, not individual segments, and employs similar terminology with SLM, as the TL contrasts are described as **new contrasts** or **similar contrasts**.[22] This model concentrates on L2 vowels as previous research has shown that there is variation in the production of vowels the learners perceive from the L1 speakers and there is variation in the production of the L2 learners' production (Colantoni et al., 2015, pp. 44–45). The native listeners are, according to the L2LP, *optimal perceivers*, and they are equipped with a *perceptual grammar*. The perceptual grammar system weights and parses the acoustic values of the input, the outcome being the perceptual representations, the allophones or phonemes.

The perceptual mappings (the end result of the perception grammar analyses of the auditory input) and the abstract and discrete sound category representations are distinct according to the L2LP[23] (Colantoni et al., 2015, p. 51). The model consists of abstract representations of both phonology and lexicon and two grammars. The perception grammar "map[s] the acoustic signal to phonological representations" and the recognition grammar "map[s] the phonological categories onto lexical representations" (Colantoni et al., 2015, p. 51). With these measures the L2 learners' difficulties and development can be described and explained. The development in L2 perception is succeeded when learning problems are solved. According to this model there are two problems that need solving. First, the perceptual problem, concerns the L1 perception grammar, which needs to be changed or adjusted to fit better with the TL input, and second, the representational problem, which involves possibly **NEW** (SC in PAM) TL categories (Escudero, 2009). NEW categories are not created when the TL sounds are similar to the L1 lexical categories, as they can be reused, however, perception grammar needs to be adjusted. (Colantoni et al., 2015) L2 **SIMILAR** (TC in PAM) sounds are phonologically equal to but phonetically different from acoustically most similar L1 sounds. In this case, the model predicts that two L2 phonemes are equated with two L1 phonemes in lexical storing (lexical equation is related to perception since the auditory properties of L2 sound tokens are similar to those of the L1 sounds). However, in this case, there are also phonetic realisations which do not match the L1 phonemes. (e.g., Escudero, 2009) NEW and SIMILAR sounds are described as follows in the L2LP model: NEW sounds scenario

22 Even though there are similarities with PAM, articulatory gestures are not the core in the L2LP. Rather, it uses acoustic information as SLM.

23 Whereas in the SLM and PAM the predictions are made according to the phonetic and phonological categories.

is that where two L2 sounds equate to one L1 sound. SIMILAR sounds scenario, on the other hand, is that where two L2 sounds equate to two L1 sounds. In the SIMILAR scenario the challenge facing the learner is the need to adjust perceptual mappings and category boundaries. Whereas in the NEW scenario, the challenge is greater, since the learner needs to create new perceptual mappings and categories; i.e., the learner needs to learn to create new categories and shift L1 boundaries in attaining optimal perception, or split an existing single L1 category (van Leussen & Escudero, 2015). In examining L2 perception in the viewpoint of the L2LP model, the first step is to lay out the optimal perception of the languages in question. (Escudero, 2009)[24]

Two considerably earlier views on differences between two sound systems still needs to be introduced to show the similarities compared to the newer models and to introduce the basis on which the newer models evolved. These views are presented by Weinreich (1953/1963) and Wiik (1965). Weinreich (1953/1963, pp. 18–19) suggests four difference types between two languages which cause learning problems. The first, **under-differentiation of phonemes**, is the case where two TL phonemes are not distinguished in the NL. Severe learning problems are caused by this difference type. **Over-differentiation of phonemes** is the second difference which describes an opposite type. Two NL phonemes belong to one phoneme in the TL, which does not cause severe learning problems. The third difference is called **reinterpretation of distinctions** where the TL contrast is differentiated by features relevant to the NL, but redundant in the TL. The final type is **phone substitution** which describes a situation where phonemes in the two languages are similarly described but differently pronounced. Wiik's (1965, pp. 15–30) description also contains four types of differences: physical, relational, distributional, and segmental differences. **Physical differences** describe a situation where one language contains a sound which does not occur in the other language. When the same physical sounds occur in the NL and the TL, but differ in classification, and thus belong to different phonemes, the difference is called **relational**. **Distributional differences** are those where both languages make use of the same two phonemes which occur in different environments. The last type is **segmental differences** describing the situation where

24 L2LP also describes a *SUBSET* scenario, comparable to UC or UU (although in their comparison van Leussen & Escudero (2015) call them uncategorised or categorised-uncategorised I think it is safe to refer to the UC and UU assimilation patterns of PAM). The SUBSET scenario, or multiple category assimilation, in L2LP illustrates a situation where a single L2 sound is perceived as two or more native language categories. This causes fewer problems than the NEW scenario since there is no need to create a new contrast in L2 according to L2LP, and PAM predicts only little difficulties in discrimination. (Escudero, 2005, p. 123,125; van Leussen & Escudero, 2015) The model has also been revised in van Leussen and Escudero (van Leussen & Escudero, 2015) as the earlier version was not adequately explaining a scenario of the L1 having more categories than the L2 in a given area or continuum.

phonetically similar set of segments occur in both languages but they are divided into a different amount of phonemic segments.

Not all early views on language comparison were as successful as others, as for example the strong version of the CAH assumes that differences between NL and TL always equals learning problems and that greatest difficulty occurs when there is greatest difference between the languages (Lado, 1957). A later, and weaker, version of CAH does not acknowledge NL interference as such. It does not predict problems, it explains the actual observed interference (e.g., Oller & Ziahosseiny, 1970; Wardhaugh, 1970)[25].

The models on second language learning or perception (or production) all share at least one characteristic, namely that they explain the problems a second language learner might encounter. Another connective factor is the supposition that NL affects the discrimination of non-native speech contrasts; how it affects is a difference between the models. The perspectives and the bases of the models differ vastly. PAM and PAM-L2 have their roots in the ecological direct-realist premise (e.g., Best & Tyler, 2007, p. 22), not on CAH, and neither does SLM, as Best and Tyler (2007) point out. As mentioned earlier, SLM was founded on the state of research obtained on various perspectives during the early state of Flege's research and the L2LP bases on the Stochastic Optimality Theory. Van Leussen and Escudero (2015) compare L2LP with PAM and SLM: In PAM and L2LP, the SC or NEW scenario, respectively, is more problematic than the TC or SIMILAR scenario. And while SLM and L2LP use the same terminology, the terms refer to different phenomena, and perhaps even more importantly, the comparison of SLM and L2LP is problematic, since SLM describes single categories while L2LP describes sound contrasts. As Tyler and colleagues (2014) point out, both PAM-L2 and L2LP acknowledge and evaluate individual differences in L2 categorisation and assimilation. Further, SLM compares individual TL and L1 categories, whereas, for example, PAM compares contrasts of TL to contrasts of L1. Although, Flege (2003) mentions that two TL allophones corresponding allophones of one L1 category are more difficult to learn than two TL allophones resembling two L1 allophones of two different phonemes. Peltola (2003, p. 30) points out that while NLM and PAM emphasise the difference between the native and foreign language category hierarchies, SLM, as a more simple model, fails to do that. Yet another connecting characteristics of these models is that they all acknowledge that plastic changes are possible, and even probable, in the L2 learners.

[25] Both use the terms strong and weak, but especially the term weak refers to a slightly different description, however, the main point in both is that the later version of CAH is not as strong as the former.

STUDIES ON SECOND LANGUAGE PERCEPTION AND LEARNING Perceptual Assimilation Model, as well as the articulatory organ hypothesis[26], state that primarily infants perceive the articulatory information, i.e., infants perceive the articulatory gestures, including the visual and proprioceptive information (Best & McRoberts, 2003). According to their view, this is necessary for the infants to become speakers of their native language, not only native listeners. This is also the case in second language learning, perception is learned first in order to allow for the development of production according to the target language (e.g., Flege, 1993, 1999; Flege et al., 1999). Curiously, Baese-Berk & Samuel (2016) have shown that mere perception training (discrimination) leads to better perception learning results than simultaneous production with the perception training. Further, producing the same sounds as in the discrimination is even worse than producing differing sounds. The extra load during training may, however, be one explanation for the reduced learning for those producing during training.

It becomes evident that after the native language has been acquired it affects the acquisition or learning of non-native phonemes, both quality (e.g., Iverson et al., 2003) and quantity (e.g., Meister & Meister, 2011). For that reason, foreign language speech sound perception learning has been studied quite thoroughly both in children and adults. The next paragraph will discuss perception learning results obtained with varying methods and research questions on different age groups and different language backgrounds.

Neural plasticity and functioning are the fundamental bases for speech perception learning. The research methods measuring the organisation and functioning of the brain are excellent tools for studying foreign language speech learning. Many foreign language speech learning studies have, for example, used EEG or MEG measuring MMN. It has been shown that foreign language speech sound learning in a classroom environment does not necessarily lead to native-like perception in children (Jost et al., 2015) or in adults (Grimaldi et al., 2014; Hisagi et al., 2016; M. S. Peltola et al., 2003). On the other hand, learning in a natural environment results in plastic changes and in native-like speech sound processing in immersion program children (Cheour, Shestakova, et al., 2002; M. S. Peltola et al.,

[26] According to the hypothesis, infants perceive the primary articulatory organs (e.g., lips, larynx) which produce speech, not so much the details of the gesture (e.g., speed, location). Hence, discrimination is more difficult when the phonetic contrast is accomplished by two different gestures by the same primary articulator (i.e., within-organ contrasts), than when there is one gesture by different articulators (i.e., between-organ contrasts). (Best & McRoberts, 2003)

2005; Shestakova et al., 2003)[27] and in adult immigrants (Winkler et al., 1999). Immersion learning is more implicit in nature than explicit classroom learning and may hence result in greater learning results[28] (certainly, there is a difference in age of exposure). The non-native teaching and far less input in a classroom learning environment compared to a natural learning environment may also account for the lesser learning results in a classroom. Learning in a natural environment improves behavioural perception and production learning in immigrant children as well, and the longer the residence in a natural environment the better the learning results are (Tsukada et al., 2005); children also outperform adults. Immersion learning, compared to classroom learning, in children has more positive effects on foreign language production learning as well (Immonen & Peltola, 2018). Also, explicit pronunciation teaching for advanced level university students majoring in the target language is beneficial (M. S. Peltola et al., 2014) as participating and passing a pronunciation course leads to native-like target language pronunciation while students not yet attending the course differ from the other students and the native speakers. The study environment at a university may, however, be less explicit in nature than classroom learning at school, as there are native speaker lecturers and target language is used much more than at school. Furthermore, the amount of native language usage while learning a second language in a natural language environment, affects the level of foreign accent; greater use of native language results in greater accent and vice versa (Flege et al., 1997). Changes, showing plasticity, through foreign language learning are hence seen in pre-attentive and attentive levels of perception and in production. Taken into consideration the massive amount of acquisition occurring, for example, during the first year of a child's life, it is not that surprising that second or foreign language acquisition or leaning later in life takes time and may not be as smooth as during infancy. The plasticity of an infants' brain is noticeably greater than in an older child or adult, and even the natural second/foreign language learning environment, not to mention school environment, is not the same as that for the early development.

27 Peltola and colleagues (2007) have, however, shown opposing results on immersion program children as well. The contradiction is explained by difference in the stimuli, as in the 2005 study they used near category boundary stimuli and in the 2007 study the stimuli were prototypical category members. Hence, the results of these two studies suggest that, "category boundaries may be enhanced earlier, but that the native-like hierarchical categories with prototypical representatives may be formed later, if at all" (M. S. Peltola et al., 2007, p. 21).

28 Even the immersion environment does not guarantee straightforward language acquisition. Rinker and golleagues (2010) showed a reduced MMN response to German vowel contrast compared to a Turkish one and to German monolinguals in 5-6 year-old Turkish-German children. The children were born and still living in Germany, had more home exposure in Turkish, and attended to German kindergarten.

It needs to be mentioned that L2 learning affects L1 as well (e.g., Kartushina, Frauenfelder, et al., 2016), although this backward influence is not studied nearly as vastly as L1 influence on L2. The effects of L2 on L1 have been shown both in phonetic perception (e.g., Mora & Nadeu, 2012) and production (Chang, 2012; Lev-Ari & Peperkamp, 2013; Mora & Nadeu, 2012; Yeni-Komshian et al., 2000). Also training[29] studies have shown foreign language affecting the production of native language (Kartushina, Hervais-Adelman, et al., 2016). It has to be mentioned, however, that as in any other type of research, not all studies have shown L2 effects on L1. This is caused by reasons similar to any L2 studies: Differing methods, participant groups, stimuli etc. The influence of L2 on L1 may be positive, negative, or neutral.

To sum up, the native language acquisition begins already during the fetal period, and during the first year of their lives, infants very rapidly acquire a large proportion, for example, of the native language speech perception and phonology. After the native language has been acquired, foreign or second language learning becomes more difficult. With varying perspectives, the models of second language learning predict learning problems on the basis of the native and target language phonological systems for foreign or second language learners. There is a vast amount of foreign language learning research with varying methods and tasks, research questions, and viewpoints, and with varying results resulting from differences in them.

2.3 Bilingualism and speech perception

Second language acquisition or learning ultimately leads to bilingualism. The definition for different bilinguals, however, varies according to many factors. In this chapter, some relevant, perhaps even historically relevant, terminology connected to bilingualism is described. However, even though the terminology might be of significance, the main purpose is to consider the different definitions and backgrounds of bilinguals relevant especially for experimental purposes in general and in relation to the present thesis. Further, in this chapter, research on bilingualism is mostly examined in studies concerning speech perception.

The definition of *bilingualism* is quite ambiguous since it refers to people who have acquired their two languages from birth to second language learners, or people who use two languages regularly. The age and manner of acquiring or learning both of the languages or the second language need to be considered when defining different types of bilingualism. The definition also varies, for example, according to fluency, competence, and proficiency, which are further affected by both age and

[29] Language learning by training and training studies will be discussed in chapter 2.4 Training and language learning.

manner of learning. According to, e.g., Francis (2005, p. 253) and Romaine (1995, p. 79), Weinreich's (1963) description of *compound* and *coordinate* bilinguals divided the different types of bilinguals according to phonological and semantic representations so that one meaning has two phonological representations in the compounds, whereas in the coordinates two meanings have their own phonological representations[30]. Later, as Francis (2005, p. 253) and Romaine (1995, p. 79) also point out, Ervin and Osgood (1954) redefined these terms to outline how the languages are learned; compound bilinguals have learned their languages at the same time in a common context, whereas coordinate bilinguals have learned them in different contexts consecutively. Furthermore, the definition of this term pair seems to be contradictive since some scientists define them the other way round as for example Diller (1970) and Kroll and Tokowicz (2005) point out and which is partly seen in Albert and Obler (1978) and Romaine (1995) as well. The definition of this term pair is perhaps out-dated, but the original thought behind it is somewhat interesting. As a synthesis of these definitions and according to Albert and Obler (1978, pp.3, 5), Romaine (1995, pp. 78–79), and Wei (2000, p. 6), it could be outlined, that the manner of learning determines, for instance, whether the bilingual is a *compound* or *coordinate* bilingual. The compound bilinguals learn the languages in the same context and use them concurrently and the languages are stored in one system and they are interdependent. This is the case at least for bilinguals from birth. The coordinate[31] bilinguals, on the other hand, learn the languages in different environments and they are stored and kept separate, and the languages are independent. This would be the case, for example, if the second language is learned at school. Between these two extremes, compound and coordinate, in a continuum is a s*ubordinate* bilingual who processes the second language via the stronger language. (Albert & Obler, 1978, pp. 3, 5; Romaine, 1995, pp. 78–79; Wei, 2000, p. 6)

Bilinguals who master both languages equivalently, i.e., have similar proficiency in both, are *balanced* bilinguals, whereas *dominant* bilinguals master one language better, which they also use more frequently (Albert & Obler, 1978, p. 5; Wei, 2000, p. 6). One might ask whether any truly balanced bilinguals exist and whether it can even be verified in any way. S*imultaneous* bilinguals acquire both languages simultaneously from birth, whereas *sequential* bilinguals acquire the second

30 For example, if the two languages were English and French, there would be one meaning for the compounds: book=livre, but two phonological representations: /buk/ and /livr/ and two meanings for the coordinates: book and livre, and two phonological representations: /buk/ and /livr/ (Romaine, 1995, p. 79).

31 Albert and Obler (1978), contrary to most other sources, state that one-parent, one-language situation results in coordinate bilingualism and learning in school through a translation method results in compound bilingualism.

language later, as a child or an adult (Escudero, 2012, p. 407; Wei, 2000, p. 7). Hence, sequential bilinguals are, for example, language learners who learn the second language at school or as immigrants in a natural setting. Compared to, for example, sequential bilinguals, simultaneous bilinguals are of course considerably less numerous as it has been estimated that simultaneous bilinguals consist less than 20 per cent of bilingual children (e.g., Grosjean, 2010, p. 178).

Bilinguals have also been defined as being two monolinguals in one person, which is the *monolingual* or *fractional view of bilingualism*. On the other hand, according to the *bilingual* or *holistic view*, the interaction of the two languages creates "a unique and specific linguistic configuration" (Grosjean, 1989, p. 6) which does not mean that a bilingual consists of two monolinguals. In fact, bilinguals' speech perception and production are not directly proportional to those of monolinguals' (e.g., Guion, 2003; Molnar et al., 2014; Sundara, Polka, & Baum, 2006; Sundara, Polka, & Genesee, 2006).

To summarise the bilingual terminology, bilinguals who acquire both languages from birth are usually fluent, competent, and highly proficient in understanding, speaking, reading, and writing both of the languages. These bilinguals may be defined as compound and simultaneous, and in some cases also balanced[32], if they end up using both languages more or less equally, but dominant, if they, for instance, in day care and school and at their studies use only one of the languages and do not use the other language that often. Language learners, on the other hand, may be defined as subordinate in the beginning of learning and coordinate at the later stages of learning, they may also be defined as dominant or sequential bilinguals. These bilinguals, termed learners, may be children who learn the second language at school or as immigrants from natives and/or at school. On the other hand, they may be adults who learn the language at language courses or as immigrants from natives, from their own children, or also at language courses. The fluency, competence, and proficiency level in understanding, speaking, reading, and writing the second language depends, for example, on the age of acquisition (AOA) and the length of residence (LOR) in the new country for the immigrants. Also, they may not be equally fluent, competent, and proficient in every modality. Naturally, such definitions as age and manner of

[32] As noted earlier, the term balanced refers to someone who has native language proficiency in both languages. It does not necessarily mean exactly equal proficiency. Further, bilinguals from birth might not be equally proficient in both languages and, for example, the school language may be the dominant one.

acquisition/learning are stable, but the balance and dominance of the languages, as well as fluency, competence, and proficiency, may change.[33]

Defining bilinguals only according to the extremes may be misleading, or even fallacious. Bilingualism defined as a continuum is more realistic and suitable in most cases. Particularly a continuum where the bilingual moves in regard to time or circumstance or context could be appropriate.

As far as the participants in an experimental study, and also in this thesis, are concerned, the most important aspects are how and when the languages are learned/acquired, what is the language history of the participants until the testing situation, and what are the fluency, competence, and proficiency levels at the time of testing.[34] In other words, the terminology is of secondary importance.

Another continuum where bilinguals find themselves, is the language mode continuum where the language activation level changes according to the interlocutors, the conversation topic, the setting, the reasons for the exchange of words, and so forth (Grosjean, 1997). The other end of this continuum is the monolingual language mode, the other is the bilingual language mode. In the monolingual mode one of the two languages is active and the other one is deactivated (as much as possible) and in the bilingual mode one language is active and the other one is partly active (ibid.).

Indeed, the background information of the bilingual participants need to be described in detail, the terminology describing the bilinguals is of secondary relevance, as seen in the research findings on bilinguals discussed below. Studies have concentrated on different kinds of bilinguals and have used varying methods, which inevitably lead to various and even contradictory results. The two languages in early, fluent bilinguals, who have been exposed to both languages before the age of 6 years, activate common brain areas when processing written language (Chee et al., 1999). Common areas are represented for both languages also in picture naming when languages are acquired before the age of 5 (Hernandez et al., 2001). Also the two languages of a bit later bilinguals, who have acquired the second language after the age of 5 years, activate identical neural areas during phonological and semantic tasks (Klein et al., 1995). Less fluent, late bilinguals, who have learned the second

33 For example in Wei's (2000) list of a varying bilingual terms balanced bilinguals are also called ambilingual, equilingual, and symmetrical bilingual. Both consecutive and successive bilinguals are those who have acquired or learned the second language after the first language has begun to developed, which is comparable to sequential bilingual. On the other hand, early bilinguals are also called ascribed bilinguals and late bilinguals are also called achieved bilinguals. This is just a small example of the vast set of terminology describing the different kinds of bilinguals.

34 Grosjean (1997) details important background information to be taken into account when testing bilinguals.

language at school after the age of 7 years, show separate activations for the two languages (Perani et al., 1996). Whereas fluent bilinguals, early (second language acquired before the age of 4 years) or late (second language acquired after the age of 10 years), show similarly distributed neural activations for both languages (Perani et al., 1998) in story listening tasks. Further, the fluent, early or late, bilinguals' L2 activate the same areas as did the L1 in the less fluent bilinguals. Perani and colleagues (1998) hence suggested that proficiency, rather than the age of acquisition, is an important determinant of the cortical representation of the L2. Kim and colleagues (1997) showed that early bilinguals, who were exposed to both languages in infancy, showed representations of both their languages in common frontal cortical areas, whereas late bilinguals, who were exposed to the second language in early adulthood, showed separated representations of their two languages within the frontal lobe language sensitive regions. However, during the silently performed sentence-generation, these different types of bilinguals showed no separation within the temporal lobe language sensitive regions. It has also been shown, that bilinguals have separate linguistic systems, with a switching mechanism[35], on which they process their languages (Garbin et al., 2011; Nakamura et al., 2010)[36]. From another perspective, non-linguistic processing (complex tones) is similar in early bilinguals and monolinguals (Nenonen et al., 2003; Ortiz-Mantilla et al., 2010) [37].

To summarise, the above studies suggest that common brain areas are activated when the background of the bilingual participant is "early acquisition and fluent". This was shown in phonological, semantic (Klein et al., 1995) and written language processing (Chee et al., 1999), as well as in language listening (Perani et al., 1998) and silent sentence-generation (Kim et al., 1997).[38] On the other hand, when the background of the bilingual is "late exposure", there are separate representations for the two languages. This was shown during listening (Perani et al., 1996) and silent

35 The change in the activation caused by the change in the processing language is not, however, necessarily a switching mechanism but simply a reaction to the language change which demands additional processing capacity.

