

**Preschool Children Learning a Foreign Vowel Through a  
Two-Day Listen-And-Repeat Training**

Liina Repo  
Minor Thesis  
Phonetics, Language Specialist Path  
School of Languages and Translation Studies  
Faculty of Humanities  
University of Turku  
June 2019

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

# UNIVERSITY OF TURKU

School of Languages and Translation Studies/Faculty of Humanities

## REPO, LIINA: Preschool Children Learning a Foreign Vowel Through a Two-Day Listen-And-Repeat Training

Minor Thesis, 46 pp., 9 appendix pp.  
Phonetics, Language Specialist Path  
June 2019

---

This thesis concentrates on second language learning during childhood. The theoretical part of the thesis examines different kinds of second language learning theories from the perspective of foreign language learning in children. The empirical part of the study utilises the theories and models presented in the theoretical part in determining the theoretical framework for the study to shed light on the questions around second language learning during childhood in the best possible way.

The aim of this thesis is to study how listen-and-repeat production training affects children's ability to perceive and produce a non-native vowel, which is theoretically as challenging as possible. The non-native vowel was embedded in a pseudo-word context. The stimuli used in the study were two semisynthetic pseudo-words, /tʌ:ti/ and /ty:ti/. Primary interest was pointed to the pseudoword /tʌ:ti/, as it contained the target vowel /ʌ/ which is phonemically irrelevant in Finnish, but relevant in Swedish. The participant group consisted of 12 Finnish preschool girls, aged 6–7. The participants participated in the study on two consecutive days which consisted of two training and two recording sessions. The recorded productions were analysed acoustically to find out the first two formant frequencies of the target vowel /ʌ/ and the non-target vowel /y/. The values were then statistically analysed.

The results revealed that the participants learned to produce the non-native vowel in two days, after four training sessions. The results support earlier research on foreign language speech sound learning by children. Children are fast learners and benefit from listen-and-repeat training in learning foreign language speech sounds.

Key words: L2 learning, children, speech production, speech perception

# Table of Contents

<b>1. Introduction</b> .....	1
<b>2. Theories of Second Language Acquisition</b> .....	3
<b>2.1 Speech Learning Model</b> .....	4
<b>2.2 Perceptual Assimilation Model</b> .....	7
<b>2.3 Native Language Magnet Theory</b> .....	10
<b>3. Children as Language Learners</b> .....	13
<b>3.1 Sensitive Period</b> .....	14
<b>3.2 Factors affecting L2 learning</b> .....	16
<b>3.3 Effects of Training</b> .....	18
<b>4. The Present Study</b> .....	20
<b>4.1 Participants</b> .....	20
<b>4.2 Stimuli</b> .....	21
<b>4.3 Procedure</b> .....	24
<b>4.4 Analysis</b> .....	25
<b>4.5 Results</b> .....	26
<b>5. Discussion</b> .....	30
<b>6. Conclusion</b> .....	38
<b>References</b> .....	41
<b>Appendices</b>	

## List of Tables

<b>Table 1</b> Average frequencies (Hz) of F1 and F2 in four sessions .....	26
<b>Table 2</b> Standard deviations of F1 and F2 in four sessions .....	27
<b>Table 3</b> Minimum, maximum, mean and standard deviation of F2 (/u/) at group level.	34

## List of Figures

<b>Figure 1</b> Spectrogram of the stimulus word /tʌ:ti/ .....	23
<b>Figure 2</b> Spectrogram of the stimulus word /ty:ti/ .....	23
<b>Figure 3</b> Average frequencies of F2 values (Hz) in four sessions .....	27
<b>Figure 4</b> The third pseudoword /tʌ:ti/ from session 1 produced by a participant .....	33
<b>Figure 5</b> The second pseudoword /ty:ti/ from session 1 produced by the same participant.....	33
<b>Figure 6</b> The fifth pseudoword /tʌ:ti/ from session 3 produced by the same participant..	34
<b>Figure 7</b> Average F2 values of /u/ in each session by each participant.....	35

## List of Abbreviations

<b>ANOVA</b>	Analysis of Variance
<b>AOA</b>	age of arrival
<b>ASL</b>	American Sign Language
<b>CPH</b>	Critical Period Hypothesis
<b>DAF</b>	delayed auditory feedback
<b>F0</b>	fundamental frequency
<b>F1</b>	first formant
<b>F2</b>	second formant
<b>F3</b>	third formant
<b>ID</b>	infant-directed
<b>L1</b>	first language
<b>L2</b>	second language
<b>LPC</b>	Linear Predictive Coding
<b>MMN</b>	mismatch negativity
<b>NLM</b>	Native Language Magnet
<b>NLM-e</b>	Native Language Magnet theory expanded
<b>NLNC</b>	Native Language Neural Commitment
<b>PAM</b>	Perceptual Assimilation Model
<b>SD</b>	standard deviation
<b>SLA</b>	second language acquisition
<b>SLM</b>	Speech Learning Model
<b>SSG</b>	Semisynthetic Speech Generation

## 1. Introduction

Language learning is a well-studied subject in the field of phonetics. Research on native (L1) and second language (L2) acquisition have been widely conducted from many different perspectives. Questions such as “How L1 and L2 learning differ from each other?” or “How can we predict difficulties that arise from the relationship between a learner’s L1 and L2?” have long been and are still central questions in phonetics. A learner’s L1 is present in learning any subsequent languages and affects the learning process in different ways. Infants have an ability to discriminate between all the sounds in the world (Kuhl 1994, 812). Over time, however, this ability diminishes as categorical perception and the child’s L1 hinder the discrimination process of small differences (*ibid.*). This is important, as it is not meaningful to perceive small differences in the productions of other people with the same L1.

Speech sounds and other phonetic phenomena in different languages have different kinds of functions. The perception and production of these differences is important in L2 learning in order to avoid misunderstandings and foreign accent, which has often been seen as negative. The difficulties in producing foreign language sounds may be due to the inability to perceive the foreign sound and its difference to L1 sounds (Flege 1988, 277). In order to produce a foreign sound correctly, it must first be perceived correctly. If a difference between two sounds is not perceived, the sounds are likely to be produced similarly. The difficulties may also rise from motoric difficulties in producing the sound or changing the already learned articulatory manners (Flege 1988, 324-5).

Speech is naturally based on properly working hearing mechanisms, but also on different kinds of sensory feedback mechanisms, such as auditory, tactile and proprioceptive feedback, as well as a feedforward mechanism. These mechanisms give the speaker information and feedback on their production enabling thus successful communication between speakers. Auditory feedback, that is both air- and bone-conducted, gives the speaker information that is based on hearing oneself. This mechanism is essential in language acquisition by infants, but its importance is reduced after the acquisition of L1 due to its slowness. Auditory feedback is too slow to monitor ongoing production in real time because the information arrives too late, though it is important mechanism in maintaining speech (Borden, Harris & Raphael 2003, 126). Interfering with auditory

feedback by delaying it can distort fluent speech by causing repeated or prolonged syllables. Delayed auditory feedback (DAF) experiments have been used in speech studies to demonstrate the importance of auditory feedback in speech perception and production (*ibid.*).

