TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

SCHOOL-AGED CHILDREN LEARNING SECOND LANGUAGE SOUNDS

The Effects of Different Learning Backgrounds on Children's Second Language Sound Production and Perception Learning

Katja Haapanen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. B OSA – TOM. 552 | HUMANIORA | TURKU 2021

TURUN YLIOPISTO
UNIVERSITY OF TURKU

SCHOOL-AGED CHILDREN LEARNING SECOND LANGUAGE SOUNDS

The Effects of Different Learning Backgrounds on Children's Second Language Sound Production and Perception Learning

Katja Haapanen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. B OSA – TOM. 552 | HUMANIORA | TURKU 2021

University of Turku

Faculty of Technology
Department of Computing
Phonetics
Doctoral programme in Mathematics and Computer Sciences (MATTI)

Supervised by

Professor, Maija S. Peltola
University of Turku

Reviewed by

Docent, Mikko Kuronen
University of Jyväskylä

University lecturer, Minnaleena Toivola
University of Helsinki

Opponent

Professor, Hannele Dufva
University of Jyväskylä

ISBN 978-951-29-8592-0 (PRINT)
ISBN 978-951-29-8593-7 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)
Painosalama, Turku, Finland 2021

UNIVERSITY OF TURKU
Faculty of Technology
Department of Computing
Phonetics
KATJA HAAPANEN: School-Aged Children Learning Second Language Sounds: The Effects of Different Learning Backgrounds on Children's Second Language Sound Production and Perception Learning
Doctoral Dissertation, 153 pp.
Doctoral Programme in Mathematics and Computer Sciences
September 2021

ABSTRACT

Earlier phonetic research on children's second language sound learning has primarily focused on naturalistic language learning environments and the comparison of child and adult learners. The aim of this thesis was to examine how 6–13-year-old Finnish children with different learning background factors learn to perceive and produce second language sounds in instructed settings and phonetic training contexts. The background factors examined in Studies I–IV of this thesis were age, language immersion education and music-oriented education.

The aim of Study I was to discover how passive auditory training affects 7–12-year-old children's non-native sound perception. The results showed that the older participants' (9–12 years) perceived the trained non-native sound after passive auditory training. Study II examined how studying in a language immersion program in elementary school affects 11–13-year-old children's second language pronunciation. The findings suggest that studying in a language immersion program results in more accurate second language pronunciation when compared to non-immersive learning. Study III investigated how music-oriented education in elementary school affects 9–11-year-old childen's ability to learn non-native sound production through auditory training. No significant effects of music-oriented education were found. Study IV aimed to discover how listen-and-repeat training affects 6–7-year-old preschoolers' non-native sound production. The results showed rapid changes in production after minimal amount of listen-and-repeat training. Overall, the results of Studies I–IV suggested that age of learning has the strongest effect on school-aged children's second language sound learning.

KEYWORDS: children, phonetics, pronunciation, second language learning, speech production, speech perception, speech sounds

TURUN YLIOPISTO
Teknillinen tiedekunta
Tietotekniikan laitos
Fonetiikka
KATJA HAAPANEN: Kouluikäiset lapset oppimassa vieraan kielen äänteitä: Erilaisten oppimistaustojen vaikutus lasten vieraan kielen äänteiden tuoton ja havaitsemisen oppimiseen
Väitöskirja, 153 s.
Matemaattis-tietotekninen tohtoriohjelma
syyskuu 2021

TIIVISTELMÄ

Aiemmat lasten vieraan kielen äänteiden oppimista selvittävät foneettiset tutkimukset ovat pitkälti keskittyneet luonnollisiin kielenoppimisympäristöihin sekä lapsi- ja aikuisoppijoiden vertailuun. Tämän tutkielman tavoite oli selvittää, kuinka erilaiset oppimisen taustatekijät vaikuttavat 6–13-vuotiaiden suomenkielisten lasten vieraan kielen äänteiden havaitsemisen ja tuoton oppimiseen. Tarkasteltavat taustatekijät olivat ikä ja kieli- tai musiikkiluokalla opiskelu.

Tutkimus koostui neljästä osatutkimuksesta. Tutkimuksessa I selvitettiin passiivisen auditorisen treenin vaikutuksia 7–12-vuotiaiden lasten vieraan kielen äänteen havaitsemiseen. Tulokset osoittivat, että vanhemmat (9–12 vuotta) osallistujat havaitsivat treenatun vieraan kielen äänteen passiivisen kuuntelun jälkeen. Tutkimuksessa II tarkasteltiin peruskoulun kieliluokalla opiskelun vaikutuksia 11–13-vuotiaiden lasten vieraan kielen ääntämiseen. Tulokset viittasivat, että kieliluokalla opiskelu johtaa parempaan ääntämistarkkuuteen kuin perinteinen luokkahuoneoppiminen. Tutkimuksen III tavoite puolestaan oli selvittää, vaikuttaako musiikkiluokalla opiskelu 9–11-vuotiaiden lasten kykyyn oppia tuottamaan vieraan kielen äännekontrasti auditorisen treenin avulla. Musiikkiluokalla opiskeluun liittyviä merkitseviä eroja tuoton oppimisessa ei löydetty. Tutkimuksessa IV selvitettiin, kuinka kuuntele-ja-toista treeni vaikuttaa 6–7-vuotiaiden esikoululaisten vieraan kielen äännekontrastin tuottoon. Tulokset osoittivat, että esikoululaiset muuttivat tuottoaan nopeasti vähäisen harjoittelun jälkeen. Kokonaisuudessaan tutkimuksen tulokset antoivat viitteitä siitä, että oppimisikä vaikuttaa kouluikäisten lasten vieraan kielen äänteiden oppimiseen voimakkaammin kuin muut tarkastellut taustatekijät.

ASIASANAT: fonetiikka, lapset, puheen havaitseminen, puheen tuottaminen, vieraan kielen oppiminen, äänteet, ääntäminen

Acknowledgements

This is it. The moment that all the hard work has been building up to, although a small part of me doubted this day would ever actually come. But now it is here and it's time to acknowledge those who made it possible.

First, I should thank the University of Turku Graduate School and the Department of Computing for funding my work and Sanako Corp. for sponsoring some of the software used in this study.

I want to thank my supervisor Maija S. Peltola for all her support. Thank you Maija, for steering me back in course whenever my mind began to wander off the goal. You knew when to push me further when I hesitated, and when to step back and let me take the wheel. Your supervision and support have allowed me to grow into the researcher I had the potential of becoming.

Second, I would like to thank docent Mikko Kuronen and university lecturer Minnaleena Toivola for their valuable comments and thoughts on my work. I would also like to thank professor Hannele Dufva for accepting the invitation to be my opponent and professor Tapio Salakoski for being my director.

My co-authors and my colleagues in the Phonetics and Learning, Age & Bilingualism –laboratory have been an important and an integral part of this journey. Thank you Antti Saloranta, Henna Tamminen, Jemina Kilpeläinen and Kimmo Peltola for the bad jokes in our coffee room and for the opportunity to discuss anything and everything relating to this peculiar branch of science called phonetics. I would also like to thank professor Paavo Alku for creating the stimuli used in this study and for his helpful comments during the manuscript process.

My sincerest thanks goes to all the children, teachers, parents and schools that voluntarily participated in this study. The willingness to explore and discover the wonders of language is a resource I hope will never run out.

For those who are closest to me, it has undoubtedly become clear through the years that this journey has not always been an easy one. Without my friends, it would have been even harder. You are my people, my second family. Thank you for being in my life.

All my gratitude goes to my parents, who have always been my biggest cheerleaders and who have encouraged me to be independent and follow my own

path, no matter what. Mom, you are my biggest inspiration. Your example sparked my interest in science and showed me I could have a future in research. Thank you for taking me to the lab and letting me calibrate your pipettes. And dad, thank you for always believing in me and supporting me.

Finally, words can hardly describe my gratitude to my husband, my partner, my best friend and companion. Antti-Jussi, you have encouraged me when I had doubts, consoled me when I was disappointed, calmed me down when I was stressed and celebrated with me when I was happy. As a wise man once said, you are not only the man I deserve, but the man I need by my side. Thank you for being you.

September 2021
Katja Haapanen

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

II Immonen, K. & Peltola, M. S. Finnish children producing English vowels – Studying in an English immersion class affects vowel production. *Journal of Language Teaching and Research*, 2018; 9(1): 27–33.

III Immonen, K., Kilpeläinen, J., Alku, P. & Peltola, M. S. 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*, 2021; 12(5): 678–687.

IV Immonen, K., Alku, P. & Peltola, M. S. Phonetic listen-and-repeat training alters 6–7-year-old children's non-native vowel production after one training session. Submitted.

1 Introduction

This thesis examines the effects of different learning backgrounds on 6–13-year-old children's second language (L2) production and perception learning. The study focuses on the phonetic aspect of L2 speech learning by investigating the perception and production of individual L2 sounds and sound contrasts. Phonetic learning of L2 sounds was selected for the purposes of this thesis, because it offers insight into how individual sounds are learned, but also because it serves as a gateway to understanding L2 speech learning in a larger context, as learning individual speech sounds leads to the production of words and sentences, making communication in an L2 possible. The thesis consists of four studies (I-IV), each of which examines a specific aspect of L2 sound learning on different age groups, and with varying behavioral and preattentive measures and phonetic training methods.

The text consists of five main sections: theoretical background, aims, methodology, results and discussion. The theoretical section introduces models of second language sound learning that form the theoretical basis for the methodology of Studies I–IV. In addition, the section focuses on previous research on the effects of age and other background factors on L2 sound learning, which are used to form the hypotheses for Studies I–IV. The research questions and hypotheses are presented in the aims section of the thesis. The methodology and results sections describe the research procedures, analyses and results of each study separately. Finally, the discussion section considers the implications and relevance of the results found in Studies I–IV.

1.1 Theoretical background

1.1.1 Models of second language sound learning

The phonetic learning of a second language has been the subject of extensive research over the years, since accurate sound perception and production is an essential aspect of successful communication in a foreign language. If speakers replace L2 sounds with perceptually similar sounds of their first language (L1), they may unintentionally produce phonemic distinctions that change word meanings. On the other hand, even subtler allophonic mispronunciations can cause L2 speakers' speech to become less intelligible. Therefore, challenges in L2 sound perception and production can cause major or minor difficulties in communication.

Earlier cross-language research on L2 phonetic learning has led to the emergence of several models of second language sound learning. These models aim to predict and explain the degree of possible difficulties that speakers face when they learn the sounds of a new language. Identifying the most probable challenges in L2 sound perception and production allows us to focus on the most difficult and communicationally essential phonetic features in L2 learning. The theoretical foundation for the research design of this thesis is based on three of these models: the Speech Learning Model (SLM; Flege, 1995), the Perceptual Assimilation Model (PAM; Best, 1994, 1995) and the Native Language Magnet Model (NLM; Kuhl, Williams, & et al, 1992; Kuhl et al., 2008). These models are used in the experiment design and stimulus selection of Studies I–IV and they are described in more detail in the following chapters.

1.1.1.1 The Speech Learning Model

The Speech Learning Model (SLM) created by Flege (1995) is one of the most cited and most established models of second language sound learning. The core proposition of the model is that the degree of perceptual similarities and dissimilarities between L1 and L2 categories determine the degree of difficulty that a speaker faces when learning L2 sounds.

The SLM posits that L2 sounds can be either identical, similar or new in relation to L1 categories (Flege, 1995). Identical sounds are L2 sounds that already exist in the speaker's L1 phonology, and therefore they are not expected to cause challenges in L2 perception or production. New L2 sounds, on the other hand, do not resemble any existing L1 categories. Therefore, new sounds are easily perceived as distinct from all L1 sounds, but the acquisition of a new L2 phonological category requires some repetition and input. The most challenging L2 sounds are similar sounds, since they resemble one or more L1 categories, but their accurate perception and production requires the formation of a new sound category close to the same area in the phonetic space.

However, not all similar sounds are equally challenging, since the degree of similarity can be very different depending on the relative perceptual distance between the L2 and L1 categories. In fact, more recent versions of the SLM propose that the three perceptual categories (new, similar and identical) form a continuum of similarity and dissimilarity, where the probability of the formation of a new category for an L2 sound increases as the perceived similarity between L2 and L1 sounds decreases (Flege, 2007). According to this prediction, similar sounds are more challenging the closer they are perceptually to an L1 category, because the formation of a new category is still required but less likely (Flege, 2007).

Another core propostition of the SLM is that sound perception precedes production. In other words, the accurate production of L2 sounds is limited by how accurately they are perceived (Flege, 2007, 1995, 1999; Flege, MacKay, & Meador, 1999a). This proposition has important implications for L2 learning research, since it includes the assumption that speakers cannot acquire new or similar L2 production patterns before they perceive the new or similar L2 categories. In addition, the SLM's hypothesis on the relationship between speech perception and production entails the implication that the accurate production of L2 sounds automatically indicates accurate perception of the produced categories.

1.1.1.1.1 SLM-r

A recent revised version of the Speech Learning Model (SLM-r; Flege & Bohn, 2021), expands and alters the original predictions made by the SLM (Flege, 1995, 2007), for example, by reviewing the significance of L2 phonetic input, L1 sound categories age and the link between perception and production. The SLM-r posits that L2 sound category formation primarily depends on three factors. These factors are the degree of perceived similarity between the L2 sound and closest L1 categories, the quality and quantity of L2 phonetic input and the precision (i.e., F1-F2 variability) of existing L1 categories. The basic principles of phonetic similarity of L2 and L1 sounds affecting L2 category formation remain the same as in the SLM (see section 1.1.1.1.). The predictions regarding quality and quantity of input expand the SLM by extending the model from naturalistic L2 environments to different L2 learning contexts.

The SLM-r states that speakers who have resided in an L2 speaking environment (in immigration contexts) for the same amount of time (length of residence, LOR) are often exposed to very different amounts of L2 input and the quality of input may vary greatly, which makes the LOR an unreliable measure of L2 input. Instead, the SLM-r considers L2 input in terms of quantity and quality by defining phonetic input as sensory stimulation received in meaningful conversations (Flege & Bohn, 2021). This definition accounts also for the differences in L2 input received in naturalistic and instructed learning settings, since phonetic input received in an L2 classroom varies greatly from input received in naturaistic contexts (both in quantity and quality).

The precision of L1 sound categories, on the other hand, refers to the within-category F1-F2 variability of an L1 sound. According to the SLM-r, L1 categories are more precise in childhood and adolescence and less precise in adulthood. This means that children's L1 categories tend to be tighter or smaller, with less variability, whereas adults' categories are more expanded. The SLM-r proposes that precise L1 categories increase the probability of L2 category formation (Flege & Bohn, 2021).

The SLM-r abandons the idea of perception preceding production. Instead, the SLM-r proposes that L2 sound perception and production coevolve in constant interaction and neither one precedes the other (Flege & Bohn, 2021). This notion poses important implications for L2 productin learning, since it entails that the perception accuracy of L2 sounds does not automatically pose an upper limit to their production.

Another notable difference between the SLM (Flege, 1995) and the SLM-r (Flege & Bohn, 2021) is the perspective on ultimate L2 attainment. While the SLM focused on the nativelike perception and production of L2 sounds by experienced learners, the SLM-r rejects the idea of nativelikeness being the ultimate goal of L2 learning. Flege and Bohn (2021) suggest that L2 learners can never be identical to native speakers of the L2, because the L1 and L2 phonetic and phonological systems automatically interact and the L2 phonetic input differs from the phonetic input native speakers receive when learning an L1. In addition, the SLM-r introduced the principle of continuous learning, according to which the phonetic systems of a speaker's L1 and L2 interact and adapt throughout life. Therefore, according to the SLM-r, L2 category formation is possible at any age, regardless of the initial age of learning/exposure to the L2 (Flege & Bohn, 2021).

1.1.1.2 The Perceptual Assimilation Model

The Perceptual Assimilation Model (PAM; Best, 1994, 1995) of cross-language speech perception and its more recent version, the Perceptual Assimilation Model of L2 speech learning (PAM-L2; Best & Tyler, 2007), state that L2 speech segments (i.e., sounds) are perceived based on their dissimilarities and similarities to L1 segments that are closest to the L2 segments in the L1 phonological space. The PAM recognizes three main patterns of perceptual assimilation in which L2 sounds can assimilate to L1 categories (Best, 1995). First, an L2 sound can be assimilated to an L1 category either as a good, acceptable or deviant exemplar of that particular category. Second, an L2 sound can be assimilated as uncategorizable, meaning that it is categorized as a speech sound but it is not a clear exemplar of any L1 category. Third, an L2 sound might not be assimilated to speech at all (i.e., it is categorized as a nonspeech sound). These three assimilation patterns form the foundation of the PAM.

However, in order to predict cross-language speech perception more accurately, the model focuses on the assimilation of L2 sound contrasts to existing L1 categories rather than just individual L2 sounds. The assimilation patterns of L2 sound contrasts derive directly from the three main assimilation patterns of L2 sounds (Best, 1994,

1995) listed above. The two sounds in an L2 contrast can assimilate to L1 categories to different degrees, depending on their perceptual similarities and dissimilarities to the L1 categories. According to the PAM, there are six possible assimilation patterns for L2 contrasts (Best, 1994, 1995):

1. *Two-category assimilation*, where two L2 sounds are assimilated to two different L1 categories.
2. *Category-goodness difference*, where two L2 sounds are assimilated to the same L1 category as an acceptable and an unacceptable exemplar (i.e., one member of the L2 contrast is perceived as a better exemplar of the L1 category).
3. *Single-category assimilation*, where two L2 sounds are assimilated to one L1 category similarly (i.e., both L2 sounds are equally acceptable or deviant exemplars of the L1 category).
4. *Both uncategorizable*, where two L2 sounds are perceived as speech sounds, but neither can be assimilated to any existing L1 category.
5. *Uncategorized vs. categorized*, where two L2 sounds are perceived as speech sounds, but only one can be assimilated to an L1 category.
6. *Nonassimilable*, where two L2 sounds are perceived as nonspeech sounds.

These six possible assimilation patterns pose different degrees of difficulty for L2 speakers in L2 sound perception (Best, 1994, 1995). In cases where perception follows the two-category assimilation, uncategorized vs. categorized or nonassimilable assimilation patterns, discrimination is expected to be easy. When an L2 contrast follows the category-goodness difference or both uncategorizable assimilation patterns, discrimination is predicted to depend on the degree of similarities and dissimilarities between the members of the L2 contrast and L1 categories. Discrimination is predicted to be the most difficult in a single-category assimilation situation, when an L2 contrast is assimilated equally to a single L1 category.

The PAM-L2 (Best & Tyler, 2007) expands the predictions of the PAM by considering L2 phonological learning from a lexical-functional perspective instead of focusing on the degree phonetic similarity between L2 and L1 sounds. The PAM-L2 suggests that an L2 sound can be assimilated to an L1 phonological category if the two sounds have similar lexical functions in both languages. For example, native English speakers who learn French as an L2 can distinguish the French uvular voiced fricative [ʁ] from the English postalveolar approximant [ɹ] phonetically, but are likely to equate both sounds phonologially to the lexical-functional category /r/ (Best & Tyler, 2007). According to the PAM-L2, this assimilation is caused by the similar lexical-functional roles of the French and English /r/ in both languages (i.e., the sounds relate similarly to surrounding categories in the phonological space). Even though speakers can phonetically discriminate the two sounds, their lexical-functional similarity causes them to be phonologially assimilated to the same category. Vice verca, assimilation to the same lexical-functional category does not automatically imply perceived phonetic similarity. Lexical-functional assimilation does not pose great challenges to L2 sound learning, since learners are likely to learn the different phonetic realizations of a phonological category even when they assimilate to the same lexical-functional category (Best & Tyler, 2007).