36 In the Garbin and colleagues' (2011) study, a speech production (picture naming) task was used to test early, high-proficient bilinguals. The Nakamura and colleagues' (2010) study used a word recognition (lexicosemantic processing) task and tested little later and relatively proficient bilinguals.

37 See Abutalebi (2008) for a more comprehensive review on the neural aspects on bilingualism.

38 In speech production, distinct language-specific neural populations seem to code the two languages, even within common language processing areas, in highly balanced early simultaneous bilinguals (Hämäläinen et al., 2018). Both languages, Finnish and Swedish, were learned before the age of five in this study.

sentence-generation (Kim et al., 1997)[39]. Further, Perani and colleagues' (1998) suggestion that proficiency is of crucial importance, rather than the age of acquisition alone, seems to be similar to the SLM-r (Flege & Bohn, 2021) proposition that LOR alone is not sufficient, but the percentage of L2 use should be acknowledged.

As was stated earlier in this chapter, a bilingual is not two monolinguals in one person or brain and, hence, bilingual speech perception is not necessarily identical to the processing of monolingual language users. For example, simultaneous bilinguals, 2–3-year-old toddlers (Kuipers & Thierry, 2015) and adults (Kuipers & Thierry, 2010), seem to have enhanced neural attention to speech compared to monolingual peers, when presented with picture primes with semantically matched or not matched spoken words. Semantic integration is, however, similar in simultaneous bilinguals and monolinguals (Kuipers & Thierry, 2010, 2015). Also, the morphology of the subcortical brain structures, which are involved in language processing, is expanded in simultaneous bilinguals compared to matched monolinguals (Burgaleta et al., 2016)[40]. It was suggested that this may be the impact of great simultaneous usage of the two languages since birth.[41] Further, monolinguals and simultaneous bilinguals seem to have similar cortical thickness, but later L2 acquisition – both in early and late bilinguals – leads to "thicker cortex in the left inferior frontal gyrus and thinner cortex in the right inferior frontal gyrus" (Klein et al., 2014). In addition, it has been shown that the pre-attentive speech perception pattern of simultaneous bilinguals differs from that of monolingual peers of both languages (Molnar et al., 2014)[42]. The bilinguals are able to detect sub-phonemic differences in both of their languages, which are not easily detected by the monolingual speakers. The researchers suggest that this helps the simultaneous bilinguals' perception in different language contexts. In addition, they suggest that

[39] Golestani (2016), focusing on fMRI studies, reviews research on phonetic learning and phonetic processing in bilinguals and sums up that overlapping brain regions are involved in phonetic processing of the first and second languages in the bilinguals. In addition, at the different stages of L2 learning, different areas and different portions of the areas are involved in the processing.

[40] They observed regional differences "in the morphology of the basal ganglia and thalamus" between simultaneous bilinguals and monolinguals. In the bilinguals, also an enlargement of bilateral putamen and thalamus, and of left pallidum and right caudate nucleus was found, compared to the monolinguals.

[41] It has also been suggested or perhaps merely accepted, for decades, that executive functioning is superior in bilinguals compared to monolinguals. However, Lehtonen and colleagues (2018), in their meta-analytic review, show that there is no such cognitive control advantage in bilinguals.

[42] The participants were simultaneous Canadian English – Canadian French bilinguals, monolingual Canadian English speakers and monolingual Canadian French speakers. They used vowel stimuli from the closed rounded area, French /y/, control /y/, English /u/, and French /u/.

the bilinguals function in "a hybrid phonological space" (ibid., 2014, p. 540) which means that when the language context is clear, the context language cues are the most salient but when there is no clear language context, the cues that are important in distinguishing acoustically similar cross-language categories are of importance.

Yet another area of bilingual research is the processing of the languages in different language contexts. The studies have revealed seemingly contradictory results. For example, the functional processing measured with MMN has shown that language context does not affect phonological processing in late, naturally in adulthood exposed bilinguals (Winkler et al., 2003). On the other hand, different language contexts revealed that advanced language students, who have learned the L2 in classroom from the age of 9, have separate native and L2 phonemic systems (M. S. Peltola & Aaltonen, 2005). Peltola and Aaltonen (ibid.) discuss that the differences between the results of these two studies may be due to stimulus selection (for more discussion on the importance of stimuls selection, see, e.g., M. S. Peltola, 2007). In the Winkler et al. study, the difference between the stimuli was quite small and they were near the category border, whereas, in the Peltola and Aaltonen study the stimuli were good category representatives and the acoustic difference between the stimuli was greater. Language context effects on speech perception have also been shown in bilinguals whose language profile is a bit more complicated. They were Spanish-English bilinguals who were dominant in Spanish from 3 to 6 years of age when exposure to Spanish was 100%. From 9 to 12 years of age they were relatively balanced and 64% of them were exposed to English and Spanish equally. However, they were English dominants from 18 to 21 years of age when 82% of them reported being more exposed to English than Spanish. (García-Sierra et al., 2012). Significant MMN responses were elicited only in the conditions where the difference between the stimuli was phonemic.

Language context affects speech production as well. Simultaneous and sequential bilinguals' speech production has also been observed in a study were three different language contexts – monolingual Finnish, monolingual Swedish and bilingual Finnish and Swedish – were used (Tamminen et al., 2017). There was a disparity in the sensitivity to the language context between the different bilinguals. The simultaneous and sequential bilinguals seem to have different control mechanisms for their languages, the former having a shared and the latter having a separate control system for their two languages. Also dominant bilinguals (Greek acquired since birth, English learned before the age of 6 years, and were dominant in English) produce syllables according to the language context (Greek or English) and alike monolinguals (Antoniou et al., 2010). This was suggested to refer to a common phonological space for both languages, as there were minor signals of interference from L1 to L2 in the most complex phonetic positions. Perception-wise, categorisation and goodness-rating judgements were also context sensitive, however, discrimination showed effects from the (L2) dominant language in a study (Antoniou et al., 2012) with similar participants as the above mentioned. The speech perception finding was suggested to indicate a common phonetic space, which is influenced by the dominant language context. When they have to make phonological judgements (like categorisation and goodness-rating), they turn to the language-specific phonological information. The language context evidently has an impact on perception and production in bilinguals. In other words, a phonetic speech sound may have two different phonemic representations depending on the language context. This is called the bilinguals' double phonemic representation or double phonemic boundary (e.g., Elman et al., 1977; García-Sierra et al., 2009). Although, due to methodological and stimulus differences not all studies showed context differences. However, the language context affects both perception and production.

Plasticity is evidently needed to acquire or learn languages. Whether the other or second language is acquired/learned from birth, later as a child, or later as an adult, or whether the attained proficiency level is high or low, plasticity is a prerequisite for functioning with two or more languages. The next chapter concentrates on training studies and on how speech perception training affects monolinguals and different bilinguals.

2.4 Training and language learning

A clear indication of human neuroplasticity is that we learn by training. In other words, neuroplasticity, functional and physical reconfiguration of the brain, enables humans to learn new skills. Naturally, also speech perception, and production, learning and training require neuroplasticity.[43] This chapter will mainly focus on various training studies, with different methodologies, on attentive and pre-attentive speech perception; however, training studies on speech production are discussed also. Another viewpoint of the chapter is to review and discuss the speech perception and production training results in studies concerning children, adults, and seniors.

Training studies with different settings Learning in a laboratory environment is obviously different from learning in a natural setting, or even from classroom learning. From the perspective of research findings, Classroom learning and laboratory training are, however, close to each other in the sense that, for example, listen-and-repeat training is used in classroom teaching. Laboratory training may not, strictly speaking, lead to native-like speech processing, but to robust learning that is linguistically functional (see e.g., Bradlow, 2008). Listen-and-repeat training has proven efficient on speech production in children (Taimi, Jähi, et

[43] See, for example, Li and colleagues (2014) for a review on neuroplasticity and second language learning.

al., 2014) and in adults (K. U. Peltola, Alku, et al., 2017; Savo & Peltola, 2019), and in speech perception, for example, in highly advanced language learners (Tamminen & Peltola, 2015).

Perception training positively affects both perception and production, however, training in both perception and production concurrently has been shown to be effective on production, but less effective on perception (Baese-Berk, 2019). Further, by increasing a two-day training by one day, the difference in the effect was alleviated (*ibid.*). There is also evidence of production training effects on perception: It can lead to more systematic category boundary formation (K. U. Peltola et al., 2020; Tamminen & Peltola, 2015), and to decreased reaction times and improved discrimination sensitivity (Tamminen & Peltola, 2015). On the other hand, it has been shown that even though identification and discrimination training can increase the accuracy of perception and production in a similar manner, there is no significant benefit from the combination of the two training methods, and they provide similar effects when using high-variability perceptual training (HVPT) (Shinohara & Iverson, 2018). Further, HVPT has proven to be effective in children (Giannakopoulou et al., 2013) and in adults (Giannakopoulou et al., 2013; Grenon et al., 2019). HVPT with a single talker paradigm also seems to improve accuracy slightly more than a multiple talker paradigm, on the other hand, multiple talker paradigm (a more natural setting) seems to result in more compact categories and in better response to unfamiliar speakers (Kartushina & Martin, 2019). Even though HVPT has proven to be an effective training method in general, it might be detrimental to individuals with low learning aptitudes compared to those with high aptitude (Perrachione et al., 2011). Further, the target language experience level does not affect the amount of benefit one may have from HVPT vowel identification training (Iverson et al., 2012). The researchers (ibid.) suggest that, during an ID task, focused attention to phonetic differences improves L2 vowel perception, which reinforces the learning in a natural language learning setting. Just a small amount of benefit from training was seen in the discrimination or production of the target language vowels. It has also been suggested that an adaptive HVPT effects in discrimination are greater in people with higher inhibitory control (Ghaffarvand Mokari & Werner, 2019). Yet another type of training study has shown that training in an adaptive adverse (stimuli presented with a multi-talker babble in the background) condition seems to be effective, compared to fixed level adverse training (Leong et al., 2018). Speech perception training with the target presented in a fixed speech sound context, compared to two different speech sound contexts, seems to be more effective, although in Fuhrmeister & Myers' (2017) study, effects were seen immediately after training for both, over-night consolidation was seen only for the fixed context group. In summary, variation in the effects of and within various types of training studies is evident. Here, again, possible reasons for the variation are, for example, differences between participants, training and testing procedures, and the modality being tested.

Other perspectives in the training studies are the generalisation of the training effects to untrained stimuli and the duration of the training effects. The generalisation of the trained stimuli to the untrained ones has been shown to vowels with novel tokens and talkers (Grenon et al., 2019) with same vowels in different words (Leong et al., 2018) and to consonants with novel words (Flege, 1995b) or novel place of articulation (Tremblay et al., 1997). Stable training effects have been shown after varying time periods, for example, one month (Kraus et al., 1995), two months (Flege, 1995b), or six months (Leong et al., 2018) after training.

Next, training effects measured with MMN are reviewed. Learning or training to perceive something new – should it be non-linguistic complex spectro-temporal patterns (tonal patterns), pure tones, or various sound patterns presented with matching visual patterns as a computer game – elicits MMN responses (Atienza & Cantero, 2001; Kujala, Karma, et al., 2001; Menning et al., 2000; Näätänen et al., 1993). Training with linguistic stimuli – syllables with varying formant transitions, syllables with varying VOTs, or word stimuli varying in duration – have also led to neural plastic changes (Kraus et al., 1995; Menning et al., 2002; Tremblay et al., 1997, 1998; Tremblay & Kraus, 2002). These studies have used some type of discrimination or identification training methods. A same-different two-alternative forced-choice discrimination training study with visual feedback using synthetic speech stimuli varying in the onset frequencies of F2 and F3 transitions (variants of /da/) was conducted by Kraus et al. (1995). The training lasted for one week and consisted of six 1-hour sessions of training between pre- and post-training tests. The MMN response duration and magnitude increased as a function of training. In addition, the behavioural discrimination improved significantly and the result was stable, measured one month later. An identification training with visual feedback using synthetic speech sound stimuli with varying VOTs (/ba/-/pa/) was used in a nine-day training study (Tremblay et al., 1997). Pre- and post-training testing took two days each and there were five training days in between them. Training sessions lasted for 20 minutes. Training effects were seen in this study, as MMN duration and area increased and discrimination and identification scores improved. In addition, the training effects were also seen when untrained stimuli (/da/-/ta/) were used, i.e., the training effects transferred from trained labial consonants to alveolar consonants. Another study by Tremblay and colleagues (1998) also used identification training with visual feedback and synthetic stimuli varying in VOT. Training effects were seen in both neural activity and behavioural learning. The baseline measurements were on the first two days, after which followed eight days of training and testing in turns. Effects in at least one of the neurophysiological measures (MMN duration, area, and onset latency) were seen immediately after first training day for the whole group. However, behavioural effects were seen either on the same day or later. Yet another training study (Menning et al., 2002) showed training effects using a forced-choice, two-alternative, self-adjusting staircase method discrimination training. German subjects trained Japanese mora-timing and there were approximately 1.5 h of training per day for ten consecutive workdays. Here as well, visual feedback was provided. Behavioural results improved in the first session already, while the MEG was measured only before and after training, and hence, the increased amplitude of the Mismatch Negativity Field (MMF) was observed ten days after the training started. To sum up, these four studies all used speech stimuli in either an identification or discrimination training and measured MMN. There were several training sessions (4–10) in the training period which varied from 7 to 10 days. The training sessions lasted for about 20 minutes to 1.5 hours per day (duration of training was not reported in one of these studies). Feedback on the answers during training was given in all the studies. The training sessions were either between pre- and post-tests or training and testing were mixed. Altogether, training had effects on perception in all of these studies. Laboratory training is indeed effective and can lead to learning.

Training effects on children, adults and seniors The next few paragraphs will concentrate on training studies separately in children, adults, and seniors. Both speech perception and production training has proven to be effective in children. For example, speech production training effects on speech production (Taimi, Jähi, et al., 2014) and on pre-attentive perception (Taimi, Alku, et al., 2014) in children were shown in studies using a listen-and-repeat training with semi-synthetic word stimuli /tyːti/ and /tʉːti/ in four sessions during two days[44]. Nine-year-old children learned to produce the difficult /tʉːti/ by day two after three training sessions (Taimi, Jähi, et al., 2014). Training effects were seen also at the pre-attentive level as MMN amplitude was larger after training (Taimi, Alku, et al., 2014). The latter study did not, however, show any behavioural changes in production or discrimination, which may be explained by the small number of subjects. Attending to music-oriented education program, on the other hand, does not benefit children in production training when compared to children in regular education program (Immonen et al., 2021). The study used a passive auditory training method, which was otherwise similar to that of Taimi and colleagues (2014). Yet another training study by Immonen and colleagues (2022) suggest that inconsistent orthographic cues do not hinder the learning of the non-native sound. In another study, 12-year-old native Dutch speakers were trained by a three-day

[44] The participants were Finnish monolingual children and the first vowels in these pseudo words had different status in their native language, /y/ is part of the Finish phonological system, whereas /ʉ/ is a non-prototypical representative of /y/ or /u/.

identification training method to learn the Finnish quantity contrast /t–tː/ (Heeren & Schouten, 2010). The study was designed to examine the development of perceptual sensitivity in the vicinity of category boundary and the influence of learner's age on learning[45]. Training was effective and the timing of learning was similar in children and adults. HVPT was used in a study on native child and adult Greek speakers using English minimal pair words varying in vowels /i/ and /ɪ/ (Giannakopoulou et al., 2013)[46]. Training was effective and identification and discrimination performance improved in both groups, although children were more plastic since improvement was greater in children compared to adults. (Later, in a comparable study, Giannakopoulou and colleagues (2017) did not show plasticity differences between children and adults.) Further, Immonen and Peltola (2017) report that language awareness and experience may differentiate children aged 9–12 from 7–8-year-old children. They showed that the younger children did not elicit an MMN response as a function of passive auditory training, whereas an MMN response was elicited for the older children after training.

In summary, these studies used speech stimuli in production, identification, and HVPT training. The number of training sessions varied from four to ten and the training periods varied from two days to two weeks. Training sessions lasted for about 4 to 30 minutes. Feedback was given in two of the four studies. Here as well, the training sessions were either mixed with testing sessions or there were pre- and post-testing separately. Here also, training was effective and behavioural perception and production, and pre-attentive effects, were observed, though not throughout the studies. Further, two of the studies compared children and adults and one of them showed equal effects on both age groups and in the other the children were more plastic than adults.

Speech perception and production training studies have also shown learning effects on adult participants. Mere listen-and-repeat training has been compared to variations of listen-and-repeat training and some of the additional cues have proven

45 Female and male speakers were used when creating the stimuli and trial-by-trial feedback was given during training. Five training sessions lasted for 15 minutes each.

46 Both testing and training stimuli were of two kinds, natural and modified. The duration of the target vowel was equalised in the pairs so that /i/was presented as short as /ɪ/ and /ɪ/ as long as /i/. Training lasted for two weeks and there was one session of training per weekday, ten sessions altogether. A visual minimal pair was presented with the auditory pair. Feedback was given and a replay was offered as an option, also an additional trial was given after a false response.

effective on production, some not[47]: Transcription cues together with auditory stimulation, but not orthographic cues, steer towards non-native target productions (K. U. Peltola et al., 2015), and the authors state that when the auditory and visual cues conflict, the visual cues have a more significant role than the auditory one. On the other hand, listen-and-repeat training and mere listening are both as effective and non-native production can be acquired even via auditory input only (K. U. Peltola, Alku, et al., 2017). The motor commands are, indeed, altered via mere auditory exposure. However, listen-and-repeat training seems to overpower active listening, where the task is not to repeat, but to count the target words (K. U. Peltola et al., 2020). It was implicated that motoric practice is a fundamental factor for improvement in production. Further, a maximal effect may be gained with the combination of explicit pronunciation instructions and articulatory training, as production changes have been detected already after one training session (Saloranta et al., 2015). Even a very short production training can slightly affect production, however, the strength of mother tongue prevails and a longer training is more beneficial (K. U. Peltola, Rautaoja, et al., 2017). This study consisted of only one training session, using the same stimuli as the above-mentioned studies, with pre- and post-training recordings during one session and it compared Finnish and American English speakers (K. U. Peltola, Rautaoja, et al., 2017).

Another listen-and-repeat training showed that advanced university students benefit from the training (Tamminen & Peltola, 2015)[48] as memory traces strengthened (existing MMN response gained amplitude as a function of training) and the consistency of the category boundary, reaction times and discrimination sensitivity improved. Saloranta and colleagues' (2017) results showed that a vowel quantity listen-and-repeat training was effective in the light of discrimination sensitivity but it had no effects on RTs. Production was not affected either, however, the difference between the trained contrast on days two and three compared to the untrained contrast were significant. Also Grenon and colleagues (2019) have shown training effects on both trained and untrained stimuli. They trained Japanese English learners to differentiate vowels according to spectral cues (formant frequencies)

47 The studies were conducted with the same protocol of two listen-and repeat training and data collection blocks on two consecutive days, with altogether four blocks of training and data collection and with the same stimuli described earlier in Taimi, Jähi et al. (2014) study. The participants in all of these studies were young adult monolingual Finnish speakers. (K. U. Peltola et al., 2015, 2020; K. U. Peltola, Alku, et al., 2017; Saloranta et al., 2015)

48 Participants were Finnish advanced students of English. Stimulus words were /fiːl/ and /viːl/ containing a feature difficult for Finns, namely voicing in a fricative sound.

rather that duration, and consequently, to form a new vowel category[49]. Training was effective on half of the participants, while others, whose performance somewhat improved, still relied on duration. Faster and slower learners have been shown to have anatomical and functional differences in the perception of foreign sounds (Golestani et al., 2007)[50]. However, when a task irrelevant phonetic feature is varied, the learning of the trained feature is hindered (Antoniou & Wong, 2016)

Kartushina with her colleagues have studied production training effects on production (Kartushina, Hervais-Adelman, et al., 2016; Kartushina & Martin, 2019) and on production and perception (Kartushina et al., 2015). They found effects on both production and perception, although production seemed to benefit more from production training than perception (ibid.). Effective production training seems to have some effects on the native language production too, since some of the measured native vowels shifted towards the foreign trained vowels (Kartushina, Hervais-Adelman, et al., 2016). They also showed that HVPT with a single talker improved accuracy slightly more than multiple talker HVPT, however, the multiple talker paradigm resulted in more compact categories and participants in this group responded better to unfamiliar speakers (Kartushina & Martin, 2019).

In summary, the above training studies, with adult participants, used speech stimuli in production, adaptive identification, and HVPT training. The training periods varied from one day to even five weeks and the number of training sessions varied from one to ten. Training sessions lasted for about four to 45 minutes. Feedback was given in some of them. Again, the training sessions were either mixed with testing sessions or there were pre- and post-testing separately. Training proved to be more or less effective in most of the studies, and the effects were seen even within untrained stimuli, in perception and production.

Like speech perception and production studies, also training studies on elderly people are few. Language oriented elderly people seem to benefit from speech production training in contrast to seniors with other than linguistic interests (Jähi et al., 2015)[51]. Also, elderly adults with mild to moderate hearing loss benefit from training as was shown in a speech perception training study (Kuchinsky et al., 2014).

49 The stimuli were English vowels /i/ and /ɪ/, which are representatives of one Japanese category, and further, Japanese differentiate vowels according to duration. Training was an adaptive identification HVPT with feedback and lasted for 2–5 weeks.

50 Adaptive identification training with feedback was used in this study. The identification adapted to the participants performance so that when there were enough correct answers, the difference between the stimuli became smaller. The training lasted 15–20 minutes per participants and the maximum number of blocks was ten. The participants were French speakers and the stimuli used were Hindi dental-retroflex contrast.

51 Also this study used the two-day listen-and-repeat training with the earlier mentioned /tyːti/ and /tʉːti/ stimuli. In addition, the study protocol included EEG registrations, ID and discrimination tests, although these results were not reported.