Tactile feedback gives motoric information to the speaker from the sense of touch. The articulatory organs are continually coming into contact with each other, for example, when the tongue touches the alveolar ridge, or the lower lip touches the upper lip. Proprioceptive feedback, on the other hand, gives motoric information on self-movement and the position of the articulatory organs without the sense of touch. Proprioceptive feedback is used, for instance, in vowel production where the tongue's position must be determined without tactile information. Tactile and proprioceptive feedback mechanisms are faster than auditory feedback, enabling the speaker to correct their production even while speaking (Borden et al. 2003, 127–128).

Feedforward mechanism provides the speaker information about the relationship between articulation and acoustics. Feedback mechanisms are too slow for natural and fluent conversation. The feedforward mechanism is thought to be a faster mechanism that is based on already learned articulatory models (Perkell 2012, 395). The feedforward mechanism controls speech anticipatorily, enabling thus natural conversation (*ibid.*). The feedforward and the three feedback mechanisms discussed above are important in controlling and monitoring speech. These mechanisms control the coordinated use of mechanism related to respiration, articulation and phonation that are needed in the production of speech. It is thought that once speech is acquired it functions under the feedforward mechanism (*ibid.*), although the feedback mechanisms are important in maintaining the feedforward system as well as in language learning situations where it is important to notice errors in one's own production.

This study aims to find out how a two-day production training affects children's ability to produce a non-native vowel, which is as challenging as possible in the light of second language acquisition theories, as well as how quickly the children learn to produce the novel sound with the chosen training method. The research setup in the present study resembles classroom teaching in that it utilises listen-and-repeat training method. The participants were preschool girls whose ability to learn the Swedish vowel contrast /y/–

/u/ was investigated. The results of the present study were compared to a study with similar research setting by Taimi et al. (2014).

This thesis consists of six chapters. Chapter 2 focuses on three theories of second language acquisition, the Speech Learning Model, the Perceptual Assimilation Model and the Native Language Magnet theory. These theories try to account for the difficulties learners face when learning an L2. Chapter 3 is concerned with children as language learners and discusses the idea of a sensitive period as well as some factors that affect L2 learning, giving then a brief overview of the literature around perception and production training studies with children and adults. Chapter 4 concentrates on the framework of the present study and begins by introducing the participants of the study as well as the used stimuli. It then continues by describing the course of procedure and analysis, moving finally to the results of the study. Chapter 5 discusses the main findings in a broader perspective and in the light of other studies. Finally, Chapter 6 concludes this thesis by summarising the most important findings as well as commenting on future research.

## **2. Theories of Second Language Acquisition**

Second language acquisition (SLA) theories intend to explain the processes that affect the learning of foreign languages. Research in SLA is related to several disciplines, such as linguistics, psychology and education, and theories of SLA try to shed light on various aspects of language learning, but no overarching theory has yet been widely accepted. Language *acquisition* may be separated from language *learning*. Krashen (1982, 10), for example, defines acquisition as “a process similar [...] to the way children develop ability in their first language” and learning as “conscious knowledge of second language, knowing the rules [...] and being able to talk about them.” Thus, the key difference between acquisition and learning is the way of processing language information: acquisition is a subconscious process, whereas learning requires conscious awareness. As Krashen notes, language acquisition is often connected to first language acquisition by children, whereas learning is connected to second language learning by adults (*ibid.*). However, adults are also capable of acquiring a second language, although commonly adults use conscious learning mechanisms whereas young children acquire

unconsciously as they have not yet developed conscious skills to even realise that they are learning. For the purposes of this thesis, the terms acquisition and learning are used interchangeably in the context of second language acquisition.

This chapter presents three central theories around SLA. These theories are connected to difficulties learners face when learning the phonology of a new language and try to explain the reasons behind these difficulties. The theories are all contrastive in nature in that they categorise L2 sounds based on their relationship to L1 and try to account for how these L1–L2 sound contrasts affect learners. First, section 2.1 presents the Speech Learning Model, after which the Perceptual Assimilation Model is discussed in 2.2 and last, 2.3 introduces the Native Language Magnet theory.

## 2.1 Speech Learning Model

The Speech Learning Model (SLM, 1995) by Flege is a model concerned with the problems of language learning. It tries to account for the differences in the learnability of phonetic segments in a foreign language. The SLM is primarily concerned with contrastive phonetic elements and proposes that phonetic similarity between L1 and L2 sounds determines both their classification and their learnability. The SLM also tries to account for the ultimate attainment of L2 pronunciation and the age-related limits on the ability to pronounce L2 sounds in a native-like manner (Flege 1995, 237–238). The classification of L2 sounds by their relationship to L1 sounds is presented below.

According to the SLM, there are three possible types of relationship between an L1 and an L2 sound systems, and the degree of difficulty in learning the L2 sounds vary depending on the relationship. Language-specific speech sounds are represented in the long-term memory as phonetic categories and the relationship of foreign sounds and native categories affect the learning of L2 sounds. Firstly, an L2 sound may be “identical” to a sound in L1. “Identical” sounds are ones that are shared by both L1 and L2, that is, they are acoustically identical in both languages (Flege, 2005). According to the SLM, “identical” sounds should not produce any difficulties for L2 learners, since they can employ the already available L1 sound category. For instance, the sound /n/ is identical in Finnish and English and, in accordance with the SLM, should not produce any

difficulties for Finnish learners of English or vice versa. However, acoustical equivalence does not necessarily mean equivalence in use. For instance, the sound /m/ in English is a syllabic consonant and can form a syllable on its own, whereas in Finnish it cannot.

Secondly, an L2 sound may be “new” in reference to the L1 sound system. “New” sounds are not present in the learner’s L1, and may thus produce learning difficulties as the learner must create a new category for the new sound (Flege 1987, 48). “New” sounds have no resemblance to L1 sounds in acoustic, auditory or articulatory level (Flege 1988, 228), though it must be born in mind that most L2 sounds do have some similarities to L1 sounds and very few are devoid of resemblance. For example, for a Finnish learner of a Khoisan language, the Khoisan click consonants could produce learning difficulties as they are a novel sound category. It is questionable whether a vowel sound can be considered to be “new,” as the same acoustic vowel space may be distributed differently in different languages (Paganus et al. 2006; Peltola 2003). Flege (1987, 48), however, proposes that the French /y/ is a “new” sound for English learners as English has no category for /y/.

Finally, an L2 sound may be “similar” to an L1 sound. “Similar” sounds have an easily identifiable counterpart in L1 but the sounds differ from each other acoustically (Flege 1987, 48). For instance, Flege (1988, 228) presents two examples of “similar” sounds: the vowel /u/ is manifested with auditorily detectable acoustic differences in French and English, and the /s/ of English and Dutch differ systematically in their acoustic qualities. Out of the three sound categories, the “similar” sounds produce maximal learning difficulties as the sound is perceived as an allophone of another sound and thus may not be perceived and produced correctly. For example, a Finnish learner of Swedish may perceive the Swedish vowel /u/ as a representative of the Finnish /y/ or /u/. This maximal learning difficulty in which a foreign /u/ sound assimilates to an L1 sound is utilised in the present study (see Chapter 4).