The predictions made by the PAM and the PAM-L2 have also been discussed in terms of their implications for classroom learning contexts (Tyler, 2019). Since the predictions of the framework of the PAM and PAM-L2 is based on empirical research on L2 sound perception by naïve learners in an immersion environment, Tyler (2019) states that its predictions cannot be directly applied to instructed L2 learning in classrooms, where the amount and type of L2 input differs greatly from naturalistic contexts. In an L2 classroom, learners are often exposed to foreign accented L2 speech and written input. The amount of phonetic input is likely to be limited and vocabulary is often practiced via the orthographic forms of words. According to Tyler (2019), these factors cause the learning of single-category assimilations and category-goodness differences even less likely in classroom contexts than in naturalistic immersion environments.

1.1.1.3 The Native Language Magnet Model: L1 sound prototypes and the perceptual magnet effect

The Native Language Magnet Model (NLM; Kuhl et al., 1992, 2008) is based on the idea that the understanding of L1 sound perception is a gateway to understanding all L2 sound learning. According to the NLM, problems in L2 sound learning are caused by L1 category prototypes that are formed during early infancy (i.e., L1 neural commitment). The L1 sound prototypes affect all L1 and L2 sound perception after the first months of life (Kuhl, 1991; Kuhl et al., 1992, 2008).

The NLM was established when Kuhl et al. (1992) discovered that speech sound perception evolves from universal perception of acoustic differences towards language specific categorical perception of sounds during the first six months of life. According to their results, newborn children are able to discriminate all acoustic differences in speech sounds, but by six months, their perception is altered by L1 input so that the perception of L1 irrelevant contrasts is diminished and the perception of L1 relevant contrasts is enhanced. They propose that this development from universal to L1 specific speech sound perception is explained by the emergence of L1 sound prototypes. These prototypes develop as a result of L1 input when the most heard variant of an L1 sound forms a neural memory trace for that particular phoneme. In other words, L1 input during infancy results in the formation of neural memory traces (prototypes) for the L1 sound variants that are heard the most in surrounding speech. The prototype represents the most prototypical allophone of an L1 category for an individual, meaning that the prototype is perceived as the best possible realization of that phoneme.

The findings of Kuhl et al. on the formation of sound prototypes and L1 neural commitment (1992, 2008) are also supported by studies on the statistical learning of L1 sounds (Maye, Werker, & Gerken, 2002; Maye, Weiss, & Aslin, 2008). These studies suggest that L1 sound acquisition during infancy (i.e., the emergence of L1 prototypes) is facilitated by the statistical distribution of sounds. In other words, infants are able to focus on sounds that occur the most in everyday speech around them (i.e., sounds that are statistically most frequent). This leads to better discrimination between the most frequent sound categories, while the discrimination of less frequent sounds deteriorates (Maye et al., 2002, 2008).

According to the NLM, after the category prototypes are formed, they act as perceptual magnets (Kuhl, 1991; Kuhl et al., 1992), causing the perception of acoustic differences to become harder near the center of a category (i.e., near the prototype) and easier near the border of a category (i.e., away from the prototype). This phenomenon is called the perceptual magnet effect (Kuhl, 1991). The NLM proposes that the magnet effect is the basis of categorical perception, because it causes allophones to assimilate to the nearest prototype. The perceptual magnet effect also explains why the discrimination of identical acoustic differences is harder within an L1 category than between L1 categories. Therefore, the concept of the perceptual magnet effect includes the proposition that L1 phoneme categories have an internal hierarchy, where acoustical differences between sound variants are perceptually different depending on the acoustic distance to the prototype.

The NLM was later expanded when Kuhl et al. (2008) discovered that infants' L1 and L2 sound perception skills at 7.5 months of age predict the development of language skills later in life. Most importantly, the results showed that better L1 sound perception in infancy was connected to faster development of L1 language skills, whereas better L2 sound perception was linked to slower advancement of L1 skills (Kuhl et al., 2008). The neural commitment to L1 categories (i.e., the formation of L1 sound prototypes) is developmentally necessary for L1 acquisition, but it causes challenges in L2 learning later, since all speech sounds are perceived through the established L1 prototypes. Therefore, when speakers learn L2 sounds that are not relevant in their L1, they perceive the L2 sounds through their existing prototypes as good or poor realizations of an L1 phoneme. If the L2 sound is already phonological in the speaker's L1, the prototype already exists and no problems arise. If an L2 sound is acoustically and perceptually close to an L1 prototype but is distinct from all L1 categories, the speaker is faced with challenges in L2 perception and production. In a situation where an L2 phoneme does not correspond to any existing L1 prototype, the succesful perception and production of the L2 sound requires the formation of a new prototype.

1.1.1.4 The Finnish vowel system and stimulus selection in the light of second language learning models

This thesis examines the topic of children's L2 sound learning from a comparative cross-linguistic perspective, in a situation where L1 Finnish speaking children learn L2 sounds. The stimuli used in the experiment designs of Studies I–IV were selected to be challenging for L1 Finnish speakers according to the predictions made by the SLM, PAM and NLM (Best, 1994, 1995; Flege, 1995, 2007; Kuhl et al., 1992). According to the SLM, the most challenging L2 sounds are perceptually similar to L1 sounds (Flege, 1995, 2007), whereas the PAM posits that perceptual similarity leads to single-category assimilation of an L2 contrast to an L1 category (Best, 1994, 1995). The NLM, on the other hand, explains the challenges in L2 sound perception and production by the absence of L2 prototypes, which leads the perceptual magnet effect to cause the L2 sounds to be perceived through existing L1 category prototypes (Kuhl, 1991; Kuhl et al., 1992). However, the common factor in these theories is that they all propose that L2 sound perception and production is most challenging when an L2 sound or contrast is perceptually close to an L1 category, but is phonemically distinct from all L1 sounds. Therefore, for the purposes of this study, we examinded Finnish children's L2 sound learning using specific L2 vowels and vowel contrasts that are not phonemic in Finnish and are thus considered theoretically difficult for Finnish speakers to acquire according to the SLM, PAM and NLM.

The vowel chart created by the International Phonetics Association (see Figure 1) presents all the vowels of the world's languages. The Finnish sound system includes eight vowel phonemes: /i/, /y/, /e/, /ø/, /æ/, /ɑ/, /o/ and /u/. In Figure 1, the Finnish vowel phonemes are circled with green.

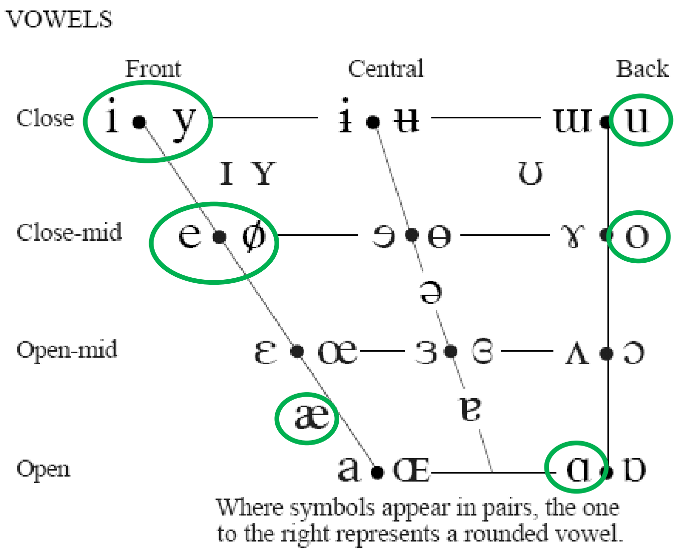

Figure 1. The official IPA vowel chart. The eight Finnish vowel phonemes are circled. The vertical and horizontal axes represent the relative F1 and F2 values of the vowels and the position of the symbols represent tongue position in the oral cavity during articulation. http://www.internationalphoneticassociation.org/content/ipa-chart, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2015 International Phonetic Association.

For the purposes of this thesis, Swedish and English vowels were selected as target L2 sounds, because both languages include vowel categories that are not part of the Finnish vowel inventory, but are easily assimilable to Finnish sound categories. In addition, both languages are studied as compulsory subjects in Finnish schools.

In Studies I, III and IV, the target L2 sound was the close rounded Swedish vowel /ʉ/, which is not phonemic in Finnish. The Swedish close rounded vowel space is divided in to three vowel categories (/y/-/ʉ/-/u/), whereas Finnish only has two close rounded vowels (/y/-/u/) in the same vowel space. In other words, the Swedish /ʉ/ is situated on the border of the Finnish /y/ and /u/ categories and the vowel contrast /y - ʉ/ is not phonemic in Finnish (see Figure 1).[1]

[1] However, the realization of the phoneme /ʉ/ varies between different dialects and regional variations. It is important to note that although /ʉ/ is phonemic in Swedish, and is therefore relevant to learn for most L1 Finnish speakers, the vowel was primarily chosen for the purposes of this study because it represents a challenging L2 sound for L1 Finnish speakers.

In Study II, the target L2 sounds were the twelve monophthong vowels of British English (BrE): /iː/, /ɪ/, /e/, /æ/, /ʊ/, /uː/, /ɔː/, /ʌ/, /ɒ/, /ɑː/, /ɜː/ and /ə/. Five of these vowels are also phonemic in Finnish (/i/, /e/, /æ/, /u/ and /ɑ/, see Figure 1). Therefore, native Finnish speakers need to learn to categorize the remaining seven BrE vowels /ɪ/, /ʊ/, /ɔː/, /ʌ/, /ɒ/, /ɜː/ and /ə/ apart from their existing L1 categories. In addition, Finnish has phonological quantity distinction, which means that vowel duration is easier for native Finnish speakers to perceive than non-native spectral differences in L2 vowels. For example, Finnish speakers might be able to discriminate between the BrE vowels /iː/ and /ɪ/, but they are likely to perceive them as long and short realizations of the same phoneme (i.e., the Finnish category /i/) and not as two spectrally different sounds.

In the light of the L2 sound learning theories, the target L2 vowels used in Studies I–IV represent theoretically difficult sounds for L1 Finnish speakers. Based on the SLM, PAM and NLM, the target vowels used in Studies I–IV (the Swedish vowel /ʉ/ and the BrE vowels /ɪ/, /ʊ/, /ɔː/, /ʌ/, /ɒ/, /ɜː/ and /ə/) can be hypothesized to be challenging for L1 Finnish speakers to perceive and produce, since they are all perceptually close to one or more Finnish vowel categories.[2]

It is important to note that the purpose of Studies I–IV was not to focus on the learning of any specific L2 sounds. The selected target sounds were used as means to collect precise, measurable data on how children from different learning backgrounds learn to produce and perceive non-native sounds in general. Therefore, the aim of this thesis is not to provide results on how Finnish children learn Swedish and English sounds, but to offer insights into how 6–13 –year-old children learn L2 phonemes.

1.1.2 Age and L2 sound learning: Studies on child and adult learners in naturalistic environments and phonetic training settings

The question of how age affects L2 sound learning is one of the most investigated topics in phonetics alongside comparative cross-linguistic research. The consensus seems to be that children are superior to adults in L2 sound learning. One of the classic hypotheses in language learning is the Critical Period Hypothesis (CPH),

[2] The vowel /ʉ/ is a rounded vowel, which makes it a marked sound according to the Markedness Differential Hypothesis (MDH; Eckman, 1977). The MDH states that the rounded /ʉ/ is more marked than unrounded vowels in the same acoustic space and therefore /ʉ/ is likely to cause more challenges for language learners. However, although /ʉ/ is considered a universally marked sound, the Finnish vowel system has two other close rounded vowels /y/ and /u/. Therefore, the rounded L2 vowel /ʉ/ could be argued to be less marked for L1 Finnish speakers.

which suggests that the ability to learn languages diminishes abruptly after puberty as the neural plasticity in the brain decreases (Lenneberg, 1967). The CPH has been widely debated over the years and some more recent studies have offered evidence against the hypothesis of abrupt decrease in L2 learning abilities after a certain age (e.g., Wang & Kuhl, 2003). However, several studies have shown that early age of learning (AOL) often results in more accurate or nativelike perception and production of L2 sounds (e.g., Baigorri, Campanelli, & Levy, 2019; Flege, Yeni-Komshian, & Liu, 1999b; Flege et al., 1999a; Munro, Flege, & Mackay, 1996; Oh et al., 2011; Piske, Flege, MacKay, & Meador, 2002; Tsukada et al., 2005).[3]

Most of the studies on the effects of AOL on L2 sound learning have focused on different immigrant populations. In addition, many of the earlier studies have focused on L2 pronunciation by examining how AOL affects accentedness in L2 speech. Studies on Italian immigrants learning L2 English in Canada and the US have shown that later AOL is related to stronger perceived L2 accent (Flege et al., 1999b; Munro et al., 1996; Piske et al., 2002) and lower L2 vowel intelligibility scores (Flege et al., 1999a) when L2 spekers' pronunciation is evaluated by L1 listeners. The results of Flege et al. also showed that early AOL correlated with higher L2 vowel discrimination accuracy (1999a).

Some of the more recent studies have focused on L2 production and perception accuracy instead of accentedness. For example, a study by Tsukada et al. (2005) compared how L1 Korean speaking children (aged 9–17 years) and adults (aged 23–41 years) with different lengths of residence (LOR) learn to perceive and produce English vowels after moving to Canada or the US. Four subject groups were formed according to the participants' age (children vs. adults) and LOR (3 vs. 5 years). The groups' perception and production accuracy were measured with discrimination and production tests at the start of the experiment and one year later. L1 English speakers served as controls in the discrimination and production tests. The discrimination results showed that the children reached higher vowel discrimination accuracy accuracy than the adults, and furthermore that the children in the 5-year LOR group performed better that the children in the 3-year LOR group. The production results showed that the child participants reached more nativelike vowel production accury than the adults (Tsukada et al., 2005).

A similar study by Oh et al. (2011) investigated how age of arrival (AOA) affects English vowel production learning. The subjects were Japanese adults (mean age 39.9 years) and children (mean age 9.9 years) who had moved to the US. In contrast

[3] A study on early and late bilingual speakers' L2 nativelikeness by Abrahamsson and Hyltenstam (2009) suggests that nativelike ultimate attainment of an L2 is practically never reached by adult learers and, interestingly, is much less common among child learners than has been thought.

to Tsukada et al. (2005), the average LOR was similar for both groups. Their English vowel production was tested with speech recordings shortly after arrival to the US and one year later. Their production accuracy was measured with acoustic analysis (vowel duration and quality) and their productions were compared to those of age-matched L1 English speakers. The results showed that the Japanese children reached higher vowel production accuracy than Japanese adults in one year of living in the US. Even though the adults produced the L2 vowels more accurately in the first recording, they showed no improvement over time and the children outperformed them in the second recording (Oh et al., 2011).

Another study by Baigorri et al. (2019) examined the effects of early and late AOL on Spanish-English bilingual speakers' perception of cross-language differences. They tested two groups of L1 Spanish speakers (aged 18–48 years) who had acquired English as an L2 after immigrating to the US. The first group consisted of early learners of English (AOL <11 years) and the second group of late learners of English (AOL >13 years). The participants' cross-linguistic perception was measured with L2-L1 vowel assimilation tests and L2 vowel discrimination tests. The results showed that both groups assimilated L2 vowels to L1 categories similarly. However, the early learner group reached higher L2 vowel discrimination accuracy than the late learners, but both groups had difficulties with discriminating certain L2 contrasts. Baigorri et al. (2019) conclude that their results suggest that early AOL leads to better perception of cross-language phonetic differences, but the L1 sound system continues to affect L2 perception even in early bilinguals.

However, these studies on child and adult learners have focused mostly on immigrant populations and naturalistic learning environments. In other words, most of the results on the positive effects of early AOL/AOA have been obtained from data that focuses on a very limited type of L2 acquisition enviroments. Based on the results of these studies (e.g., Baigorri et al., 2019; Flege, Munro, & Mackay, 1995a; Flege et al., 1999b; Oh et al., 2011; Tsukada et al., 2005), it is evident that children often reach more nativelike L2 perception and/or production when their progress is measured after moving to a new L2 environment. However, various important social and societal factors cannot be controlled when L2 acquisition is studied in immigration settings. For example, one of the factors that should be studied and taken into account more is whether it is easier for children than adults to be immersed in a new L2 environment, because attending school or daycare forces them to use the target L2, as it is usually their only way to communicate with other speakers. Furthermore, it seems that early AOL does not automatically lead to nativelike perception of L2 sounds and that LOR affects L2 sound discrimination accuracy even in child learners (Tsukada et al., 2005). In addition, since the amount of L1 use has also been found to affect L2 sound learning (Flege, Frieda, & Nozawa, 1997), its possible effects on differences between child and adult learners cannot be entirely discarded. Therefore, it is important to examine results from studies that have focused on child and adult learners in different L2 learning settings to see whether similar age differences can be found outside immigration contexts.

One way to investigate age effects on L2 phonetic learning is to use different perception and production training paradigms in controlled laboratory conditions. A study by Giannakopoulou et al. (2013) compared how Greek children (aged 7–8 years) and adults (aged 20–30 years) respond to high-variability phonetic training (HVPT) of English vowels. The training effects were measured with phoneme discrimination and identification tasks before and after training. Interestingly, the results showed that both groups benefitted from the perceptual training, but the learning effects were more pronounced for child learners. This result seems to offer further evidence against the CPH and the proposition that the developmental window for L2 learning closes after puberty. However, Giannakopoulou et al. suggest that their results imply enhanced plasticity for spoken language at the child learners' developmental stage (2013).

Another study investigated how listen-and-repeat training affects 7–10-year-old Finnish children's production of a non-native vowel contrast (Taimi, Jähi, Alku, & Peltola, 2014). The children participated in four short listen-and-repeat training sessions on two consecutive days (two training sessions per day). The acoustic stimuli were two two-syllable pseudowords with the non-native vowel contrast /y - ʉ/ embedded in the first syllable. The acoustic stimuli used in the experiment were the same as used in Studies III and IV of this thesis. The children's production accuracy was measured with pre- and post-tests as well as two intermittent recordings between trainings (four recording sessions in total). The results showed clear training effects on the production of the L2 vowel. The children changed their production of the L2 vowel /ʉ/ significantly after three training sessions (Taimi et al., 2014). This result indicates that 7–10-year-old children can learn to produce an L1 irrelevant L2 sound contrast quickly with phonetic training.

The study by Taimi et al. (2014) only tested children, but for the purposes of this thesis, it is interesting to compare the results to findings from other studies that have used similar training paradigms on adults. For example, Peltola, Rautaoja, Alku and Peltola (2017) tested L1 Finnish and L1 English speaking adults with a similar listen-and-repeat training paradigm. The procedure included one training session with baseline and endpoint production tests. The results showed no significant changes in production after one training session (Peltola, K. U. et al., 2017). However, a later study showed that a two-day listen-and-repeat training protocol with the same acoustic stimuli did improve L1 Finnish speaking adults' L2 vowel production and identification (Peltola, K. U., Tamminen, Alku, Kujala, & Peltola, 2020). The results of Peltola et al. (2017, 2020) indicate that training effects on adult learners may depend on the amount of phonetic training.

Studies have also examined the effects of phonetic training on the perception of the Finnish consonant quantity /t – t:/ in Dutch 12-year-old children (Heeren & Schouten, 2010) and adults (Heeren & Schouten, 2008). The two studies used the same experiment design with discrimination and identification tests before and after training. The training consisted of trials of the same identification task as in the pre- and posttests. The amount of training differed slightly between the child and adult participants, as the children received a fixed amount of training and the adults continued training until they reached a certain performance level (Heeren & Schouten, 2010, 2008). However, the children received a total amount of training that equaled the median amount of training completed by the adults (Heeren & Schouten, 2010). The results showed that both groups' identification scores improved slightly, which means that their category boundary of the /t – t:/ distinction tended to move towards native Finnish listeners' category boundary. However, there were no significant changes in the children's perceptual sensitivity in the discrimination tests and the adults showed higher overall discrimination scores than the children (Heeren & Schouten, 2010, 2008).