This word recognition training study used spoken word stimuli presented in background noise. Training was self-paced and lasted approximately 8.5 weeks with 20 sessions on average and participants received both auditory and visual feedback. In the actual word recognition task, they also recorded eye movements and RTs. Language learning, multilingualism, and cognitive training are considered to benefit cognitive function in older adults. Since cognitive abilities decline along with aging[52], language and speech perception are affected as well. Antoniou, Gunasekera and Wong (2013) contemplate these issues, among other things, in their review and recommend foreign language training as a cognitive training activity for elderly adults. They also argue that despite the fact that language training in the older age does not necessarily lead to bilingualism, the older adult brain is still plastic and trainable. Hence, more speech perception training studies on the elderly are needed.

It could be summarised that a combined set of different methodologies would lead to maximal learning and training effects would be extensive. However, a large set of various training types, including “acoustic exaggeration, varying talkers, visible articulation and adaptive listening, does not guarantee a native-like category boundary even when identification improves, however, the effect may transfer to novel stimuli (Zhang et al., 2009). This large scale training was also seen as increased neural sensitivity and efficiency measured with MEG. To conclude, there is a vast number of different speech perception and production training studies, all of which could not be included in this chapter. A tremendous amount of training research has been conducted on children and on adults, many of them have compared children and adults. However, speech perception or production training studies on elderly people are needed.

2.5 Aging and language learning

This chapter briefly introduces some aging related issues concerning language processing and speech perception. Brain structure changes and compensatory effects on different levels of language processing are introduced first. Then, the effects of aging on cognitive functions and hearing are discussed.

The aging brain structure changes and these changes are, for example, shrinkage, decrease in white matter integrity, depletion of dopamine, a neurotransmitter, which is vital in learning (e.g., Wise, 2004). To some extent the age-related declining in the brain and in various processes are compensated to increase functional brain activity.

52 Aging and its effects on language learning and processing are discussed in more detail in Chapter 2.5 Aging and language learning.

However, the compensatory networks may not be as efficient as the original networks. (Park & Reuter-Lorenz, 2009)[53]

Studies on different language and speech processing levels have shown different, or compensatory, processing in the elderly compared to younger adults within native language. Compared to younger adults, the compensatory mechanisms or processing has been evident in spoken word processing in noise (Wong et al., 2009) and in phonological word retrieval in rhyme judgement tasks (Geva et al., 2012). Also, aging may affect semantic processing (Dennis & Cabeza, 2011) and may not affect spoken syntactic complexity (Nippold et al., 2014). Performance in verbal fluency may (Mougias et al., 2019), or may not (Machado et al., 2009), be influenced by age among different senior age groups. Again, compared to younger adults, older people take more time in re-orienting toward a relevant speaker when they hear concurrent speech (Getzmann et al., 2015). However, despite the declining in the cognition as a result of aging, knowledge and expertise are not significantly affected by aging (Park & Reuter-Lorenz, 2009).

Although there are aging related changes, the older brain is still plastic and language learning has been shown to improve language related functions, but it also improves cognitive functions (Antoniou et al., 2013). Language learning, with its beneficial effects on cognitive functions[54], should be encouraged among elderly people, and language training methods suitable for seniors should be developed. Moreover, the bilingual advantage on executive functions is seen also in the elderly people like in younger adults[55], and the age-related decline in certain executive processes has been shown to be slower within the bilingual elderly people compared to monolinguals (Bialystok et al., 2004)[56]. Improving cognitive functions is important also for slightly declined hearing, as it may worsen by the reduced cognitive abilities (Anderson & Kraus, 2013). Training on native language speech perception in the elderly, who have a mild to moderate hearing loss, seems effective

53 The statement that both decline in sensory processing and cortical activation and compensation as an increased activation of more general cognitive areas take place in cognitive aging is the core of the decline-compensation hypothesis (e.g., Wong et al., 2009).

54 As a mere aside comment, a fairly resent meta-analysis (Lehtonen et al., 2018) on bilingual executive functions showed that bilingualism does not lead in any advantage in cognitive control functions in adults.

55 As mentioned in Chapter 2.3 Bilingualism and speech perception, there is evidence that bilingualism does not lead in any advantage in cognitive control functions in adults (Lehtonen et al., 2018).

56 The participants were early balanced bilinguals and monolinguals, younger adults were 30–54 years of age and the older adults were 66–88 years of age. The bilinguals were Tamil-English bilinguals living in India and the monolinguals were English speakers living in Canada.

(Kuchinsky et al., 2014). Kuchinsky and colleagues (2014) showed, that after training, identification of words in noise and discrimination of speech from background noise was improved, when compared to similar aged controls.

The elderly also perceive temporal features in speech less precisely in comparison with younger adults (Strouse et al., 1998; Tremblay et al., 2002), or children (Bellis et al., 2000), as shown with VOT studies on both behavioural and psychophysiological discrimination. Extracting fine temporal details is difficult for the elderly, compared to younger adults, also in non-speech (Ostroff et al., 2003). As mentioned in the previous chapter, foreign language learning or training studies on phonological processing in speech perception, or production, in the elderly are scarce. Hence, knowledge on the effects of aging on language learning is limited. Interest in language learning, however, seems to benefit the elderly in foreign language speech production training more than interest in other recreational activities (Jähi et al., 2015)[57].

The age-related effects on the brain and different cognitive functions, as well on different levels of language processing are evident. However, the elderly brain is still plastic and new skills may be learned.

2.6 Summary: speech perception and learning

Brain plasticity enables us to acquire and learn new skills, new languages, and our native languages. In addition, plasticity enables us to learn to perceive foreign language speech. After acquiring the native language that has started already in the fetal period and continued to develop during the first years of the child's life, the second or foreign language speech perception learning becomes more difficult. Foreign language learning models provide different viewpoints and predictions for the problems the second language learners encounter during the learning process of the foreign language speech sounds. The comparison of the native and target language phonological systems is nevertheless the common factor in the various foreign language speech learning models. Foreign language speech perception learning studies are numerous with varying viewpoints and methods. Speech perception and production training constitutes one set of different methods in the foreign language speech learning studies. Both children and adults' learning is vastly studied and compared. On the other hand, speech sound learning studies among the elderly are quite few. However, the older brain is still plastic and learning is by no

[57] Language orientation has been shown to be beneficial also in children in production learning (Immonen & Peltola, 2018) and in adult simultaneous bilinguals and language learners, compared to monolinguals, in vowel perception (Tamminen & Peltola, 2019).

means impossible. Further, the maintenance of the cognitive functions is important in the older age and foreign language learning offers an excellent possibility for that.

Research on bilingualism and bilingual speech perception is extensive. Part of the research concentrates on foreign or second language learners while the other part is related to bilinguals who acquire both languages very early on in their lives. There is a wide range of varied and overlapping terminology describing bilinguals. The use of the terminology is sometimes even contradictory. However, the description of the language background, the language history, and the present linguistic state of the person are far more interesting and important, from the research point of view at least.

Speech perception theories and models have described different aspects of speech perception processes from many different viewpoints. Lack of invariance is the centre for some of them; some consider it not relevant at all. Some theories see the acoustic signal as the object of perception. Some consider the intended or actual gesture as object of perception, while other theories do not examine the object of perception at all. Some theories concentrate on the within category perception while others study the across category perception.

The link between speech perception and production emerged in several contexts. PAM, for example, suggests that infants perceive primarily the articulatory gestures that include visual and proprioceptive information to become native language speakers, not only listeners (Best & McRoberts, 2003). Similar pattern is suggested for second language learning: perception learning precedes in order for the production to develop (e.g., Flege, 1993, 1999; Flege et al., 1999). The latest version of the SLM, the SLM-r, state that there is no precedence but there is a bidirectional connection between perception and production (Flege & Bohn, 2021). Also Baese-Berk (2019) states that there has to be some kind of a link between speech perception and production. If there was no connection, her finding, where production training disturbed perception, would be surprising. She further speculates that if the two would be identical, the disturbance would also be unlikely. Hence, perception and production are separate but linked to each other. She also points out that perception and production processes accomplish different representations during learning, which is seen in the temporal difference in the effects of concurrent perception and production training. In addition, training only perception is less demanding than the joint training, and hence, transfer of the perception training effect to production is possible. (Baese-Berk, 2019)

The next section presents the aims of the present thesis. The following two sections after that describe the methods used in the studies and present an overview of the five studies of the thesis. The discussion and the concluding remarks are presented in the final two sections of the thesis.

3 Aims of the present thesis

The aims of this thesis were twofold. First, the interest was in the bilingual speech processing, speech perception and the functioning of their phonological systems (see 1) and 4) below). The second aim was to examine how different factors such as manner and age of learning affect the formation of memory traces (see 2) and 3) below). The specific aims were the following:

1) To assess whether the neural processing of speech is differently organised in different bilinguals, namely simultaneous and sequential[58]. (Studies I and II)
2) To evaluate whether speech processing becomes bilingual-like by training and whether the training effect is permanent or not, observed with both behavioural and neurophysiological measures. (Studies III and IV)
3) To examine how different background factors – manner and age of learning – affect the formation of memory traces through training. (Studies I, III, IV and V)
4) To evaluate whether perceptual processing differs when operating with two phonological systems, acquired simultaneously, compared to operating with only one monolingual phonological system. (Study II)

[58] See discussion on the bilingualism terminology in Chapter 4 Methods.

4 Methods

Studies I and II share the same stimuli and procedures, as do **Studies III, IV and V**, and hence the methods chapter is organised accordingly. Even though the participants are not the same in **Studies I and II**, and in **Studies III, IV and V**, the same grouping is provided in the description below. Speech perception between Simultaneous and Sequential bilinguals is compared in **Study I** and between Simultaneous bilinguals and Monolinguals in **Study II**. **Studies III, IV, and V** are speech perception training studies, **Study III** concentrating on young adult Monolinguals, **Study IV** compares young adult Monolinguals and Sequential bilinguals, and **Study V** concentrates on elderly Monolinguals.

It is noteworthy that the participants in the original **Studies I and II** were referred to as Balanced and Dominant bilinguals. However, according to the description of different bilingual types referred to in Chapter 2.3, these participants could also be described as simultaneous and sequential bilinguals. Finnish most definitely is the dominant language for the second language learners, but also, they have acquired their languages sequentially. The other participants have acquired their languages simultaneously but it is actually very difficult, and, to be fair, unnecessary, to measure whether they are truly balanced. The status of Balanced bilinguals was, however, established by the self-reported proficiencies on both languages (see Table 1 in Chapter 4.1.1). Hence, from now on, the bilinguals from birth are referred to as Simultaneous bilinguals and the second language learners are referred to as Sequential bilinguals for the sake of coherence with the other studies[59]. The terms are by no means considered as synonyms even though here they are used as they were.

[59] The terms Balanced and Dominant bilinguals are, however, used in the overview of **Studies I and II** in accordance with the original publications.

4.1 Participants

In all the studies, the voluntary subjects were tested right handed with a modified Edinburgh Handedness Inventory (Oldfield, 1971), and they had normal hearing which was tested with an audiometer using perceptually relevant frequencies (250 Hz, 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz). The subjects had no diagnosed neurological illnesses or medication affecting the central nervous system. All subjects were given written and oral information about the study in question, and a written or oral consent was obtained from them before testing. A written consent was signed by the parents in case a participant was under-aged. Altogether data from 66 participants from 142 sessions are reported in this thesis.

4.1.1 Participants in Studies I and II

In **Studies I** (Simultaneous vs. Sequential bilingual study) and **II** (Simultaneous bilingual vs. Monolingual study), there were three groups of subjects. The first group[60] consisted of 12 (age range 16–31, mean age 20.3 years, 7 females) Finnish-Swedish bilinguals (Simultaneous bilinguals). They had acquired Finnish and Finland-Swedish from birth in a one-language–one-parent manner. None of these subjects had ever lived in Sweden. The subjects made self-evaluations of their language proficiency and, according to these evaluations, they had a high proficiency in both their native languages, see Table 1. Daily usage of both languages was possible since Finnish and Swedish are the official languages in Finland and the socioeconomic status of these languages is equal. Public services are provided in Finnish as well as in Swedish.

The second group consisted of 10, although the number of participants was not mentioned in the original journal publication, (age range 20–24, mean age 20.2 years, 6 females) advanced Finnish university students of Swedish (Sequential bilinguals). They had a high command in Swedish which was ensured by a highly demanding entrance examinations through which they entered at the Department of Scandinavian Language at the University of Turku. Regular exposure to the Swedish language was high because only Swedish is used at the Department. The subjects had studied Swedish as their major at least two years at the time of testing. The average age of exposure (AOE) was 12.6 years and they had studied Swedish at school for 6.4 years on average (range 3–10 years). See Table 1 for self-evaluations of Swedish proficiency.

[60] The group numbering deviates from the one in the original publications.

The third group, Monolinguals[61], were 10 (age range 17–42, mean age 26.7 years, 7 females) native speakers of Finnish. They had studied Swedish at school for 5.9 years on average (range 3–9 years) (Swedish was an obligatory subject at the primary and secondary schools as well as at the upper secondary school at the time of testing), and they reported quite low proficiency in Swedish, see Table 1. Also, they reported very low usage percentage (from none to 10 %; average 2.65 %).

All the subjects in these three groups were from the same dialectal area, in near vicinity of Turku. Hence, their Finland-Swedish dialect was the same. **Study I** concentrated on comparing Simultaneous and Sequential bilinguals' speech perception, whereas **Study II** compared Simultaneous bilinguals' speech perception on that of the Monolingual group.

61 The term monolingual is probably almost as difficult as bilingual. Is a monolingual person someone who has not acquired/learned another language from birth, or early in life, or as an immigrant, or as a classroom language learner? Nowadays there are probably less and less monolinguals regardless of the criteria for the definition. There are possibly less monolinguals in the world than bilinguals if the definition for a bilingual is anything other than "both languages acquired from birth".

Table 1. Self-reports on language proficiency; how well speaking, understanding, reading, and writing is mastered in Finnish and Swedish by the Simultaneous bilinguals and in Swedish by the Sequential bilinguals and the Monolinguals. Self-reports on Finnish proficiency were not obtained from the Sequential bilinguals or the Monolinguals.

SIMULTANEOUS BILINGUALS

FINNISH	excellent	well	satisfactory	badly	not at all
speaking	12	0	0	0	0
understanding	12	0	0	0	0
reading	12	0	0	0	0
writing	10	2	0	0	0

SWEDISH	excellent	well	satisfactory	badly	not at all
speaking	12	0	0	0	0
understanding	12	0	0	0	0
reading	12	0	0	0	0
writing	11	1	0	0	0

SEQUENTIAL BILINGUALS

SWEDISH	excellent	well	satisfactory	badly	not at all
speaking	1	4	5	0	0
understanding	5	3	2	0	0
reading	2	8	0	0	0
writing	0	8	2	0	0

MONOLINGUALS

SWEDISH	excellent	well	satisfactory	badly	not at all
speaking	0	0	5	5	0
understanding	0	3	5	2	0
reading	1	2	3	4	0
writing	0	1	5	4	0

4.1.2 Participants in Studies III, IV and V

There was a different set of participants in **Studies III, IV, and V: Study III** concentrated on young adult Monolinguals, **Study IV** compared Monolingual and Sequential bilinguals, and finally, **Study V** was conducted to study senior Monolinguals. The first group[62] consisted of 12 Monolingual native Finnish speakers, who had no noteworthy skills in the English language and after upper secondary school they had not studied any languages. They had studied English for 8.8 years on average (range 5–11 years). At the time they went to school English was usually taught from the third grade (age 9) onwards in Finland. Nonetheless, the participants are considered monolinguals. They reported hearing English daily 'little' (1), 'quite a lot' (8) or 'a lot' (3). Self-reported language proficiency in English is shown in Table 2. From the 12 subjects one had to be excluded from the identification (ID) and goodness rating (GR) tests and another from the discrimination test. Hence, the grouping was as follows: in the ID and GR tests there were 11 subjects (age range 18–32, mean age 23.1 years, 7 females), in the discrimination test there were 11 subjects as well (age range 18–32, mean age 23.4 years, 6 females), and finally, in the electrophysiological test there were 12 subjects (age range 18–32, mean age 23.4 years, 7 females). Eight subjects (age range 19–27, mean age 22.6 years, 5 females) participated in the follow up study (7 in the ID and GR tests). All of the exclusions were due to technical problems or unreadable data due to incorrect use of the answer buttons.

The second group consisted of 11 (age range 20–28, mean age 23.6 years, 6 females) advanced Finnish university students majoring at the Department of English at the University of Turku. They had been studying English for 9.7 years on average (range 8–10 years) at school and for 3.5 years on average (range 1–7 years) at university, and had never participated in exchange programmes. The participants reported hearing English daily 'a lot' (7) or 'quite a lot' (4). Self-reported language proficiency in English is shown in Table 2. They can be classified as sequential bilinguals.

There were 11 Finnish monolingual Seniors in the third group, although, one subject had to be excluded from the study altogether and another subject had to be excluded from the behavioural tests because of technical problems and unreadable data as a result of incorrect use of the answer buttons. Thus, the age range was 61–69 years (9 subjects, mean age 64.3 years, 4 females) in the behavioural tests and 61–71 years (10 subjects, mean age 65 years, 5 females) in the electrophysiological studies. All subjects were retired at the time of testing and had been off the working life for at least one year. These subjects had no other language identities than Finnish.

[62] The group numbering deviates from the one in the original publications.

They had been studying English for 3.3 years on average (range 0–7 years) at school. The participants reported hearing English daily 'quite a lot' (4) or 'little' (6). Self-reported language proficiency in English is presented in Table 2.

Table 2. Self-reports on language proficiency; how well speaking, understanding, reading, and writing is mastered in English by the Monolingual native Finnish speakers, Finnish Sequential Finnish-English bilinguals, and senior Monolingual native Finnish speakers. Self-reports on Finnish proficiency were not obtained.

MONOLINGUALS[a]					
ENGLISH	excellent	well	satisfactory	badly	not at all
speaking	2	4	2	1	0
understanding	3	6	0	0	0
reading	6	2	2	0	0
writing	1	5	3	1	0
SEQUENTIAL BILINGUALS					
ENGLISH	excellent	well	satisfactory	badly	not at all
speaking	7	3	1	0	0
understanding	10	1	0	0	0
reading	9	1	1	0	0
writing	5	5	0	1	0
SENIOR MONOLINGUALS					
ENGLISH	excellent	well	satisfactory	badly	not at all
speaking	0	1	5	1	3
understanding	0	2	5	0	3
reading	0	1	6	0	3
writing	0	0	6	1	3

a Two participants did not give an answer at all and one participant did not give an answer on speaking and understanding

Because of the Finnish school system, it is practically impossible to classify anyone as a monolingual in the strictest classification. For example, English and Swedish, the language settings on which the research in this thesis are built on, are taught at school from the early age. However, as shown in Table 2 above, the monolinguals' English or Swedish language proficiency does not reach the level of the target language university students. The senior participants had studied the target language (English) even less than the young adults, since during the time they went to school, German was studied perhaps more than English.

4.2 Stimuli

Two different sets of synthesised (HLsyn software, version 1.0. Sensimetrics, Inc.) stimuli were used in the five studies. Vowel stimuli were used in **Studies I** and **II**, whereas word stimuli, with consonants as target, were used in **Studies III–V**.

4.2.1 Stimuli in Studies I and II

The vowel continuum, used in **Studies I** (Simultaneous vs. Sequential bilingual study) **and II** (Simultaneous bilingual vs. Monolingual study), consisted of 18 synthesised closed rounded isolated vowels from /y/ to /u/. Second formant (F2) values ranged from 703 Mel to 1553 Mel (606–2077 Hz) in 50 Mel steps. The values for F1, F3, and F4 were kept constant (250 Hz, 2600 Hz, and 3500 Hz, respectively). The fundamental frequency (F0) contour started from 112 Hz, reached its maximum 132 Hz by 100 ms and descended to 92 Hz by the end of the stimulus in order to imitate natural speech. The amplitude of the stimuli was smoothened at the onset and at the offset by a 30 ms ramp. The duration of the stimuli was 350 ms.

Finnish language has two closed rounded vowels – /y/ and /u/ – whereas Finland-Swedish has three – /y/, /ʉ/, and /u/[63]. This vowel area was selected because it offers an interesting difference in the phonology of the two languages. The bilinguals have to be able to identify, discriminate and produce three vowels according to one of the languages and, on the other, they have to ignore one category and at the same time function with two somewhat differently distributed vowels, when perceiving and producing the other language. For native speakers of Finnish learning Swedish, this vowel area is bound to cause problems, as the Swedish /ʉ/ is located at the boundary area of their native language /y/ and /u/. The Swedish /ʉ/ is, hence, **Similar** with the native vowels (Flege, 1987a) or the Swedish contrast /y/ – /ʉ/ is **assimilated unequally into one native category**, /y/ as a good representative, /ʉ/ as poor (Best & Strange, 1992; Best & Tyler, 2007).

63 It is noteworthy that Finland-Swedish and Sweden-Swedish closed vowels are not identical, especially the F2 value of the /ʉ/ vowel is lower in Finland-Swedish than in Sweden-Swedish (Asu et al., 2009; Ewald et al., 2017).

4.2.2 Stimuli in Studies III, IV and V

Studies III (young adult Monolingual training study)**, IV** (young adult Monolingual vs. Sequential bilingual training study) and **V** (elderly Monolingual training study) used synthesized words as stimuli. The stimulus continuum consisted of 15 variants of the English words /fiːl/ *'feel'* and /viːl/ *'veal'*. Hence, only the voice onset time (VOT)[64] of the first sound of the stimuli varied – from entirely voiceless to completely voiced fricative in 14 ms steps. From 197 ms onwards (i.e., the vowel and lateral part) the stimuli were identical and the duration of the entire stimulus was 499 ms.

Unlike the English phoneme system, Finnish does not make the /f/ – /v/ distinction in its phonological system. Only the voiceless fricative /f/[65] is part of the Finnish phoneme repertoire. Finnish does not differentiate fricatives with voicing, and hence, the difference between /f/ and /v/ is based on a new feature resulting in possible learning difficulties. According to SLM (Flege, 1987a), the English /v/ is **Similar** to the Finnish /f/, and according to PAM (Best & Strange, 1992) or PAM-L2 (Best & Tyler, 2007), the English /f/ and /v/ assimilate unequally to the Finnish /f/, which is expected to cause learning problems.