Rather than being distinct categories, the SLM’s division of foreign sounds into “identical,” “new,” and “similar” can be better thought as a continuum. Depending on the contrast and the size of phonetic difference between the L1 and L2 sounds, learning difficulties may vary from minimal to maximal. When the size of phonetic difference is minute or non-existent (“identical”), learning difficulties are minimal. When the size of difference grows (“similar”), learning difficulties start to arise and may produce even

maximal difficulties. The “similar” sounds may vary in their similarity and thus their perception may vary as well. Flege (1997, 12) notes that both “similar” and “new” sounds differ acoustically from L1 sounds but there is both qualitative and quantitative difference in the amount of phonetic similarity between these two types of L2 sounds and the sounds of L1. “New” sounds represent the other end of the continuum, as the size of phonetic difference grows substantial and there is no phonetic counterpart for the L2 sound in the L1 sound system.

In addition to categorising L2 sounds and their relations to L1 sounds, the SLM also discusses the ultimate attainment of L2 pronunciation and the age-related limits on this attainment. The SLM emphasises the amount of L2 experience as L1–L2 sound contrasts become less likely to be identified by L2 learners when age and language experience increases. For instance, Flege (1995, 243) notes that a learner with Spanish L1 should more likely to be able to establish a phonetic category for the English /æ/ than /i/. This is in accordance with SLM’s categorisation as the English /æ/ could be thought as “new,” whereas the English /i/ would be a “similar” sound. Furthermore, Flege states that an 8-year-old Spanish learner of English would be more likely to establish a category for the English /æ/ than a Spanish 16-year-old (*ibid.*). The SLM predicts that older L2 learners will not be as successful in learning “similar” sounds because they equate the L2 sounds with their L1 (Flege 1997, 14). Flege proposes that the reason behind this could be that adults may take greater advantage of higher-order syntactic and semantic information than young children (*ibid.*).

As a reaction against the Critical Period Hypothesis, according to which there is an optimal time for language learning (see discussion in 3.1), Flege presents empirical evidence from his own studies as well as observations from other researchers’ studies that let him conclude that although foreign accents are very common, they are not inevitable and thus even adult learners can attain native proficiency of an L2 (Flege 1995, 236). The amount of experience in L2 affects the proficiency. Flege, Frieda and Nozawa (1997) found out that actively using one’s L1 affects the L2 production accuracy. Though misarticulations are related to motoric difficulties, the roots of a foreign accent, according to Flege, lie in the perception of the L2 (Flege 1987, 236).

## 2.2 Perceptual Assimilation Model

The Perceptual Assimilation Model (PAM, 1995) by Best and Strange is a model of speech perception that operates on gestural level. Phonetic gestures, as defined by Fowler and Rosenblum (1991, 102), are “organized movements of one or more vocal-tract structures that realize phonetic dimensions of an utterance.” In speech perception, the articulatory gestures are perceived, whereas the acoustic cues are simply means by which the gestures are perceived (*ibid.*) The perception of gestures is a core premise of PAM, whereas SLM (see 2.1) operates on acoustic-phonetic cues that are stored in memory as phonetic categories. As with SLM, however, the emphasis in PAM is on phonetic level. However, unlike SLM, PAM does not exclude phonological analysis from the framework: “a common gestural domain for both phonetic details and phonological structure, in which the constellations of language-specific gestural details *are* the phonological elements of the language” (Best 1995, 183, emphasis in original).

As with SLM, PAM postulates that the perceived L2 sounds are filtered through the native language. Any non-native segment is perceived according to its similarity to or discrepancy from a native segment closest to it in the native phonological space (Best 1995, 193). The universal phonetic domain and phonological spaces are defined by the vocal tract spatial layout as well as the characteristics of articulatory gestures, providing the scope within which the similarity is evaluated (Best 1995, 193). In this sense, the L2 sounds are perceptually assimilated to the native phonological system and its categories. PAM supports the view that perceptual learning can occur in adulthood but highlights the differences between adult and child learners in their abilities to perceive foreign sounds.

PAM presents three patterns of perceptual assimilation in which an L2 sound can be assimilated: 1) to a particular L1 category, 2) as uncategorizable speech sound and 3) as non-speech sound (Best 1995, 194–195). In the first case, sounds that are assimilated to an L1 category may be perceived as a good, acceptable or notably deviant exemplar of the L1 category (Best 1995, 194). Uncategorizable speech sounds are assimilated within the L1 phonological space but they are not perceived as exemplar of any specific native category (*ibid.*). On the other hand, sounds that cannot be assimilated into the native phonological space are not heard as speech at all (Best 1995, 195). In relation to these assimilation patterns, PAM also discusses the assimilation patterns of L2 contrasts that

“follow predictably from the assimilation of each member of the contrast” (*ibid.*) Six different pairwise assimilation patterns (1) “Two-Category Assimilation,” (2) “Category-Goodness Difference,” (3) “Single-Category Assimilation,” (4) “Both Uncategorizable”, (5) “Uncategorised versus Categorised” and (6) “Non-assimilable” and predictions of discrimination levels for each discussed in PAM are presented below.

The first assimilation pattern is “Two-Category Assimilation” in which two L2 sounds can assimilate to two L1 sound categories, and their discrimination should be excellent (Best 1995, 195). Best provides an example of the Hindi voiced retroflex stop /d/ and breathy-voiced dental stop /dʰ/ that are likely to be assimilated to the English voiced alveolar stop /d/ and the voiced dental fricative /ð/ (Best 1991, 14). Next, “Category-Goodness Difference” is a pattern in which two L2 sounds can assimilate to one L1 category so that one of the sounds will be considered as a better exemplar of an L1 category whereas the other sound will be considered as a poor exemplar of that same category (Best 1995, 195). Discrimination is expected to be from moderate to very good, depending on the difference between category goodness of the sounds (*ibid.*) For instance, the English near-close back rounded (or lax) /ʊ/ and the close back rounded vowel /u:/ may assimilate to the Finnish category of /u/ so that the /ʊ/ represents a poor instance of the Finnish /u/. The third pattern “Single-Category Assimilation” predicts that two L2 sounds can assimilate to one L1 category equally well or poorly causing the highest degree of difficulty in discrimination (Best 1995, 195). Best gives an example of the Thompson Salish ejective velar /k/ and uvular /q/ that are both likely to assimilate to the English /k/ as poor exemplars (Best 1991, 14).

In the fourth pattern, “Both Uncategorizable,” two L2 sounds fall within phonetic space but are not perceived as exemplars of any native category, and their discrimination is expected to be from poor to good, depending on their proximity to the L1 categories and each other (Best 1995, 195). For instance, according to Bundgaard-Nielsen, Best and Tyler (2011) the Australian English /əʌ/ – /o:/ pair is categorised as uncategorizable with reference to Japanese L1. The fifth pattern, “Uncategorised versus Categorised,” predicts that one L2 sound assimilates to one L1 category whereas the other L2 sound falls within phonetic space but not within any L1 category (Best 1995, 195). For instance, in the English /s/ – /θ/ contrast, /θ/ has no obvious counterpart in Japanese, whereas /s/ is assimilated to Japanese /s/ (Guion et al., 2000). Discrimination of this pattern is expected

to be very good (Best 1995, 195). The final pattern is “Non-assimilable” in which two L2 sounds may be too discrepant from the L1 phonological categories to be even perceived as speech sounds, and are therefore non-assimilable (*ibid.*). Discrimination of this contrast is expected to range from good to very good (*ibid.*). For example, the click consonants of Zulu could be thought as non-assimilable for native English learners as their phonetic-articulatory features do not correspond well to English phonemes (Best, McRoberts, & Sithole 1988, 5).