Research comparing child and adult L2 learners in instructed classroom environments is extremely limited. A recent study by Kopečková, Dimroth and Gut (2019) examined German children (aged 9–10 years) and adults' perception and production of Polish sibilants during an instructed Polish course. The participants' perception (phoneme discrimination task) and production (sentence imitation task) of the Polish sibilants was tested at two time points during the course (14 hours of teaching in total). Phoneme discrimination was tested after 4.5 hours and 11.5 hours, and sentence imitation after 9 and 13.5 hours. The results showed no differences in production between the child and adult learners. However, the adults discriminated the L2 sibilants more accurately than the children at both testing times. Kopečková et al. (2019) propose that child learners are not necessarily faster or better at perceiving L2 phonemes in instructed learning settings.

These earlier studies show that early AOL often seems to lead to superior results in L2 sound perception and production (e.g., Baigorri et al., 2019; Flege et al., 1995a; 1999b; Giannakopoulou et al., 2013; Oh et al., 2011; Tsukada et al., 2005). However, some results indicate that adults can also learn L2 perception and production with the right amount of training and input (Heeren & Schouten, 2010, 2008; Kopečková et al., 2019; Peltola, K. U. et al., 2020) and that adults may actually outperform children in instructed learning of L2 sound perception (Kopečková et al., 2019). Therefore, age effects on L2 sound learning seem to be more complicated than the fast deterioration after puberty predicted by the CPH (Lenneberg, 1967). Especially research comparing different child learners remains limited, and the concept of age in L2 sound learning needs to be expanded outside the classic child-adult distinction before we can truly understand how age affects L2 learning in children and adults from different learning backgrounds.

1.1.3 Different background factors and L2 sound learning: Studies on language immersion education and musical experience

Studies have shown that age is not the only significant factor in L2 sound learning. For example, the amount of L1 use (Flege et al., 1997), length of residence (Tsukada et al., 2005) and musical experience and abilities (e.g., Delogu, Lampis, & Belardinelli, 2006; Marie, Delogu, Lampis, Belardinelli, & Besson, 2011; Milovanov, Huotilainen, Välimäki, Esquef, & Tervaniemi, 2008) have been found to affect L2 sound perception and production to some extent.

Studies comparing learners with early and late AOLs have offered evidence that earlier AOL is likely to lead to more nativelike L2 perception and production patterns than late AOL (e.g., Flege et al., 1995a, 1999b; Giannakopoulou et al., 2013; Oh et al., 2011; Tsukada et al., 2005). However, these studies have focused almost solely on naturalistic L2 environments or phonetic training paradigms. There is very little research on how L2 classroom environments (e.g., immersion education) or other background factors (e.g., musical experience) could affect childrens L2 sound learning. Therefore, in order to understand L2 sound learning better, we have to examine which factors, besides age, may be significant in L2 perception and production learning.

Some studies have investigated how children learn L2 sounds in different immersion education settings, though research on this issue remains relatively limited. For example, a study by Peltola, Kuntola, Tamminen, Hämäläinen and Aaltonen (2005) examined how early L2 immersion education affects the preattentive discrimination of L2 vowel contrasts in children. The study tested two groups of Finnish children (aged 5–7 years). The first group consisted of monolingual children with no L2 experience and the second group consisted of children who had received early L2 exposure in a French immersion daycare. The children's preattentive discrimination of two vowel contrasts (one phonological in French and the other phonological in Finnish) was tested by measuring the mismatch negativity (MMN) responses with electroencephalography (EEG). The results showed that both contrasts (L1 and L2) elicited similar MMN responses in the immersion group, while the L1 contrast elicited a larger response than the L2 contrast in the monolingual group (Peltola, M. S. et al., 2005). This finding suggests that early L2 exposure in an immersion daycare leads to enhanced preattentive discrimination of L2 sound contrasts. However, a later study by Peltola, Tuomainen, Koskinen and Aaltonen (2007) offered contrasting evidence on the effects of L2 immersion education on preattentive L2 sound perception. They tested how two groups of Finnish children (aged 14 years) with different L2 experience perceived L1 and L2 vowel contrasts. The two groups consisted of monolingual Finnish children (controls) and children who had participated in early or partial English immersion education. The groups' MMN responses to Finnish and English vowel contrasts was measured with EEG. The results showed that the most salient L2 contrast elicited a MMN response in both groups, whereas the L1 contrast elicited a response only in the monolingual group (Peltola, M. S. et al., 2007). Contrary to earlier results (Peltola, M. S. et al., 2005), the findings of Peltola et al. (2007) indicate that L2 immersion education in a classroom setting might not lead to the formation of nativelike memory traces of L2 sounds.

A study by Darcy and Krüger (2012) examined how early dual-language education in daycare affects the interaction of children's L1 and L2. They tested two groups of sequential bilingual and monolingual participants (aged 9–12 years). The sequential bilinguals were L1 Turkish speaking children, who had started to learn L2 German between the ages of 2–4 in a dual-langauge (Turkish and German) daycare. The monolinguals were L1 German speaking children. The groups' perception and production of German vowel contrasts were measured with discrimination and picture naming tasks. The stimuli used in the discrimination task were CVC syllables with German vowels embedded in bilabial /p/ and velar /k/ consonant contexts. In the production task, children were asked to name common German words (that included the same target vowels as in the discrimination experiment) pictured in memory game cards. The results showed that the sequential bilinguals' discrimination of German vowels was not nativelike, in other words their discrimination sensitivity differed from the L1 German children's discrimination results. However, the results of the production tests showed that the sequential bilinguals reached mostly nativelike accuracy in their productions of the same German vowel contrasts (Darcy & Krüger, 2012).

These results (Darcy & Krüger, 2012; Peltola, M. S. et al., 2005, 2007) offer rather contradictory information on the effects of early immersion education on L2 perception and production. On one hand, it seems that early L2 exposure in immersion daycare may lead to enhanced perception (Peltola, M. S. et al., 2005) and production (Darcy & Krüger, 2012) of L2 sounds. On the other hand, some of these findings indicate that immersion style education does not necessarily result in improvements in L2 perception (Darcy & Krüger, 2012; Peltola, M. S. et al., 2007). The results of Darcy and Krüger (2012) are particularly interesting, since they are in contrast with the proposition that L2 sound perception precedes production (Flege, 1995, 1999; Flege et al., 1999a). However, Darcy and Krüger (2012) state that the vowel discrimination task used in their study may not reflect the children's perceptual abilities fully. This could be due to the artificial and abstract nature of the task, or because the sequential bilingual children may have encoded the vowel contrasts lexically first, causing them to perform better in the production task with real words rather than separate syllables. Nevertheless, the effects of L2 immersion education on L2 sound perception and production remain unclear and more research is needed to draw any definite conclusions on the topic. Furthermore, since previous studies have focused mainly on immersion daycare, research on different immersion education settings in school is needed to discover how older school-aged children learn to perceive and produce L2 sounds in immersion education.

Contrary to immersion education settings, the relationship between musical and linguistic processing has been widely researched from varying perspectives in recent years. Results from studies on the effects of different musical background factors on L2 perception have suggested that there is some interplay between music and L2 sound processing. For example, a study by Milovanov et al. (2008) investigated the relationship between L2 pronunciation skills and musical aptitude in 10–12-year-old children. They used preattentive MMN measurements and different behavioral tests to examine whether Finnish children with superior L2 production skills differ from children with less advanced L2 production skills in their musical aptitude scores. The children were first tested with a Seashore musical aptitude test as well as chord and L2 (English) phoneme discrimination pre-tests. After the initial testing, the children practiced English pronunciation at home for eight weeks. After the training period, the children's L2 pronunciation skills were tested with an English production test. The participants were then divided into two groups (advanced and less advanced) according to their L2 production skills. The experiment concluded with chord and L2 phoneme discrimination post-tests and preattentive MMN measurements with EEG (using the chord and phoneme stimuli from the discrimination tests). The results showed that the children in the advanced L2 production group obtained higher musical ability scores for timbre, pitch discrimination, sense of tonality and sense of rhythm. Furthermore, the MMN results showed that the children with advanced L2 skills had more pronounced sound-change evoked activation for the chord stimuli than the children with less advanced L2 skills. The researchers propose that, in the light of these results, musical and linguistic skills could at least partly be based on shared neural mechanisms (Milovanov et al., 2008).

Findings from several other studies have also suggested some interplay between musical expertise or ability and linguistic processing. For example, a study by Marie et al. (2011) investigated how French musicians and non-musicians perceive tonal and segmental variations in Mandarin Chinese words. Behavioral discrimination tests and event-related potential (ERP) measurements with EEG were used to measure perception accuracy. The behavioral results showed that the musicians discriminated both tonal and segmental variations more accurately (with lower error rates) than non-musicians. However, both groups showed higher error rates for the tonal variations than the segmental variations. In addition, the ERP results showed that the tonal variations elicited an earlier response in musicians than non-musicians. Delogu et al. (2006) also investigated the perception of lexical tones by testing whether melodic ability affects tone discrimination in L1 Italian speakers. Their results showed that the speakers with high melodic ability discriminated lexical tones more accurately than speakers with lower melodic ability. However, all speakers performed better overall in discriminating phonological differences than tonal differences.

Some studies have found a connection between rhythm perception and L2 learning. For example, a study by Bhatara, Yeung and Nazzi (2015) found positive correlations between L1 French speakers' rhythm perception and L2 experience, and rhythm perception and music training. They propose that L1 French speakers' perception of rhythm is related to the amount of music training and L2 experience. However, their results showed no correlations with melody perception. In addition, a later study by Boll-Avetisyan, Bhatara, Unger, Nazzi and Höhle (2016) found that musical and L2 experience affect L1 French speakers' rhythmic grouping preferences of L2 German syllables.

Studies on music and language offer evidence that there may be some connections between musical expertise, melodic abilities and the perception of tonal and segmental variations (Delogu et al., 2006; Marie et al., 2011), as well as between rhythm perception, music experience and L2 input (Bhatara et al., 2015; Boll-Avetisyan et al., 2016). From a broader perspective, there are also some indications that L2 production skills and musical aptitude are connected, and that shared neural mechanisms could be involved in the development of musical and linguistic skills (Milovanov et al., 2008). However, the vast majority of previous research has focused on different measurable aspects of musical ability and L2 perception. Therefore, the relationship between participating in musical activities and L2 sound perception and production learning remains unclear. In addition, most of the studies on music and language have tested adult subjects (e.g., Bhatara et al., 2015; Boll-Avetisyan et al., 2016; Delogu et al., 2006; Marie et al., 2011). More research is needed on how general musical activities (not related to musicality or musical abilities) might affect L2 sound learning in different age groups, including children.

2 Aims of the study

This chapter introduces the two main aims of the thesis as well as the specific research questions and hypotheses of Studies I–IV. First, the aims of Studies I and IV focus on the effects of age on children's L2 phonetic learning and they are presented in section 2.1. Studies II and III exmine the effects of other learning background factors on children's L2 sound learning and their aims are introduced in section 2.2.

2.1 The effects of age on children's L2 sound production and perception learning

Age of learning (AOL) has been found to affect L2 phonetic learning, especially after immigration to a new L2 environment (e.g. Flege et al., 1999b; Oh et al., 2011; Tsukada et al., 2005). Child learners usually reach a more nativelike level of L2 perception and production faster when their performance is compared to adults who have lived in the same L2 environment for an equal amount of time (Baigorri et al., 2019; Oh et al., 2011; Tsukada et al., 2005). Age has also been found to be a significant factor in some perceptual L2 training studies that have compared child and adult learners of an L2 (e.g. Giannakopoulou et al., 2013). However, some studies have found the amount of L1 and L2 use to be even more important than age in L2 phonetic learning (e.g. Flege, Munro, & MacKay, 1995b; Flege et al., 1997). In addition, some earlier studies have found that early L2 immersion education might or might not improve children's L2 sound perception (Peltola, M. S. et al., 2005, 2007) and that perceptual training of phonological contrasts does not necessarily lead to improvements in L2 discrimination accuracy in children (Heeren & Schouten, 2010). Taken together, these earlier results indicate that children might have a developmental bias to acquiring non-native speech sounds (e.g., Giannakopoulou et al., 2013). However, since most of the research on the effects of age on L2 learning have compared child and adult learners or have focused on immigrant populations, there is not enough information on how children of different ages learn L2 sounds outside a naturalistic L2 environment, in different instructed learning settings.

Studies I and IV investigated how Finnish children respond to two different types of phonetic training settings of a non-native sound contrast. Study I investigated L2 sound perception and Study IV examined L2 sound production. The shared aim in these studies was to see whether young children are as naturally efficient L2 learners as earlier research (e.g. Giannakopoulou et al., 2013; Oh et al., 2011; Tsukada et al., 2005) has suggested. We wanted to test whether the proposition that earlier is better is true in laboratory and classroom conditions where young monolingual children are trained with synthetic or semisynthetic phonetic stimuli.

The overall hypothesis concerning Studies I and IV was that the child participants' age would affect training results. In other words, we hypothesized that the children in Studies I and IV would benefit from their young age and enhanced neural plasticity in phonetic training of L2 sound perception and production. The more detailed research questions and hypotheses for Studies I and IV are presented below.

2.1.1 Research questions and hypothesis of Study I

Study I investigated the effects of passive auditory exposure on 7–12-year-old children's attentive and preattentive perception of non-native vowels. The aim was to see whether it would be possible to create a new memory trace for an L2 vowel category by training the participants with different synthetic variants of the L2 category.

For the purposes of Study I, we wanted to use a training paradigm that would mimic natural exposure to speech sounds in real life. Earlier research has suggested that the effective acquisition of L1 phonemes in early infancy is the result of statistical learning. This means that the statistical distribution information of the different sounds that infants hear around them allows them to focus on the most frequent sounds (Maye et al., 2002, 2008). Therefore, the training paradigm of Study I was designed so that the most prototypical variant of the trained L2 vowel was presented statistically the most during training. The research questions of Study I were:

1. Does children's neural plasticity enable the emergence of a new memory trace for a non-native vowel through unimodally distributed auditory training of the vowel category?
2. Does auditory training result in changes in children's non-native sound perception on the behavioral and pre-attentive levels?
3. Are the training effects different near the trained L2 prototype and category boundary?

The hypothesis was that due to the children's neural plasticity, the passive auditory training would result in the formation of a new memory trace for the trained non-native vowel. Furthermore, the second hypothesis was that if a new memory trace for the non-native vowel were formed, the perceptual magnet effect (Kuhl, 1991) would cause perception to become more difficult near the trained prototype and easier near the new category boundary. This hypothesis was based on the theoretical predictions made by the NLM (Kuhl et al., 1992, 2008). In addition, we hypothesized that the training effects would be reflected pre-attentively in the mismatch negativity (MMN) responses and behaviorally in the discrimination sensitivity and reaction times. The MMN is an event-related potential (ERP) measured with EEG. The MMN is elicited pre-attentively whenever the subject's brain perceives a difference (Näätänen et al., 1997; Näätänen, 2001; Näätänen, Paavilainen, Rinne, & Alho, 2007). The MMN is widely used in phonetic studies to measure preattetive non-native sound perception in children and adults (e.g. Peltola, M. S. et al., 2007; Peltola, M. S., Tamminen, Toivonen, Kujala, & Näätänen, 2012; Shestakova, Huotilainen, Čeponienė, & Cheour, 2003; Winkler et al., 1999) and it was therefore chosen for the purposes of Study I.

2.1.2 Research questions and hypothesis of Study IV

The aim of Study IV was to examine Finnish preschoolers' ability to learn non-native sound production in an experimental situation, because the majority of previous research on child learners of an L2 has been conducted in bilingual or language immersion environments. The phonetic listen-and-repeat training method was selected because it has traditionally been used to teach foreign language pronunciation in classrooms. The experiment was designed to answer the following two research questions:

1. Can 6–7-year-old Finnish children learn to produce a difficult non-native vowel contrast during four short sessions of listen-and-repeat training?
2. If the children's production of the L2 sound contrast changes as a function of training, how fast and in what direction does the change occur?

The first hypothesis was that 6–7-year-old children's production would change as a function of listen-and-repeat training. This expectation was based on previous studies that have shown that children learn L2 pronunciation succesfully in bilingual environments (Darcy & Krüger, 2012; Oh et al., 2011; Tsukada et al., 2005). The second hypothesis was that the children's production of the non-native vowel /□/ would change towards the acoustic stimulus. Previous findings on the positive effects of phonetic training on children's L2 sound perception accuracy (Giannakopoulou et al., 2013) and the connection of L2 sound perception and production (e.g. Flege, 1995, 1999; Flege et al., 1999a) suggest that phonetic training could also benefit children's L2 production learning. Most importantly, similar listen-and-repeat training paradigms have been found to lead to production changes also in adult learners (Peltola, K. U. et al., 2020) and 7–10-year-old children (Taimi et al., 2014). Therefore, one of the objectives was to compare the results of Study IV to earlier findings by Peltola et al. (2020) and Taimi et al. (2014) from older subject groups to see whether age affects training results.

The third hypothesis on how rapidly the change in production would occur was twofold: The change could be very fast due to the children's age and plasticity. On the other hand, the listen-and-repeat method could be cognitively too challenging and require much concentration from 6–7-year-old children, who have not yet started school. Despite earlier findings on children's success in L2 perception and production learning (Giannakopoulou et al., 2013; Oh et al., 2011; Taimi et al., 2014; Tsukada et al., 2005), we expected that the classroom-like nature of the training task in Study IV could affect how rapidly training effects towards the L2 vowel category would emerge.

2.2 The effects of learning backgound on children's L2 sound production learning

Studies II and III examined how language immersion education and music-oriented education in elementary school affect L2 production learning. The majority of studies on children's L2 learning have compared child and adult learners or focused on bilingual children (e.g., Giannakopoulou et al., 2013; Oh et al., 2011; Ramon-Casas, Swingley, Sebastián-Gallés, & Bosch, 2009; Tsukada et al., 2005). Therefore, there has been very little research on how monolingual children from the same age group with different learning backgrounds learn L2 sounds. It seems that since children are regarded as naturally efficient L2 learners (e.g., Giannakopoulou et al., 2013), there has been little focus on whether other background factors aside from age can cause differences between child learners of an L2. Therefore, the aim of studies II and III was to investigate how language immersion education and music-oriented education in elementary school affect children's L2 production. The overall hypothesis was that children's linguistic or musical learning background would affect L2 production learning.

Study III focused on 9–11-year-old children from music-oriented and regular fourth grades. The study used a phonetic training protocol to examine whether the children's musical background would affect training results. Study II, on the other hand, was an intervention study that tested how 11–13-year-old children from an English immersion education program and a Finnish speaking class produce English vowels. Since the children in the immersion program had begun their English studies earlier and received more English input than the children in the Finnish speaking class, the two groups differed not only in the manner of learning, but also in the amount of input and age of learning. However, the focus of Study II was not directly on the effects of age on L2 learning, but rather on the overall effects of language immersion education on L2 production learning.

2.2.1 Research questions and hypothesis of Study II

The purpose of Study II was to discover how language immersion education in elementary school affects Finnish children's L2 pronunciation. The research questions were:

1. Does daily L2 exposure received in an English immersion education program in elementary school affect 11–13-year-old Finnish children's production of British English vowels?
2. Do children from an English immersion education program produce British English vowels more accurately than children attending a regular Finnish speaking class?

Since the children in the English immersion program had studied English from an earlier age and received more English input in school, the primary hypothesis was that attending an English immersion education program would affect children's English pronunciation. More precisely, we hypothesized that 11–13-year-old Finnish children from an English immersion education program would produce British English vowels more accurately than their peers from a regular Finnish speaking class when tested with a listen-and-repeat protocol. The hypotheses were based on earlier research on the benefits of early L2 exposure on L2 perception and production (e.g., Darcy & Krüger, 2012; Flege et al., 1999b; Oh et al., 2011). The aim of the study was not to measure exact nativelikeness but possible differences between the two groups that had different amounts of English exposure in school.