4.3 Procedure and data analyses

4.3.1 Tasks and measures

Behavioural perception was measured with tasks such as forced choice, self-paced identification (ID) and goodness rating (GR) to locate category boundaries, to see how systematic the boundaries were and to reveal the hierarchy of the category members as well as oddball discrimination tasks measuring reaction time (RT) and discrimination sensitivity (*d'*). Electroencephalogram (EEG) was used to measure pre-attentive event-related potentials (ERPs) such as mismatch negativity (MMN) and N1 to reveal memory trace formation. Combining the phonetic tests with the

[64] Conventionally VOT refers to the voice onset time in stop consonants (see e.g., Abramson & Whalen, 2017; Suomi, 1980, p. 60). However, the term has been used for fricatives as well (e.g., Abramson & Whalen, 2017; Massaro & Cohen, 1976, 1977). Massaro and Cohen (1976) point out that VOT in fricatives may be defined as the time between the onset of frication and the onset of voicing when the frication duration is the same in the voiced and voiceless members, similar to the current stimuli.

[65] The voiceless fricative appears only in relatively recent loanwords. Word-initial /f/ in loanwords was replaced by /ʋ/ in Finnish in the older times and the sequence /hʋ/ replaced the word-internal /f/. Many Finnish dialects still lack /f/, and it is still replaced by /ʋ/ or /hʋ/. However, /f/ occurs in the dialects which have been in contact with Swedish. (Suomi et al., 2008, p. 35)

psychophysiological measures is a convenient way to see the language acquisition/learning effects on the plastic memory traces (M. S. Peltola, 2001). The experiments were conducted according to the guidelines defined by the ethical committee of the University of Turku. All the experiments were conducted at Centre for Cognitive Neuroscience at the University of Turku. Stimuli were presented and behavioural data collected with Presentation software (NeuroBehavioral Systems). EEG data was recorded with Synamps amplifier.

The MMN is a pre-attentive change detection response, recently described as a regularity-violation response, not just a sound-change response (Näätänen et al., 2019, p. 2), elicited by a change in a string of similar sounds. It elicits 150–250 ms after a change onset in a stimulus string. The MMN indexes auditory discrimination and reflects the memory traces to feature and temporally integrated events. No attention is needed for MMN elicitation. (Kujala et al., 2007; e.g., Kujala & Näätänen, 2010; Näätänen et al., 2011). The N1 response indexes detection of separate auditory stimulus features, not discrimination as MMN, and peaks after stimulus onset, offset, or a change in stimulus energy at about 100 ms (e.g., Kujala & Näätänen, 2010; Näätänen et al., 2011). The MMN and N1 reflect different types of memory traces (Näätänen et al., 2011). Preconscious and conscious auditory processing can be studied with these responses and plastic changes caused by learning are seen as an enhancement of the responses (e.g., Kujala & Näätänen, 2010). A repetitive stimulation forms short-term memory traces in the auditory cortex which last for approximately 10 s (e.g., Näätänen & Escera, 2000), and hence, the traces can be formed and tested during laboratory experiments[66]. The early developed (Cheour et al., 1998; Dehaene-Lambertz & Baillet, 1998) language-specific (Näätänen et al., 1997) memory traces, as well as the learning and training induced memory traces, are obviously also detectable with MMN measurements. They are, thus, excellent measures of native and foreign language memory traces and sound discrimination accuracy.

4.3.2 Procedure and analyses in Studies I and II

Forced choice, self-paced ID tests were used in **Studies I** (Simultaneous vs. Sequential bilingual study) **and II** (Simultaneous bilingual vs. Monolingual study). Altogether 180 vowel stimuli (all the 18 different stimuli appeared ten times) were randomly presented to the subjects via headphones (Sennheiser HD 25). Both the Simultaneous and Sequential bilinguals performed the ID tests two times – once in Finnish, once in Swedish with Finnish or Swedish instructions, respectively. The

66 Very rapid cortical plasticity can be seen already in minutes for new words within the last 25% of a testing block (Shtyrov et al., 2010).

Monolinguals naturally did the test only in Finnish. In the test with Finnish instructions, the subjects were told they would hear Finnish vowels which they were asked to label as Finnish /y/ or /u/ by pressing the appropriate, labelled button on a numpad. In the other test, instructions were given in Swedish, and now the subjects were told they would hear Swedish vowels and they were asked to categorise them as Swedish /y/, /ʉ/, or /u/. In the Finnish context, the experimenter was a native Finnish speaker who always spoke only Finnish with the subjects, whereas in the Swedish context the experimenter was bilingual and always spoke Swedish. In other words, the setting was established as monolingual as possible. The order of the Finnish and Swedish ID tests was counterbalanced between the subjects in both subject groups. A short familiarisation, during which the subjects heard all the stimuli once in a random order and labelled them, started each ID task. This data was not included in the actual data. Also no feedback was provided. These ID tests were introduced to the participants in order to see the category representation, i.e., where category boundaries lie and the systematicity of the boundary areas, in the closed rounded vowel area in Finnish and Finland-Swedish in different language contexts.

The ID data was subjected to logit transformation analysis (using SPSS) in order to obtain the cross-over point in the answers where the distribution of the answers was 50%, pointing out the category boundary locations. The analyses also provided a steepness value for the boundaries indicating the consistency of the boundary area. On the basis of these individual analyses we selected individual stimuli for each subject for the EEG registrations. In other words, we selected a stimulus pair two steps away from each other so that they belonged to different categories in the subject’s Finnish ID results and fell within the Swedish /ʉ/ category for the same subject. This made sure that, for each individual participant, the stimuli represented phonologically contrastive sounds in Finnish while both were representatives of one category in Swedish, the contrast being phonologically irrelevant. For the Monolingual group the stimulus pair was naturally selected only on the basis of the Finnish ID results. For the Simultaneous bilinguals the stimuli used for /y/ ranged from number 4 to 11 and for /u/ from 6 to 13 (on average the stimuli used were 8 and 10), for the Sequential bilinguals the range for /y/ was from 7 to 10 and for /u/ from 9 to 12 (on average 9 and 11), and finally, for the Monolinguals the range for /y/ was from 7 to 10 and for /u/ from 9 to 12 (on average 9 and 11). The quite massive deviation between the individual stimuli confirms the need to use individually selected stimuli.

Pre-attentive perception in the two bilingual groups and the monolinguals was observed via measuring the MMN responses in different language contexts. The stimuli were randomly presented in an oddball paradigm consisting of 783 standard and 120 deviant stimuli resulting in 13.3% deviant probability and the inter stimulus interval (ISI) was 550 ms. The stimulus from the /y/ category served as standard and the /u/ category representative served as deviant. EEG was recorded with a 21 channel electrocap (Electro-Cap International, Inc) from the scalp with Sn electrodes using Synamps amplifier (model 5083). Sampling rate was 250 Hz and bandwidth was 0.5–70 Hz. Eye movements were monitored by two electro-oculogram (EOG) electrodes attached below and near the outher canthus of the right eye. Impedance was kept under 5 kΩ. The EEG data was digitally filtered off-line by a 1–30 Hz bandpass filter and artefact criterion was set at ±100µV. The epochs were 600 ms long including a 100 ms pre-stimulus baseline period. Average waveforms for standards and deviants were separately computed and difference waveforms were created by subtracting the response to the standard stimulus from that to the deviant stimulus. The responses for the standards after the deviants were excluded from the analyses and the minimum amount of accepted deviant stimulus trials was 80. For further mean amplitude analyses, which was determined from Fz, Cz, F3, F4, C3 and C4 electrodes (MMN is most prominent at fronto-central areas (e.g., Kujala et al., 2007; Näätänen et al., 2007)), two consecutive 50 ms time windows were selected, namely 180–230 ms and 230–280 ms. A long, 150–300 ms, time window was selected from the Cz electrode for the latency analyses. The time windows for the mean amplitude analyses were selected around the peak maxima of the grand average waveforms and the time window for the latency analysis was selected so that the peak maximum was shown in all the different language contexts in all groups. Hence, the time windows were not selected prior to data analyses, but rather the selection was performed data driven.

The bilingual participants were tested twice and there was at least a week between the two sessions, the monolinguals were tested only once. The first session consisted of the ID tests and the first EEG measurement. The second EEG measurement was conducted during the second session. The order of the Finnish and Swedish EEG measurements was counterbalanced between the subjects so that for half of them the first session was in Finnish and the second in Swedish. If the first ID test was performed in Finnish and the second in Swedish, then the first session continued with Swedish EEG recordings and the Finnish EEG recording was during the second session a week later, and vice versa.

4.3.3 Procedure and analyses in Studies III, IV and V

In **Studies III** (young adult Monolingual training study)**, IV** (young adult Monolingual vs. Sequential bilingual training study) **and V** (elderly Monolingual training study), we used a combined forced choice, self-paced ID and GR test. Fifteen stimuli were randomly repeated 8 times resulting in 120 presented stimuli. The subject first listened to the stimulus, then labelled it either as /fiːl/ or /viːl/ after which rated its goodness in a 1–7 scale where 1 was poor and 7 excellent. The subjects were told they would hear English words *'feel'* and *'veal'*. A short familiarisation block with all the stimuli once in a random order started the ID and GR task. No feedback was given at any point. The ID data were similarly analysed as the data in **Studies I and II**. The goodness rating answers were averaged for each stimulus on the continuum.

Two stimuli (7 and 10) from the stimulus continuum were selected for the oddball discrimination task on the basis of an earlier pilot ID test where native speakers of English labelled these stimuli as representatives of different categories. The representative of the /fiːl/ category was the standard and the /viːl/ was the deviant. Inter-stimulus interval was 1000 ms and there were 130 standards and 20 deviants resulting in 0.13 deviant probability. Participants were asked to press an answer button as soon as they heard the deviating stimulus. In the beginning of the discrimination block was a short familiarisation. There was no feedback in this test either. Reaction times for the deviant stimuli were calculated from the onset of the deviant and answers within ±3 standard deviation were incorporated in the analysis. The discrimination data contains hit, miss, false alarm, and correct rejection answers and the discrimination sensitivity (*d'*) values are calculated as follows: $d' = z(H) - z(F)$, where H = hits, F = false alarms[67] (Macmillan & Creelman, 1991, pp. 9–10).

The same stimulus pair was used in the MMN registration as in the discrimination test. Pre-attentive perception in young monolingual, sequential bilingual adults and older monolingual adults was observed before, during and after training by measuring the MMN and N1 responses. The EEG recording and analyses were carried out otherwise with the same specifications as **Studies I and II** but with a 650 ms ISI and a 550 ms epoch including a 50 ms pre-stimulus period. The baseline correction period started at 50 ms before the onset of the stimulus and ended at the onset of the stimulus (**Studies III and IV**) or at 71 ms after stimulus onset where the difference between the stimuli started (**Study V**).

The time windows selected for the analysis in **Studies III, IV and V** were as follows: Two consecutive time windows of 300–340 ms and 340–380 ms were selected on the basis of the maximum amplitudes in the grand average difference MMN waveforms in Fz and Cz electrodes in **Study III**. A time window of 300–340 ms was selected in **Study IV** for the MMN response. Different time windows were selected for the two groups for the N1 response; 205–225 ms for the monolinguals and 190–210 for the bilinguals. In **Study V** a 300–360 ms time window was used in

[67] "[The] *d'* is defined in terms of *z*, the inverse of the normal distribution function…The *z*-transformation converts a hit or false-alarm rate to a *z*-score, that is, to standard deviation units. A proportion of .5 is converted into a *z*-score of 0, larger proportions into positive *z*-scores, and smaller proportions into negative ones." (Macmillan & Creelman, 1991, pp. 9–10)

the mean amplitude MMN analysis in electrodes Fz, F3, and F4. Separate latency analysis was not necessary in these studies. The difference between the stimuli started at 71 ms, and hence, the late time windows. The time windows for the mean amplitude analyses were selected around the peak maxima of the grand average MMN and N1 waveforms in the different sessions in all groups. Hence, the time windows vary between the different studies as no pre-analyses selections were made.

Studies III, IV and V were training studies in which a simple, self-paced listen-and-repeat training was carried out in three days. The stimuli were the same as in the discrimination task and in the EEG measurements and they were presented so that every other stimulus was /fiːl/ and the other /viːl/. They were repeated 30 times each resulting in 60 stimuli altogether in one training session. There were four training sessions altogether and no feedback was provided during training. The participants were instructed to listen to the stimulus presented via headphones and to repeat it very carefully and then to press a button to continue to the next stimulus. The productions were recorded but this data is not included in this thesis. The training sessions lasted for few minutes each depending on the individuals' pace.

These training studies were carried out in three consecutive days so that the first day consisted of baseline measurements on the ID and GR task, the discrimination and RT task and the MMN registration, after which the first listen-and-repeat training session was carried out. The second day started with training and carried on with ID and GR task, continued with discrimination and RT task and MMN registration ending on the third training session. The third and last day was otherwise identical with the second one but there was no training at the end. The order of the discrimination and RT task and the MMN recording was counterbalanced between subjects in each group. The monolingual group was also tested once more about a year from the first testing to see whether training effects were permanent or not.

4.4 Statistical analyses

4.4.1 Statistical analyses in Studies I and II

The ID data was statistically analysed by a Group (2: Simultaneous bilinguals, Sequential bilinguals) × Context language (2: Finnish, Finland-Swedish) Repeated measures analysis of variance (ANOVA) separately for the boundary location and steepness variables. Further post hoc tests were performed to find out the cause of possible interactions. The same analysis design was carried out for the MMN latency data. The MMN amplitude data was analysed by a Group (2: Simultaneous bilinguals, Sequential bilinguals) × Context language (2: Finnish, Finland-Swedish) × Time window (2: 180–230 ms, 230–280 ms) × Electrode (6: Fz, Cz, F3, F4, C3 and C4) ANOVA. Further post hoc tests were carried on to detect the cause for any interactions. These six electrodes were chosen for the statistical analysis as the MMN is most prominent at fronto-central areas (e.g., Kujala et al., 2007; Näätänen et al., 2007), and the aim was to see whether an MMN is elicited or not and in what manner the responses are comparable between the conditions and groups, not, for example, to conduct source localisation analysis of the brain activity.

4.4.2 Statistical analyses in Studies III, IV and V

In **Study III** (young adult Monolingual training study), the category boundary location and steepness data were separately analysed as a function of Session (3). The GR data was analysed by a Session (3: first, second, third) × Stimulus (5: 3, 7, 9, 10, 13) ANOVA; the five stimuli for this analysis were selected as they were the training stimuli (7 and 10), the best exemplars of the categories (3 and 13) and the boundary stimulus with the worst rating (9). RTs and *d'* were also separately subjected to a Repeated measures ANOVA. To determine whether the MMN responses significantly differed from zero, a one-sample t-test was performed. A Session (3: first, second, third) × Time window (2: 300–340 ms, 340–380 ms) × Electrode (6: Fz, Cz, F3, F4, C3, C4) Repeated measures ANOVA was carried on to analyse the MMN responses. Further post hoc tests were carried on to find out the cause for interactions when appropriate.

In **Study IV** (young adult Monolingual vs. Sequential bilingual training study) the analyses were carried as follows: The baseline and third session of the Monolinguals were compared to the follow-up session in order to find out whether the training effects were permanent. Further, the Monolingual follow-up was compared to the Sequential bilinguals' baseline session to see whether the possible permanent training effects were comparable to the Sequential bilinguals. The within group comparisons were carried out with repeated measures ANOVAs and the between group analyses were conducted by multivariate ANOVAs. Pairwise comparisons were carried out when necessary. Separate analyses for the category boundary, steepness, *d'*, and RT were carried out with a (Group (2) ×) Session (2) analyses. The GR data, on the other hand, was analysed with a (Group (2) ×) Session (2) × Stimulus (5) analysis. The goodness ratings for the prototypes (Group 1: stimulus 3 and stimulus 13; Group 2: stimulus 2 and stimulus 13) of the categories, the trained stimuli (7 and 10) and the border (9) were used in the analyses. The ERP analyses were carried out with a (Group (2: monolinguals, bilinguals) ×) Session (2: baseline or third session, follow-up) × Electrode (6: Fz, Cz, F3, F4, C3, C4) analyses separately for MMN and N1.

In **Study V** (elderly Monolingual training study), the category boundary location and steepness data were analysed by comparing Sessions 1 and 3. The GR data were analysed with Session (2: first, third) × Stimulus (4: trained stimuli 7 and 10, the category boundary area stimuli 8 and 9) Repeated measures ANOVA. The RT and *d'* data were also analysed separately comparing the two sessions. MMN response amplitudes were analysed with a Session (2: first, third) × Electrode (3: Fz, F3, F4) Repeated measures ANOVA. The fronto-central or frontal electrodes were chosen for the statistical analysis as the MMN is most prominent at these areas (e.g., Kujala et al., 2007; Näätänen et al., 2007), and the aim was to see whether an MMN is elicited or not and in what manner the responses are comparable between the conditions and groups, not, for example, to conduct source localisation analysis of the brain activity.

5 Overview of the results of the original studies

This thesis consists of five studies focusing on the foreign language perception learning and bilingualism. On one hand, the thesis concentrates on foreign language perception learning through formal classroom learning in university language students, or sequential bilinguals, and through laboratory training in young and older adults and in sequential bilinguals. The other focus is on different bilinguals' speech processing and perception. Behavioural perception paradigms demanding attentive responses and pre-attentive electrophysiological measurements are the main methodology in this thesis. Four of the studies have been published in peer-reviewed journals, one of them has been submitted. The studies were all carried out at the University of Turku, Finland.

Studies I and II tested whether linguistic context affects behavioural and neurophysiological processing of vowels and whether different types of bilinguals – balanced and dominant, or more precisely, simultaneous and sequential bilinguals – differ from each other (M. S. Peltola et al., 2012) and whether the dominant, or sequential, bilinguals differ from monolinguals (Tamminen et al., 2013). **Studies III, IV and V** were conducted to find out whether a listen-and-repeat training has effects on behavioural identification and discrimination and on neurophysiological language processing in young monolingual adults (Tamminen et al., 2015) and in older adults (Tamminen et al., 2021). A follow-up study (**Study IV**) examined whether the training effects in the young adults were permanent and comparable to young dominant, or sequential, bilinguals (Tamminen et al., submitted).

5.1 Different kinds of bilinguals – Different kinds of brains: The neural organisation of two languages in one brain (Study I)

The roles of the two languages of Balanced and Dominant bilinguals differ since they can both be considered as native languages of the Balanced bilinguals, whereas only one language is native language in the case of the Dominant bilinguals with the other being a second language. Although the proficiency level of the second language in the Dominant bilinguals may reach the same level as the Balanced bilinguals proficiency level, the age of exposure differentiates the two bilingual types in any case. As was seen in Table 1 in Chapter 4.1.1, the self-reported proficiency level in Swedish is high in Dominant bilinguals, but not as high as in Balanced bilinguals. **Study I** concentrated on finding out whether language context affects speech processing of Balanced and Dominant bilinguals differently – whether the neural processing of speech is differently organised in different bilinguals.

According to the second language speech sound acquisition theories mentioned in Chapter 2.2, native and second language phonological systems have certain similarities and differences which cause varying amount of problems. In **Study I**, a set of closed rounded vowels were used. The area is divided differently in Finnish and Finland-Swedish since it covers only two vowels in Finnish and three in Swedish. The bilinguals have to manage the two phonological systems so that when they use Finnish, the same acoustic area covers only /y/ and /u/ and when they use Finland-Swedish it covers /ʉ/ as well, dividing the area differently. See also Chapter 4.2.1.

Both bilingual groups were tested twice, in a Finnish and in a Swedish language context. An identification test and an EEG registration measuring MMN in both languages were carried out. The hypothesis was that the groups would behaviourally perceive the stimulus continuum according to the context language. However, the Swedish /ʉ/ may interfere the Finnish category perception. An intertwined phonological system, where both languages are represented, would show interference from one language to another. The language context defines the phonological relevance of the two vowels presented in the pre-attentive discrimination measurement. The contrast was phonologically relevant in the Finnish context and an MMN response should elicit. However, in the Swedish context the contrast is irrelevant, as the vowels are representatives of the same category, and it is possible that no response elicits. In this case the phonological systems would be separate and the elicitation of a native memory trace would not be automatic. If the Swedish context elicits an MMN, the systems would be intertwined.

Both bilingual groups placed category boundaries at same locations, since there were no statistically significant differences between the groups in either language. The consistency of the category boundary was, however, different between the groups. The boundary area was less systematic in the Balanced group compared to the Dominant group.

The MMN latency was shorter in the Dominant bilinguals. The latency was 200 ms in the Finnish context and 223 ms in the Swedish context in the Dominant bilinguals, whereas the same latencies were 254 ms and 253 ms in the Balanced bilinguals, respectively. Also the MMN amplitude was different in the groups depending on the language context. The MMN response elicited contextindependently in the Balanced bilinguals whereas the response was context dependent in the Dominant bilinguals.

The ID results showed clearly that the Swedish phoneme repertoire interferes the Finnish category boundary area in the Balanced bilinguals in the attentive processing. This was shown as a less consistent boundary in the Finnish context. Even though the Dominant bilinguals have a high command in Swedish, the Swedish sound system does not interfere their native language perception at the attentive level.

Both pre-attentive and attentive processing indicates the same. The Finnish and Swedish sound systems are separate in the Dominant bilinguals but intertwined in the Balanced bilinguals. This was shown in that the MMN response latencies were greater in the Balanced bilinguals than in the Dominant ones, which further strengthens the interpretation of the intertwined system in the Balanced bilinguals. The intertwined system consists of more phonological categories, than one separate system, and hence the processing is slower. The difference between the two types of bilinguals' different processing of their two languages was shown also in the MMN amplitudes. Linguistic context did not have an effect on Balanced bilingual's speech sound perception, as they elicited similar MMN amplitudes in both contexts. Whereas Dominant bilinguals perceived speech sounds according to the language context, which was shown by existing/non-existing MMN responses according to the context[68]. Even the native-language contrast was ignored in the L2 context. It should, however, be kept in mind that there were only twelve simultaneous and ten sequential bilinguals tested. The small number of participants may have impacted the results.

The García-Sierra and colleagues' (2012) study referred to in Chapter 2.3 had similar results as ours regarding the Dominant bilinguals as responses were elicited in the conditions where the difference between the stimuli was phonemic. The description of the bilinguals in that study was complicated as they shifted from Spanish dominant to English dominant through a fairly balanced stage. However, by comparing the results to our results, it is fairly safe to say that the participants were dominant bilinguals who have two separate phonological systems for the two languages.