PAM also focuses on the weakening of speech perception in non-native context, arguing that infants’ recognition of articulatory gestures is the cause of change in speech perception. Infants’ difficulty in perceiving non-native sound contrasts is predicted by the similarity of the articulatory gestures between L1 and L2 categories (Best 1995, 194). PAM, as it was devised by Best in 1995, focuses on the speech perception of inexperienced and naïve listeners who are defined as “functional monolinguals, that is, not actively learning or using an L2, and are linguistically naïve to the target language” (Best & Tyler 2007, 16). By comparison, according to Best and Tyler, L2 learners are in the “process of *actively learning* an L2 to achieve functional, communicative goals, that is, not merely in a classroom for satisfaction of educational requirements” (*ibid.*, emphasis in original).

PAM has been revised in 2007 to extend the theory to more experienced L2 learners. The revised theory, PAM-L2 (Best & Tyler 2007), highlights how extensive L2 experience may change a language-specific phonological system. As noted above, PAM addresses perceptual assimilation at the phonetic level like SLM, but also at the phonological level. Equivalence between L1 and L2 phonological category does not automatically imply equivalence at the phonetic level (Best & Tyler 2007, 28). As an example, Best and Tyler discuss in the revised version of PAM the French and English /r/ and how English learners of French recognise its phonetic version [ʀ] as phonologically equal to the English [ɹ], despite their perceptible differences (*ibid.*).

### 2.3 Native Language Magnet Theory

The Native Language Magnet theory (NLM, 1992) by Kuhl is a speech perception and an L2 learning model that is based on categorical perception of speech and the perceptual magnet effect. NLM is a theory that connects innate factors and early experience in language and describes how these interact in the development of speech perception. The basic idea behind NLM is that the brain becomes sensitive for the sounds of L1 very early on, by the age of 6 months, and the exposure to L1 thus alters the infant's perception well before the acquisition of word meaning or linguistic contrast, for instance (Kuhl 1991). Due to the infant's sensitiveness to L1, prototypes start to develop for the sounds of L1 (Kuhl 1994, 813–814) The prototypes represent the sound variants that are heard most frequently and are exceptionally good instances that represent the cores of phonetic categories. These prototypes, according to Kuhl, function as “perceptual magnets” that attract the nearby instances of the same category (*ibid.*). That is to say, the closer the L2 sound is to the L1 magnet, the more difficult it becomes to perceive the distinction between the sounds. Consequently, the perception of distinctions is easier near the boundaries of the phonetic categories where the effect of the magnet is smaller.

Kuhl specifies three phases for the development of early speech perception. In the first phase, infants are able to differentiate between all human speech sounds, and the ability originates from general auditory processes rather than any speech-specific mechanisms (Kuhl 1991). The second phase is related to linguistic experience as infants' sensitivity to sounds change when experience in L1 sounds accumulates, activating prototypes that begin to act as perceptual magnets and increasing the perceived similarity between members of the L1 sound category (Kuhl et al. 2008, 982) In the final phase, the perceptual magnet effect, or the distortion of perception, facilitates L1 phonetic abilities and reduces L2 phonetic abilities (*ibid.*).

In relation to L2 language learning, categorical perception together with the perceptual magnet effect creates challenges for language learners when the sounds of the L2 do not correspond with the prototypes of L1. The sounds of L2 are categorized to fit within the native categories which creates difficulties in the perception and production of L2 sounds. For instance, the phonemes /ɾ/ and /l/ are distinguished in English, but not in Japanese.

Thus, for a Japanese L2 learner of English, both sounds fit in the Japanese liquid /r/ category and differentiating them may be challenging.

NLM has been revised to better describe the developmental transition from infants' ability to discriminate between all the speech sounds in the world to adults' difficulty in discriminating non-native phonetic sounds. The revised theory, Native Language Magnet theory expanded (NLM-e, 2008), operates on five principles that guide the model: (1) "Distributional patterns and infant-directed speech are agents of change," (2) "Language exposure produces neural commitment that affects future learning," (3) "Social interaction influences early language learning at the phonetic level," (4) "The perception–production link is forged developmentally" and (5) "Early speech perception predicts language growth" (Kuhl et al. 2008, 982–985).

According to the first principle, infants' phonetic perception is altered first, because the distribution of language specific speech sounds heard by infants causes language-specific representation to form and second, because infant-directed (ID) speech exaggerates phonetic differences making them more discriminable (Kuhl et al. 2008, 982–983). The first principle predicts that the transition from universal phonetic perception to language-specific phonetic perception derives thus from infants' hearing their L1 sounds more frequently and being able to distinguish between sounds more easily when hearing ID speech, which leads to attuning to the L1.

The second principle addresses linguistic exposure and its effects on L2 learning. The brain's early language experience affects later abilities to learn new phonetic schemes (Kuhl et al. 2008, 983). This process has been termed as *native language neural commitment* (NLNC), according to which initial language experience physically affects the brain's neural tissue causing difficulties in later perception of language input (*ibid.*).

The third principle predicts that natural and complex social interaction situations affect the infant's phonetic skills and infants may need a social tutor in order to learn (Kuhl et al. 2008, 984). Research has shown that infants learn to perceive phonetic contrasts better from complex natural language learning situations, as opposed to hearing the same language input from television or audiotape (Kuhl, Tsao, & Liu 2003). Thus, human interaction in a natural setting provides infants with better language input that facilitates language learning.

The fourth principle has to do with the relationship between speech perception and production. According to NLM-e, the link between perception and production of speech is formed through perceptual experience with language which causes sensory learning which, in turn, guides the development of motor patterns (Kuhl et al. 2008, 984–985). As NML-e predicts strong relationship between perceptual mapping through experience with vocal play and articulatory movements, it resembles earlier theories arguing for close linkage between perception and production, such as the Motor Theory<sup>1</sup> (Liberman et al. 1967) and direct realism<sup>2</sup> (Fowler 1986), although the relationship is seen as developmental, rather than automatic or innate.

The fifth principle states that the ability to discriminate between foreign language phonetic contrasts corresponds with the degree of the brain's "initial [...] state – 'open' and uncommitted to native language speech patterns" (Kuhl et al. 2008, 985). Early perception of L1 phonetic units predicts accelerated language development whereas good perception of non-native phonetic units predicts slower language developments (Kuhl et al. 2005). In other words, early language development results from the brain's attunement to the native language phonetic system which leads to better discrimination of sound contrasts relevant for the L1, in the expense of irrelevant sound contrasts.