2.2.2 Research questions and hypothesis of Study III

Study III examined whether studying in a music-oriented education program affects children's L2 sound production learning through phonetic auditory training. Earlier studies have provided evidence that musical experience and L2 phonetic processing might be connected (e.g., Bhatara et al., 2015; Boll-Avetisyan et al., 2016; Marie et al., 2011; Milovanov et al., 2008). However, the majority of those studies have focused on adult subjects (Bhatara et al., 2015; Boll-Avetisyan et al., 2016; Marie et al., 2011). Therefore, the question of whether children benefit from musical experience in L2 phonetic learning remains unclear. There were three main research questions in Study III:

1. Do 9–11-year-old children attending a music-oriented education program in elementary school have a sensitivity to acoustic variation that is transferred to the trainability of non-native sound contrasts?
2. Does auditory training of an L2 vowel contrast without any production training or articulatory instructions result in changes in 9–11-year-old children's L2 vowel production?
3. Do 9–11-year-old children from a music-oriented and a regular fourth grade respond differently to auditory training of an L2 sound contrast?

The first hypothesis was that musical background would affect chldren's sensitivity to acoustic variation in speech sounds, which would be reflected in the training results. The second hypothesis was that children from a music-oriented class would learn to produce the trained L2 vowel contrast faster through auditory training than children from a non-music class. Therefore, the hypotheses included the assumption that mere auditory training would result in some changes in L2 sound production at least in the musical group. These predictions were based on earlier findings on the interplay and overlap between musical experience and L2 phonetic processing (Bhatara et al., 2015; Boll-Avetisyan et al., 2016; Marie et al., 2011; Milovanov et al., 2008).

2.3 Summary of the aims and objectives of the studies

The purpose of these studies was to examine how different background factors affect children's second language sound learning. The background factors that we were interested in were age, L2 immersion education and music-oriented education. There were two primary aims: First, to see how children of different ages learn non-native sounds (studies I and IV). Second, to see how children from the same age group with different linguistic or musical backgrounds learn non-native sounds (studies II and III).

Studies I and IV investigated the effects of age on children's non-native sound perception and production learning through phonetic training. Studies II and III investigated whether L2 immersion education or music-oriented education would result in differences in L2 pronunciation and production learning between children of the same age.

3 Materials and methods

This chapter introduces all the materials and methods (participants, stimuli, procedure and analysis) used in Studies I–IV. The materials and methods of each study are first presented separately in sections 3.1–3.4 and then summarized in section 3.5.

3.1 Study I

3.1.1 Participants

Seven monolingual Finnish children (aged 7;7–12 years, mean age 9;2, 4 females, age reported as years;months) were tested in Study I. All participants were right handed and of normal hearing. None of the participants had a history of speech difficulties. The younger participants (aged 7;7–8;11) had no experience with foreign languages, whereas the older children (aged 9–12 years) had already started to study English in school on the third grade at the age of nine. All children and their parents gave written informed consent prior to the experiment.

3.1.2 Methods

3.1.2.1 Stimuli

The stimuli were 34 synthetic vowels that represented the Swedish category /ʉ/ from close rounded vowel continuum /y/ - /ʉ/ - /u/. The Swedish /ʉ/ is situated between the Finnish /y/ and /u/ categories, and it is therefore perceptually difficult to identify and discriminate from the native sound categores for Finnish speakers (e.g., Best, 1994; 1995; Flege, 1995). The 34 stimuli were selected based on an earlier study with native Swedish speakers to ensure that they accurately represented the /ʉ/ category (Peltola, M. S. et al., 2012). The stimuli were synthesized in 20 Mel steps in the F1 (range 231–474 Hz) and F2 (range 740–1544 Hz) continuums using HLsyn software (see Table 1). Stimulus duration was 350 ms. The 30 ms amplitude ramps were set at 0–30 ms and 320–350 ms. The stimuli were presented with a 500 ms inter-stimulus interval (ISI) in all stimulus blocks during training and testing.

From the 34 stimulus vowels, 33 were used in the auditory training so that the most prototypical examplar of /ʉ/ from the center of the category was presented the most and the surrounding variants were presented in a decreasing amount according to their acoustic distance from the prototype (see Figure 2). The training session consisted of three blocks, which included 1864 repetitions of the 33 stimuli in total. Two of the three training blocks were 13 minutes and one 27 minutes long. This training design was based on the theoretical predictions made by the NLM (Kuhl et al., 1992), which propose that the category prototype is the most frequently heard variant of a phoneme. The prototype acts as a perceptual magnet (Kuhl, 1991), causing discrimination to become more difficult near the center of the category (i.e. near the prototype) and easier around category boundaries (i.e. further away from the prototype). For more details on the perceptual magnet effect, see chapter 1.

Two standard and deviant stimulus pairs were selected for the EEG recordings and discrimination tasks. The stimulus pairs were used to form two test blocks (the Prototype and the Boundary blocks, see Table 1). The prototypical Swedish vowel /ʉ/ from the center of the synthesized stimulus space served as the standard and its non-prototypical variant as the deviant in the Prototype block. The Boundary block included a non-prototypical /ʉ/ near the boundary of Swedish /y/ and /ʉ/ categories as the standard stimulus and a Finnish /y/ as the deviant stimulus. The vowel /y/ was not included in the training. The EEG blocks (Prototype and Boundary) included 914 standards and 140 deviants (deviant probability 13.3 %) and lasted approximately 17 minutes. The discrimination blocks consisted of 130 standards and 20 deviants (deviant probability 13.3%) and lasted approximately three minutes.

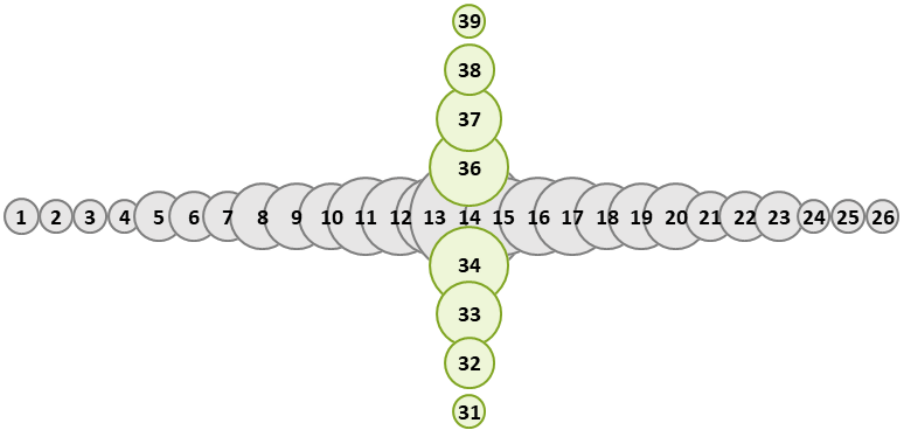

Figure 2. The synthesized vowel space for Swedish /ʉ/. The vertical dimension reflects the F1 and the horizontal dimension the F2 of each stimulus. The size of the spheres reflects the amount of repetitions during training. The unimodal distribution was selected for training to imitate natural formation of new vowel categories, when the most heard variant of a sound becomes the category prototype.

Table 1. The stimulus vowels from the synthesized Swedish /ʉ/ category. Stimulus number 14 from the center of the served as the prototypical /ʉ/. Stimulus number 1 was situated across the Swedish /ʉ/ - /y/ category boundary, i.e. it represented the Swedish vowel /y/.

STIMULI					**TRAINING**		**TEST BLOCKS**
	F1	**Mel steps**	**F2**	**Mel steps**	**Repetitions**	**%**	
1	300		1544	1313	0	-	Deviant (Boundary)
2	300		1505	1293	17	0.9	
3	300		1466	1273	17	0.9	Standard (Boundary)
4	300		1428	1253	17	0.9	
5	300		1390	1233	38	2.0	
6	300		1354	1213	38	2.0	
7	300		1318	1193	38	2.0	
8	300		1282	1173	63	3.4	
9	300		1247	1153	63	3.4	
10	300		1213	1133	63	3.4	
11	300		1179	1113	90	4.8	
12	300		1146	1093	90	4.8	Deviant (Prototype)
13	300		1114	1073	90	4.8	
14	**300**	**402**	**1082**	**1053**	**200**	**10.7**	Standard (Prototype)
15	300		1051	1033	90	4.8	
16	300		1020	1013	90	4.8	
17	300		989	993	90	4.8	
18	300		960	973	63	3.4	
19	300		931	953	63	3.4	
20	300		902	933	63	3.4	
21	300		874	913	38	2.0	
22	300		846	893	38	2.0	
23	300		819	873	38	2.0	
24	300		792	853	17	0.9	
25	300		766	833	17	0.9	
26	300		740	813	17	0.9	
51	374	482	1082		17	0.9	
52	355	462	1082		38	2.0	
53	336	442	1082		63	3.4	
54	318	422	1082		90	4.8	
55	282	382	1082		90	4.8	
56	265	362	1082		63	3.4	
57	248	342	1082		38	2.0	
58	231	322	1082		17	0,9	
					Total: 1846	100 %	

3.1.2.2 Procedure

The experiment was conducted on two consecutive days in a sound attenuated room in the phonetics laboratory. The participants were monitored via webcam from an adjoining room. The procedure lasted approximately two to three hours each day.

The first day started with a hearing test using an audiometer and a handedness test using the Edinburgh handedness inventory (Oldfield, 1971). The experiment began with a baseline EEG recording measuring the MMN with the Prototype and Boundary blocks. After the EEG, a baseline discrimination task with reaction time (RT) and discrimination sensitivity (d') measurements was conducted with the same Prototype and Boundary stimulus pairs. The first day concluded with passive auditory training, where the participants heard the 33 different variants of the synthesized /ʉ/ category 1846 times in a unimodally distributed paradigm (see Table 1 and Figure 2). The second day of the experiment proceeded in a reverse order, in other words the day began with the training, followed by the discrimination task and concluded with the final EEG. The course of the experiment is described in Table 2.

Table 2. The experiment procedure in Study I.

DAY 1	DAY 2
EEG (MMN) -Prototype block -Boundary block → *counterbalanced*	**Auditory training** -13+13+27 minutes → 1864 repetitions of the 33 stimuli
Discrimination (RT + d') -Prototype block -Boundary block → *counterbalanced*	**Discrimination (RT + d')** -Prototype block -Boundary block → *counterbalanced*
Auditory training -13+13+27 minutes → 1864 repetitions of the 33 stimuli	**EEG (MMN)** -Prototype block -Boundary block → *counterbalanced*

The EEG was recorded from the scalp with 32 active electrodes (actiCAP with actiCHamp amplifier and actiPOWER battery) using Brain Vision software. The sampling rate was set at 500 Hz and the electrode impedance was kept below 20 k Ω during recording. Horizontal eye movements were monitored with F7 and F8 electrodes and vertical eye movements were registered with two electrooculogram (EOG) electrodes placed above and below the left eye. The stimuli were presented pseudorandomly in an oddball paradigm via headphones while the children watched a silent movie to divert attention from the stimuli. The Prototype and Boundary blocks were counterbalanced between participants.

During the discrimination task, the children heard the stimuli via headphones and they were instructed to push a button every time they perceived a difference between the stimuli. The order of the Prototype and Boundary blocks were counterbalanced between participants.

During passive auditory training, the 33 vowel stimuli were presented in a randomized order using Sanako Lab 100 software. The training session lasted approximately one hour and it consisted of three training blocks with short breaks in between. The purpose of the training design was to simulate natural exposure to new speech sounds, i.e. a situation where the most frequently heard variant of a sound becomes the category prototype (e.g., Kuhl, 1991; Kuhl et al., 1992).

3.1.2.3 Analysis

Most of the participants in Study I did not complete the discrimination task. Therefore, the behavioral discrimination data was not analyzed further.

The EEG data was analyzed with BrainVision Analyzer (version 2.0) software. The first and second standard stimuli after each deviant were automatically excluded from the analysis by the software. The EEG data filtered with a 1–30 Hz bandpass filter and automatic artefact rejection ($\pm 100\,\mu$V). Eye movements were filtered from the data with ICA-correction. Separately averaged waveforms for standard and deviant stimuli from a 550 ms time window (including a -100 ms baseline correction) were computed for individual participants.

The EEG data from the four younger participants (aged 7–8 years) showed overall ERP positivity and distrurbances in the data, which were interpreted as signs of possible neural maturational differences between the younger and older subjects. Another possibility was that the younger children’s data was affected by tiredness and restlessness during recordings. Closer examination revealed that there was too much disturbance in the younger children’s data for further analysis. As there was a clear difference in the brain activities of the younger (7–8 years) and older (9–12 years) children, the participants were divided into two groups according to their age at this stage of analysis.

Next, grand averaged waveforms for the three older children were computed by subtracting the strandard responses from the deviant responses at each electrode site. The mean amplitudes were measured from 30 ms time windows (Prototype block: 225 ms–255 ms, Boundary block: 275 ms–305 ms) that were selected around the maximum peak amplitudes in the grand averaged waveforms.

The mean amplitudes from electrodes Cz and Fz from the three older (aged 9–12 years) participants were statistically analyzed using IBM SPSS (version 22) software. The analysis was conducted using the same statistical methods that have previously been used in similar MMN studies (Peltola, M. S. & Aaltonen, 2005; Taimi et al., 2014). The statistical analysis begun with normality tests. Next, the MMN mean amplitudes for the Prototype and Boundary stimulus blocks within the selected time windows at Cz and Fz electrode cites were subjected to one-tailed one sample t-tests. The one-tailed one sample t-tests were performed to see whether the MMN responses for the Prototype and Boundary blocks differed from zero before or after training.

3.2 Study II

3.2.1 Participants

The participants in Study II were 32 children (aged 11–13 years) from a Finnish elementary school. The participants were divided into two groups (Early learners and Controls) according to their English learning background.

The participants' language backgrounds were not strictly controlled because the aim of the study was to test two representative groups of Finnish schoolchildren from a regular Finnish school. However, all participants answered to a language background questionnaire prior to the experiment in order to exclude possible Finnish-English bilinguals from the data. In addition, all children and their parents gave a written informed consent before participating.

The Early learner group included 17 participants (aged 12;4-13;2, mean age 12;10, 10 females) from an English immersion class, where all school subjects were taught in British English. Most of the participants had attended an English immersion class all through elementary school for 6 years. All the children were Finnish monolinguals, but almost all of them had some contact with the English language outside school through friends or family. However, none of the children in the Early learner group had English-speaking parents or spoke English at home.

The Control group included 15 children (aged 11;9–12;7, mean age 12;1, 7 females) from a regular Finnish speaking class from the same school. The Control group had studied English as a separate subject for four years, since the third grade. The children had studied English for two to three hours per week, as dictated by the Finnish National Agency for Education's national core curriculum.

3.2.2 Methods

3.2.2.1 Stimuli

The auditory stimuli in Study II were the natural productions of a native British English (BrE) male speaker. The stimuli were 23 English words containing 12 BrE monophtong vowels in voiceless and voiced consonant contexts (see Table 3). Each word was produced by the native speaker seven times (161 tokens in total). All seven repetitions of each word were included in the experiment procedure in order to maintain natural variance in the vowel qualities. The stimuli were recorded during an earlier study (Peltola, M. S., Lintunen, & Tamminen, 2014).

Table 3. The stimulus words containing the twelve monophtong BrE vowels. The words were produced by a native BrE male speaker. The written words are listed in the table for clarity.

	voiceless	**voiced**
/iː/	/hiːt/ - heat	/hiːd/ - heed
/ɪ/	/hɪt/ - hit	/hɪd/ - hid
/e/	/bet/ - bet	/bed/ - bed
/æ/	/hæt/ - hat	/hæt/ - had
/ʊ/	/fʊt/ - foot	/hʊd/ - hood
/uː/	/huːt/ - hoot	/huːd/ - who'd
/ɔː/	/bɔːt/ - bought	/bɔːd/ - board
/ʌ/	/hʌt/ - hut	/hʌd/ - hud
/ɒ/	/tɒt/ - tot	/tɒd/ - todd
/ɑː/	/hɑːt/ - heart	/hɑːd/ - hard
/ɜː/	/hɜːt/ - hurt	/hɜːd/ - heard
/ə/		/hɑːdə/ - harder

3.2.2.2 Procedure

An intervention method instead of a training paradigm was used in the experiment in order to see how English immersion education might have affected children's pronunciation. The procedure was a simple listen-and-repeat task where the participants heard seven tokens of the 23 English stimulus words and repeated them on tape after the acoustic model. The experiment was conducted in the school library during school hours. The participants were instructed to listen carefully to the words and repeat what they heard. They received no feedback on their pronunciation during testing. The acoustic data was collected using a portable laboratory consisting of a laptop and a headset.

The stimuli were presented automatically in a pseudorandomized order with an ISI of 3 seconds. The pseudorandomization ensured that none of the words appeared twice consecutively during the task. The participants heard the acoustic stimuli without any written prompts to avoid any orthographical effects on their pronunciation. There were two self-paced breaks during testing and the experiment lasted approximately 10–12 minutes per participant.

3.2.2.3 Analysis

The children's productions were acoustically analyzed using Praat software version 5.3.01 (Boersma, 2001). The first (F1) and second (F2) formants were measured from the steady state phase of the vowels using the Linear Predictive Coding (LPC) Burg algorithm.[4] Individual average F1 and F2 values for each vowel were calculated from the seven repetitions of each word. The vowel /ə/ was excluded from statistical analysis due to creakiness and and poor acoustic quality caused by the vowel's word final position in the stimulus.

The average F1 and F2 values of the remaining 11 vowels in the 22 words with voiced and voiceless consonant contexts were subjected to statistical analysis using IBM SPSS Statistics software (version 22). A repeated mesures Analysis of Variance (ANOVA) was conducted on the average formant values from both groups. The statistical analysis proceeded systematically so that each step of analysis was justified and required by the significant findings from the previous stage of analysis. Appropriate analyses and tests were performed for all the significant main effects and interactions. The statistical analysis of the average formant values began with a repeated mesures ANOVA with the between-subject factor defined as Group (Early learner, Control) and the within-subject factors defined as Vowel (/iː/, /ɪ/, /e/, /æ/, /ʊ/, /uː/, /ɔː/, /ʌ/, /ɒ/, /ɑː/, /ɜː/), Context (Voiced, Voiceless) and Formant (F1, F2). Based on the initial findings, a Group (2) × Vowel (11) × Formant (2) repeated measures ANOVA was performed separately for the voiceless and voiced words to see how the children's production of the vowels developed in the two contexts. Finally, a Group (2) × Vowel (11) analysis was conducted separately for the F1 and

[4] In Studies II–IV, the acoustic analysis was performed with frequency settings appropriate for the higher F0 of child speakers. Because the participants were relatively young, no great differences in the F0 of female and male speakers were observed (which is typical for adult speakers). Therefore, the fact that there were more female participants in some of the studies should not affect the acoustic results largely.

F2 values from the voiceless and voiced words in order to discover how each formant developed in the two voicing contexts.

3.3 Study III

3.3.1 Participants

The participants in Study III were 23 monolingual Finnish children (aged 9–11 years) from regular and music-oriented fourth grades. The children were divided into two groups according to their musical experience. The participants who attended a music-oriented education program were assigned to the Music group and the children who attended a regular education program were assigned to the Non-music group. This allowed us to investigate how the musical experience and exposure received in a music-oriented education program might affect children’s ability to learn to produce non-native sounds through auditory training.