[68] The Sequential bilinguals were context sensitive as opposed to the results by Winkler and colleagues (2003) discussed in Chapter 2.3 Bilingualism and speech perception. The stimuli in both studies were near category boundary representatives, but the present study used individually selected stimuli that represented different categories.

5.2 Phonological processing differences in bilinguals and monolinguals (Study II)

The type of bilingualism affects speech perception as the phonological systems in different kinds of bilinguals are organised differently. The intertwined sound system in Balanced bilinguals is seen as interference at both attentive and pre-attentive levels compared to the separate systems in Dominant bilinguals in **Study I**. **Study II** concentrated on finding out whether speech processing of Balanced bilinguals differs from that of Monolingual Finns who only process language through one sound system. The monolingual Finns were tested in a same manner as the different types of bilinguals in **Study I**, but only in a Finnish language context.

Both the MMN latency and amplitude were different in the two groups. The MMN latency was greater and the amplitude smaller in the Balanced bilinguals compared to the Monolinguals. Here, as well as in **Study I**, the most probable explanation for the difference in the speech processing is the intertwined phonological system with which processing takes more time than with a phonological system of only one language. However, the participants consisted of only ten monolinguals and twelve simultaneous bilinguals, and hence, the results should be cautiously examined.

Palomar-García and colleagues (2015) also studied differences between bilingual and monolingual language processing and found that Spanish-Catalan bilinguals process their native language (Spanish) differently than Spanish monolinguals. They used fMRI and the tasks were a passive listening task and a picture-naming task, and even though there were no differences in passive listening, there were differences when they were naming objects. They concluded that participation in the left-lateralised brain areas (e.g., left middle temporal gyrus, left HG), usually involved in this kind of processing, was reduced and engagement of other areas (e.g., precuneus, right STG, dorsal ACC) was increased[69]. In other words, a wider brain area is activated in native language processing in the bilinguals and is different to that of monolinguals.

In conclusion, access to individual items in the more extensive, intertwined phonological system of the balanced bilinguals takes time. Monolinguals only have one phonological system with less exemplars which makes the access to individual items quicker. The two languages of a balanced bilingual interfere each other when processing only one language.

[69] For this thesis the specific brain areas are not important, rather organisational comparison or functional differences, but the locations are mentioned if they are of relevance in the studies referred to.

5.3 Phonetic training and non-native speech perception – New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures (Study III)

Automatically responding language-specific memory traces (Näätänen et al., 1997) are vital in speech sound perception. The native language representations evolve in the early childhood (Cheour et al., 1998). However, during the learning process of a new language, new neural representations can evolve for the foreign language speech sound categories as well (e.g., M. S. Peltola et al., 2005, 2012; M. S. Peltola & Aaltonen, 2005; Winkler et al., 1999).

Study III aimed at finding out how phonetic training of non-native speech sounds affects speech perception. The second aim was to find out whether new memory traces are generated for the foreign language speech sounds. Speech perception was measured with attentive behavioural tasks namely identification, goodness rating, reaction time and discrimination sensitivity. Further, MMN responses were measured to investigate how the pre-attentive perception reflects non-native speech processing and whether memory traces for the non-native speech sounds evolved. Young adult monolingual Finns were tested within three consecutive days with speech perception measurements. The training was carried out within these three days using a listen-and-repeat training method.

The hypothesis was that the perception of the foreign voicing contrast in fricatives, which is unfamiliar for Finns as Finnish does not differentiate sounds by voicing, would be challenging during the baseline measurements. The /v/ category was hypothesised to be smaller than /f/ in the ID test and to receive fewer good category representative ratings. Further, discrimination was hypothesised to be slower and harder, and the MMN response to be non-existent or very small at the category boundary. The training was hypothesised to result in increased /v/ category size and changes in the hierarchy of the boundary and /v/ category representatives. Also, at the category boundary, discrimination sensitivity was expected to increase while the RTs were expected to decrease as a function of training. Further, the MMN amplitude for the category boundary stimuli, consistent with the other expectations, should increase if new memory traces are evolved.

The three-day phonetic training was effective as shown by both behavioural measures and the pre-attentive MMN response. The category boundary location shifted towards the centre point of the trained stimuli in the last session, resulting in larger /v/ category. In other words, the category become phonologically more relevant. The maximum effect in the category boundary steepness was reached already after two training sessions. Also, discrimination sensitivity and RTs improved throughout training. And finally, the MMN amplitude was enhanced as the new category boundary was established during the training. The young adult monolinguals learned to perceive the difficult contrast during the short training period of only three days consisting of four training sessions without any feedback. These behavioural and neural changes indicate that new memory traces have developed. The goodness rating remained unchanged, showing no within category hierarchy effects. The fact that the GR of /f/ category did not change is not that surprising as the sound itself is probably identical with the native Finnish /f/, even though it was presented in an unfamiliar context. The unchanged GR of the /v/ was not an expected result, however, it takes probably more time than three days to learn category hierarchy. It has been suggested that memory trace formation may precede the development of a hierarchically structured category (M. S. Peltola et al., 2007). It probably also needs more variety in the stimuli and a high variability training approach (see e.g., Bradlow, 2008) could be more effective for a native-like hierarchy to be established within, and between, categories. The time course was hence more suitable for the other measures than GR to show effects. Also, a greater number than twelve participants could have been more beneficial.

The amplitude enhancement in the MMN response, which took place after only two training sessions, indicates the formation of a memory trace. This further facilitates the discrimination of the sounds (Kujala & Näätänen, 2010). The category boundary consistency, discrimination sensitivity, reaction times and MMN amplitude all changed after two training sessions without any feedback even though the voicing contrast is highly difficult for Finns. In accordance with the suggestion that the MMN indicates accuracy in perception and neural plasticity (Kujala & Näätänen, 2010), the attentive and pre-attentive discrimination changed simultaneously (for similar results see e.g., Amenedo & Escera, 2000; Kujala, Kallio, et al., 2001).

To sum up, a similar listen-and-repeat method, which is widely used at foreign language classes in schools, i.e., four few minute listen-and-repeat training sessions without any feedback, was used in this study. This training was powerful enough to show effects in both attentive and pre-attentive perception and the formation of new memory traces.

5.4 Training non-native speech sounds results in permanent plastic changes – Hard-wiring new memory traces takes time (Study IV)

Speech perception training effects have been shown in numerous studies (e.g., Kraus et al., 1995; Menning et al., 2002; Tremblay et al., 1997, 1998). Studies have also shown training effects to maintain for at least one month in behavioural performance (Kraus et al., 1995) and to increase for up to 72 hours as shown by an increase in the

MMN (Atienza et al., 2004). Further, the memory traces for the native language formed during infancy are strong enough to maintain rather stable after a long-lasting deafness (Salo et al., 2002).

Study IV was conducted to find out whether the training induced memory traces seen in the monolinguals in **Study III** persist a year later. In addition to this, the monolinguals' learning effects were compared to advanced learners of the target language, i.e., sequential bilinguals. Both monolinguals and bilinguals were tested with the same procedure as the monolinguals a year before, but without training and only once. The hypothesis was that the laboratory training induced memory traces may not be as strong as those induced by classroom learning during several years. However, the earlier induced effect could be detectable in the follow-up similarly to the study by Salo et al. (2002) showing native language memory traces after long deafness.

The MMN response induced in the original three-day training did not differ from that in the follow-up. The response was not significantly diminished within the year in between. Further, the monolinguals' MMN response in the follow-up did not differ from the bilinguals' baseline MMN. The permanent memory traces are, hence, comparable to those of the advanced language learners. The behavioural perception between the two groups did not differ when the monolingual follow-up and the bilingual baseline were compared. The baseline category steepness was, however, different between the groups as the bilinguals were more systematic in placing the boundary than the monolinguals. Within the monolinguals, there were no category boundary location or steepness differences between baseline and follow-up. The goodness rating analysis showed that the stimuli were rated differently in the monolinguals' baseline and follow-up. The difference was localised in the stimulus that represented the prototype of the /fiːl/ category. The follow-up ID and GR results were unfortunately analysed only from seven participants as only eight participants agreed to the follow-up testing and there were technical problems with one ID and GR result. The only behavioural discrimination difference was found in the reaction times between the baseline and the follow-up in the monolinguals as the RTs decreased.

What was unexpected was the change in the N1 response. The N1 remained similar during the three-day training, however, the amplitude of the response was significantly larger a year later. In addition to this, the monolinguals' follow-up N1 response was similar to the baseline response of the advanced students.

Change in the monolinguals' neural processing, the training induced memory traces, were persistent as shown by the MMN response. Further, the training induced permanent memory traces are also highly similar to the advanced learners' memory traces. Fairly similar effects of stability have been shown with discrimination training: Kraus and colleagues' (1995) training effects in the behavioural performance remained for the one-month follow-up and Atienza and colleagues' (2004) training increased the MMN response for 72 hours after the final training. The behavioural performance was similar in both groups in **Study IV**, and hence, the behavioural results support the MMN result. Only the baseline category boundary consistency differed between the monolinguals and the bilinguals, indicating that the bilinguals were more systematic than the monolinguals. Foreign sound identification was probably hindered by the native language in the monolinguals in the beginning causing less stable identification. The higher rating of the prototypical /f/ may be explained by the fact that, even though it occurred in foreign context, it is identical to the native sound (Kuhl, 1991).

The change in the N1 response in the monolinguals and the similarity of it compared to the bilinguals was striking. No further training was needed for the response to increase. This may be an indication of an ongoing nature of the learning process. Further, this could be the mechanism underlying effortless acquisition of foreign speech. An increase in the sensitivity in perception is a probable cause for the increased N1 amplitude in the monolinguals. This might denote a mechanism enabling the speech processing system to be open for further learning, for deep-rooted automatic learning. Saloranta and colleagues' (2020) research showed changes in the perception of an untrained duration contrast, as a week from the last training session, an N1 response was elicited. It was suggested that the increment might reflect sensitivity towards the duration as such since the N1 is not a linguistic response. The participant groups were twelve monolinguals, of whom only eight participated in the follow-up testing, and eleven sequential bilinguals. A greater number of participants could have made the findings more reliable.

The MMN and N1 responses reflect different memory traces (Näätänen et al., 2011) and plasticity shown by learning may be seen in the enhancement of these components (e.g., Kujala & Näätänen, 2010). Auditory discrimination may be indexed with the MMN response and the detection, not discrimination, of separate auditory stimulus features may be indexed by the N1 (e.g., Kujala & Näätänen, 2010; Näätänen et al., 2011). The MMN hence reflects the training and learning related memory traces for the foreign, or second, language, as it generates to feature and temporally integrated events (ibid.). The sensitivity growth to the previously learned and the openness of the learning process are reflected by the N1 which generates for attention-catching properties of an event (ibid.).

5.5 Aging and non-native speech perception: A phonetic training study (Study V)

Speech perception and foreign language learning are affected by aging related cognitive decline. The cognitive processes needed in speech perception and learning are both conscious and subconscious and aging may affect these processes differently. Studies comparing elderly and young adults have shown compensatory processing in elderly during various native language linguistic tasks (see Chapter 2.5). Perception of temporal features in the elderly is also reduced compared to young adults (Strouse et al., 1998; Tremblay et al., 2002). However, foreign or second language perception learning research concerning the elderly is scarce.

Study V was carried out to see whether the elderly benefit from a listen-and-repeat training and are able to learn to perceive the non-native contrast. The same procedure as in **Studies III and IV** was used. The hypothesis was that training would not be as effective for behavioural perception as it was in the young adults in **Study III** since perceiving temporal features is reduced in the elderly (Strouse et al., 1998; Tremblay et al., 2002). The pre-attentive perception was also hypothesised not to be affected by the training as much as it did the monolingual young adults in **Study III** or advanced language learners (Tamminen & Peltola, 2015).

Training was effective in the identification as the category boundary location was significantly shifted. The baseline boundary lay at the same location as the /v/ category stimulus used in the training and the post-training boundary located between the training stimuli resulting in a phonologically more relevant category. The consistency of the category boundary remained unaffected. However, the stimuli were rated differently in the sessions. The difference was found in the post-training session between the new boundary stimulus and the /v/ category training stimulus. At this point, the new boundary stimulus was rated poorer than the /v/ category representative. The RTs, discrimination sensitivity, or the MMN response remained unaffected.

The shift in the category boundary was similar to that seen in the young adult monolinguals in **Study III**. However, the elderly showed changes in the GR unlike the young adults. The GR changes suggest change in the hierarchy as the elderly rated the baseline category boundary stimulus and the post-training boundary stimulus differently in the post-training session, the new boundary was rated poorer than the former boundary. Hence, training resulted in shifted category boundary and change in the hierarchy in goodness estimates.

As the used measures constitute a perceptual continuum so that the ID and GR end covers the most attention demanding perception, *d'* and RTs are indicators of less attentive processing closer to the pre-attentive processing revealed by the MMN. In the elderly, the training effects were seen in the ID and GR. This task needs most attention, effort, and probably also some linguistic knowledge, of the tasks used in this study. Further, there was no time limit for completing it. These characteristics were probably particularly beneficial for the perception in the elderly. Although there were no training related changes in the pre-attentive discrimination, learning took place at the behavioural level, similar to the linguistically oriented elderly (Jähi et al., 2015). Further, although both the young adults and the seniors are considered to be monolinguals, the baseline English knowledge was probably quite different (e.g., see Table 2 in Chapter 4.1.2). Also in this study, the number of participants, ten whose MMN data was analysed and 9 whose behavioural data was analysed, could have been greater.

In summary, the only training related changes in the elderly were shown at the behavioural level, not at the pre-attentive level. This was opposite to the young adults. It seems that experience based linguistic knowledge benefited the elderly as the most attention demanding tasks showing training related changes.

6 General discussion

The thesis has its basis on speech perception and the functioning of the phonological systems in simultaneous and sequential bilinguals. Second interest was to find out how the memory trace formation is affected by different factors. More specifically, the aim was to look into the **neural functioning and processing of speech in simultaneous and sequential bilinguals**, to evaluate whether **non-native training shapes speech processing to become bilingual-like** and whether the effect is **permanent or not**, to examine how **different background factors** such as manner and age of learning **affect memory trace formation as a result of training**, and finally, to evaluate whether **simultaneously acquired two phonological systems versus one phonological system lead to different perceptual processing**. Each of these aims are discussed separately in the following paragraphs in the light of earlier research.

6.1 Findings in plasticity, learning, age and bilingualism

The results in **Study I** undoubtedly showed differences in **neural functioning and speech processing between the simultaneous and sequential bilinguals** (see Figure 1 on page 83). The simultaneous bilinguals were not able to switch off the context irrelevant language as the MMN elicited in both Finnish and Swedish language contexts to the Finnish contrast. Contrary, the sequential bilinguals switched off the context irrelevant language, the native language, and the MMN elicited only in the Finnish language context to the Finnish contrast. Further, the simultaneous bilinguals elicited the MMN responses with a greater latency than the sequential bilinguals did. This implies that the **simultaneous bilinguals' two phonological systems are intertwined**, and hence, they are **not able to switch off the context irrelevant language and the processing takes more time**. In contrast, it is suggested that the **sequential bilinguals have separate phonological systems for the two languages which is manifested by an ability to inhibit the context irrelevant language, even if it is the native language**. Further, the **processing was quicker** than in the simultaneous bilinguals. The fact that the language context triggers the languages differently in the different bilinguals, may imply that the pre-attentively responding native language speech sound processing in the sequential bilinguals is not perhaps that secured and may be ignored by a neural choice to turn the native language off. The behavioural identification results support the pre-attentive finding: The Finnish category boundary was not as consistent in the simultaneous as in the sequential bilinguals – the Swedish phonology disturbs the Finnish identification as the two languages are intertwined. Contrary, in picture naming (word production) the L1 seems to interfere the L2 of sequential bilinguals (Hoshino & Thierry, 2011). Although, this might be due to the difference in the task: both early and late bilinguals are suggested to perform at native-like level in L2 phonological tasks involving pre-lexical processes, such as identification of isolated phonemes (like in **Study I**), but there is a decrease in the performance in lexical processes such as selecting the appropriate words (Sebastián-Gallés & Díaz, 2012). The behavioural identification hints also towards the double phonemic boundary effect (e.g., Casillas & Simonet, 2018; García-Sierra et al., 2009; Gonzales & Lotto, 2013), albeit that was not explicitly tested in **Study I**.

As noted in Chapter 2.3, earlier studies with various linguistic tasks on early and late bilinguals have shown varying speech processing results (e.g., Chee et al., 1999; Garbin et al., 2011; Kim et al., 1997; Klein et al., 1995; Perani et al., 1996). The disparity between the aforementioned studies results probably from the different linguistic tasks, different methods, and different types of bilinguals (except Kim and colleagues (1997)). **Study I** in the present thesis tested two different kinds of bilinguals with precisely the same procedure to measure speech sound perception and resulted showing that the **language context triggers the languages differently in the different bilinguals**. The functional separateness of the phonological sound systems in the sequential bilinguals may be a result of inhibition or distinct cortical brain areas (Garbin et al., 2011; Kim et al., 1997; Perani et al., 1996) responsible of the processing of the two languages. A confusing language context[70] does not elicit a MMN response for the native language contrast, whereas a L2 context elicits a response for the same contrast, which represents the same vowels in the L2 but are poorer representatives of the categories (M. S. Peltola & Aaltonen, 2005). It also seems that early bilinguals and relatively proficient sequential bilinguals, but not less proficient learners, are able to perceive language-specifically in a mixed language context, where the standard stimulus functions as a hint of the context-language (Casillas & Simonet, 2018). The fact that an intertwined phonological system is suggested for the simultaneous bilinguals incorporates the idea that there are separate speech sound categories for both languages, not within-category variants, similarly

[70] One group of sequential bilinguals received instructions that they would hear either their native language or their L2 vowels and the other group of sequential bilinguals received instructions that they would hear L2 vowels. (M. S. Peltola & Aaltonen, 2005)

as suggested by Molnar and colleagues (2014). Partly supporting the findings of **Study I**, García-Sierra and colleagues (2012) showed context-sensitivity in bilinguals who were definitely not simultaneous but probably somewhere in between simultaneous and sequential bilinguals as their language profile had changed quite a bit. The early bilinguals' (although not simultaneous as the L2 was acquired at the age of five or later) languages are also suggested to be interactive as their native language phonology and phonotactic constraints are accessible during L2 comprehension (Freeman et al., 2016). This hints towards the intertwined system as well. Highly fluent, very early bilinguals also seem to accept tokens pronounced incorrectly as correct words[71], as they have developed a procedure with dual-mapping between accurate phonetic encoding and accurate lexical representations (Samuel & Larraza, 2015). This acceptance was found not to be a failure in acoustic-phonetic processing, nor was it because of them having two different exemplar-based lexical representations for words. It could be that the intertwined phonological system permits/allows more variation and incorrectness in the pronunciation.

In summary, the suggested two intertwined phonological systems are constantly active and hence responding even in irrelevant language context. For the same reason, the reacting takes more time compared to only one active system. The suggested two separate phonological systems, on the other hand, activate separately, when needed, and according to the language context. Therefore, the response is quicker compared to the intertwined system. Further, even the native language can be ignored altogether. In other words, **the intertwined and separate systems, and the fact that the sequential bilinguals can ignore the L1, suggest that the simultaneous and sequential bilingual speech perception processing is differently organised**.

The second aim of the thesis was to evaluate whether **non-native training shapes speech processing to become bilingual-like** and whether the training effects are **permanent or not**. As **Study III** showed, new memory traces were elicited as a result of a listen-and-repeat training. Training of the theoretically challenging (e.g., Best & Strange, 1992; Best & Tyler, 2007; Colantoni et al., 2015; Escudero, 2005; Flege, 1987b, 2007; Flege & Bohn, 2021; Tyler, 2019) speech sound contrast did not just elicit a MMN response indicating plasticity and the formation of new memory traces (e.g., Näätänen et al., 2019) in the Monolinguals, but the phonological significance of the training was seen also in the behavioural perception. The present training was production training only with not extra load from any additional training

[71] The participants were L1 Spanish very early Spanish-Basque bilinguals (Basque learned at age 3), and stimuli were real Basque words and non-words where one phonetic feature of one segment in a real Basque word was changed, e.g., the place of articulation in a nasal sound.

procedure disturbing the perception (as in Baese-Berk & Samuel, 2016, see Chapter 2.2). The training effects on the behavioural perception were seen in particular in the shifting of the phoneme category boundary location, as the post-training boundary located almost in the middle between the trained stimuli. However, the within category hierarchy was not affected by the training. Adjustment in the category hierarchy may need more time and perhaps a different type of training. The HVPT with varying stimuli along the whole continuum could probably lead into the formation of a hierarchical structure of a learned phoneme category. Although the learning of foreign language speech sounds and the related memory trace formation after the acquisition of the mother tongue is difficult (Kuhl et al., 2008), it is definitely possible. An authentic setting is typically the most ideal learning environment and native-like perception has been shown in immigrant (Tsukada et al., 2005; Winkler et al., 1999) and immersion program (Cheour, Shestakova, et al., 2002; M. S. Peltola et al., 2005; Shestakova et al., 2003) studies. Even though learning in classroom may not lead to native-like perception and memory trace formation (Grimaldi et al., 2014; Hisagi et al., 2016; Jost et al., 2015; M. S. Peltola et al., 2003), foreign speech sound perception and production training with various methods have shown perceptual learning effects (e.g., Grenon et al., 2019; Kartushina et al., 2015; Kraus et al., 1995; Menning et al., 2002; K. U. Peltola et al., 2020; Saloranta et al., 2020; Tamminen & Peltola, 2015; Tremblay et al., 1997, 1998; Tremblay & Kraus, 2002).

Study IV further demonstrated the **persistent MMN response a year later indicating that the training effect was indeed permanent**. The MMN response was similar a year after training compared to the response immediately after the initial training sessions. It was hypothesised that the memory traces formed during laboratory training a year before would not be as strong as the Sequential bilinguals' memory traces that are formed as a result of long-lasting classroom learning. Especially when, as implied in a study by Flege and MacKay (2004), even the native language seems to need constant use to remain intact. The result of **Study IV**, however, showed that **speech processing had become bilingual-like** (see Figure 1) as the Monolinguals' follow-up MMN response was similar to the Sequential bilinguals. The neurophysiological findings were supported by the behavioural measures – identification and discrimination were similar a year after training in the monolinguals compared to the sequential bilinguals. Similar permanent nature of the memory traces of native language has been shown after long deafness (Salo et al., 2002). Consistent with the result in **Study IV**, earlier training studies have shown that the training effects in behavioural tasks can last for a month (Kraus et al., 1995) or for a year even (Escudero & Williams, 2014). The amount of imperceptible English language input in the native Finnish context seemed to be sufficient to sustain the memory trace acquired through listen-and-repeat training.