The theories presented in this chapter rely on the idea that difficulties in foreign language learning are caused by the differences between the L1 and L2 sound systems. SLM, PAM and NLM do not fully agree on what kinds of differences there are or how these differences affect L2 learning. In the light of these theories, children who have already acquired their L1 are considered to be equal to adults in language learning: children are faced with the same language learning difficulties as adults. According to SLM, the learning difficulties caused by similar sounds in L1 and L2, for instance, are considered to be the same for learners of all age. Also, the same assimilation patterns in PAM apply to children as well as adults. Only NLM takes into account the brain's maturational stage and the learner's age. Overall, theories of SLA usually do not consider the learner's age in L2 learning even though a child and an adult usually differ from each other in many

---

<sup>1</sup> The Motor Theory hypothesizes that speech is perceived by identifying the actual vocal tract gestures with which the sounds are produced, or the intended gestures, rather than by identifying the acoustic sound patterns generated by them.

<sup>2</sup> The direct realist theory of speech perception is part of a general direct realist theory that postulates that the object of speech perception are actual vocal tract movements or gestures.

respects. The next chapter focuses on children as language learners and considers the effect of age in language learning situations.

### **3. Children as Language Learners**

This chapter focuses on children as language learners and examines the common belief that younger equals better, that is, children are considered to be better language learners than adults. It is widely accepted that learning an L2 is easier at an early age and that attaining native level proficiency is more likely if one starts to learn a new language at early childhood (see e.g. Singleton & Ryan 2004 for a review). Native-like proficiency is also possible in adulthood, but native-like pronunciation in particular seems to be very hard to attain in later life. Evidence for the view that children are better language learners comes mainly from studies that compared groups of adults that differed in their age of exposure to a foreign language in immigration situations. For example, Munro, Flege and MacKay (1996) studied English vowel productions of native Italians who had arrived in Canada at 2–23 years old. Similarly, MacKay et al. (2001) examined English consonant productions of four groups of bilinguals who differed in their age of arrival (AOA) in Canada. Yeni-Komshian, Flege and Liu (2000) studied the pronunciation proficiency and overall degree of foreign accent in Korean and English bilinguals who had moved to the USA and whose AOA ranged from 1 to 23 years. These studies show that participants who were introduced to the L2 in early childhood typically showed more native-like proficiency in comparison with participants who were introduced to the L2 in later adolescence or adulthood.

Children as language learners are considered in this chapter from three perspectives. First, section 3.1 discusses the concept of sensitive period in language learning and the effect of age in language attainment. Next, section 3.2 examines other factors affecting foreign language learning, such as learning environment and motivation, and compares the effect of these factors in adults and children. Finally, section 3.3 provides an overview of the literature around the effects of speech perception and production training in children and adults.

### 3.1 Sensitive Period

The optimal age to learn languages has long been debated by researchers. This optimal age, or time frame, is known as critical, or sensitive, period during which language learning is thought to be easier than in later life. In language acquisition theory, this optimal time frame is known as the Critical Period Hypothesis (CPH). The idea was first introduced in 1959 by Penfield and Roberts who claim that “the human brain becomes progressively stiff and rigid after the age of nine” ([1959] 1981, 236).

The CPH was then popularized by Lenneberg (1967) who claims that there is a critical age for first language acquisition that lasts from early infancy to puberty. The CPH is a theory according to which there is a non-linear relationship between the ability to acquire a language and the age of the learner. The ability to learn languages is reduced at puberty due to the brain’s loss of neurological plasticity (Lenneberg 1967, 158). Due to the loss of plasticity, the brain’s ability to change is reduced and “the brain behaves as if it had become set in its ways” (*ibid.*).

The CPH claims that the ideal time window from early infancy to puberty is the time when a native language must be acquired in a linguistically rich environment in order to secure normal and balanced language acquisition (Lenneberg 1967). If infants are not provided adequate language input inside this time frame, they will never fully achieve a command of their L1. Some support for the critical period comes from the cases of people deprived from language input and natural social interaction until about puberty. One of the most famous of such cases is Genie, who was kept away from social contact, and thus lacked language exposure for the most of the first 13 years of her life. Although Genie learned to communicate to some extent, she was unable to fully acquire a language (for a full account on Genie, see Curtiss 1977). Genie’s case can thus be seen as supporting evidence for the CPH. However, as Johnson and Newport (1989, 62) note, the question whether Genie’s language difficulties resulted only from lack of social and linguistic input during her early years is under debate due to her abnormal upbringing.

Johnson and Newport also discuss a study on critical period effect in the acquisition of American Sign Language (ASL). The results show that there is a linear decline in the performance of the subjects (native learners, early learners and late learners of ASL) with increasing age of exposure (Johnson and Newport 1989, 62–63). The conclusion that

critical period exists for acquiring ASL is further elucidated, though not without reservations:

This study thus provides direct evidence that there is a decline over age in the ability to acquire a first language. It also tells us, however, that Lenneberg's portrayal is at least partially incorrect in two regards. First, the results show a continuous linear decline in ability, instead of a sudden drop-off at puberty as his hypothesis implies. [...] Second, it should be noted that, while the postpubescent learners did not reach as high a level of proficiency as the native or early learners, language had not become totally unlearnable for them. This rules out any extreme interpretation of the critical period hypothesis. (Johnson and Newport 1989, 63)

As can be concluded from the cases discussed above, it is possible to acquire enough language after puberty to even allow limited verbal communication.

The CPH has so far been discussed in relation to first language acquisition. The CPH, however, has also been extended from L1 acquisition to L2 learning. Adults are generally claimed to be inferior to children in learning new languages, as was discussed above. This applies especially in relation to foreign speech sounds, as adult learners rarely achieve native-like pronunciation, in spite of having better conscious learning mechanisms and thus progressing faster in the beginning. The processes of acquiring a first language and learning a subsequent one are different, however. As Paradis (2004, 60) notes, speakers process their later learned L2 and their native language differently, and the L2 production and comprehension are rarely at the level of L1 production and comprehension. Even if the production of an L2 learner would be observably identical to the production of an L1 speaker, the learner has to resort to a very different language learning and processing mechanism as the acquisition of implicit competence is not possible anymore, or at least very time-consuming and difficult (*ibid.*).

Research does not agree on the timing of the critical period in L2 learning, as estimates range from very young to puberty. The debate on timing derives from differences in research focus: research on L2 phonological development may conclude that the critical period ends very early on, whereas research on L2 syntactic development may show that it ends much later (see e.g. Johnson & Newport, 1989; Ruben, 1999). The timing of critical period may also be affected by the research subjects: for example, if 12-year-old children are studied, it can be concluded that critical period does or does not extend to

this age. Gathering subjects from many different age ranges would be too laboursome. Newport (1990) proposes that the critical age, or the age when critical period ends, is connected to the maturational stage of the learner, rather than any specific age. Based on their study on American children and young adults learning Mandarin Chinese tones, Wang and Kuhl (2003) note that L2 learning is not a strictly timed developmental process with fixed ending periods. Seliger (1978, 18) also concludes that rather than a fixed critical period, there are multiple critical periods that correlate with localisation and the gradual loss of plasticity.