The Music group included 11 children (aged 9;10–10;9 years, mean age 10;4, 10 females) who attended a music-oriented fourth grade. At the time of testing, the children were on their second year of the music-oriented program.The children in the Music group had taken a musicality test before being admitted to the music-oriented program and they participated in daily musical activities in school.

The Non-music group included 12 children (aged 10;1–11;2, mean age 10;7, 10;6, 10 females) who attended a regular fourth grade. The children in the Non-music group had one compulsory music lesson per week.

All the participants reported having normal hearing. They completed a language and music background questionnaires prior to testing. In addition, the children and their parents gave an informed written consent before participating in the experiment.

3.3.2 Methods

3.3.2.1 Stimuli

The auditory stimuli in Study III were two semi-synthetic pseudo words /ty:ti/ and /tʉ:ti/. The Swedish close rounded vowel contrast /y/ - /ʉ/ was embedded in the first syllable of the stimulus words. The Semisynthetic Speech Generation method (SSG, (Alku, Tiitinen, & Näätänen, 1999) was used to create the stimuli based on the natural speech productions of a 24-year-old Finnish-Swedish bilingual male speaker. The F1 value for the non-native vowel /ʉ/ in the word /tʉ:ti/ was synthesized at 338 Hz and the F2 value was set at 1258 Hz. The F1 and F2 values for the native vowel

/y/ in the stimulus word /ty:ti/ were 269 Hz and 1866 Hz respectively. Therefore, the primary acoustic difference between the stimuli was the F2 value of the first syllable vowels /y/ and /ʉ/ (Figure 3).

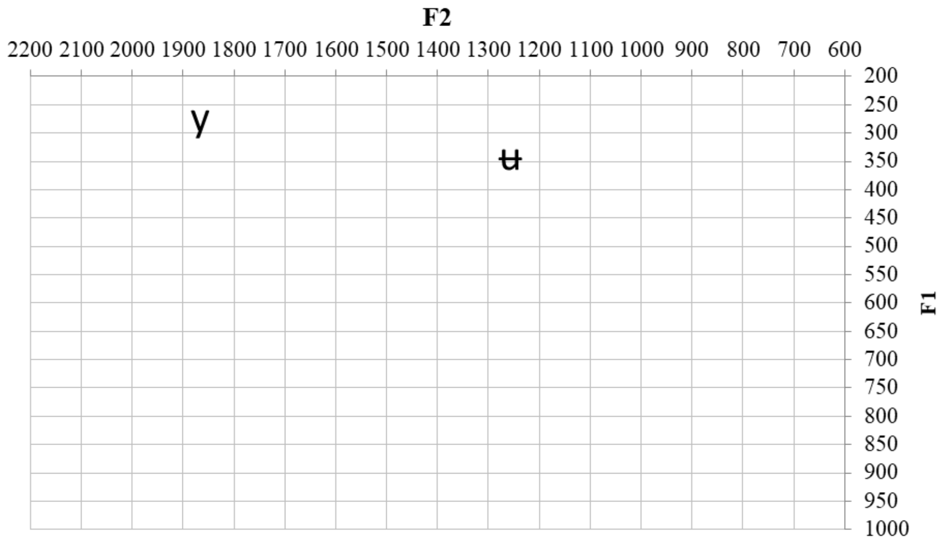

Figure 3. A vowel chart of the semisynthetic vowels /y/ and /ʉ/ used in the stimulus words. The Swedish contrast /y/ - / ʉ / is not phonemic in Finnish. The position of the symbols reflects the position of the tongue during articulation. The F1 (vertical axis) and F2 (horizontal axis) values of the vowels are based on the voice of an adult male L1 Finnish-Swedish speaker. The child participants were not expected to reach the exact formant values of the stimulus vowels because of the physiological differences in the vocal tracts of adult and child speakers.

3.3.2.2 Procedure

Study III used a two-day training paradigm with alternating listen-and-repeat production recordings and auditory training sessions. There were four training and four recording sessions in total on two consecutive days. The procedure lasted approximately 15 minutes per participant.

The experiment was conducted during school hours in a quiet room using a laptop and a headset with an external sound card. Sanako Study Recorder (version 8.22.0.0) software was used to present the auditory stimuli during training and recording sessions. The stimuli were presented automatically with an ISI 3 seconds so that every other stimulus was /ty:ti/ with the native vowel /y/ and every other stimulus was /tʉ:ti/ with the non-native vowel /ʉ/.

During training sessions, the participants heard both auditory stimuli 30 times without repeating them. The training sessions did not include any production because the aim of the study was to test the children's auditory sensitivity to non-native speech sounds and to see whether they are able to change their own production patterns by merely listening to an L2 vowel contrast.

During recording sessions, the participants listened and repeated both stimulus words ten times. The participants were instructed to listen carefully to the stimuli and repeat what they heard. They received no feedback on their productions during the experiment. The Sanako Study Recorder software was used to record the participants' productions. The experiment procedure is described in more detail in Table 4.

Table 4. The experiment procedure of Study III.

	DAY 1		**DAY 2**
LISTEN AND REPEAT	1st Recording session (baseline) 10 x /tʉ:ti/ 10 x /ty:ti/ →*Recorded*	*LISTEN*	3rd Training session 30 x /tʉ:ti/ 30 x /ty:ti/ →*Not recorded*
LISTEN	1st Training session 30 x /tʉ:ti/ 30 x /ty:ti/ →*Not recorded*	*LISTEN AND REPEAT*	3rd Recording session 10 x /tʉ:ti/ 10 x /ty:ti/ →*Recorded*
LISTEN AND REPEAT	2nd Recording session 10 x /tʉ:ti/ 10 x /ty:ti/ →*Recorded*	*LISTEN*	4th Training session 30 x /tʉ:ti/ 30 x /ty:ti/ →*Not recorded*
LISTEN	2nd Training session 30 x /tʉ:ti/ 30 x /ty:ti/ →*Not recorded*	*LISTEN AND REPEAT*	4th Recording session 10 x /tʉ:ti/ 10 x /ty:ti/ →*Recorded*

3.3.2.3 Analysis

Acoustic analysis of the production data obtained in the four recording sessions was carried out using Praat software (Boersma, 2001) version 6.0.43. As in Study II, the F1 and F2 values of the first syllable vowels were extracted from the steady state phase of the vowel using the same LPC Burg algorithm. The participants' individual average F1 and F2 values for /y/ and /ʉ/ from the ten productions within each recording session were then calculated.

The average F1 and F2 values for both vowels in all four recording sessions were subjected to statistical analysis using IBM SPSS Statistics software (version 25.0.0.1). The analysis proceeded systematically with appropriate analyses and tests based on the significant findings. The statistical analysis began with a repeated measures ANOVA performed for the average formant values with the between-subject factor defined as Group (Music, Non-music) and the within-subject factors defined as Session (first, second, third, fourth), Word (/ty:ti/, /tʉ:ti/) and Formant (F1, F2). The purpose of the initial ANOVA was to see whether the productions of the two groups differed in any way across recording sessions. Next, a Group (2) × Session (2) × Word (2) × Formant (2) repeated measures ANOVA was performed for one session pair at a time. The first recording session was compared separately to the second, third and fourth sessions to see how the children's productions developed across time compared to the baseline. Finally, paired samples t-tests for both vowels' F1 and F2 values in the three session pairs were performed to discover how the formant values developed compared to the baseline.

3.4 Study IV

3.4.1 Participants

The participants in Study IV were 16 monolingual Finnish preschoolers (aged 6–7;4, mean age 6;8, 13 females). The children answered to a language background questionnaire with the help of their parents prior to testing. None of the participants knew any other languages besides Finnish and all reported having normal hearning. All the participants and their parents gave written informed consent before the experiment.

3.4.2 Methods

3.4.2.1 Stimuli

The stimuli were the same semisynthetic pseudowords /ty:ti/ and /tʉ:ti/ that were used in Study III. For more details, see chapter 3.3.2.1.

3.4.2.2 Procedure

The experiment procedure in Study IV was a modification of the two-day training paradigm used in Study III. The main difference was that, in Study IV, the training paradigm was a listen-and-repeat task instead of a listening task. In other words, in Study IV, the participants listened and repeated the stimuli during training and recording sessions. The experiment included two trainings and two recordings per day on two consecutive days, which means that there were four trainings and four recording sessions in total. The procedure lasted approximately 15 minutes per day.

The experiment was conducted in a quiet room in a preschool using a portable laboratory consisting of a laptop and a headset with an external sound card. During training and recording sessions, the stimuli were presented automatically (ISI 3 seconds) in a fixed order so that every other word was /ty:ti/ and every other word was /tʉ:ti/. The Sanako Study Recorder software (version 8.22.0.0) was used to present the stimuli and to record the participants' productions. The participants were instructed to listen carefully to the stimuli and repeat them aloud during training and recording sessions. They received no feedback on their productions during testing. The procedure is described in more detail in Table 5.

Table 5. The experiment procedure in Study IV

	DAY 1		DAY 2
LISTEN AND REPEAT	1st Recording session (baseline) 10 x /tʉ:ti/ 10 x /ty:ti/ →*Recorded*	*LISTEN AND REPEAT*	3rd Training session 30 x /tʉ:ti/ 30 x /ty:ti/ →*Not recorded*
LISTEN AND REPEAT	1st Training session 30 x /tʉ:ti/ 30 x /ty:ti/ →*Not recorded*	*LISTEN AND REPEAT*	3rd Recording session 10 x /tʉ:ti/ 10 x /ty:ti/ →*Recorded*
LISTEN AND REPEAT	2nd Recording session 10 x /tʉ:ti/ 10 x /ty:ti/ →*Recorded*	*LISTEN AND REPEAT*	4th Training session 30 x /tʉ:ti/ 30 x /ty:ti/ →*Not recorded*
LISTEN AND REPEAT	2nd Training session 30 x /tʉ:ti/ 30 x /ty:ti/ →*Not recorded*	*LISTEN AND REPEAT*	4th Recording session 10 x /tʉ:ti/ 10 x /ty:ti/ →*Recorded*

3.4.2.3 Analysis

The participants' productions from the four recording sessions were acoustically alalyzed using Praat software version 6.0.43 (Boersma, 2001). As in Studies II and III, the F1 and F2 values of the first syllable vowels were measured from the steady state phase of the vowels using the LPC Burg Algorithm. After the formant values for /y/ and /ʉ/ were extracted from all the productions, the individual average formant values for both vowels from the ten repetitinon within each session were calculated.

The average F1 and F2 values of /y/ and /ʉ/ were subjected to statistical analysis using IBM SPSS Statistics software (version 25.0.0.1). The statistical analysis was performed systematically so that each stage of analysis was followed by appropriate tests that were justified and required by any significant findings. First, a repeated measures ANOVA with the factors defined as Session (first, second, third, fourth), Word (/ty:ti/, /tʉ:ti/) and Formant (F1, F2) was performed on the average formant data. The purpose of the initial ANOVA was to see whether the participants' production changed significantly in any direction across time. In other words, the participants served as their own controls, since the average formant values of the vowels were compared between the four recording sessions to see whether there was a change as a function of training compared to the baseline recording. In the next stage of analysis, the words /ty:ti/ and /tʉ:ti/ were subjected separately to a Session (4) × Formant (2) repeated measures ANOVA to see whether the children's production of the two words developed differently across sessions. The sessions were then examined in pairs by subjecting them to a Session (2) × Word (2) × Formant (2) repeated measures ANOVA. The first recording (basline) was compared in turns to the second, third and fourth sessions to see whether there was a change in production in any of the sessions as a function of training compared to the baseline recording. Next, both words were investigated separately within each session pair with a Session (2) × Formant (2) repeated measures ANOVA. Finally, paired samples t-tests were performed for the F1 and F2 values of /ʉ/ in the target word /tʉ:ti/ in all three session pairs to see how the two formants developed in the second, third and fourth sessions compared to the baseline.

3.5 Summary of the materials and methods used in the studies

The thesis is comprised of four studies that used different phonetic training paradigms and production and perception measures to examine childen's L2 sound learning from different perspectives. Study I examined L2 sound perception using behavioral discrimination tasks and EEG measurements focusing on the MMN response. Studies II, III and IV focused on different aspects of L2 pronunciation by measuring production accuracy with acoustic analysis of speech recordings.

The stimuli used in the studies varied from synthetic vowels (Study I) to semisynthetic pseudowords (studies III and IV) and natural speech tokens (Study II). In addition, the experiment methods included different passive auditory trainings (studies I and III) as well as listen-and-repeat tasks (studies II and IV). Combining these varying stimuli, methods and training paradigms allowed us to obtain precise information on how age, L2 immersion education and music-oriented education affect children's L2 sound perception and production learning.

All the studies were conducted with permission from the Ethics Committee of the University of Turku. The participants in all four studies were volunteers and none of the children or their parents received any compensation for participating.

4 Results

This chapter presents the results of Studies I–IV and views them in the light of the original hypotheses. The results of each study are first presented separately in sections 4.1–4.4 and then summarized in section 4.5.

4.1 Results of Study I

The results of Study I showed that the passive auditory training resulted in changes in the older (9–12-year-old) participants' preattentive perception of the trained L2 category prototype. There were no changes in the perception of the L2 category boundary. In other words, an MMN response was elicited in the older children after training only for the most trained variant of the L2 category, but not for the less trained variants near the category boundary. The EEG data from the younger (7–8-year-old) children showed a very different positive MMN pattern from the older children, and the data included too much disturbance for further analysis. No results were obtained from the behavioral discrimination tests since most of the children were not able to complete the task according to the instructions. In other words, the discrimination task was too challenging for these particular age groups.

Our hypothesis that the children's neural plasticity would enable the formation of a new memory trace for an L2 vowel through passive auditory training was only partly confirmed, as only the older children showed signs of enhanced MMN responses to the trained L2 prototype after training. The second hypothesis was that the formation of a new memory trace would cause discrimination to become more difficult near the trained prototype and easier near the category boundary. This hypothesis, which was based on the theoretical framework of the NLM (Kuhl et al., 1992), was not confirmed. Instead, it seemed that the familiarity of the trained sounds affected the results. Training effects emerged only for the /ʉ/ prototype, in other words the variant of the /ʉ/ category that was presented the most during training. Furthermore, we hypothesized that training effects would be reflected pre-attentively in the mismatch negativity (MMN) responses and behaviorally in the discrimination sensitivity and reaction times. This hypothesis was not confirmed nor rejected, as no results were obtained from the discrimination task.

The statistical analysis of the three older participants' (aged 9–12 years) MMN mean amplitudes revealed that the MMN response elicited by the Prototype condition did not differ significantly from zero before training in the Fz and Cz electrode locations within the selected 225 ms –255 ms time window. After training, the same contrast elicited an MMN response that differed significantly from zero in both electrode locations (see Figure 4). The older children's MMN responses for the Boundary trial did not differ significantly from zero before or after training.

The results of Study I showed that 9–12-year-old children learn to perceive a non-native vowel through passive auditory training. The unimodal distribution of the training stimuli resulted in changes only in the most frequently presented variant of the trained L2 category, which indicated that the children learned to perceive the allophone of the L2 sound that was the most familiar to them after training.

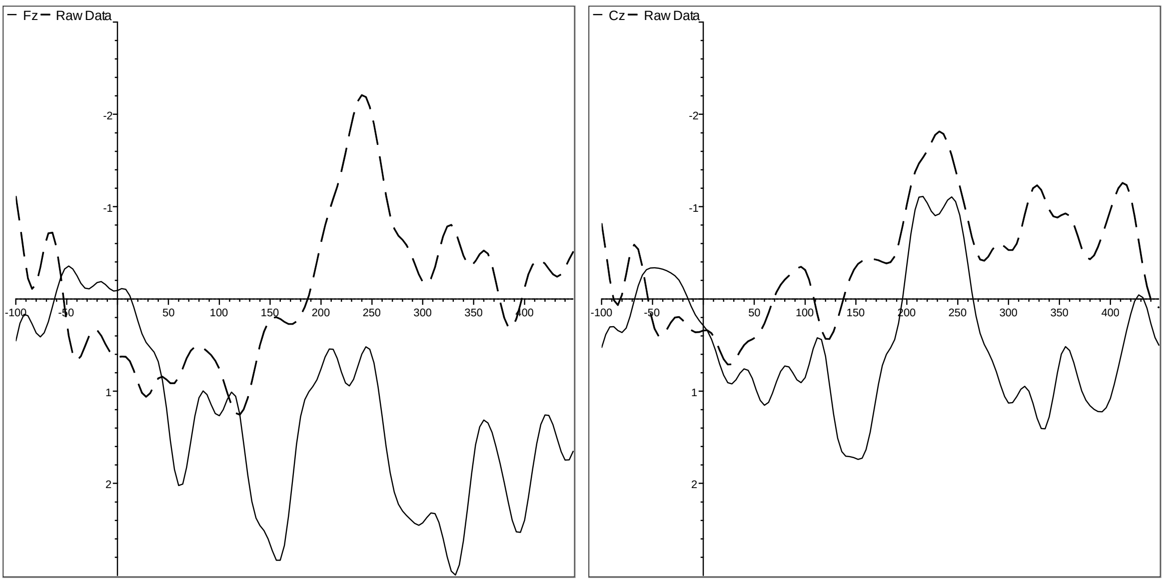

Figure 4. The grand averaged difference waveforms for the Prototype contrast before (solid line) and after (dashed line) training at the Fz and Cz electrode locations from the older (aged 9-12 years) children.

4.2 Results of Study II

The results of Study II revealed that the two groups (Early learners and Controls) tested in the study produced the eleven BrE monopthong vowels differently. In other words, the results of Study II indicated that daily L2 exposure received in an English immersion education program in elementary school affects 11–13-year-old Finnish children's production of BrE vowels. Furthermore, the results showed that the Early learners from the English immersion program tended to produce the BrE vowels more accurately, closer to the native model. Therefore, our hypothesis that studying in an English immersion education program would result in more accurate production of BrE vowels than studying English as a separate subject in a regular Finnish speaking class was confirmed.

The statistical analysis revealed that there was a significant difference between the productions of the Early learner and Control groups. Analysis of the vowel formant values (F1 and F2) revealed that the vowel formant values produced by the two groups were different. The difference in production was located in the vowels' F2 values. Further analyses revealed that the groups produced the BrE vowels with different F2 values in the voiced context words. There was no significant difference between the groups' productions in the voiceless context words.

Closer exmamination of the vowel formant values produced by the two groups revealed that the Early learners tended to produce the vowels with lower F2 values in the voiced context (see Table 6). This means that their productions were closer to the stimulus values produced by the native speaker. In addition, comparing the F1 and F2 values of the children's productions (Table 6) showed that the differences between the groups seemed to be most pronounced for the the BrE vowels /ɪ/, /ɒ/, /ɔ/ and /ɜ/, which were easily falsely assimilable to Finnish vowel categories. These vowels are not phonological in Finnish, but are perceptually very similar to Finnish vowels /i/, /ɑ/, /o/ and /ø/. In addition, the difference between the groups seemed to be most clearly reflected in the BrE lax vowel /ɪ/. The Control group produced the vowel /ɪ/ with considerably higher F2 values than the Early learner group (Table 6) and they seemed to assimilate /ɪ/ to the Finnish /i/. These findings were in keeping with the predictions of second language learning theories such as the SLM (Flege, 1995) and the PAM (Best, 1994, 1995). In other words, the formant values listed in Table 6 seemed to support the proposition that L2 sounds that are similar to L1 sound categories cause the most difficulties for L2 learners.

The results of Study II showed that the L2 experience of the children from the English immersion class was reflected in their English pronunciation. Their production of BrE vowels differed significantly from the vowels produced by the Control group who were enrolled in a Finnish speaking class. Both the age of acquisition and the manner of learning were concluded to explain the difference between the groups' pronunciation.

Table 6. The average formant values in the 11 BrE vowels in voiceless and voiced consonant contexts produced by the Early learner and Control groups. The stimulus formant values produced by a native BrE speaker are offered as reference.