The **ongoing nature of the learning process of the monolinguals was strikingly shown by the increased N1 response in the follow-up session**. Even more surprising was the fact that the N1 response, **reached the same level with the three-day listen-and-repeat training and the continued environmental input during the year in the Monolinguals as the extensive studies and excellent proficiency of the Sequential bilinguals**. This further strengthens the finding that training induced learning and classroom learning result in similar speech processing. The increased N1 response may indicate a more increased sensitivity (see also e.g., Brattico et al., 2003; Tremblay et al., 2001) which enables the speech processing apparatus to be alert for further learning. This denotes some sort of mechanism behind effortless speech acquisition. In this manner, the deep-rooted learning is possible and occurs automatically. In other words, enhanced sensitivity caused by training, as well as by prolonged studies, is seen in the increment of the N1 response. In similar lines, a speech sound duration training study by Saloranta and colleagues (2020) demonstrated an N1 response elicitation a week from the last training session for an untrained linguistic contrast. This was suggested to be an indication of the brain becoming more sensitive for the contrast.

The listen-and-repeat training method, which is widely used in schools to practice and learn challenging non-native language speech sound contrasts, is evidently effective as new memory traces evolve in such a short time and with a fairly brief training. Training effects may transfer to stimuli that are not trained within the training paradigm. For example, the trained effect may transfer to CV syllable stimuli with different place of articulation in the consonant (Tremblay et al., 1997). Even though the transfer effect was not tested in the studies of this thesis, it is reasonably fair to argue that the gained effect would generalise to other contrasts. For example, for a native Finnish speaker the English /s/ – /z/ contrast is probably equal to the /f/ – /v/ contrast, and equally difficult to learn (Best & Strange, 1992; Best & Tyler, 2007; Colantoni et al., 2015; Flege, 1987b; Flege & Bohn, 2021; Tyler, 2019).

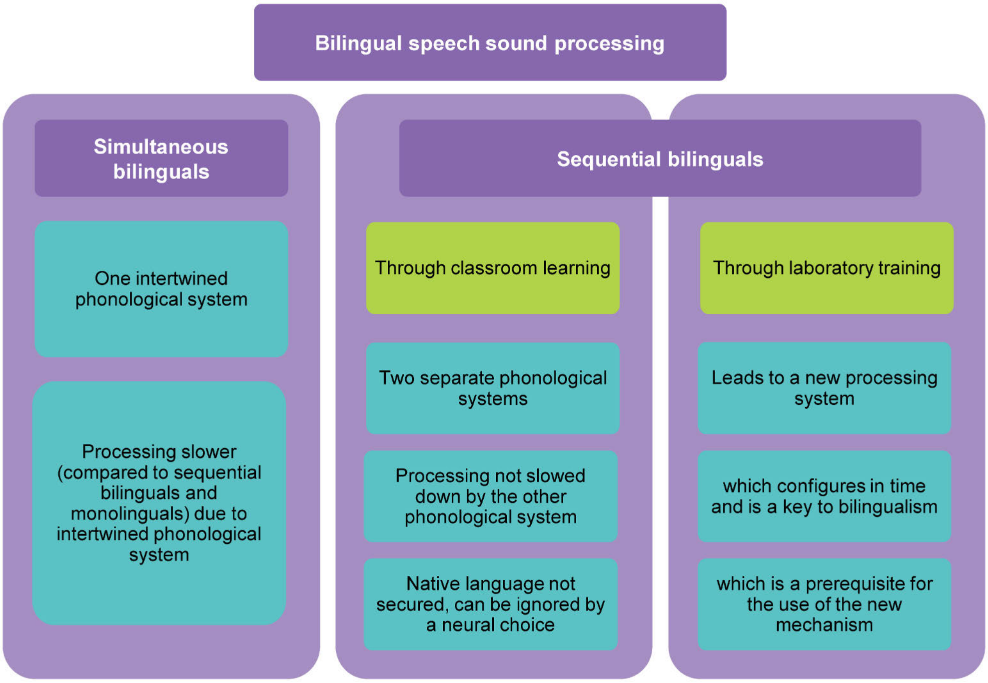

Figure 1. Bilingual speech sound processing in simultaneous bilinguals and in sequential bilinguals with different learning routes.

We then move on to the third aim – **how different background factors affect the formation of memory traces through training**. The background factors were **manner and age of learning**. Manner of learning refers to learning through classroom teaching or learning through an intensive laboratory training. Age of learning refers to three situations: young adults who have been learning the language in question from elementary or secondary school, young adults learning through laboratory training, and elderly people learning through laboratory training. Let us first look at the manner of learning. **Studies I and IV** introduced the young adult sequential bilinguals who have been trained at school learning the language through classroom teaching and **Studies III and IV** introduced the young adult monolinguals who learned through laboratory training. As was seen earlier, **learning through classroom teaching seems to be effective enough for memory trace formation** (**Studies I and IV**, Sequential bilinguals) (see Figure 2 on page 85), and further, even the native language phonological contrast is ignored in the L2 context if the contrast is phonologically irrelevant in the L2 (**Study I**, Sequential bilinguals). This functioning of the brain, that the native language can be ignored, is a clear indication of deep-rooted learning of the L2 and it might also indicate that the pre-attentive processing is prone to neural choice according to the language context irrespective of the native language. However, the attentive native language speech processing is not affected by the high proficiency in the L2 in the sequential bilinguals, as the native language identification is not affected by the second language sound category similarly as in Simultaneous bilinguals. L2 effects on the mother tongue identification and discrimination have, however, been shown, for example, in highly proficient bilinguals exposed to the L2 at the age of 4–5 years when entering to school (Mora & Nadeu, 2012). In this case, the age of exposure may be a key difference as the bilinguals in **Study I** exposed to the L2 approximately at the age of thirteen. Further, as was seen above the **laboratory training induced memory traces** (**Study III**) **which are also permanent** (**Study IV**) **appear to be comparable to those formed through classroom learning** in the sequential bilinguals (**Study IV**) (see Figure 2). **Neural plasticity hence enables learning and memory trace formation both through classroom learning and through laboratory training**.

The other background factor, age of learning, showed that **both classroom training that has started from a young age** (and which led to sequential bilingualism) (**Studies I and IV**) **and laboratory training in adulthood** (**Study III**) **induce memory traces**, even though continuous learning from a young age is different from laboratory training. However, the **laboratory training in elderly did not lead to memory trace formation** (**Study V**) (see Figure 2). There are a few explanations for the different age effects on the memory trace formation. Even the native language linguistic processing is affected by aging (e.g., Dennis & Cabeza, 2011; Getzmann et al., 2015; Geva et al., 2012; Wong et al., 2009). Also, the perception of temporal features in speech is weakened (Strouse et al., 1998; Tremblay et al., 2002) and re-orienting toward a relevant speaker takes more time (Getzmann et al., 2015) compared to younger adults. Even though aging has some undesirable effects on language processing, learning and training improve language related functions and cognitive functions, and hence, it seems rational to design training procedures that are more suitable for the elderly. The training could be slower in rate, there could be more training sessions and/or more training per session, and guidance between training sessions. For example, tasks requiring memorisation, rote and speed may not be the most suitable alternatives (ibid.). In addition, the acoustic complexity and acoustic difference between the stimuli should be taken into consideration in the training procedures for the elderly. One possible explanation for the lack of training effects on the pre-attentive perception is that, compared to younger people, the elderly may have difficulties in extracting fine temporal detail in speech (Bellis et al., 2000; Strouse et al., 1998; Tremblay et al., 2002), not to mention the theoretical difficulty (e.g., Best & Strange, 1992; Flege, 1987b) of the stimuli. The better-suited training methods for the elderly are important as the older adult brain is plastic and functions related to language can be improved by language learning, and importantly, because training also improves cognitive functions (Antoniou et al., 2013). In addition, the improvement of cognitive functions is important, as the slightly declined hearing sensitivity can worsen by reduced cognitive abilities (Anderson & Kraus, 2013). The neural plasticity is not at the same level in the elderly compared to the young adults, and hence, the memory traces did not form in the elderly with the same training method as in the young adults. The elderly, nevertheless, showed training effects in the most attention demanding behavioural task that requires effort and prior linguistic knowledge.

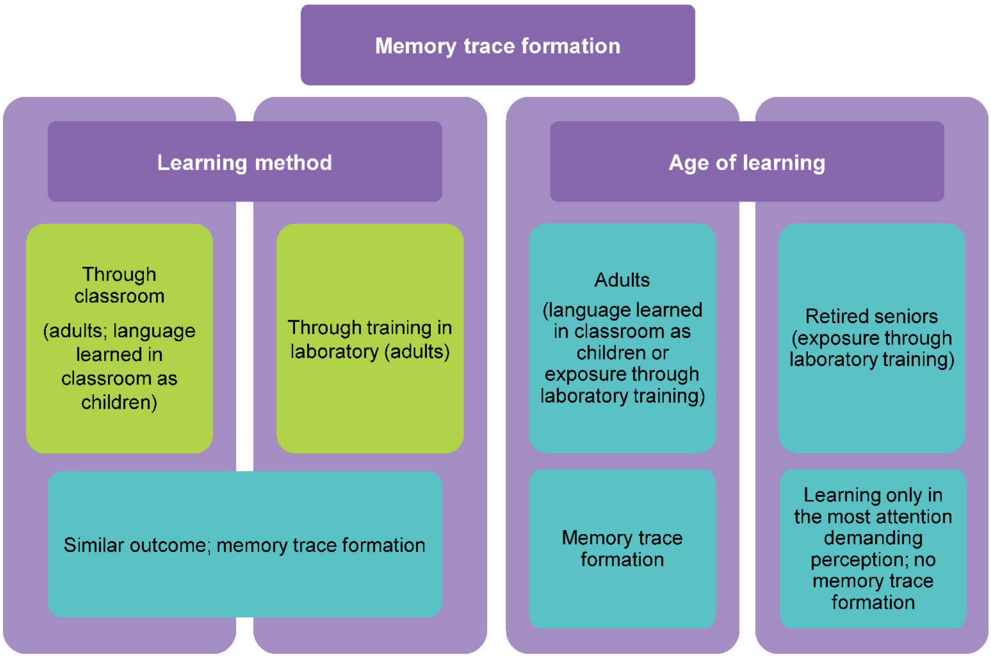

Figure 2. Memory trace formation as an effect of manner and age of learning.

The last aim in this thesis was to compare the neural processing of simultaneous bilinguals and monolinguals – to evaluate whether perceptual processing differs when operating with two intertwined phonological systems compared to operating with only one monolingual system. Although bilinguals and monolinguals process non-linguistic stimuli similarly (Nenonen et al., 2003; Ortiz-Mantilla et al., 2010), Study II showed a clear result, that the neural processing of speech sounds is different between the simultaneous bilinguals and the monolinguals (see Figure 1). It was suggested that, if the Simultaneous bilinguals could be able to switch off the language that is not required in the current context, a similar MMN response could be elicited as in the Monolinguals. On the other hand, in the case of both languages being active all the time, the target language processing might be affected by the non-target one. The larger latency and smaller amplitude of the MMN response in the bilinguals suggest, again (as in Study I), that the two intertwined languages of the bilinguals affect speech processing. The amount of phonological categories is larger in the bilinguals and to retrieve the correct match takes more time than in monolinguals with only one set of speech sound categories. The interference from the other native language of the bilinguals is seen also in the slightly reduced MMN amplitude. The cortical thickness in simultaneous bilinguals and monolinguals is shown to be similar and it is suggested that acquiring languages simultaneously does not have any additional effect on the development of the brain (Klein et al., 2014). Yet, the functional processing seems to be different between the simultaneous bilinguals and monolinguals.

Both an increased deviation of the stimuli (e.g., Näätänen et al., 2005) and the training of non-native contrasts (Menning et al., 2002) are shown to decrease the MMN latency. However, two native languages and the greater number of phonological categories in the simultaneous bilinguals, as opposed to one native language in the monolinguals, increases the MMN latency unlike training. Access to the individual phonological items and memory traces takes more time, as suggested also by Study I. Similar slowing effect was shown for example in a lexical learning study (Davis et al., 2009) where the newly-learned and existing similar-sounding words, constituting a more extensive lexical inventory, slowed down lexical access and memory trace retrieval. In addition, the interference of the other language in the simultaneous bilinguals may explain the tentative reduced MMN amplitude (a non-native-like response) result, implying the activeness of both languages during the processing of only one language.

The main findings of this thesis are: The different acquisition and learning routes lead to differences in the neural processing of speech perception when the two languages are acquired from birth like in simultaneous bilinguals or when the L2 is learned later in life like in sequential bilinguals or when only one language is acquired from birth like in monolinguals. In addition, neural plasticity enables similar speech sound processing via non-native laboratory training and classroom learning, and the effect of the laboratory training is permanent. However, the older age affects memory trace formation, as the laboratory training that induces memory traces in young adults does not elicit memory traces in the elderly.

6.2 Future directions and study adjustments

The methodologies employed in the studies of this thesis were carefully designed; however, a few methodological observations emerged when reflecting the results. The procedure in the training studies (**Studies III, IV, and V**) offered training other than the listen-and-repeat training within the course of the testing itself. There was exposure to the whole stimulus continuum in the ID and GR task and to the training stimuli within the discrimination task and the psychophysiological measures. It could be argued that the whole research protocol affected the results providing the additional exposure and that the training results were elicited through mere exposure to the stimuli. The follow-up study for the young adults, however, proved the training paradigm to be effective as the changes in the behavioural and psychophysical perception were long lasting. In other words, training with a small amount of repetition without feedback, a similar training used in schools, resulted in new permanent memory traces and improvement in speech perception. Also, it was unfortunate that only eight participants agreed to take part in the follow-up and, due to technical problems, the ID and GR data was analysed only from seven participants. Furthermore, the sample sizes of all the studies could have been larger for the results to be more reliable.

As usual, new ideas for research paradigms and procedures manifest along the data collection and analysis and while pondering the results. The language context affects the simultaneous and sequential bilinguals differently as was shown in **Study I** where both groups were tested in two monolingual contexts, the native languages of the simultaneous bilinguals and the L1 and L2 of the sequential bilinguals. Sequential bilinguals also react differently in a confusing language context compared to L2 context (M. S. Peltola & Aaltonen, 2005). A natural continuum would be a mixed language context presented to matching simultaneous and sequential bilinguals as in **Study I**. Native speaker researchers of each language communicating only with their native language would create the mixed context. Casillas & Simonet's behavioural perception study (2018) created the language context with the non-word stimuli where the non-target syllable created the language context, whereas communication took place only in one language. The study showed that early bilinguals (both simultaneous and sequential) and proficient, but not less proficient, learners perceived the words according to the context, displaying the double phonemic boundary. The mixed language context created by communication in both languages testing the speech perception processing at the pre-attentive level would constitute a perfect continuum to **Study I**, and **Study II** as well. This paradigm would show, first, whether speech processing functions similarly in a mixed context as in fixed, one language context, and, second, whether the different types of bilinguals indicate different functional speech processing in mixed context similarly as in fixed language context.

Further, a different kind of training could benefit the goodness rating and the category hierarchy development as it may need a longer period than three days to develop. Also, a HVPT (e.g., Bradlow, 2008) with exposure to a multitude of varying stimuli would probably be a better suited task in order to create a native-like category hierarchy. However, the current training studies used the production training as the training method, and hence, there was no need for another one.

In order for the results of the training studies in this thesis to be comparable, the same training paradigm was used for the elderly and for the young adults. However, another future research would be to establish better-suited training paradigms for the elderly. As suggested earlier, they would probably benefit from a different type of training than the young adults and also the stimuli need to be considered and chosen carefully. In addition, the foreign language learning and training research on speech perception and phonological processing in the elderly needs to be increased.

As second language teaching in Finland starts at the first grade when children are 7 years of age, it would be interesting to see how the listen-and-repeat training affects them. As the young brain is very plastic, the memory trace formation would most probably be at least as fast as in the young adults. However, the behavioural perception might not proceed at the same rate as in the young adults, as children are somewhat unsophisticated and immature language learners as some helpful cognitive skills such as the ability to abstract, generalise, infer, and classify, are not fully acquired (Grosjean, 2010, p. 185; McLaughlin, 1992). Further, at least in controlled experimental conditions, children often perform worse than adults (except often in pronunciation children outperform adults (McLaughlin, 1992).

The last future research proposition concerns the production data gathered from the training studies (**Studies III, IV, and V**). That data were delimited outside the scope of this thesis as the focus was on speech perception and the production training merely functioned as a training method. However, the data should be analysed in the future. The three groups – young naïve Finnish monolinguals, young Finnish proficient target language students, and elderly naïve Finnish monolinguals – provide a broad series of research questions: What is the starting point for each group? Do the groups differ in the pronunciation during the first training session? Do the different groups evolve in pronouncing the stimuli during the course of the training? How do the groups compare to each other?

7 Conclusions

There were two main aims for this thesis: The first was to study bilingual speech processing. The second was to investigate memory trace formation in the light of different background factors.

Earlier research, with varying methods and stimuli, has provided a large and varying picture of how the two languages of different bilinguals are organized in the brain. The present studies concentrated on both simultaneous and sequential bilinguals who showed differences in neural functioning and in neural mechanisms: The **simultaneous bilinguals have one intertwined phonological system where both languages are active all the time which further slows the processing down** – compared to both sequential bilinguals and monolinguals. Whereas the **sequential bilinguals' two phonological systems are separate, and even the dominant native language can be ignored** in a language context irrelevant for it. This enables quicker processing compared to the intertwined system. These two types of bilinguals hence have different mechanisms for speech processing; one is pre-attentively more flexible enabling both languages at the same time and the other activates only one language at a time. Further, the **inactivation of the native language in the sequential bilinguals implies that the pre-attentive native language speech sound processing is not that secured as a neural choice may turn off the native language**. In addition, as a result of listen-and-repeat training, familiar from classroom teaching, the brain is able to process language in two different manners as the new memory traces evolve. The **new processing mechanism accomplished through training is a key to (sequential) bilingualism**.

Also earlier training research has provided a considerable amount of varying results with, for example, different amount of training and varying methods including feedback. The studies in this thesis investigated how manner and age of learning affect the formation of memory traces. The **training method that is familiar from classroom teaching**, namely listen-and-repeat training, **is very effective as memory traces evolved in just three days** after a fairly small amount of repetitions. Further, the training **effects were permanent and neural processing was comparable to that of sequential bilinguals trained in classroom**. In other words, laboratory training and classroom training result in similar speech perception processing. However, where **age of learning did not affect memory trace formation within the young adults,** who had trained either in classroom since childhood or in laboratory in adulthood, as training resulted in similar processing, **the elderly did not benefit from the laboratory training**.

The findings indicate that **different types of bilinguals process speech perception differently** and **simultaneous bilinguals speech processing is different from monolinguals'** speech processing. The findings also suggest that **training in classroom and training in laboratory lead to similar functioning of memory traces**. And further, **the elderly brain does not show similar plasticity as young adults** concerning training. Hence, **neural and behavioural plasticity in speech perception** enable **learning**, exhibits differently according to **age**, and is necessary for **bilingualism**.

Abbreviations

ANOVA	analysis of variance
AOA	age of acquisition
AOE	age of exposure
CAH	Contrastive Analysis Hypothesis
CP	categorical perception
CPH	Critical Period Hypothesis
d’	discrimination sensitivity
EEG	electroencephalogram
EOG	electro-oculogram
ERP	event-related potential
F0	fundamental frequency
F1, F2, F3	first, second, third formant
FLMP	Fuzzy Logical Model of Perception
fMRI	functional magnetic resonance imaging
FTE	Full Time Equivalent
GR	goodness rating test
HVPT	high-variability perceptual training
ID	identification test
ISI	inter stimulus interval
L1	first language, mother tongue
L2	second language, also second native language
L2LP	Second Language Linguistic Perception
LOR	length of residence
MEG	magnetoencephalography
MMF	mismatch negativity field
MMN	mismatch negativity
NLNC	native language neural commitment
NLM	Native Language Magnet model
NLM-e	Native Language Magnet model, expanded
PAM	Perceptual Assimilation Model
SC	Single-Category assimilation

TC	Two-Category assimilation
CG	Category-Goodness difference in assimilation
UC	Uncategorised-Categorised
UU	Uncategorised- Uncategorised
NA	Non-Assimilable
PAM-L2	Perceptual Assimilation Model of Second Language Speech Learning
PBE	phoneme boundary effect
PET	positron emission tomography
PME	perceptual magnet effect
RT	reaction time
SLM	Speech Learning Model
SLM-r	Revised Speech Learning Model
TL	target language
VOT	voice onset time

List of References

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, *63*, 75–86. https://doi.org/10.1016/j.wocn.2017.05.002

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica*, *128*(3), 466–478. https://doi.org/10.1016/j.actpsy.2008.03.014

Albert, M. L., & Obler, L. K. (1978). *The Bilingual Brain. Neurophysiological and Neurolinguistic Aspects of Bilingualism*. Academic Press, Inc.

Amenedo, E., & Escera, C. (2000). The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *European Journal of Neuroscience*, *12*(7), 2570–2574. https://doi.org/10.1046/j.1460-9568.2000.00114.x

Anderson, S., & Kraus, N. (2013). Auditory Training: Evidence for Neural Plasticity in Older Adults. *Perspectives on Hearing and Hearing Disorders. Research and Research Diagnostics*, *17*, 37–57. https://doi.org/10.1044/hhd17.1.37

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, *38*(4), 640–653. https://doi.org/10.1016/j.wocn.2010.09.005

Antoniou, M., Gunasekera, G. M., & Wong, P. C. M. (2013). Foreign language training as cognitive therapy for age-related cognitive decline: A hypothesis for future research. *Neuroscience and Biobehavioral Reviews*, *37*(10), 2689–2698. https://doi.org/10.1016/j.neubiorev.2013.09.004

Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics*, *40*(4), 582–594. https://doi.org/10.1016/j.wocn.2012.05.005

Antoniou, M., & Wong, P. C. M. (2016). Varying irrelevant phonetic features hinders learning of the feature being trained. *The Journal of the Acoustical Society of America*, *139*(1), 271–278. https://doi.org/10.1121/1.4939736

Asu, E. L., Schötz, S., & Kügler, F. (2009). The acoustics of Estonian Swedish long close vowels as compared to Central Swedish and Finland Swedish. In P. Branderud & H. Traunmüller (Eds.), *FONETIK 2009, The XXIIth Swedish Phonetics Conference* (pp. 54–59).