The term *sensitive period* has been used as less rigid kind of critical period. The difference between these terms is subtle. Theorists believing in critical period believe that without adequate input in the optimal time window, children cannot develop the abilities they should have developed during that time. On the other hand, theorists believing in sensitive period believe that without adequate input in the optimal time window, children are going to have problems in later life but they do not think the inability to develop is permanent (Hurford 1991). Some researchers have opted to use the term critical period to denote the optimal time window for acquiring an L1 and the term sensitive period for the optimal time window for learning an L2 (e.g. Patkowski 1980).

### **3.2 Factors affecting L2 learning**

As has been discussed in the previous section, early L2 learners may be more likely to achieve native-like proficiency in their L2, especially in terms of pronunciation. However, factors other than age have been claimed to be as significant as, or even more important, as age in determining L2 learning success. For example, as was mentioned in 2.1, the degree of using one's L1 affects the pronunciation of an L2 even though the AOA would be the same (Flege, Frieda and Nozawa, 1997). Robertson (2002) claims that factors such as personal motivation, learning environment and anxiety may be more significant in L2 learning. However, these factors are closely linked to the learner's age and cannot be completely separated. This section focuses on social psychological factors, namely learning environment, motivation and affective factors, and their effect on L2 attainment in children and adults.

Environment may be a significant factor in determining L2 learning success. Often times L2 learning in childhood takes place in natural environment or language setting, such as in foreign language day care centres or schools. In these kinds of situations, children are exposed to a foreign language as it would be their L1. Thus, learning takes place as unconscious acquiring rather than conscious studying and learning. Adults, however, often study their L2 consciously in a classroom setting. This creates a completely different learning situation compared to the natural social setting children are often provided with. Unlike children, adults who have emigrated to a different country may not have to use the language of the country they are residing in. For instance, Snow and Hoefnagel-Höhle (1978) studied various aspects of L2 Dutch perception and production of native English adults and children of 3–15 years of age. After six months of residing in the Netherlands, adults showed a somewhat more native-like performance than younger children, but the children were likely to outperform the adults after 10-11 months (*ibid.*). Snow and Hoefnagel-Höhle's study was conducted to assess the CPH (see section 3.1) and the results, according to them, failed to support it (Snow & Hoefnagel-Höhle 1978, 1125). These results could be interpreted to mean that the children were more successful in learning the L2. However, all the subjects were learning Dutch by “picking it up at school or at work, with little or no formal instruction” (Snow & Hoefnagel-Höhle 1978, 1115). Children need to use the L2 for everyday activities in day-care and schools, whereas adults may resort to English, which is widely spoken by Dutch adults. Consequently, the learning environment is a significant factor in L2 learning and may have a great impact on the learner's level of L2 proficiency.

In addition to learning environment, motivation may be another factor affecting learner's achievement in learning an L2. Motivation is the internal drive that gives an incentive to pursue a course of action and later to sustain the process. According to Dörnyei and Clément (2001), other factors than motivation are not enough on their own to ensure achievement in L2 learning and thus motivation is a significant factor in L2 learning. However, motivation alone does not automatically guarantee success in learning. Csizer and Dörnyei (2005, 20) provide an extreme example where students may pursue a learning task with great enthusiasm, but learn nothing due to inadequate instruction. Studies suggest that motivation declines with age in school context and that many external factors such as teaching methods and materials and the classroom atmosphere may have an effect on learner's motivation (e.g. Williams & Burden 1999; Dörnyei 1994; Chambers

1998). Two types of motivation have generally been distinguished: *intrinsic* and *extrinsic* motivation. Intrinsic motivation refers to doing something because it arouses interest and brings enjoyment. Extrinsic motivation, on the other hand, refers to doing something in hopes of an external reward or in fear of punishment. Ryan and Deci (2000) note that older learners are more likely to have an extrinsic motivation for learning an L2, for academic or economic success, for instance, whereas younger learners are motivated by intrinsic desires to participate actively with their peers in a more naturalistic acquisition setting.

L2 learning involves numeral other factors, such as attitudes towards the L2 and the language community, language aptitude, anxiety, personality and self-esteem. These factors may account for different degrees of success in L2 attainment. Positive attitudes towards learning an L2 are linked to achievement (for further discussion, see Gardner 1985, 39–50). Anxiety may also explain why younger learners are generally more successful in language learning at school. Studies have shown that anxiety is negatively related to L2 attainment (e.g. Horwitz, Horwitz, and Cope 1986). Individual features, such as introvertedness or low self-confidence may also work as a barrier in learning an L2 (Twyford 1987, 5; Dewaele and Furnham 1999).

### **3.3 Effects of Training**

This section examines the effects of perception and production training in adults and children in language learning contexts. The present study investigates children's ability to perceive and produce a non-native sound after production training (see Chapter 4), which has been shown to facilitate the learning of a novel sound category. This section attempts to provide a brief summary of the literature relating to perception and production training in adults and children.

Research has revealed that adult L2 learners can authentically produce new L2 sounds that have no counterpart in their native language and that auditory training has been successful in enhancing adult listeners' discrimination of non-native sounds. For example, Peltola et al. (2015) utilised a listen-and-repeat training protocol, similar to the present study (see Chapter 4), with an additional visual cue, either orthography or

transcription. Their study revealed that misleading orthographical information hindered production learning. Saloranta et al. (2015) used the same stimulus and training setting with additional production instructions. Their study revealed direct change in production. Logan, Lively and Pisoni (1991) have also trained native Japanese adults to identify English /r/ – /l/ contrast and Jamieson and Morosan (1986) trained French adults to identify English /θ/ – /ð/ contrast. These studies have shown that training has improved the discrimination of non-native speech sounds. A similar effect was also found with adults in a study using behavioural tests and mismatch negativity (MMN, an electrophysiological response to a stimulus): three-day perceptual training resulted in behavioural and neural changes, indicating that a new memory trace was formed for the non-native sound (Tamminen et al. 2015). The effect of training on suprasegmental level has also been demonstrated by Wang et al. (1999) who trained American adults to discriminate Mandarin tones and found out that the learners' ability to identify the tones improved significantly after training.

The effect of training has also been studied with children. Rvachew (1994), for instance, studied preschoolers with phonological impairment and found out that the participants learned to articulate the target sound better when traditional speech therapy was accompanied with additional perception training. Perceptual reorganisation has been shown in children in training setting (Giannakopoulou, Uther & Ylinen 2013) as well as in early language immersion setting (Peltola et al. 2005). Listen-and-repeat training has also been shown to alter the production patterns of children in a study with comparable research setting to the present study (Taimi et al. 2014). The results of Taimi et al. are further discussed in Chapter 5. Wang and Kuhl (2003) have also extended the training procedure used in Wang et al. (1999) to the examination of Mandarin tone identification by children.

This chapter has discussed children as language learners and examined the affect of age in L2 learning. Age is a significant factor in L2 learning but not the only meaningful one. The idea of a sensitive period has been extended to L2 learning, as children generally tend to outperform adults in L2 attainment, especially in relation to pronunciation. The process of learning a foreign language is affected by several individual factors, such as learning environment, motivation, anxiety and attitudes toward to language being learned as well

as the language community. The following chapter is concerned with the procedures and methods, as well as the principal findings of the present study.