VOWEL	CONTEXT	FORMANT	**NATIVE**	**EARLY**	**CONTROL**
/iː/	/hiːt/	F1	269	398	428
		F2	2334	2773	2831
	/hiːd/	F1	269	397	418
		F2	2311	2749	2827
/ɪ/	/hɪt/	F1	378	457	454
		F2	2160	2572	2706
	/hɪd/	F1	333	443	449
		F2	2204	2587	2712
/e/	/bet/	F1	613	621	632
		F2	1916	2134	2230
	/bed/	F1	502	623	624
		F2	1964	2171	2228
/æ/	/hæt/	F1	950	897	915
		F2	1577	1826	1908
	/hæd/	F1	884	872	880
		F2	1511	1833	1890
/ʊ/	/fʊt/	F1	411	478	468
		F2	857	1299	1262
	/hʊd/	F1	351	454	449
		F2	967	1180	1220
/uː/	/huːt/	F1	280	432	444
		F2	1398	1415	1404
	/huːd/	F1	272	422	428
		F2	1256	1358	1272
/ɔː/	/bɔːt/	F1	403	515	548
		F2	678	1005	1087
	/bɔːd/	F1	381	504	552
		F2	604	967	1029
/ʌ/	/hʌt/	F1	721	766	789
		F2	1079	1383	1433
	/hʌd/	F1	728	724	766
		F2	1114	1459	1483
/ɒ/	/tɒt/	F1	590	584	623
		F2	1134	1188	1285
	/tɒd:/	F1	507	579	614
		F2	856	1194	1271
/ɑː/	/hɑːt/	F1	651	720	751
		F2	985	1229	1259
	/hɑːd/	F1	632	694	720
		F2	968	1214	1230
/ɜː/	/hɜːt/	F1	491	605	655
		F2	1415	1823	1785
	/hɜːd/	F1	461	580	620
		F2	1480	1824	1774

4.3 Results of Study III

The results of Study III indicated that attending a music-oriented education program in elementary school does not lead to additional sensitivity to acoustic variation that would be transferred to the trainability of non-native sound contrasts. However, auditory training of an L2 vowel contrast does lead to changes in 9–11-year-old children's L2 production regardless of their musical background. Our hypothesis was that children from a music-oriented education program would learn to produce the non-native vowel contrast through auditory training faster than children from a regular Finnish speaking fourth grade. This hypothesis was not confirmed.

The results of the statistical analysis revealed that the Music and Non-music groups responded similarly to the auditory training. There were no significant differences between the groups across sessions. Both groups changed their production of the non-native vowel /ʉ/ after the first training, in other words by the second recording session (Figure 6). The change in production was later further reflected in the lowering of the F2 value in the vowel /ʉ/ in the third recording session (Figure 7), after three trainings. This implies that the children changed their pronunciation of the non-native sound already in the second recording session, but they still needed more training to produce the vowel contrast consistently in their own speech. The changes in production remained throughout the experiment. There were no significant changes in the production of the native vowel /y/.

The average formant values for /y/ and /ʉ/ produced by the Music and Non-music group are shown in Table 7 and Figure 5. The development of the average F2 values for /y/ and /ʉ/ across sessions are depicted in Figures 6. The primary acoustic difference between the frontal vowel /y/ and the central vowel /ʉ/ is the F2 value, which is higher for /y/ than for /ʉ/. The F2 value reflects tongue backness during articulation, and since /ʉ/ is articulated with the tongue further back in the oral cavity, the F2 value is lower than for /y/. Therefore, the finding that both groups started to produce the non-native vowel /ʉ/ with lower F2 values indicated that the children were able to perceive the acoustic difference between the stimuli and then adapt their own production accordingly.

Some aspects of musical experience and aptitude have been found to overlap with L2 phonetic learning in adult learners (Bhatara et al., 2015; Boll-Avetisyan et al., 2016; Ghaffarvand Mokari & Werner, 2018; Marie et al., 2011). However, in Study III we did not measure musical aptitude or musicality. Instead, we were interested in whether participating in music-oriented education in elementary school would affect L2 sound learning. Our results indicated that studying in a music-oriented class in elementary school did not significantly improve the 9–11-year-old children's L2 production learning in this experiment. The results could intdicate that the 9–11-year-old children tested in Study III were at a developmental stage where the benefits of linguistic plasticity and age outweighed the possible benefits of musical experience on L2 sound learning.

Table 7. The average F1 and F2 values (Hz) for the non-native vowel /ʉ/ and the native vowel /y/ in each recording session. Both stimulus words containing the vowels were repeated ten times in each recording session by all speakers. The overall standard deviations are reported in parentheses.

VOWEL	GROUP	FORMANT	**SESSION 1**	**SESSION 2**	**SESSION 3**	**SESSION 4**
/ʉ/	Non-music	F1	454 *(34)*	461 *(41)*	481 *(28)*	474 *(36)*
		F2	1814 *(540)*	1783 *(523)*	1701 *(509)*	1685 *(508)*
	Music	F1	445 *(31)*	449 *(28)*	456 *(58)*	439 *(35)*
		F2	1718 *(383)*	1624 *(362)*	1502 *(403)*	1493 *(416)*
	Both groups	F1	450 *(32)*	455 *(35)*	469 *(46)*	457 *(39)*
		F2	1768 *(463)*	1706 *(450)*	1606 *(462)*	1593 *(466)*
/y/	Non-music	F1	460 *(33)*	458 *(35)*	474 *(30)*	470 *(30)*
		F2	2148 *(176)*	2104 *(188)*	2103 *(162)*	2127 *(210)*
	Music	F1	438 *(32)*	441 *(29)*	438 *(45)*	437 *(51)*
		F2	2232 *(158)*	2215 *(144)*	2135 *(242)*	2149 *(230)*
	Both groups	F1	449 *(34)*	450 *(33)*	457 *(41)*	454 *(44)*
		F2	2188 *(169)*	2157 *(174)*	2118 *(200)*	2137 *(215)*

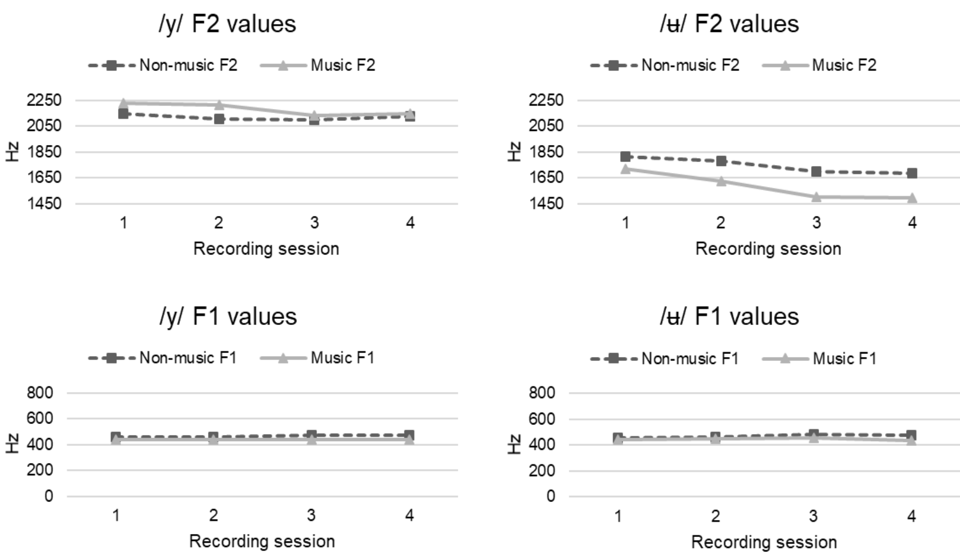

Figure 5. The average F1 and F2 values for /y/ and /ʉ/ across sessions.

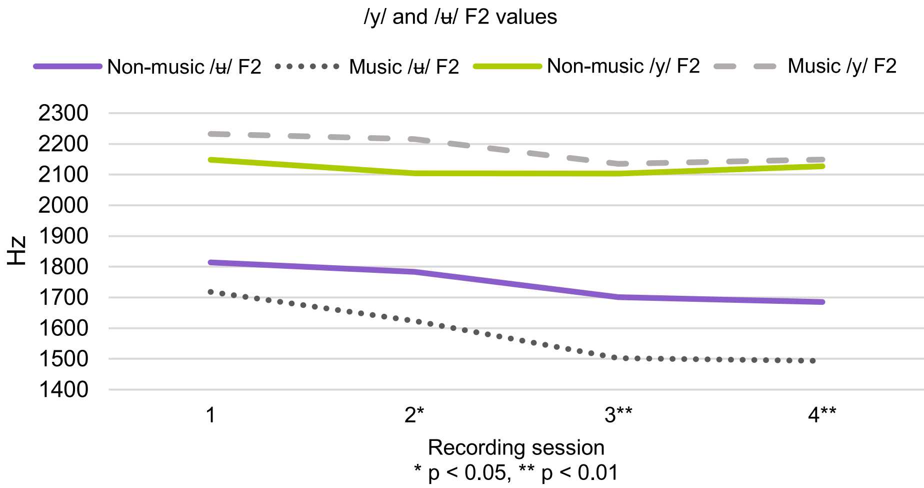

Figure 6. The average F2 values for /y/ and /ʉ/ produced by both groups across sessions. A Group (2) × Session (4) × Word (2) × Formant (2) repeated measures ANOVA did not reveal any group differences, but comparison of the three session pairs revealed a main effect of Session in the second, third and fourth sessions when compared to the baseline. Significant between-session changes are marked with asterisks (* p < 0.05, * p < 0.01).

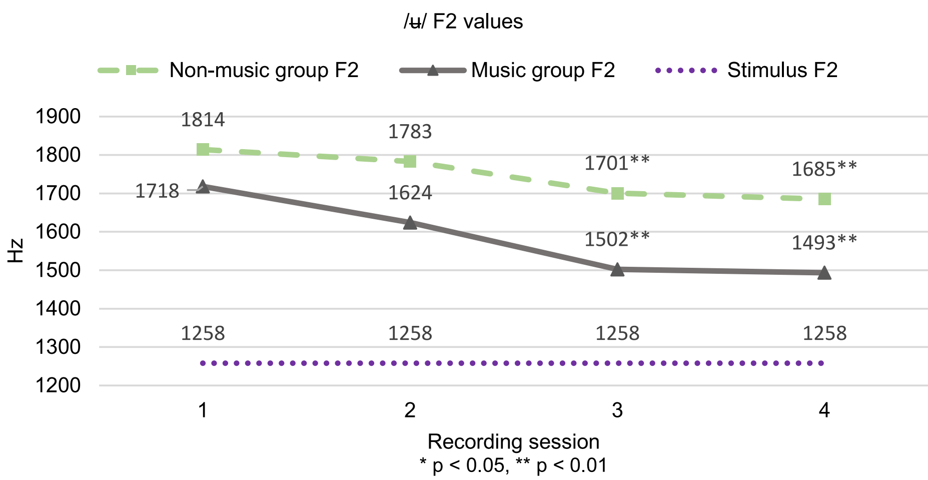

Figure 7. The average F2 values (Hz) for /ʉ/ produced by the Music and Non-music group across recording sessions.The dashed line indicates the F2 value in the acoustic stimulus /tʉ:ti/. Paired samples t-tests revealed that the /ʉ/ F2 values lowered significantly by the third session compared to the baseline. Significant between-session changes are marked with asterisks. No significant changes emerged for the vowel /y/ or the F1 values in /ʉ/.

4.4 Results of Study IV

The results of Study IV revealed that 6–7-year-old children learned to produce the difficult non-native vowel contrast /y/ - /ʉ/ through listen-and-repeat training. The statistical analyses revealed that the participants changed their production of the non-native vowel /ʉ/ already after the first training, in other words by the second recording session, and the change in production remained throughout the experiment (Table 8 and Figure 8). The training effects were immediate and persistent. Furthermore, the change in production was situated in the F2 value of /ʉ/, which lowered significantly towards the acoustic model after the first training session (Figure 9). There were no significant changes in the production of the native vowel /y/. The results confirmed our hypothesis that 6–7-year-old children's production would change as a function of listen-and-repeat training and that the direction of the change would be towards the acoustic target stimulus.

The fact that the participants changed their production of the non-native vowel towards the acoustic model immidiately after the first training was surprising, since we hypothesized that the children might be too young to concentrate adequately on the listen-and-repeat task. The results of Study IV indicate that the benefits of age at the developmental stage of 6–7-year-old children allows them to learn a difficult non-native vowel contrast through listen-and-repeat training even faster than we expected. The fact that the children started to produce /ʉ/ with lower F2 values after the first training shows that they were able to distinguish the relevant acoustic information from the stimuli and produce the same acoustic difference in their own speech. In addition, the fact that there were no changes in the children's production of the native vowel /y/ shows that they were able to perceive and produce the two vowels as separate sounds. Table 8 and Figure 8 depict the development of the average formant values in the children's productions across all four recording sessions.

The results of Study IV were compared to earlier studies that investigated the effets of a similar listen-and-repeat training method on L2 sound production by 7–10-year-old children (Taimi et al., 2014) and adults (Peltola, K. U. et al., 2017, 2020). The results by Taimi et al. (2014) showed that 7–10-year-old children changed their production of an L2 vowel after three training sessions. The results by Peltola et al., on the other hand, showed that one day of training did not lead to changes in adults' production of an L2 vowel (2017), but that two days of training did result in slight improvements in adult speakers' production accuracy (2020). Comparing the results of Study IV to these earlier findings demonstrates that 6–7-year-old children change their L2 sound production faster through listen-and-repeat training than 7–10-year-old children or adults.

Table 8. The average formant values for /y/ and /ʉ/ produced by the participants across recording sessions. Each stimulus word was repeated ten times by each participant during a recording session. The overall starndard deviations are reported in parentheses.

VOWEL	FORMANT	**SESSION 1**	**SESSION 2**	**SESSION 3**	**SESSION 4**
/y/	F1	485 *(42)*	488 *(48)*	473 *(60)*	481 *(59)*
	F2	2280 *(259)*	2201 *(256)*	2235 *(232)*	2240 *(219)*
/ʉ/	F1	486 *(42)*	489 *(46)*	480 *(59)*	484 *(60)*
	F2	2258 *(280)*	2077 *(342)*	2065 *(298)*	2041 *(312)*

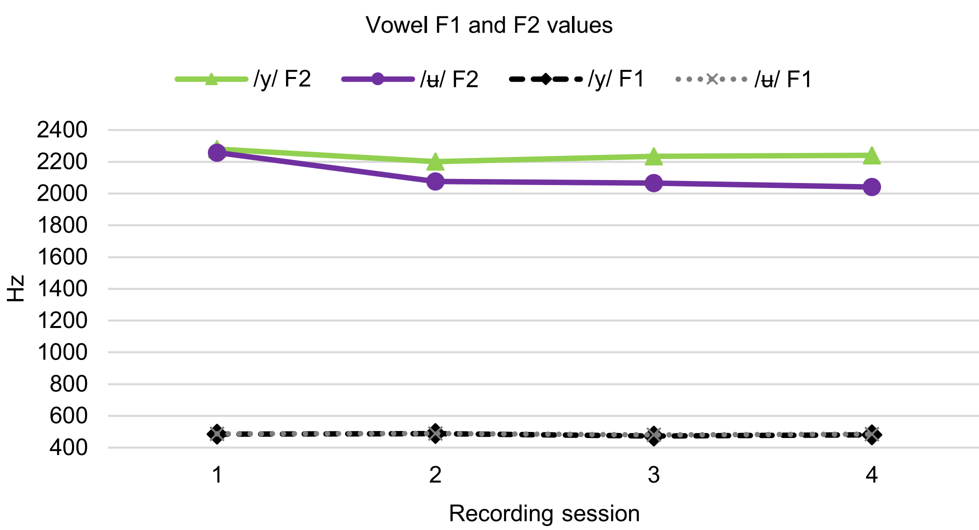

Figure 8. The average F1 and F2 values for /y/ and /ʉ/ produced by the participants across sessions.

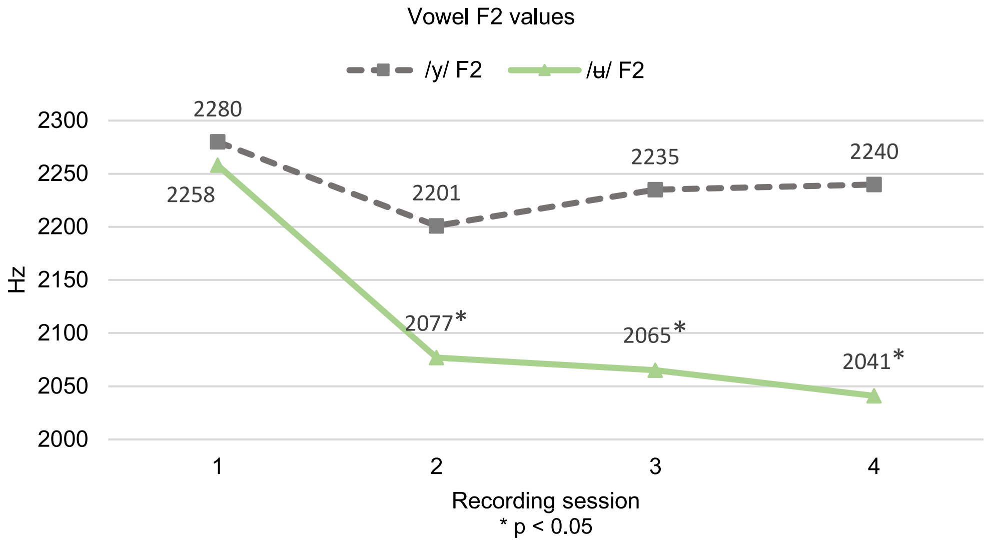

Figure 9. The average F2 values of the first syllable vowels produced by the participants across recording sessions. Separate Session (2) × Formant (2) repeated measures ANOVAs for the two words revealed the main effect of Session for the vowel /ʉ/ in all three session pairs when sessions 2, 3 and 4 were compared to the baseline. Significant between-session are marked with an asterisk (* $p < 0.05$). Subsuquent paired samples t-tests revealed significant lowering of the /ʉ/ F2 values in the same session pairs. No significant findings emerged for /y/ or the F1 of /ʉ/.

4.5 Summary of the results

The results of Study I indicated that passive auditory training can enhance 9–12-year-old children's, but not 7–8-year-old children's, L2 sound prototype perception when measured preattentively with MMN. The results of Study II showed that L2 immersion education in elementary school affects 11–13-year-old children's production of L2 vowels. The results of Study III, on the other hand, showed that studying in a music-oriented education program does not enhance 9–11-year-old children's ability to learn L2 vowel production through auditory training. Finally, the results of Study IV indicated that 6–7-year-old preschoolers learn to produce an L2 vowel after just one session of phonetic listen-and-repeat training.

The first overall hypothesis concerning the effects of age on children's L2 learning (Studies I and IV) was that the children would benefit from their young age and enhanced neural plasticity in phonetic training of L2 sound perception and production. This hypothesis was confirmed in the case of L2 production (Study IV), but only partly confirmed in the case of L2 perception (Study I). The second larger hypothesis was that children’s linguistic or musical learning background would affect L2 production learning (Studies II and III). This hypothesis was also only partly confirmed, since studying in a language immersion program was found to affect L2 sound production (Study II), but studying in a music-oriented education program did not have an effect (Study III).