Atienza, M., & Cantero, J. L. (2001). Complex sound processing during human REM sleep by recovering information from long-term memory as revealed by the mismatch negativity (MMN). *Brain Research*, *901*(1–2), 151–160. https://doi.org/10.1016/S0006-8993(01)02340-X

Atienza, M., Cantero, J. L., & Stickgold, R. (2004). Posttraining sleep enhances automaticity in perceptual discrimination. *Journal of Cognitive Neuroscience*, *16*(1), 53–64. https://doi.org/10.1162/089892904322755557

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, and Psychophysics*, *81*(4), 981–1005. https://doi.org/10.3758/s13414-019-01725-4

Baese-Berk, M. M., & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, *89*, 23–36. https://doi.org/10.1016/j.jml.2015.10.008

Bellis, T. J., Nicol, T., & Kraus, N. (2000). Aging affects hemispheric asymmetry in the neural representation of speech sounds. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *20*(2), 791–797. https://doi.org/10632608

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In H. C. Nusbaum & J. C. Goodman (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). MIT Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Press.

Best, C. T., & McRoberts, G. W. (2003). Infant Perception of Non-Native Consonant Contrasts that Adults Assimilate in Different Ways. *Language and Speech*, *46*(2–3), 183–216. https://doi.org/10.1177/00238309030460020701

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775–794. https://doi.org/10.1121/1.1332378

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360. https://doi.org/10.1037/0096-1523.14.3.345

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*, 305–330.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). John Benjamins.

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, *19*(2), 290–303.

Binder, J. (2000). The new neuroanatomy of speech perception. *Brain*, *123*(12), 2371–2372. https://doi.org/10.1093/brain/123.12.2371

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America*, *66*(4), 1001–1017.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowl environments. *Journal of the Acoustical Society of America*, *67*(2), 648–662.

Boersma, P. (1998). *Functional phonology*. Holland Academic Graphics.

Bradlow, A. R. (2008). *Training non-native language sound patterns* (J. G. Hansen Edwards & M. L. Zampini (eds.); pp. 287–308). John Benjamins Publishing Company.

Brandmeyer, A., Desain, P. W. M., & McQueen, J. M. (2012). Effects of native language on perceptual sensitivity to phonetic cues. *NeuroReport*, *23*(11), 653–657. https://doi.org/10.1097/WNR.0b013e32835542cd

Brattico, E., Tervaniemi, M., & Picton, T. W. (2003). Effects of brief discrimination-training on the auditory N1 wave. *NeuroReport*, *14*(18), 2489–2492. https://doi.org/10.1097/01.wnr.0000098748.87269.a1

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180. https://doi.org/10.1159/000261913

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, *33*(3), 433–461. https://doi.org/10.1017/S0272263111000040

Burgaleta, M., Sanjuán, A., Ventura-Campos, N., Sebastian-Galles, N., & Ávila, C. (2016). Bilingualism at the core of the brain. Structural differences between bilinguals and monolinguals revealed by subcortical shape analysis. *NeuroImage*, *125*, 437–445. https://doi.org/https://doi.org/10.1016/j.neuroimage.2015.09.073

Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics*, *7*(3), 207–239. https://doi.org/10.1017/S0142716400007542

Casillas, J. V., & Simonet, M. (2018). Perceptual categorization and bilingual language modes: Assessing the double phonemic boundary in early and late bilinguals. *Journal of Phonetics*, *71*, 51–64. https://doi.org/https://doi.org/10.1016/j.wocn.2018.07.002

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, *40*(2), 249–268. https://doi.org/10.1016/j.wocn.2011.10.007

Chee, M. W. L., Caplan, D., Soon, C. S., Sriram, N., Tan, E. W. L., Thiel, T., & Weekes, B. (1999). Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron*, *23*(1), 127–137. https://doi.org/10.1016/S0896-6273(00)80759-X

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351–353. https://doi.org/10.1038/1561

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. (2002). Speech sounds learned by sleeping newborns. *Nature*, *415*(6872), 599–600.

Cheour, M., Shestakova, A., Alku, P., Ceponiene, R., & Näätänen, R. (2002). Mismatch negativity shows that 3–6-year-old children can learn discriminate non-native speech sounds within two months. *Neuroscience Letters*, *325*, 187–190. https://doi.org/10.1016/S0304-3940(02)00269-0

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech: Theory and practice*. Cambridge University Press.

Crystal, D. (1997). *The Cambridge encyclopedia of language* (2nd ed.). Cambridge University Pres.

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *J Cogn Neurosci*, *21*(4), 803–820. https://doi.org/10.1162/jocn.2009.21059

Dehaene-Lambertz, G., & Baillet, S. (1998). A phonological representation in the infant brain. *Neuroreport*, *9*(8), 1885–1888. https://doi.org/10.1097/00001756-199806010-00040

Denes, P. B., & Pinson, E. N. (1963). *The speech chain: The physics and biology of spoken language*. Bell Telephone Laboratories.

Dennis, N. A., & Cabeza, R. (2011). Age-related dedifferentiation of learning systems: An fMRI study of implicit and explicit learning. *Neurobiology of Aging*, *32*(12), 2318.e17-2318.e30. https://doi.org/10.1016/j.neurobiolaging.2010.04.004

Díaz, B., Baus, C., Escera, C., Costa, A., & Sebastián-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(42), 16083–16088. https://doi.org/10.1073/pnas.0805022105

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, *1*(2), 121–144.

Diehl, R. L., Kluender, K. R., & Walsh, M. A. (1990). Some auditory bases of speech perception and production. In W. A. Ainsworth (Ed.), *Advances in speech, hearing and language processing* (pp. 243–267). JAI Press Ltd.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech Perception. *Annual Review of Psychology*, *55*(1), 149–179. https://doi.org/10.1146/annurev.psych.55.090902.142028

Diller, K. C. (1970). "Compound" and "coordinate" bilingualism: A conceptual artifact. *Word*, *26*(2), 254–261.

Elman, J. L., Diehl, R. L., & Buchwald, S. E. (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America1*, *62*(4), 971–974. https://doi.org/10.1121/1.381591

Ervin, S. M., & Osgood, C. E. (1954). Second language learning and bilingualism. *Journal of Abnormal and Social Psychology*, *Supplement*, 139–146.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Utrecht University & LOT.

Escudero, P. (2009). The linguistic perception of SIMILAR L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in perception* (pp. 151–190). Mouton de Gruyter.
Escudero, P. (2012). Speech processing in bilingual and multilingual listeners. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 406–417). Oxford University Press.
Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*, 551–585. https://doi.org/10.1017/S0272263104040021
Escudero, P., & Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition*, *133*(2), 408–413. https://doi.org/10.1016/j.cognition.2014.07.002
Ewald, O., Asu, E. L., & Schötz, S. (2017). The formant dynamics of long close vowels in three varieties of Swedish. *Proceedings of Interspeech 2017*, 1412–1416. https://doi.org/10.21437/Interspeech.2017-1134
Faris, M. M., Best, C. T., & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America*, *139*(1), EL1–EL5. https://doi.org/10.1121/1.4939608
Flege, J. E. (1987a). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47–65.
Flege, J. E. (1987b). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. https://doi.org/10.1016/S0095-4470(19)30537-6
Flege, J. E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders, a review* (pp. 224–401). Ablex.
Flege, J. E. (1992). The intelligibility of English vowels spoken by British and Dutch talkers. In R. D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (pp. 157–232). John Benjamins Publishing Company.
Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America*, *93*(3), 1589–1608. https://doi.org/10.1121/1.406818
Flege, J. E. (1995a). 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). York Press. https://doi.org/10.1111/j.1600-0404.1995.tb01710.x
Flege, J. E. (1995b). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, *16*, 425–442.
Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. S. Meyer & N. O. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319–355). Mouton de Gruyter. https://doi.org/10.1037/0096-1523.31.5.912
Flege, J. E. (2005). Origins and development of the Speech Learning Model. In *Keynote lecture presented at the 1st ASA Workshop on L2 Speech Learning*.
Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 353–381). Mouton de Gruyter.
Flege, J. E. (2021). New methods for second language (L2) speech research. In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 119–156). Cambridge University Pres.
Flege, J. E. (1999). The relation between L2 production and perception. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of The XIVth International Congress of Phonetic Sciences* (Vol. 2, Issue August, pp. 1273–1276).
Flege, J. E., Aoyama, K., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r) Applied. In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 84–118). Cambridge University Pres.

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). Cambridge University Press.

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*, 169–186. https://doi.org/10.1006/jpho.1996.0040

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving Vowels in a Second Language. *Studies in Second Language Acquisition*, *26*(1), 1–34. https://doi.org/10.1017/S0272263104261010

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. https://doi.org/10.1121/1.428116

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.

Fowler, C. A. (2003). Speech production and perception. In I. B. Weiner (Ed.), *Handbook of psychology* (pp. 237–266). John Wiley & Sons. https://doi.org/10.1002/0471264385.wei0409

Fowler, C. A., & Rosenblum, L. D. (1991). The perception of phonetic gestures. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 102–117). Lawrence Erlbaum Associates.

Francis, W. S. (2005). Bilingual semantic and conceptual representation. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism: Psycholonguistic Approaches* (electronic, pp. 250–267). Oxford University Press. https://login.ezproxy.utu.fi/login?url=http://site.ebrary.com/lib/uniturku/Doc?id=10233717

Freeman, M. R., Blumenfeld, H. K., & Marian, V. (2016). Phonotactic constraints are activated across languages in bilinguals. In *Frontiers in Psychology* (Vol. 7, p. 702). https://www.frontiersin.org/article/10.3389/fpsyg.2016.00702

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The Identification and Discrimination of Synthetic Vowels. *Language and Speech*, *5*(4), 171–189. https://doi.org/10.1177/002383096200500401

Fuhrmeister, P., & Myers, E. B. (2017). Non-native phonetic learning is destabilized by exposure to phonological variability before and after training. *The Journal of the Acoustical Society of America*, *142*(5), EL448–EL454. https://doi.org/10.1121/1.5009688

Garbin, G., Costa, A., Sanjuan, A., Forn, C., Rodriguez-Pujadas, A., Ventura, N., Belloch, V., Hernandez, M., & Ávila, C. (2011). Neural bases of language switching in high and early proficient bilinguals. *Brain and Language*, *119*(3), 129–135. https://doi.org/10.1016/j.bandl.2011.03.011

García-Sierra, A., Diehl, R. L., & Champlin, C. (2009). Testing the double phonemic boundary in bilinguals. *Speech Communication*, *51*(4), 369–378. https://doi.org/10.1016/j.specom.2008.11.005

García-Sierra, A., Ramírez-Esparza, N., Silva-Pereyra, J., Siard, J., & Champlin, C. A. (2012). Assessing the double phonemic representation in bilingual speakers of Spanish and English: An electrophysiological study. *Brain and Language*, *121*(3), 194–205. https://doi.org/10.1016/j.bandl.2012.03.008

Getzmann, S., Falkenstein, M., & Wascher, E. (2015). ERP correlates of auditory goal-directed behavior of younger and older adults in a dynamic speech perception task. *Behavioural Brain Research*, *278*, 435–445. https://doi.org/10.1016/j.bbr.2014.10.026

Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., & Warburton, E. A. (2012). The effect of aging on the neural correlates of phonological word retrieval. *Journal of Cognitive Neuroscience*, *24*(11), 2135–2146. https://doi.org/10.1162/jocn_a_00278

Ghaffarvand Mokari, P., & Werner, S. (2019). On the role of cognitive abilities in second language vowel learning. *Language and Speech*, *62*(2), 260–280. https://doi.org/10.1177/0023830918764517

Giannakopoulou, A., Brown, H., Clayards, M., & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ*, *5*(e3209). https://doi.org/10.7717/peerj.3209

Giannakopoulou, A., Uther, M., & Ylinen, S. (2013). Enhanced plasticity in spoken language acquisition for child learners: Evidence from phonetic training studies in child and adult learners of English. *Child Language Teaching and Therapy*, *29*(2), 201–218. https://doi.org/10.1177/0265659012467473

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Golestani, N. (2016). Neuroimaging of phonetic perception in bilinguals. *Bilingualism: Language and Cognition*, *19*(4), 674–682. https://doi.org/10.1017/S1366728915000644

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*, *17*(3), 575–582. https://doi.org/10.1093/cercor/bhk001

Gonzales, K., & Lotto, A. J. (2013). A Bafri, un Pafri: Bilinguals' Pseudoword Identifications Support Language-Specific Phonetic Systems. *Psychological Science*, *24*(11), 2135–2142. https://doi.org/10.1177/0956797613486485

Grenon, I., Kubota, M., & Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. *Journal of Phonetics*, *72*, 17–34. https://doi.org/10.1016/j.wocn.2018.10.005

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, *25*(4), 577–588. https://doi.org/10.1037/0012-1649.25.4.577

Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P., & Brattico, E. (2014). Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: a behavioral and ERP study. *Frontiers in Human Neuroscience*, *8*, 1–14. https://doi.org/10.3389/fnhum.2014.00279

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*, 3–15.

Grosjean, F. (1997). Processing mixed language: issues, findings, and models. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic perspectives* (pp. 225–254). Lawrence Erlbaum Associates.

Grosjean, F. (2010). *Bilingual: Life and Reality*. Harvard University Press.

Guion, S. G. (2003). The vowel systems of Quichua-Spanish bilinguals. *Phonetica*, *60*(2), 98–128. https://doi.org/10.1159/000071449

Hämäläinen, S., Mäkelä, N., Sairanen, V., Lehtonen, M., Kujala, T., & Leminen, A. (2018). TMS uncovers details about sub-regional language-specific processing networks in early bilinguals. *NeuroImage*, *171*(December 2017), 209–221. https://doi.org/10.1016/j.neuroimage.2017.12.086

Hawkins, S. (1999a). Looking for invariant correlates of linguistic units: Two classical theories of speech perception. In J. M. Pickett (Ed.), *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology* (pp. 198–231). Allyn & Bacon.

Hawkins, S. (1999b). Reevaluating assumptions about speech perception: Interactive and integrative theories. In J. M. Pickett (Ed.), *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology* (pp. 232–288). Allyn & Bacon.

Heald, S. L. M., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, *8*(MA), 1–15. https://doi.org/10.3389/fnsys.2014.00035

Heeren, W. F. L., & Schouten, M. E. H. (2010). Perceptual development of the Finnish / t-t:/ distinction in Dutch 12-year-old children: A training study. *Journal of Phonetics*, *38*(4), 594–603. https://doi.org/10.1016/j.wocn.2010.08.005

Hernandez, A. E., Dapretto, M., Mazziotta, J., & Bookheimer, S. (2001). Language switching and language representation in Spanish-English bilinguals: an fMRI study. *NeuroImage*, *14*(2), 510–520. https://doi.org/10.1006/nimg.2001.0810

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(May), 393–402.

Hickok, G., & Poeppel, D. (2016). Neural basis of speech perception. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 299–310). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00025-0

Hisagi, M., Shafer, V. L., Miyagawa, S., Kotek, H., Sugawara, A., & Pantazis, D. (2016). Second-language learning effects on automaticity of speech processing of Japanese phonetic contrasts: An MEG study. *Brain Research*, *1652*(October), 111–118. https://doi.org/10.1016/j.brainres.2016.10.004

Hoshino, N., & Thierry, G. (2011). Language selection in bilingual word production: Electrophysiological evidence for cross-language competition. *Brain Research*, *1371*, 100–109. https://doi.org/10.1016/j.brainres.2010.11.053

Immonen, K., Kilpeläinen, J., Alku, P., & Peltola, M. S. (2021). Does studying in a music-oriented education program affect non-native sound learning? — Effects of passive auditory training on children's vowel production. *Journal of Language Teaching and Research*, *12*(5), 678–687. https://doi.org/dx.doi.org/10.17507/jltr.1205.06

Immonen, K., Peltola, K. U., Tamminen, H., Alku, P., & Peltola, M. S. (2022). Orthography does not hinder non-native production learning in children. *Second Language Research*.

Immonen, K., & Peltola, M. S. (2017). Children learning a foreign vowel contrast - The effects of passive auditory exposure on L2 category perception. *Linguistica Lettica*, *25*, 385–400.

Immonen, K., & Peltola, M. S. (2018). Finnish Children Producing English Vowels — Studying in an English Immersion Class Affects Vowel Production. *Journal of Language Teaching and Research*, *9*(1), 27–33.

Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *The Journal of the Acoustical Society of America*, *99*(2), 1130–1140.

Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: do they arise from a common mechanism? *Perception & Psychophysics*, *62*(4), 874–886. https://doi.org/10.3758/BF03206929

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. https://doi.org/10.1016/S0010-0277(02)00198-1

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(01), 145–160. https://doi.org/10.1017/S0142716411000300

Jähi, K., Peltola, M. S., & Alku, P. (2015). Does interest in language learning affect the non-native phoneme production in elderly learners? *Proceedings of the 18th International Congress of Phonetic Sciences*. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0234.pdf

Jost, L. B., Eberhard-Moscicka, A. K., Pleisch, G., Heusser, V., Brandeis, D., Zevin, J. D., & Maurer, U. (2015). Native and non-native speech sound processing and the neural mismatch responses: A longitudinal study on classroom-based foreign language learning. *Neuropsychologia*, *72*, 94–104. https://doi.org/10.1016/j.neuropsychologia.2015.04.029

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*(5), 630–645. https://doi.org/10.1006/JMLA.1994.1030

Kartushina, N., Frauenfelder, U. H., & Golestani, N. (2016). How and when does the second language influence the production of native speech sounds: A literature review. *Language Learning*, *66*(Suppl. 2), 155–186. https://doi.org/10.1111/lang.12187

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, *138*(2). https://doi.org/10.1121/1.4926561

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2016). Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training. *Journal of Phonetics*, *57*, 21–39. https://doi.org/10.1016/J.WOCN.2016.05.001

Kartushina, N., & Martin, C. D. (2019). Talker and acoustic variability in learning to produce nonnative sounds: Evidence from articulatory training. *Language Learning*, *69*(1), 71–105. https://doi.org/10.1111/lang.12315

Kewley-Port, D. (1982). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. *Journal of the Acoustical Society of America*, *72*(2), 379–389.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *73*(1), 322–335.

Kim, K. H. S., Relkin, N. R., Lee, K.-M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, *388*(6638), 171–174. https://doi.org/10.1038/40623

Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., Zhang, K., & Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*(3), 220–224. https://doi.org/10.1111/1467-9280.02435

Klein, D., Milner, B., Zatorre, R. J., Meyer, E., & Evans, A. C. (1995). The neural substrates underlying word generation: A bilingual functional-imaging study. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(7), 2899–2903. https://doi.org/10.1073/pnas.92.7.2899

Klein, D., Mok, K., Chen, J.-K., & Watkins, K. E. (2014). Age of language learning shapes brain structure: A cortical thickness study of bilingual and monolingual individuals. *Brain and Language*, *131*, 20–24. https://doi.org/https://doi.org/10.1016/j.bandl.2013.05.014

Kraus, N., McGee, T., Carrell, T. D., King, C., Tremblay, K. L., & Nicol, T. (1995). Central auditory system plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience*, *7*(1), 25–32. https://doi.org/10.1162/jocn.1995.7.1.25

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism: Psycholonguistic Approaches* (electronic, pp. 531–553). Oxford University Press. https://ebookcentral.proquest.com/lib/kutu/detail.action?docID=281414

Kronrod, Y., Coppess, E., & Feldman, N. H. (2012). A unified model of categorical effects in consonant and vowel perception. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 629–634). Cognitive Science Society. https://cognitivesciencesociety.org/past-conferences/

Kuchinsky, S. E., Ahlstrom, J. B., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2014). Speech-perception training for older adults with hearing loss impacts word recognition and effort. *Psychophysiology*, *51*(10), 1046–1057. https://doi.org/10.1111/psyp.12242.Speech-perception

Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107. https://doi.org/10.3758/BF03212211

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11850–11857. https://doi.org/10.1073/pnas.97.22.11850

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews. Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979–1000. https://doi.org/10.1098/rstb.2007.2154

Kuhl, P. K., 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. https://doi.org/10.1207/s15473341lld0103&4_2

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the “perceptual magnet effect.” In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 121–154). York Press.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138–1141.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, *100*(401), 2425–2438. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3651031/

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (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*, *100*(15), 9096–9101. https://doi.org/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 moths of age. *Science*, *255*(5044), 606–668.

Kuipers, J. R., & Thierry, G. (2010). Event-related brain potentials reveal the time-course of language change detection in early bilinguals. *NeuroImage*, *50*(4), 1633–1638. https://doi.org/10.1016/j.neuroimage.2010.01.076

Kuipers, J. R., & Thierry, G. (2015). Bilingualism and increased attention to speech: Evidence from event-related potentials. *Brain and Language*, *149*, 27–32. https://doi.org/10.1016/j.bandl.2015.07.004

Kujala, T., Kallio, J., Tervaniemi, M., & Näätänen, R. (2001). The mismatch negativity as an index of temporal processing in audition. *Clinical Neurophysiology*, *112*(9), 1712–1719. https://doi.org/10.1016/S1388-2457(01)00625-3

Kujala, T., Karma, K., Ceponiene, R., Belitz, S., Turkkila, P., Tervaniemi, M., & Näätänen, R. (2001). Plastic neural changes and reading improvement caused by audiovisual training in reading-impaired children. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(18), 10509–10514. https://doi.org/10.1073/pnas.181589198

Kujala, T., & Näätänen, R. (2010). The adaptive brain: A neurophysiological perspective. *Progress in Neurobiology*, *91*(1), 55–67. https://doi.org/10.1016/j.pneurobio.2010.01.006

Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology*, *74*(1), 1–19. https://doi.org/10.1016/j.biopsycho.2006.06.001

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. The University of Michigan Press.