## **4. The Present Study**

The previous chapters have addressed theories and research on L2 learning and children, focusing especially on the view that young children are better language learners than adults. This chapter presents a study that is based on the theoretical dimensions discussed in the previous chapters. As was discussed in Chapter 1, the present study aims to find out how production training affects children's ability to perceive and produce a non-native vowel sound when the learning situation is in theory as challenging as possible, as well as how quickly the children learn to produce the new vowel with the chosen production training method. This chapter is concerned with the framework of the study. It is divided into five parts that address the data, methodology, as well as the results of the study. First, section 4.1 presents the participants of the study, after which section 4.2 discusses the stimuli. Next, section 4.3 gives an overview of the procedure and the testing situation itself. Section 4.4 describes the analysis process and how the analysis was performed. Finally, section 4.5 presents the results of the study.

### **4.1 Participants**

Altogether 12 preschool girls volunteered to participate in the study. The mean age of the participants was 6;4 years (range 6;0 – 7;4). All participants had only lived in Finland and spoke Finnish as their first language. None of them reported using any other languages at home or having knowledge of other languages than Finnish, apart from knowing few words in English or Swedish. One participant reported having participated in English language immersion. However, this was not meaningful for the study, as having no prior knowledge of Swedish was the main criteria for participants. All participants reported having normal hearing, though three of them reported having problems with the production of either /r/, /l/ or /s/. These participants were not excluded

from the analysis as the stimuli do not contain these particular sounds and thus problems in their production would not affect the study. Information regarding the participants' language skills as well as other background were gathered with a questionnaire (see Appendix 1). Written consent was obtained from all participants and their parents prior to testing.

All participants were enrolled in the Finnish pre-primary education system. Participating in pre-primary education is compulsory in Finland and the education begins when the child turns six years of age (Finnish National Agency for Education 2018a). All the participants in this study are girls for this study to be comparable to Taimi et al. (2014). However, the subject group represents a sample of Finnish-speaking monolingual preschool-aged children with very little or no language deficits. The study was conducted with the permission of the Ethics Committee of the University of Turku.

## 4.2 Stimuli

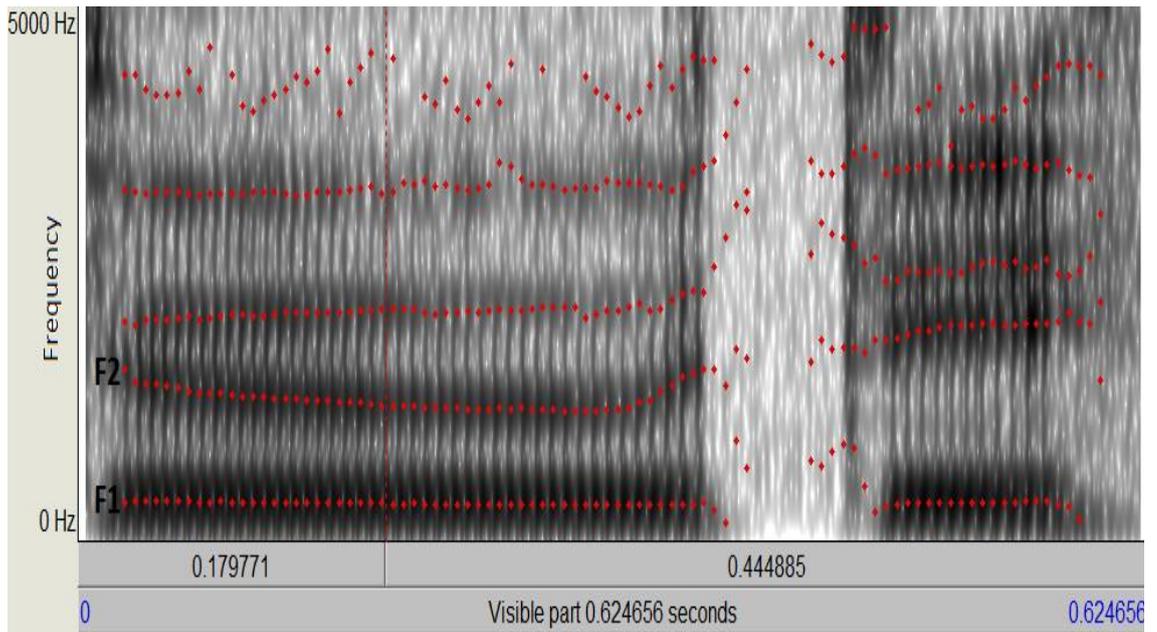
The stimuli used in this study consisted of two semi-synthetic pseudowords /tʉ:ti/ and /ty:ti/ in which the Swedish close central rounded vowel /ʉ/ and the close front rounded vowel /y/ acted as the target and the non-target vowels, respectively. The process of generating the stimuli together with the technical details of the stimuli are presented after a short overview of the vowels /ʉ/ and /y/ and their relation to the Finnish vowel system.

Both of the vowels /ʉ/ and /y/ are close vowels, that is, they are produced with the tongue positioned close to the roof of the mouth. The main difference between these vowels is the degree of backness (the tongue's position relative to the back of the mouth) and, to a lesser extent, lip rounding. The close rounded vowel space in Finnish is only divided into two categories, the close front rounded vowel /y/ and the close back rounded vowel /ʉ/. In Swedish, however, the same space is divided between three vowels: /y/, /ʉ/ and /u/. This can be a cause of learning difficulties, as the vowel /ʉ/ is acoustically located between the Finnish /y/ and /u/ and is thus, according to SLM (Flege 1987) and PAM (Best 1991), likely to be identified as a poor example of either /y/ or /u/ (see discussion in 2.1 and 2.2). Since the close central rounded vowel /ʉ/ is not phonologically relevant