Based on these findings, age was found to be an important factor affecting children’s L2 sound learning. The 6–7-year-old children changed their production as a function of listen-and-repeat training even faster than we had expected (Study IV). In addition, the effect of age was found to be even stronger than the effect of musical background in 11–13-year-old children, since training effects were immediate regardless of the participants’ musical background (Study III). However, young age did not automatically quarantee better results in L2 sound perception learning, since language awareness might play a role in the formation of new memory traces for L2 sounds (Study I). The positive effects of language immersion education on L2 sound production found in Study II could also be explained by earlier age of learning. Even though we primarily focused on language immersion education as a linguistic background factor in Study II, the better pronunciation accuracy of the immersion students is most probably at least partly explained by the fact that they had begun their L2 studies at a younger age. The effects of age seem to explain the results of Study II better than the type of L2 input received in immersion education, since our other results showed that 6–7 and 9–11-year-old children learned to produce a difficult L2 sound after only a few minutes of semisynthetic L2 input (Studies III and IV).

Therefore, overall, the results of Studies I–IV indicate age to be the strongest factor affecting school-aged children’s L2 sound learning. Our findings show that 6–13-year-old children can learn L2 sounds efficiently both through experimental training paradigms and early classroom input. The theoretical and practical implications of these findings and the reliability of the results are discussed in the next chapter.

5 Discussion

This chapter discusses different aspects of the results presented in Chapter 4. The theoretical implications of each study (I–IV) are discussed separately and overall in section 5.1. The overall practical implications of the studies are then discussed in section 5.2. The reliability of the results and the possible directions for future research are finally discussed in sections 5.3 and 5.4.

5.1 Theoretical implications

The majority of earlier research on children's L2 phonetic learning has focused on comparing child and adult learners in naturalistic L2 acquisition settings. Early AOL has been found to be one significant predictor of succesfull L2 production and perception learning after immigration to a new L2 environment (Baigorri et al., 2019; Flege et al., 1999b; Oh et al., 2011; Tsukada et al., 2005). Positive effects of early AOL on L2 category perception learning have also been found in some phonetic training studies (e.g., Giannakopoulou et al., 2013) while others show contradictory results (Heeren & Schouten, 2010, 2008). Research on instructed L2 learning remains scarce, even though the amount and quality of L2 input is considerably different in instructed learning situations compared to naturalistic L2 environments (Flege & Bohn, 2021; Tyler, 2019). Some results suggest that early AOL in instructed classroom settings or language immersion education can lead to accurate L2 sound production (Darcy & Krüger, 2012) but not perception (e.g., Darcy & Krüger, 2012; Kopečková et al., 2019; Peltola, M. S. et al., 2007). The effects of various musical factors on L2 phonetic learning, on the other hand, have been studied more extensively (e.g., Bhatara et al., 2015; Boll-Avetisyan et al., 2016; Delogu et al., 2006; Marie et al., 2011; Milovanov et al., 2008), but these previous studies have focused almost solely on adult L2 learners. Therefore, there is a need for wider research into different factors affecting children's L2 phonetic learning in order to gain better understanding of the mechanisms underlying L2 sound perception and production learning in children.

The aim of this thesis was to respond partly to this need by examining how children of different ages with different learning backgrounds learn to perceive and produce L2 sounds in instructed learning environments and phonetic training conditions. Studies I and IV investigated how 7–12 and 6–7-year-old monolingual children learn L2 sound perception and production when they are trained with auditory and listen-and-repeat training paradigms using synthetic or semisynthetic phonetic stimuli. The aim was to test whether the proposition of early AOL leading to enhanced learning results is true in laboratory and classroom-like training conditions. Study II focused on the effects of language-immersion education on L2 sound production in 11–13-year-old children, and Study III examined the effects of music-oriented education on 9–11-year-old children's L2 sound production. The aim was to discover wheter different learning backgrounds (i.e. L2 immersion education and music-oriented education) would cause differences in L2 pronunciation and production learning between children of the same age. The theoretical implications of the results of each study are first discussed individually and then, most importantly, as a whole in relation to earlier research and theoretical frameworks on children's L2 sound learning.

5.1.1 Study I

The results of Study I did not confirm our hypotheses regarding the formation of a new category prototype and the perceptual magnet effect (Kuhl, 1991; Kuhl et al., 1992) for an L2 sound category through auditory training (see section 4.1). However, the hypothesis that the auditory training would lead to the formation of a new memory trace for the L2 vowel was true for the older children (aged 9–12). They showed a familiarity effect for the most trained variant of the L2 vowel category, but no training effects were found in the younger (aged 7–8) children.

Although no results were obtained from the behavioral discrimination tests, and our hypotheses were only partly confirmed, the preattentive MMN results of Study I may offer some insight into children's L2 sound perception learning. The fact that training effects were only found in the 9–12-year-old children may indicate that the training procedure was too abstract and tiring for the 7–8-year-olds, causing fatigue and thus leading to disturbance in the EEG data. Another factor to be considered is the level of language awareness of the participants and the nature of the training protocol. As the training was designed to mimic the statistical distribution of sounds in naturalistic language environments during infancy (Maye et al., 2002, 2008), no explicit instructions or context were provided before testing and individual synthetic vowels were selected as stimuli to ensure controllability in the MMN protocol. It could be that the passive training protocol together with the lack of explicit instructions or context caused the 7–8-year-olds to be unaware that they were actually listening to speech sounds when they heard the stimuli. The MMN is a preattentive response that is evoked by changes in auditory stimuli and it should not be affected by awareness or attentive cognitive processes (Näätänen, 2000, 2001).

However, some studies have shown that language context and explicit knowledge on the language being tested can affect MMN elicitation for speech sound contrasts in advanced L2 speakers (Peltola, M. S. & Aaltonen, 2005; Peltola, M. S. et al., 2012). The older children (aged 9–12) had already started their first L2 lessons in school and may have benefitted from more matured and advanced language awareness in comparison to the younger (aged 7–8) participants, who had no L2 experience. The enhanced MMN for the P-NP contrast in the older children could therefore be explained by their level of language awareness and a familiarity effect caused by the unimodal distribution of the auditory training stimuli. This means that because the L2 prototype was presented the most during training, the contrast between the prototype and non-prototype became easier to perceive than the other stimuli.

5.1.2 Study II

The results of Study II showed that L2 immersion education affects 11–13-year-old children's L2 vowel production. The children from the English immersion education program (Early learners) produced the eleven BrE monophthong vowels differently than the children who had studied English as a separate school subject (Controls). This finding confirmed our hypothesis that the daily exposure to BrE in the immersion classroom could be reflected in the Early learners' L2 vowel production.

The statistical analysis revealed that the significant difference between the Early learner and Control groups' productions was the use of the second formant values in the production of the BrE vowel qualities. The formant values presented in the results (Table 6) show that the groups' production of the vowel F2 values differed especially in the lax vowel /ɪ/, which is close to the Finnish /i/ category. The vowel /ɪ/ was expected to cause challenges in perception and production according to the SLM (Flege, 1995; Flege & Bohn, 2021) and the PAM (Best, 1994, 1995; Best & Tyler, 2007). As can be seen in Table 6, the Control group produced the vowel /ɪ/ in the words *hit* and *hid* and the vowel /i:/ in the words *heat* and *heed* with similar F2 values. This finding supports findings from earlier studies showing that L1 Finnish speakers tend to rely on duration cues rather than spectral cues when discriminating the /i:/ - /ɪ/ contrast (Ylinen et al., 2010). Interestingly, the difference in the F2 values produced by the two groups was only significant in the voiced context words. This could be explained by the phonotactic rules of Finnish, which state that only vowels or the consonants /t/, /n/, /s/, /l/ or /r/ can appear in word final position (Suomi, Toivanen, & Ylitalo, 2008) and therefore the production of the voiced stop /d/ would probably have required additional articulatory effort from the Control group with less BrE experience. This may have drawn attention away from the vowel quality in the voiced context words. Another possible explanation could be that the pre-lenis lengthening phenomenon of BrE resulted in a longer steady state phase in the children's vowel formant values, allowing group differences to be more accurately reflected in the acoustic measurements. However, no duration data was extracted from the productions during acoustic analysis, so no definite conclusions on the possible effects of vowel duration can be drawn.

The two groups' learning backgrounds differed in terms of both AOL and the manner of learning (L2 immersion vs. instructed classroom learning). The Early learners had an average of two to three years more L2 experience than the Controls, since most of them had entered the immersion program already in the first grade and the Control group had started their English studies in the third grade. The difference in the groups' productions is therefore likely to be at least partly caused by the earlier AOL and the longer L2 exposure of the Early learners. However, the quality of L2 phonetic input needs to be considered when discussing the theoretical implications of the results. The Early learners learned English in an immersion classroom setting, where the L2 is used in all communication and teaching situations in the classroom and the teacher used BrE pronunciation consistently in their speech. The Control group studied English for 2–3 hours a week as a separate school subject in an instructed setting. Therefore, the quality, as well as quantity, of English input received in school was very different for the two groups in these two learning settings. As the quality of L2 input has been found to be one of the factors affecting L2 sound learning (Flege & Bohn, 2021), the quality of spoken input in the classroom could also partly explain the results of Study II.

5.1.3 Study III

The results of Study III showed that both groups of 9–11-year-old children benefitted from auditory phonetic training of an L2 sound contrast and changed their production towards the acoustic model. The finding that musical experience (i.e., studuing in a music-oriented education program) did not significantly affect training results was surprising, because it was not in line with the initial hypothesis. Our hypothesis was that the Music group's musical background could be reflected in enhanced auditory sensitivity to spectral differences in vowels, as previous research has suggested significant connections between different aspects of musicality or musical experience and L2 processing (e.g., Bhatara et al., 2015; Marie et al., 2011; Milovanov et al., 2008). The findings of Study III seemed to be contradictory with these earlier studies on the interplay between music and language.

One possible explanation could be that the 9–11-year-old participants were simply at a linguistically sensitive age, as suggested by the CPH (Lenneberg, 1967), and therefore the benefits of age (i.e. pre-puberty) and developmental linguistic sensitivity outweighed the possible effects of musical experience on L2 production learning. However, the CPH's proposition that puberty marks an abrupt decrease in linguistic sensitivity has been widely challenged, and more recent research suggests that there is not an abrupt decline in L2 learning abilities after a certain age (i.e. (Wang & Kuhl, 2003). For example, the SLM-r (Flege & Bohn, 2021) has replaced the CPH and the original age hypothesis of the SLM (Flege, 1995) with the hypothesis that L2 sound learning is affected by L1 category precision.

According to the SLM-r, L1 categories tend to be more precise (with less F1-F2 variability) in childhood and less precise in adulthood, as L1 categories become more established. The precision hypothesis proposes that high L1 category precision is often related to better discrimination of L1-L2 phonetic differences (Flege & Bohn, 2021), because there is more perceptual distance between the precise L1 categories in the acoustic phonetic space. The precision hypothesis could explain the findings of Study III. The fact that the two groups responded similarly to training and no significant group-differences emerged indicates that both groups were able to produce the subtle spectral difference in the L1 irrelevant contrast /y/ - /ʉ/ after listening to the acoustic stimuli. If the 9–11-year-old children in Study III had precise L1 categories with relatively large between-category distances in the acoustic phonetic space, their auditory sensitivity to L2 contrasts was probably so high that any additional sensitivity gained in the music-oriented program was not reflected in the production of the auditory training paradigm. This could be a plausible explanation also considering the earlier studies on musical experience and L2 perception, which focused on adult learners (Bhatara et al., 2015; Marie et al., 2011). The results may have been different if the participants' music and L2 sound perception was measured with preattentive measures, as in the study by Milovanov et al. (2008). As the tentative group-difference seen in the /ʉ/ F2 values in Figure 7 is subtle, it could reach significance in a larger sample size. However, the fact that both groups changed their production after the first training shows that any tentative differences in the Music and Non-music groups' productions were not reflected in how rapidly the groups changed their production towards the acoustic model.

The results of Study III indicate that the 9–11-year-old participants were at a developmental stage, where the benefits of age outweighed any possible benefits of musical experience on L2 vowel production learning in this particular training paradigm. The processes and factors underlying the fast training effects in this age group found in the production measurements remain hypothetical without further research, but we propose that the explanation could lie in the developmental stage of the L1 phonologial system at this age (i.e. L1 category precision and/or L1 category establishment; Flege & Bohn, 2021).

5.1.4 Study IV

The results of Study IV showed that listen-and-repeat training changed 6–7-year-old children's production of an L2 vowel after just one training session. We hypothesized the participants to respond to the training by changing their production of the L2 vowel, but the rapidity with which the change occurred was unexpected. This result supports earlier research on the positive effects of early AOL on L2 sound production and perception (e.g., Baigorri et al., 2019; Flege et al., 1999a, 1999b; Oh et al., 2011; Piske et al., 2002; Tsukada et al., 2005). However, most of these studies have focused on naturalistic L2 environments, and very limited amount of research has focused on children's L2 production learning in instructed classrooms (Kopečková et al., 2019) or phonetic training settings (Giannakopoulou et al., 2013). Our findings offer additional information on 6–7-year-old children's L2 sound production learning and show that they can learn to produce L2 sounds rapidly with a listen-and-repeat method that has traditionally been used in classrooms. This indicates that the benefits of early AOL are not limited to phonetic L2 acquisiton in naturalistic L2 environments and that on 6–7-year-old children are able to learn L2 production efficiently also with limited L2 sound exposure in a classroom-like training setting.

The findings are particularly interesting when compared to earlier results from studies that used the same listen-and-repeat training method and the same stimuli on 7–10-year-old children (Taimi et al., 2014) and adults (Peltola, K. U. et al., 2017, 2020). The 7–10-year-old monolingual Finnish children tested by Taimi et al. (2014) changed their production of the L2 vowel towards the acoustic model after three training sessions, on the second day of the two-day experiment. The monolingual L1 Finnish adults or L1 English speaking adults tested by Peltola et al. (2017) did not change their production of the L2 vowel after one listen-and-repeat training session in a one-day paradigm and the results showed strong effects of the participants' L1. However, a later study (Peltola, K. U. et al., 2020) showed slight production improvements in monolingual L1 Finnish adults after four training sessions in a two-day training paradigm. The amount of training per session was identical in all of these studies (i.e. the stimulus words /ty:ti/ and /tʉ:ti/ were repeared 30 times each per session). This discovery shows that the 6–7-year-old preschoolers tested in Study IV were able to respond to the training and adapt their articulation of the L2 sound contrast faster than the 7–10-year-olds (Taimi et al., 2014) or adults (Peltola, K. U. et al., 2017, 2020). Since the only major difference between the monolingual Finnish subject groups in these studies was the age of the participants, the results indicate that the child speakers in Study IV (6–7 years) benefitted from their younger age (i.e. earlier AOL) in the listen-and-repeat paradigm. Since the age-difference between the child learners in Study IV and Taimi et al. (2014) is relatively small, the difference between the groups' results seems rather large to be explained simply by the CPH or the general enhanced plasticity of the younger children in Study IV. Therefore, we suggest that similar to the results of Study III, the age-related differences found in these four studies using a listen-and-repeat method could suggest differences in L1 category establishment and/or precision between age groups. This explanation is in line with the L1 category precision hypothesis of the SLM-r (Flege & Bohn, 2021).

5.1.5 Overall theoretical implications

Taken together, the results of Studies I–IV show that 6–13-year-old children are able to learn L2 sound production and perception efficiantly through phonetic training paradigms and in immersion classroom settings. Age was found to affect learning results more than other backround factors, but the results indicated that quality and quantity of L2 phonetic input might also affect children's L2 sound learning. The initial hypotheses were primarily based on earlier research comparing children and adults or early and late learners of an L2 (e.g., Flege et al., 1995a; Giannakopoulou et al., 2013; Oh et al., 2011; Tsukada et al., 2005). The findings of Studies I–IV seem to support the earlier findings implicating positive effects of early AOL on L2 sound learning. However, some of our results cannot be explained simply by age of learning or enhanced plasticity for speech sounds. In order to understand the findings of this thesis and its theoretical implications, the possible processes underlying the effects of early AOL as well as the quality and quantity of L2 phonetic input need to be discussed further.

First, AOL has traditionally been considered a linear predictor of L2 sound learning. For example, the original SLM (Flege, 1995) proposed that L2 category formation is more likely for early than late learners of an L2 (the age hypothesis). In other words, according to the age hypothesis, the perception of cross-language phonetic differences and the formation of L2 phonetic categories become progressively less likely as monolingual speakers' L1 categories develop during childhood and adolescence. As already discussed earlier, several results from earlier studies seem to support the age hypothesis (e.g., Flege et al., 1995a; 1995b; 1999b). However, as Flege and Bohn (2021) point out, the age hypothesis fails to distinguish the speakers' L1 phonological development from their overall neurocognitive development.

The recently published revised version of the model, the SLM-r (Flege & Bohn, 2021), has abandoned the original age hypothesis and replaced it with the L1 category precision hypothesis. The category precision hypothesis states that the more precise L1 categories are at the time of first exposure to an L2, the easier it is for listeners to perceive phonetic differences between L1 and L2 sounds and the more likely it is for new L2 categories to be formed. Category precision is hypothesized to be highest in early childhood and lower in adulthood. The precision of an L1 category is defined as “the variability of acoustic dimensions measured in multiple productions of a phonetic category” (Flege & Bohn, 2021). In other words, an L1 phonetic category with high precision has relatively little F1-F2 variability, whereas a category with lower precision has more F1-F2 variability. According to the SLM-r, the category precision is not only connected to age and phonological development, but also to individual differences in auditory processing and acuity (Flege & Bohn, 2021).

The development of the L1 phonetic system linked to the precision of L1 categories may account for some of the age-related findings of Studies I–IV. The developmental stage of L1 cagetories could be the explanining factor underlying the fast training effects found in Studies III and IV as well as the age-differences found between the participants of Study IV and the older age groups in earlier studies (Peltola, K. U. et al., 2017, 2020; Taimi et al., 2014). The fact that both the Music and the Non-music group (Study III) responded rapidly to the auditory training and changed their production of the L2 vowel could indicate that the L1 vowel categories of the participants (9–11 years) were still developing and relatively precise. The acoustic distance between the precise L1 categories could have allowed them to perceive the spectral difference in the stimuli effortlessly and therefore the musical experience of the groups resulted in no differences in L2 vowel production in this training paradigm. If the L1 category precision hypothesis is accepted, the immediate training effects found in 6–7-year-old children (Study IV) are not surprising. The Immersion group in Study II may have also benefitted from L1 category precision due to their earlier AOL compared to the Control group with higher AOL. As mentioned previously, the SLM-r states that L1 category precision is significant at the time of first exposure to an L2. It therefore seems logical that the category precision may have been higher for the Immersion group in the first grade at the age of 7 (i.e., when they started to learn English in the immersion classroom) than for the Control group at the age of 9 in the third grade. This could partly explain the differences in the groups’ L2 vowel production. Overall, the category precision hypothesis seems a plausible explanation for most of our findings. Effects of L1 category development on L2 sound perception has also been proposed by the NLM (Kuhl et al., 1992) and the neural commitment to L1 categories has been found to be connected to reduced ability to perceive L2 contrasts that are irrelevant in the L1 (Kuhl et al., 2008). In addition, the vowel formant frequencies extracted from the participants’ productions in Studies III and IV (Tables 7 and 8), show relatively small standard deviations for the L1 relevant vowel /y/ (compared to the L2 vowel /ʉ/), which could indicate high precision of the /y/ category.

However, the age and the L1 category precision hypotheses fail to account for the age-related findings of Study I, which showed that unimodally distributed auditory exposure to L2 vowel category variants led to the emergence of an MMN response for the L2 prototype in the 9–12-year-old, but not in the 7–8-year-old children. If we were to accept the age or the category precision hypothesis as the explanation for the finding of Study I, the training effects should have been equally or even more pronounced for the 7–8-year-olds. In addition, the results of Study II cannot be attributed to AOL and category precision at the time of first exposure, because age was not the only distinguishing factor between the Immersion and Control groups. The quality and amount of L2 input received in an immersion classroom differ from those received in a standard classroom setting, where the target L2 is only studied as a separate subject. Therefore, the quality and quantity of L2 input need to be discussed in relation to the results of this thesis.