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, *144*(4), 394–425. https://doi.org/10.1037/bul0000142

Leong, C. X. R., Price, J. M., Pitchford, N. J., & van Heuven, W. J. B. (2018). High variability phonetic training in adaptive adverse conditions is rapid, effective, and sustained. *PLOS ONE*, *13*(10), e0204888. https://doi.org/10.1371/journal.pone.0204888

Lev-Ari, S., & Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics*, *41*(5), 320–331. https://doi.org/10.1016/J.WOCN.2013.06.002

Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex*, *58*, 301–324. https://doi.org/https://doi.org/10.1016/j.cortex.2014.05.001

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. https://doi.org/10.1037/h0020279

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36. https://doi.org/10.3758/BF03193857

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, *243*(4890), 489–494. https://doi.org/10.1126/science.2643163

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), F1–F10. https://doi.org/10.1111/1467-7687.00275

Machado, T. H., Fichman, H. C., Santos, E. L., Carvalho, V. A., Fialho, P. P., Koenig, A. M., Fernandes, C. S., Lourenço, R. A., Paradela, E. M. de P., & Caramelli, P. (2009). Normative data for healthy elderly on the phonemic verbal fluency task - FAS. *Dementia & Neuropsychologia*, *3*(1), 55–60. https://doi.org/10.1590/S1980-57642009DN30100011

Mack, M., & Blumstein, S. E. (1983). Further evidence of acoustic invariance in speech production: The stop-glide contrast. *The Journal of the Acoustical Society of America*, *73*(5), 1739–1750.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Press Syndicate of the University of Cambridge.

Massaro, D. W. (1994). Psychological aspects of speech perception. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics1* (pp. 219–263). Academic Press, Inc.

Massaro, D. W., & Cohen, M. M. (1976). The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. *Journal of the Acoustical Society of America*, *60*(3), 704–717.

Massaro, D. W., & Cohen, M. M. (1977). Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction. *Perception & Psychophysics*, *22*(4), 373–382.

McLaughlin, B. (1992). *Myths and misconceptions about second language learning: What every teacher needs to unlearn*. National Center for Research on Cultural Diversity and Second Language Learning, Center for Applied Linguistics, Santa Cruz, CA.

Meister, L., & Meister, E. (2011). Perception of the short vs. long phonological category in Estonian by native and non-native listeners. *Journal of Phonetics*, *39*, 212–224. https://doi.org/10.1016/j.wocn.2011.01.005

Menning, H., Imaizumi, S., Zwitserlood, P., & Pantev, C. (2002). Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. *Learning & Memory*, *9*, 253–267. https://doi.org/10.1101/lm.49402

Menning, H., Roberts, L. E., & Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *NeuroReport*, *11*(4), 817–822. https://doi.org/10.1097/00001756-200003200-00032

Molnar, M., Polka, L., Baum, S., & Steinhauer, K. (2014). Learning two languages from birth shapes pre-attentive processing of vowel categories: Electrophysiological correlates of vowel discrimination in monolinguals and simultaneous bilinguals. *Bilingualism: Language and Cognition*, *17*(3), 526–541. https://doi.org/10.1017/S136672891300062X

Moon, C. M., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica, International Journal of Paediatrics*, *102*(2), 156–160. https://doi.org/10.1111/apa.12098

Mora, J. C., & Nadeu, M. (2012). L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism*, *16*(4), 484–500. https://doi.org/10.1177/1367006911429518

Mougias, A., Christidi, F., Synetou, M., Kotrotsou, I., Valkimadi, P., & Politis, A. (2019). Differential effect of demographics, processing speed, and depression on cognitive function in 755 non-demented community-dwelling elderly individuals. *Cognitive & Behavioral Neurology*, *32*(4), 236–246. https://doi.org/10.1097/WNN.0000000000000211

Näätänen, R., & Escera, C. (2000). Mismatch negativity: clinical and other applications. *Audiol Neurootol*, *5*(3–4), 105–110. https://doi.org/13874

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology*, *42*(1), 25–32. https://doi.org/10.1111/j.1469-8986.2005.00256.x

Näätänen, R., Kujala, T., & Light, G. (2019). *Mismatch Negativity: A Window to the Brain* (1st ed.). Oxford University Press. http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=2075193&site=ehost-live

Näätänen, R., Kujala, T., & Winkler, I. (2011). Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology*, *48*(1), 4–22. https://doi.org/10.1111/j.1469-8986.2010.01114.x

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. In *Nature* (Vol. 385, pp. 432–434). https://doi.org/10.1038/385432a0

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*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *NeuroReport*, *4*(5), 503–506.

Nakamura, K., Kouider, S., Makuuchi, M., Kuroki, C., Hanajima, R., Ugawa, Y., & Ogawa, S. (2010). Neural control of cross-language asymmetry in the bilingual brain. *Cerebral Cortex*, *20*(9), 2244–2251. https://doi.org/10.1093/cercor/bhp290

Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2003). Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Cognitive Brain Research*, *16*(3), 492–495. https://doi.org/10.1016/S0926-6410(03)00055-7

Nippold, M. A., Cramond, P. M., & Hayward-Mayhew, C. (2014). Spoken language production in adults: Examining age-related differences in syntactic complexity. *Clinical Linguistics & Phonetics*, *28*(3), 195–207. https://doi.org/10.3109/02699206.2013.841292

Nygaard, L. C., & Pisoni, D. B. (1995). Speech Perception: New Directions in Research and Theory. In J. L. Miller & P. D. Eimas (Eds.), *Speech, Language, and Communication* (2nd ed., pp. 63–96). Academic Press. https://doi.org/10.1016/B978-012497770-9.50005-4

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Oller, J. W., & Ziahosseiny, S. M. (1970). The contrastive analysis hypothesis and spelling errors. *Language Learning*, *20*(2), 183–189.

Ortiz-Mantilla, S., Choudhury, N., Alvarez, B., & Benasich, A. A. (2010). Involuntary switching of attention mediates differences in event-related responses to complex tones between early and late Spanish-English bilinguals. *Brain Research*, *1362*, 78–92. https://doi.org/10.1016/j.brainres.2010.09.031

Ortiz-Mantilla, S., Hämäläinen, J. A., Realpe-Bonilla, T., & Benasich, A. A. (2016). Oscillatory dynamics underlying perceptual narrowing of native phoneme mapping from 6 to 12 months of age. *The Journal of Neuroscience*, *36*(48), 12095–12105. https://doi.org/10.1523/JNEUROSCI.1162-16.2016

Ostroff, J. M., McDonald, K. L., Schneider, B. A., & Alain, C. (2003). Aging and the processing of sound duration in human auditory cortex. *Hearing Research*, *181*(1–2), 1–7. https://doi.org/10.1016/S0378-5955(03)00113-8

Palomar-García, M.-Á., Bueichekú, E., Ávila, C., Sanjuán, A., Strijkers, K., Ventura-Campos, N., & Costa, A. (2015). Do bilinguals show neural differences with monolinguals when processing their native language? *Brain and Language*, *142*, 36–44. https://doi.org/10.1016/j.bandl.2015.01.004

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. *Annual Review of Psychology*, *60*, 173–196. https://doi.org/10.1146/annurev.psych.59.103006.093656

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(37), 15145–15150. https://doi.org/10.1073/pnas.1302159110

Peltola, K. U., Alku, P., & Peltola, M. S. (2017). Non-native speech sound production changes even with passive listening training. *Linguistica Lettica*, *25*, 158–172.

Peltola, K. U., Rautaoja, T., Alku, P., & Peltola, M. S. (2017). Adult learners and a one-day production training – Small changes but the native language sound system prevails. *Journal of Language Teaching and Research*, *8*(1), 1–7.

Peltola, K. U., Tamminen, H., Alku, P., Kujala, T., & Peltola, M. S. (2020). Motoric training alters speech sound perception and production — Active listening training does not lead into learning outcomes. *Journal of Language Teaching and Research*, *11*(1), 10–16.

Peltola, K. U., Tamminen, H., Alku, P., & Peltola, M. S. (2015). Non-native production training with an acoustic model and orthographic or transcription cues. *Proceedings of the 18th International Congress of Phonetic Sciences*. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0236.pdf

Peltola, M. S. (2001). Second language acquisition, phonetic tests and event-related brain potentials: Combining methods. In R. Hiltunen, K. Battarbee, M. Peikola, & S.-K. Tanskanen (Eds.), *English in Zigs and Zags* (pp. 153–161). University of Turku.

Peltola, M. S. (2003). *The attentive and preattentive perception of native and non-native vowels: The effect of second language learning*. University of Turku.

Peltola, M. S. (2007). Speech sound perception and neural representations. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of the Phonetic Sciences* (pp. 123–126).

Peltola, M. S., & Aaltonen, O. (2005). Long-term memory trace activation for vowels depends on the mother tongue and the linguistic context. *Journal of Psychophysiology*, *19*(3), 159–164. https://doi.org/10.1027/0269-8803.19.3.159

Peltola, M. S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters*, *352*(1), 25–28. https://doi.org/10.1016/j.neulet.2003.08.013

Peltola, M. S., Kuntola, M., Tamminen, H., Hämäläinen, H., & Aaltonen, O. (2005). Early exposure to non-native language alters preattentive vowel discrimination. *Neuroscience Letters*, *388*(3), 121–125. https://doi.org/10.1016/j.neulet.2005.06.037

Peltola, M. S., Lintunen, P., & Tamminen, H. (2014). Advanced English learners benefit from explicit pronunciation teaching: An experiment with vowel duration and quality. *AFinLa-E*, *6*, 86–98.

Peltola, M. S., Tamminen, H., Toivonen, H., Kujala, T., & Näätänen, R. (2012). Different kinds of bilinguals - Different kinds of brains: The neural organisation of two languages in one brain. *Brain and Language*, *121*(3), 261–266. https://doi.org/10.1016/j.bandl.2012.03.007

Peltola, M. S., Tuomainen, O., Koskinen, M., & Aaltonen, O. (2007). The effect of language immersion education on the preattentive perception of native and non-native vowel contrasts. *Journal of Psycholinguistic Research*, *36*(1), 15–23. https://doi.org/10.1007/s10936-006-9030-y

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., Fazio, F., & Mehler, J. (1996). Brain processing of native and foreign languages. *NeuroReport*, *7*(15–17), 2439–2444.

Perani, D., Paulesu, E., Sebastian-Galles, N., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S. F., Fazio, F., & Mehler, J. (1998). The bilingual brain: proficiency and age of acquisition of the second language. *Brain*, *121*(10), 1841–1852. https://doi.org/10.1093/brain/121.10.1841

Pereira Reyes, Y. (2014). Perception and production of English vowels by Chilean learners of English: Effect of auditory and visual modalities on phonetic training [UCL (University College London)]. In *Doctoral thesis*. https://discovery.ucl.ac.uk/id/eprint/1417190

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472. https://doi.org/10.1121/1.3593366

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception & Psychophysics*, *30*(3), 217–227. https://doi.org/10.3758/BF03214276

Rinker, T., Alku, P., Brosch, S., & Kiefer, M. (2010). Discrimination of native and non-native vowel contrasts in bilingual Turkish-German and monolingual German children: Insight from the Mismatch Negativity ERP component. *Brain and Language*, *113*(2), 90–95. https://doi.org/10.1016/j.bandl.2010.01.007

Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., & Kuhl, P. K. (2005). Neural patterns to speech and vocabulary growth in American infants. *Neuroreport*, *16*(5), 495–498. https://doi.org/10.1097/00001756-200504040-00015

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, *8*, 162–172. https://doi.org/10.1111/j.1467-7687.2005.00403.x

Romaine, S. (1995). *Bilingualism* (2nd ed.). Blackwell Publishers.

Rosch, E. (1977). Human categorization. In N. Warren (Ed.), *Advances in cross-cultural psychology* (pp. 1–49). Academic Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Erlbaum.

Rosen, S., & Howell, P. (1987). Auditory, articulatory, and learning explanations of categorical perception in speech. In S. Harnard (Ed.), *Categorical perception: The groundwork of cognition* (pp. 113–160). Cambridge University Press.

Ross, E. D. (1984). Right hemisphere's role in language, affective behavior and emotion. *Trends in Neurosciences*, *7*, 342–346. https://doi.org/10.1016/S0166-2236(84)80085-5

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Salo, S., Peltola, M. S., Aaltonen, O., Johansson, R., Lang, A. H., & Laurikainen, E. (2002). Stability of memory traces for speech sounds in cochlear implant patients. *Logopedics Phoniatrics Vocology*, *27*, 132–138.

Saloranta, A., Alku, P., & Peltola, M. S. (2017). Learning and generalization of vowel duration with production training: Behavioral results. *Linguistica Lettica*, *25*, 67–87.

Saloranta, A., Alku, P., & Peltola, M. S. (2020). Listen-and-repeat training improves perception of second language vowel duration: Evidence from mismatch negativity (MMN) and N1 responses and behavioral discrimination. *International Journal of Psychophysiology*, *147*, 72–82. https://doi.org/10.1016/j.ijpsycho.2019.11.005

Saloranta, A., Tamminen, H., Alku, P., & Peltola, M. S. (2015). Learning of a non-native vowel through instructed production training. *Proceedings of the 18th International Congress of Phonetic Sciences*. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0235.pdf

Samuel, A. G., & Larraza, S. (2015). Does listening to non-native speech impair speech perception? *Journal of Memory and Language*, *81*, 51–71. https://doi.org/https://doi.org/10.1016/j.jml.2015.01.003

Savo, S., & Peltola, M. S. (2019). Arabic-speakers Learning Finnish Vowels: Short-term Phonetic Training Supports Second Language Vowel Production. *Journal of Language*, *10*(1), 45–50. https://doi.org/http://dx.doi.org/10.17507/jltr.1001.05

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26*(2), 100–107. https://doi.org/10.1016/S0166-2236(02)00037-1

Sebastián-Gallés, N., & Díaz, B. (2012). First and second language speech perception: Graded learning. *Language Learning*, *62*(Suppl. 2), 131–147. https://doi.org/10.1111/j.1467-9922.2012.00709.x

Shafer, V. L., Yu, Y. H., & Datta, H. (2010). Maturation of speech discrimination in 4- to 7-yr-old children as indexed by event-related potential mismatch responses. *Ear & Hearing*, *31*(6), 735–745.

Sharma, A., & Dorman, M. F. (1998). Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential. *The Journal of the Acoustical Society of America*, *104*(1), 511. https://doi.org/10.1121/1.423252

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America*, *106*(2), 1078–1083. https://doi.org/10.1121/1.428048

Sharma, A., & Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *The Journal of the Acoustical Society of America*, *107*(5 Pt 1), 2697–2703. https://doi.org/10.1121/1.428655

Shestakova, A., Huotilainen, M., Ceponiene, R., & Cheour, M. (2003). Event-related potentials associated with second language learning in children. *Clinical Neurophysiology*, *114*(8), 1507–1512. https://doi.org/10.1016/S1388-2457(03)00134-2

Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. *Journal of Phonetics*, *66*, 242–251. https://doi.org/10.1016/j.wocn.2017.11.002

Shtyrov, Y., Nikulin, V. V., & Pulvermüller, F. (2010). Rapid cortical plasticity underlying novel word learning. *The Journal of Neuroscience*, *30*(50), 16864–16867. https://doi.org/10.1523/JNEUROSCI.1376-10.2010\r30/50/16864

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51–66). McGraw-Hill Book Company.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3–45.

Strange, W. (1999). Perception of consonants: From variance to invariance. In J. M. Pickett (Ed.), *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology* (pp. 166–182). Allyn & Bacon.

Strange, W. (2007). Cross-language phonetic similarity of vowels: Theoretical and methodological issues. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (pp. 35–55). John Benjamins Publishing Company. http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=622748

Strouse, A., Ashmead, D. H., Ohde, R. N., & Grantham, D. W. (1998). Temporal processing in the aging auditory system. *Journal of the Acoustical Society of America*, *104*(4), 2385–2399.

Sundara, M., Polka, L., & Baum, S. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition*, *9*(1), 97–114. https://doi.org/10.1017/S1366728905002403

Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-/ in monolingual and bilingual acquisition of English. *Cognition*, *100*(2), 369–388. https://doi.org/https://doi.org/10.1016/j.cognition.2005.04.007

Suomi, K. (1980). *Voicing in English and Finnish stops : a typological comparison with an interlanguage study of the two languages in contact*. University of Turku.

Suomi, K., Toivanen, J., & Ylitalo, R. (2008). *Finnish sound structure: Phonetics, phonology, phonotactics and prosody*. University of Oulu.

Taimi, L., Alku, P., Kujala, T., Näätänen, R., & Peltola, M. S. (2014). The effect of production training on non-native speech sound perception and discrimination in school-aged children: An MMN and behavioural study. *Linguistica Lettica*, *2*, 114–129.

Taimi, L., Jähi, K., Alku, P., & Peltola, M. S. (2014). Children Learning a Non-native Vowel – The Effect of a Two-day Production Training. *Journal of Language Teaching and Research*, *5*(6), 1229–1235. https://doi.org/10.4304/jltr.5.6.1229-1235

Tamminen, H., Kujala, T., Näätänen, R., & Peltola, M. S. (2021). Aging and non-native speech perception: A phonetic training study. *Neuroscience Letters*, *740*, 135430. https://doi.org/https://doi.org/10.1016/j.neulet.2020.135430

Tamminen, H., Kujala, T. & Peltola, M. S. Training non-native speech sounds results in permanent plastic changes – Hard-wiring new memory traces takes time. Submitted.

Tamminen, H., & Peltola, M. S. (2015). Non-native memory traces can be further strengthened by short term phonetic training. *Proceedings of the 18th International Congress of Phonetic Sciences*. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0285.pdf

Tamminen, H., & Peltola, M. S. (2019). Speech sound perception in monolinguals, bilinguals and learners – Language background affects identification and discrimination differently. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2258–2262). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Tamminen, H., Peltola, M. S., Kujala, T., & Näätänen, R. (2015). Phonetic training and non-native speech perception - New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures. *International Journal of Psychophysiology*, *97*(1), 23–29. https://doi.org/10.1016/j.ijpsycho.2015.04.020

Tamminen, H., Peltola, M. S., Toivonen, H., Kujala, T., & Näätänen, R. (2013). Phonological processing differences in bilinguals and monolinguals. *International Journal of Psychophysiology*, *87*(1), 8–12. https://doi.org/10.1016/j.ijpsycho.2012.10.003

Tamminen, H., Rautaoja, T., & Peltola, M. S. (2017). Two types of bilinguals - Two types of production contexts. *Linguistica Lettica*, *25*, 109–124.

Tomaschek, F., Truckenbrodt, H., & Hertrich, I. (2011). Processing German vowel quantity: Categorical Perception or Perceptual Magnet Effect? *ICPhS*, 2002–2005.

Tremblay, K. L., & Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *Journal of Speech, Language, and Hearing Research*, *45*, 564–572.

Tremblay, K. L., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *Journal of the Acoustical Society of America*, *102*(6), 3762–3773. https://doi.org/10.1121/1.420139

Tremblay, K. L., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *NeuroReport*, *9*(16), 3557–3560.

Tremblay, K. L., Kraus, N., McGee, T., Ponton, C., & Otis, A. B. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. *Ear and Hearing*, *22*(2), 79–90. https://doi.org/10.1097/00003446-200104000-00001

Tremblay, K. L., Piskosz, M., & Souza, P. (2002). Aging alters the neural representation of speech cues. *NeuroReport*, *13*(15), 1865–1870.

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*(4), 2285–2294. https://doi.org/10.1121/1.2338290

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. E. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*, 263–290. https://doi.org/10.1016/j.wocn.2004.10.002

Tyler, M. D. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. Hjortshøj Sørensen (Eds.), *A sound approach to language matters - In Honor of Ocke-Schwen Bohn* (pp. 607–630). Department of English, School of Communication & Culture, Aarhus University.

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, *71*(1), 4–21. https://doi.org/10.1159/000356237

van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology*, *6*(August), 1–12. https://doi.org/10.3389/fpsyg.2015.01000

Virtala, P., Partanen, E., Tervaniemi, M., & Kujala, T. (2018). Neural discrimination of speech sound changes in a variable context occurs irrespective of attention and explicit awareness. *Biological Psychology*, *132*(July 2017), 217–227. https://doi.org/10.1016/j.biopsycho.2018.01.002

Wardhaugh, R. (1970). The Contrastive Analysis Hypothesis. *TESOL Quarterly*, *4*(2), 123–130. https://doi.org/10.2307/3586182

Wei, L. (2000). Dimensions of bilingualism. In L. Wei (Ed.), *The Bilingualism Reader* (pp. 3–25). Routledge.

Weinreich, U. (1963). *Languages in contact: Findings and problems* (2nd ed.). Mouton & Co.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, *52*(1), 349–355.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganisation during the first year of life. *Infant Behavior and Development*, *7*, 49–63. https://doi.org/10.1016/S0163-6383(02)00113-3

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, *50*, 509–535. https://doi.org/10.1146/annurev.psych.50.1.509

Wiik, K. (1965). *Finnish and English vowels: A comparison with special reference to the learning problems met by native speakers of Finnish learning English*. University of Turku.

Winkler, I., Kujala, T., Alku, P., & Näätänen, R. (2003). Language context and phonetic change detection. *Cognitive Brain Research*, *17*(3), 833–844. https://doi.org/10.1016/S0926-6410(03)00205-2

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, Ii., Csépe, V., Ilmoniemi, R. J., & Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*(5), 638–642. https://doi.org/10.1111/1469-8986.3650638

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews Neuroscience*, *5*, 483–494. https://doi.org/10.1038/nrn1406

Wong, P. C. M., Jin, J. X., Gunasekera, G. M., Abel, R., Lee, E. R., & Dhar, S. (2009). Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia*, *47*(3), 693–703. https://doi.org/10.1016/j.neuropsychologia.2008.11.032

Yeni-Komshian, G. H., Flege, J. E., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean–English bilinguals. *Bilingualism: Language and Cognition*, *3*(2), 131–149. https://doi.org/10.1017/s1366728900000225

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., Kawakatsu, M., Tohkura, Y., & Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *NeuroImage*, *46*(1), 226–240. https://doi.org/10.1016/j.neuroimage.2009.01.028

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. *NeuroImage*, *26*(3), 703–720. https://doi.org/10.1016/j.neuroimage.2005.02.040

Zurif, E., & Swinney, D. (1994). The neuropsychology of language. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 1055–1074). Academic Press, Inc.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2022

ISBN 978-951-29-8925-6 (PRINT)
ISBN 978-951-29-8926-3 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)