In Study I, the stimuli were isolated synthetic vowels. While the distribution of the stimuli in the training task was designed to mimic natural exposure to speech sounds during infancy (Maye et al., 2002; 2008), the use of instructions before testing was not initially thought to affect perceptual results. However, considering the findings of Peltola and Aaltonen (2005) and Peltola et al. (2012) on the effects of language context on sound category perception, it is possible that without explicit instructions and a clear context for the isolated stimulus vowels, the 7–8-year-old children were not aware they were listening to speech sounds. The 9–12-year-old participants had experience with instructed L2 learning and therefore had probably more language awareness, which could have allowed them to perceive the stimuli as vowels even without a clear language context. In other words, without an explicit context created with instructions, the perception of the isolated vowel stimuli (i.e., quality of the phonetic input) may have required some awareness on the segmental nature of speech sounds from the participants. This notion is supported by the fact that the participants were not given many explicit instructions on the stimuli or the training paradigm in order not to affect the preattentive MMN results. The participants were told briefly that they would hear different vowels, but the information was not explicitly highlighted or repeated during the experiment.

In Study II, the stimuli were the natural productions of a native BrE speaker, but the quality and quantity of input received in the classroom was not controlled and differed between the groups. The Immersion group heard and used English every day, since English was the primary language of education in the immersion classroom. The teacher reported using consistent BrE pronunciation. The Control group heard and used English only in English lessons a couple of hours a week. According to Tyler (2007), the L2 input received in an instructed classroom differs significantly from the phonetic input received in naturalistic L2 contexts, which makes the learning of L2 sound categories less likely in instructed settings. The input in an immersion classroom, where all communication and education happens in the target L2, can be expected to differ from the input received in a standard instructed classroom, where most of the communication and instructions are given in the L1 and the production of an L2 is often practiced through the written forms.

To summarize, the results of Studies I-IV indicate that the developmental stage of the L1 sound system linked to AOL and the precision of L1 categories affects 6–13-year-old children's L2 sound learning, which is in accordance with previous theoretical models of cross-language sound learning (Flege & Bohn, 2021; Kuhl et al., 1992, 2008). In addition, L2 input quality may affect L2 sound production learning in instructed classroom learning and experimental training paradigms, as suggested by earlier research (Flege & Bohn, 2021; Tyler, 2019). The significance of the amount of L2 phonetic input remains tentative, but it should be considered as one possible factor in L2 sound production learning in different classroom contexts.

5.2 Practical implications

The results of Studies I–IV offer some interesting practical implications for children's L2 sound perception and production learning.

First, the finding that even passive auditory exposure can lead to the formation of a new memory trace for an L2 sound (Study I) and changes in L2 sound production (Study III) indicates that hearing L2 speech affects L2 sound perception and production learning in children. This shows that child learners of an L2 benefit from hearing L2 sounds and speech, and that pronunciation learning can be supported by offering L2 auditory input.

Taken together, the results of Studies II, III and IV support the idea of early L2 learning. Regardless of the exact processes underlying the effects of age found in the current studies, it seems evident that early age of learning leads to fast L2 production learning even in instructed classroom settings and experimental training paradigms.

Children seem to benefit from L2 teaching in an immersion classroom when learning L2 pronunciation accuracy. As discussed in section 5.1.2, the difference between the groups' pronunciation of BrE vowels in Study II is most probably a combination of different ages of learning as well as the amount and quality of BrE input received in the classroom. Therefore, it is important to note that even though these results in part offer grounding for early L2 teaching in schools, the findings from Studies I and II show that quality and quantity of L2 input in the classroom and in training situations need to be considered as well. The importance of phonetic input quality and quantity are further supported by recent literature on L2 learning in naturalistic and instructed environments (Flege & Bohn, 2021; Tyler, 2019).

The results of Studies III and IV show that both listening to L2 sounds and repeating after an acoustic model can result in changes in L2 sound production. Based on this finding, it would seem that successful L2 production learning involves both perceptual and productional processes, which may co-evolve and interact.

Therefore, it seems that neither can be separated from L2 pronunciation development and that children benefit from practicing both when learning to produce a challenging L2 sound contrasts. This information ties together with the notion of the importance of the amount and quality L2 phonetic input in L2 sound learning and can be applied to practice by offering children opportunities to produce and to listen to L2 sounds in all learning settings.

5.3 Reliability and validity of the results

The results of Studies I–IV were obtained with methods (EEG, production recordings) and measures (MMN, F1, F2) that are well established and commonly used in phonetic research. The group sizes in studies II–IV were comparable to other L2 production and perception studies in the field (e.g., Baigorri et al., 2019; Heeren & Schouten, 2010; Kopečková et al., 2019; Taimi et al., 2014) and therefore it is unlikely that individual differences would greatly affect the results. The reliability of the results of Study III could, however, be strengthened by increasing group size. On closer inspection, the production data in Study III showed a tentative group-difference in the F2 values of /ʉ/, but it did not reach significance in the statistical analysis. Additional data could be collected to test whether the slight difference in the Music and Non-music groups’ productions seen in Figure 7 would reach significance with larger group sizes. However, the fact that musical experience did not result in significant differences between the groups in the current study shows that the effect of studying in a music-oriented program was very robust, since it did not outweigh the benefits of AOL and L1 category precision linked to the devepolmental stage of the children’s L1 sound system in Study III. In addition, the despite the tentative group-difference in the averege /ʉ/ F2 values, the data shows that both groups changed their production similarly towards the acoustic model (Figure 7), which indicates that they responded similarly to the auditory training. The number of participants in Study I was relatively small due to difficulties in recruiting voluntary participants, but there are previously published MMN studies on L2 sound perception with similar group sizes (Peltola, M. S. et al., 2007; Taimi, Alku, Kujala, Näätänen, & Peltola, 2014). However, the reliability and validity of the results of Study I could be verified with additional data from the same age group.

The results of Study II showed a group-difference between the Immersion and Control groups in the F2 values of the BrE vowels, but the difference was significant only in the voiced context words. As discussed briefly in section 5.1.2, this discovery could be explained by the phonotactic rules of Finnish or the pre-lenis lengthening/pre-fortis clipping phenomena of British English. English words produced by a native BrE male speaker were used as stimuli in Study II, because the experiment used an intervention method that tested the effects of language immersion education on L2 pronunciation. Therefore, the vowel length was slightly longer in the voiced context words. To test whether the lengthening of vowel sounds before voiced consonants (i.e. pre-lenis lengthening) affected the children's production of formant frequencies in the the voiced context words (e.g., /hɪd/ vs. /hɪt/), additional data from similar participant groups could be collected. Using the same stimuli with the vowel duration manipulated so that both the voiced and voiceless context words had the same vowel duration would eliminate any possible effects of vowel duration on vowel quality.

The natural word stimuli (Study II) and the semisynthetic pseudoword stimuli (Studies III and IV) used in the experiments were all originally produced by adult male speakers and the L2 vowel formant frequencies in the stimuli were those typical of an adult male voice. It should be noted that the children tested in these studies were not expected to reach the exact formant frequencies of the stimulus vowels. Instead, our analyses focused on the direction of the changes or the differences in the participants' productions. The results of Studies II–IV show that the child participants did not produce any of the L2 vowels with the same F1 and F2 values that were present in the stimuli. This observation was expected and it does not affect the validity of the results. In addition, the average formant data from Studies III and IV show that the participants in these studies did not produce the novel L2 vowel /ʉ/ as the backed L1 vowel /u/, because the F2 values remained considerably higher (Tables 7 and 8) than the F2 values typical for a Finnish /u/ (Iivonen, 2012).

Studies II and IV used a listen-and-repeat training paradigm, which means that the children heard the auditory stimulus and then repeated it during the three-second inter-stimulus interval. When evaluating the reliability of the results, we should discuss the possibility whether the children tested in Studies II and IV could have imitated the acoustic model. However, as discussed above, the stimuli in both experiments were produced by adult male speakers, which would make it harder for the 6–7-year-old and 11–13-year-old children to imitate the stimuli. The fact that there was a difference between the Immersion and Control groups' productions in Study II shows that the groups did not produce the L2 vowels similarly after the acoustic model and that their L2 educational background affected their pronunciation. If the participants had simply imitated the native model, a difference between the two groups' productions would not be expected, as they were from the same age group. In Study IV, the formant frequency data shows that the 6–7-year-old children were able to produce the two different vowel qualities in the L2 contrast systematically apart after training. This indicates that they had to discern the spectral difference in the stimuli and adapt their articulation accordingly, which indicates phonetic production learning. In addition, the listen-and-repeat training includes the speakers receiving constant and repetitive auditory feedback from their own productions, which means that they received two kinds of phonetic input during the experiment. Therefore, the results of Studies II and IV cannot be explained by imitation of the auditory stimuli.

The use of natural production from age-matched L1 child speakers as stimuli for Studies II and IV would have allowed us to compare the participants' productions directly to the acoustic model. However, the aim of the studies was not to measure ultimate attainment (or nativelikeness), but to examine the changes in production after training. In addition, the use of an age-matched speaker in the stimuli could have encouraged the participants to imitate the acoustic model or, at least, it would have made imitation easier and more likely.

Finally, none of the studies included follow-up measurements or recordings, and therefore no conclusions on the longterm stability of the training effects found in Studies I, III and IV can be drawn.

5.4 Possibilities for further research

Based on the findings of this thesis, future research on children's L2 sound learning could focus on how different kinds of explicit instructions and feedback affect L2 perception and production learning in different age groups. The current studies used no feedback and no explicit information on the L2 sounds present in the stimuli, so using more explicit methods would allow us to discover how childen from different age groups and learning backgrounds respond to highly instructed laboratory L2 settings. For example, the listen-and-repeat method used in Study IV was selected because it is commonly used in L2 classrooms. Adding different types of feedback and/or explicit instructions to a listen-and-repeat task should be investigated further, especially to explore the method's possible longterm effects on children's pronunciation and to offer practical tools for practicing L2 speech production in classrooms. In addition, the results of Study I should be investigated further by using explicit instructions before and during training, so that the non-native vowel stimuli would have more context and the role of language awareness on the formation of L2 sound memory traces in this kind of perceptual training paradigm could be investigated further. In addition, recreating studies II–IV with perceptual measurements would allow us to understand better the connection between perception and production in auditory and listen-and-repeat training paradigms. The effects of L1 category precision and the developmental stage of the L1 sound system on children's L2 sound production and perception learning also need to be clarified further.

Overall, children's L2 sound learning in instructed environments and different phonetic training paradigms has remained scarce over the years, and all phonetic research focusing on children's L2 perception and production learning outside naturalistic L2 environments would offer valuable insight into the processes underlying L2 sound learning in different age groups.

6 Conclusions

The individual conclusions of the studies in this dissertation (studies I–IV) are as follows:

- First, the statistical distribution of L2 category variants seems to affect preattentive L2 category perception learning. Auditory input leads to a familiarity effect, causing memory trace formation for the most familiar (statistically most frequent) L2 sound variant and not for the less frequent variants. The benefits of passive perceptual training of individual L2 sounds might be affected by the level of the children's language awareness and the linguistic context provided during training.
- Second, early L2 exposure in an immersion classroom environment leads to more accurate L2 vowel production than studying an L2 as a separate school subject later. The benefits of early exposure found in this study are probably a combination of early AOL and the quality and quantity of L2 input.
- Third, 9–11-year-old children may have a sensitivity to acoustic differences in L2 vowels due to L1 category precision that is transferred to the trainability of L2 vowel production through auditory training. Music-oriented education in elementary school might not enhance training effects, at least not in the groups tested in this study.
- Fourth, 6–7-year-old children benefit from early AOL linked to L1 category precision in L2 sound production learning, and they can learn to produce a challenging L2 vowel contrast after just one session of listen-and-repeat training.

The overall conclusions of this dissertation deriving from these individual conclusions are as follows:

- From the learning background factors examined in this dissertation, age of learning has the strongest effect on 6–13-year-old children's L2 sound perception and production learning.

- The positive effects of age found in Studies II-IV and the familiarity effect found in Study I could be at least partly explained by L1 category precision or a less established developmental stage of L1 sound categories.
- 6–13-year-old children learn L2 sound production efficiently through listen-and-repeat and auditory training paradigms.
- New memory traces for L2 sounds are formed for the most frequently heard variant of an L2 category but not for the less frequent allophones of the same category (familiarity effect).
- Quality of L2 phonetic input might affect children's L2 phonetic learning in classroom environments and phonetic training settings.

Abbreviations

ANOVA	Analysis of Variance
AOA	Age of arrival
AOL	Age of learning
BrE	British English
CPH	Critical Period Hypothesis
d’	Discrimination sensitivity
EEG	Electroencephalography
EOG	Electrooculogram
ERP	Event-related potential
F0	Pitch
F1	First formant
F2	Second formant
HVPT	High-variability Phonetic Training
ISI	Inter-stimulus interval
L1	First language, mother tongue
L2	Second language
LPC	Linear Predictive Coding
MMN	Mismatch negativity
NLM	Native Language Magnet model
NP	Non-prototype
P	Prototype
PAM	Perceptual Assimilation Model
PAM-L2	Perceptual Assimilation Model of Second Language Speech Learning
RT	Reaction time
SLM	Speech Learning Model
SLM-r	Revised Speech Learning Model

List of References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning, 59*(2), 249–306.

Alku, P., Tiitinen, H., & Näätänen, R. (1999). A method for generating natural-sounding speech stimuli for cognitive brain research. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 110*(8), 1329–1333.

Baigorri, M., Campanelli, L., & Levy, E. S. (2019). Perception of American–English vowels by early and late Spanish–English bilinguals. *Language and Speech; Lang Speech, 62*(4), 002383091880693-700. doi:10.1177/0023830918806933

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words, 167*, 224.

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 speech research* (pp. 171–206) York Press, Baltimore.

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception. In O. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning: In honour of james emil flege* (pp. 13–34). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Bhatara, A., Yeung, H. H., & Nazzi, T. (2015). Foreign language learning in french speakers is associated with rhythm perception, but not with melody perception. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 277–282. doi:10.1037/a0038736

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

Boll-Avetisyan, N., Bhatara, A., Unger, A., Nazzi, T., & Höhle, B. (2016). Effects of experience with L2 and music on rhythmic grouping by french listeners. *Bilingualism: Language and Cognition, 19*(5), 971–986.

Darcy, I., & Krüger, F. (2012). *Vowel perception and production in turkish children acquiring L2 german* doi:https://doi.org/10.1016/j.wocn.2012.05.001

Delogu, F., Lampis, G., & Belardinelli, M. O. (2006). Music-to-language transfer effect: May melodic ability improve learning of tonal languages by native nontonal speakers? *Cognitive Processing, 7*(3), 203–207.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27*(2), 315–330.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229–273) Timonium, MD; York Press.

Flege, J. E. (1999). The relation between L2 production and perception. Paper presented at the *Proceedings of the XIVth International Congress of Phonetics Sciences,* 1273–1276.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole, & J. Hualde (Eds.), *Laboratory phonology 9* (pp. 353–380). Berlin: Moyton de Gruyter.

Flege, J. E., & Bohn, O. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge, UK: 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*(2), 169–186.

Flege, J. E., MacKay, I. R., & Meador, D. (1999a). Native italian speakers' perception and production of english vowels. *The Journal of the Acoustical Society of America, 106*(5), 2973–2987.

Flege, J. E., Munro, M. J., & Mackay, I. R. A. (1995a). Effects of age of second-language learning on the production of english consonants. *Speech Communication, 16*(1), 1–26. doi:10.1016/0167-6393(94)00044-B

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995b). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America, 97*(5 Pt 1), 3125–3134. doi:10.1121/1.413041

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999b). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78–104.

Ghaffarvand Mokari, P., & Werner, S. (2018). Perceptual training of second-language vowels: Does musical ability play a role? *Journal of Psycholinguistic Research, 47*(1), 95–112. doi:10.1007/s10936-017-9517-8

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.

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*, 594–603.

Heeren, W. F. L., & Schouten, M. E. H. (2008). Perceptual development of phoneme contrasts: How sensitivity changes along acoustic dimensions that contrast phoneme categories. *The Journal of the Acoustical Society of America, 124*(4), 2291-2302. doi:10.1121/1.2967472

Iivonen, A. (2012). Kielten vokaalit kuuloanalogisessa vokaalikartassa. *Puhe Ja Kieli,* (1), 17–43.

Kopečková, R., Dimroth, C., & Gut, U. (2019). Children's and adults' initial phonological acquisition of a foreign language. *Journal of Second Language Pronunciation, 5*(3), 374–401.

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.

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. doi:10.1098/rstb.2007.2154

Kuhl, P. K., Williams, K. A., & et al. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*(5044), 606-608. Retrieved from http://search.proquest.com/docview/213553055?accountid=14774

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: John Wiley & Sons, Inc.

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in mandarin chinese. *Journal of Cognitive Neuroscience, 23*(10), 2701–2715.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*(1), 122–134. doi:10.1111/j.1467-7687.2007.00653.x

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101–B111. doi://dx.doi.org/10.1016/S0010-0277(01)00157-3

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P. A., & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. *Brain Research, 1194*, 81–89.

Munro, M. J., Flege, J. E., & Mackay, I. R. A. (1996). The effects of age of second language learning on the production of english vowels. *Applied Psycholinguistics; Applied Psycholinguistics, 17*(3), 313–334. doi:10.1017/S0142716400007967

Näätänen, R. (2000). Mismatch negativity (MMN): Perspectives for application. *International Journal of Psychophysiology, 37*(1), 3–10. doi:https://doi.org/10.1016/S0167-8760(00)00091-X

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*(1), 1–21. doi:10.1111/1469-8986.3810001

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., . . . Luuk, A. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*(6615), 432–434.

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. doi://dx.doi.org/10.1016/j.clinph.2007.04.026

Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2011). A one-year longitudinal study of english and japanese vowel production by japanese adults and children in an english-speaking setting. *Journal of Phonetics, 39*(2), 156–167.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The edinburgh inventory. *Neuropsychologia, 9*(1), 97–113.

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, 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.

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.

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: Soveltavan Kielitieteen Tutkimuksia,* (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. doi://dx.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.

Piske, T., Flege, J. E., MacKay, I. R., & Meador, D. (2002). The production of english vowels by fluent early and late italian-english bilinguals. *Phonetica, 59*(1), 49–71. doi:56205 [pii]

Ramon-Casas, M., Swingley, D., Sebastián-Gallés, N., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology, 59*(1), 96–121.

Shestakova, A., Huotilainen, M., Čeponienė, R., & Cheour, M. (2003). Event-related potentials associated with second language learning in children. *Clinical Neurophysiology, 114,* 1507–1512.

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 to non-native speech sound perception and discrimination in school-aged children: An MMN and behavioural study. *Linguistica Lettica., 2014*(22), 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.

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*(3), 263–290.

Tyler, M. D. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. In A. M. Nyvad, M. Hejná, A. Højen, A. Bothe Jespersen & M. Hjortshøj Sørensen (Eds.), *A sound approach to language Matters—In honor of ocke-schwen bohn* (pp. 607–630). Aarhus: Dept, of English, School of Communication & Culture, Aarhus University.

Wang, Y., & Kuhl, P. K.Evaluating the "critical period" hypothesis: Perceptual learning of mandarin tones in american adults and american children at 6, 10 and 14 years of age. Paper presented at the *Proceedings of 15th ICPhS,* 1537–1540.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., . . . Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36*(5), 638–642.

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2010). Training the brain to weight speech cues differently: A study of finnish second-language users of english. *Journal of Cognitive Neuroscience, 22*(6), 1319–1332.

TURUN YLIOPISTO
UNIVERSITY OF TURKU

Painosalama, Turku, Finland 2021

ISBN 978-951-29-8592-0 (PRINT)
ISBN 978-951-29-8593-7 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)