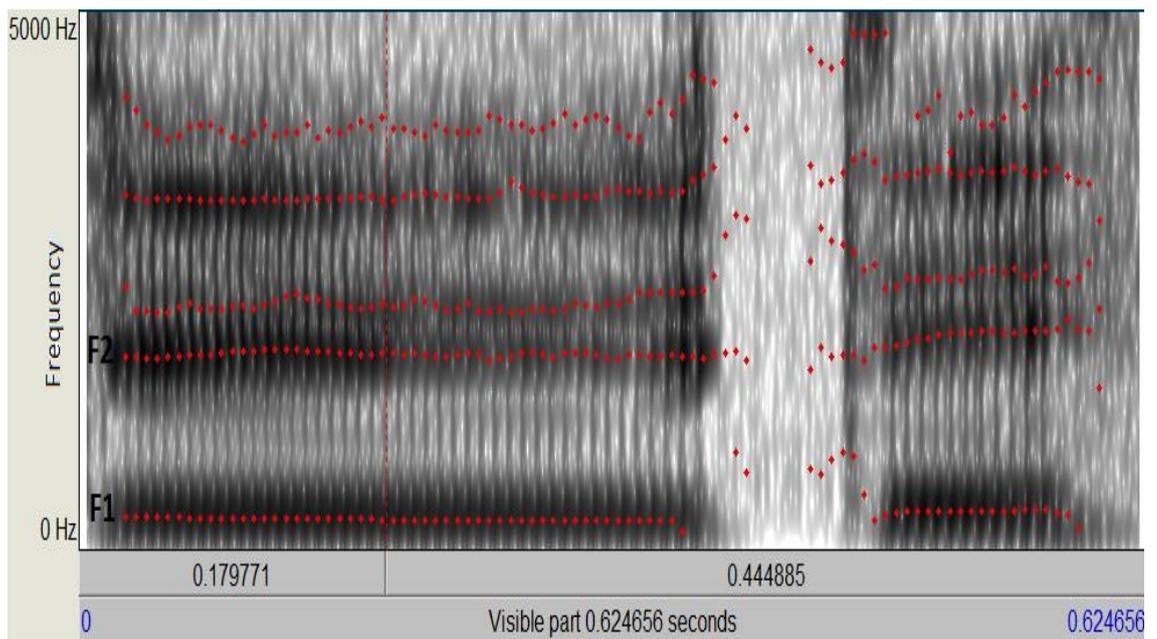
in Finnish but the close front rounded vowel /y/ is, /tʉ:ti/ was chosen as the target word and /ty:ti/ as the non-target word. This would ensure a theoretically very challenging learning situation for the Finnish monolingual participants.

The stimuli for this study were created with a Semisynthetic Speech Generation method (SSG). The SSG is based on a glottal pulse extracted from a real speech signal and an artificial process that models the vocal tract filtering effect (Alku, Tiitinen, & Nääätänen 1999, 1332). This allows the adjusting of formants (resonant frequencies of the vocal tract) without losing the naturalness of the speech signal (*ibid.*). For a detailed analysis of the speech generation method, see Alku et al. (1999).

Using the SSG method, both stimuli were synthesised to have the same prosodic and acoustic features, differing only by the formant frequencies of the first vowel. The second formant (F2) is the main factor in differentiating between /y/ and /ʉ/. As was mentioned above, the main difference between these vowels is in the backness: /y/ is a front vowel and /ʉ/ is a central vowel. The values of the first formant (F1) are connected to the open–close dimension, whereas F2 values are connected to the front–back dimension. Since both /y/ and /ʉ/ are close vowels differing in their backness, their F1 values should be very similar, whereas their F2 values should differ. The backness of a vowel lowers the frequency of the second formant, as can be seen in Figures 1 and 2 which present spectrogram images of the stimuli. Spectrograms are visual representations that show the frequencies of a sound, as well as their intensity, as they vary with time.



**Figure 1.** Spectrogram of the stimulus word /tʌ:ti/.



**Figure 2.** Spectrogram of the stimulus word /ty:ti/.

In the stimuli, the formant values for /ʌ/ were  $F1 = 338$  Hz,  $F2 = 1258$  Hz,  $F3 = 2177$  Hz, and for /y/  $F1 = 269$  Hz,  $F2 = 1866$  Hz and  $F3 = 2518$  Hz, the main difference being in the  $F2$  values.  $F1$  and  $F2$  are indicated in the images. As can be seen from Figures 1 and 2, the relative distance between  $F1$  and  $F2$  is smaller in Figure 1 (/tʌ:ti/) than in Figure 2 (/ty:ti/). The formant values in the stimuli are typical for a male adult speaker. Both

stimuli were 624 milliseconds long in duration and had the same fundamental frequency of 126 Hz. The inter-stimulus-interval (ISI) was three seconds. For a more detailed inspection of the stimuli and their creation, see Taimi et al. (2014).

### 4.3 Procedure

The procedure of this study was a two-day listen-and-repeat training exercise. The participants participated in the study on two consecutive days during which recording sessions and training sessions alternated. The tests were performed during the preschool day in the premises of a day care centre, in as quiet a room as possible. The participants were told that they would hear different kinds words, and, in both sessions, they were instructed to repeat what they had heard after the model. The participants were instructed that they could take breaks between the sessions should they need to or discontinue the study if they wished to. The tests were carried out in June 2017 and April 2018 as a part of another project.

The first day began with a baseline recording which was followed by a training session, another recording and another training session. The second day also consisted of four sessions in total but, in turn, began with a training session and ended with a recording session. Altogether the study consisted of four training sessions and four recordings. This pretest-posttest design enables the comparison of later productions to the baseline production. With this comparison, it is possible to discover what effect the training has on the production. In both the recording and training sessions, the stimuli were presented and registered with Beyerdynamic MMX 300 headphone-microphone headset with Asus Xonar U3 portable soundcard and Sanako Study Recorder software, version 8.22.0.0. A portable laboratory consisting of a laptop computer was used to collect the data.

The training block consisted of 30 pseudowords /tʌ:ti/ and 30 pseudowords /ty:ti/, altogether 60 stimuli, whereas the recording block consisted of 10 pseudowords /tʌ:ti/ and 10 pseudowords /ty:ti/, altogether 20 stimuli. The stimulus words /tʌ:ti/ and /ty:ti/ were presented in an alternating fashion. Thus, in total, the recording sessions and the training sessions contained 320 words to be repeated by each participant. The experiment lasted for 15-20 minutes per day per participant.

Prior to the first recording session on the first day as well as the first training session on the second day, the participants were presented with a familiarisation session consisting of six words: three target words and three non-target words that were repeated one after another. The familiarisation session was to provide a situation for setting the volume to a comfortable level as well as to help the participants to get adjusted to the task.

#### **4.4 Analysis**

The obtained acoustic data produced by the participants in the recording sessions were analysed with Praat software (Boersma & Weenik 2018), version 6.0.22. From each recorded word, the frequencies of the first, second and third formant, as well as the fundamental frequency (F0) were measured from the steady-state phase of the first vowel using the Linear Predictive Coding (LPC) Burg algorithm of the Praat software. The temporal mid-point of the vowel was visually located from the Praat spectrogram and the formant frequencies suggested by Praat were noted down. Mean values for the formant measurements in every session were then calculated for each participant for both the target and the non-target vowel. Similarly, standard deviations for the measurements in every session were also calculated. F0 values were obtained and analysed to ensure that there would not be any outliers distorting the formant results. F0 values, however, were not statistically analysed.

Due to technical problems, the productions of one participant had to be excluded from the analysis completely. If Praat's suggestions for formant frequencies did not seem to agree with the underlying spectrogram, the formant values for the word in question were excluded from further analysis. Two such words were excluded. Furthermore, altogether 11 words from four participants were excluded from the analysis due to non-technical reasons, such as a participant not repeating the word.

To find out whether the productions of the target and the non-target vowels change due to training, both mean values and standard deviations of F1 and F2 were statistically analysed with IBM SPSS Statistics, version 25.0.0.1, using Repeated Measures Analysis of Variance (ANOVA). Further analysis was conducted with paired samples T-tests. The statistical analysis focused on F1 and F2 values, as these two values are alone sufficient

for distinguishing vowel categories. Moreover, special attention was paid to the F2 values, since the second formant is the main factor in differentiating between /y/ and /ʉ/, as was discussed in 4.2. The next section presents the results of the analysis.

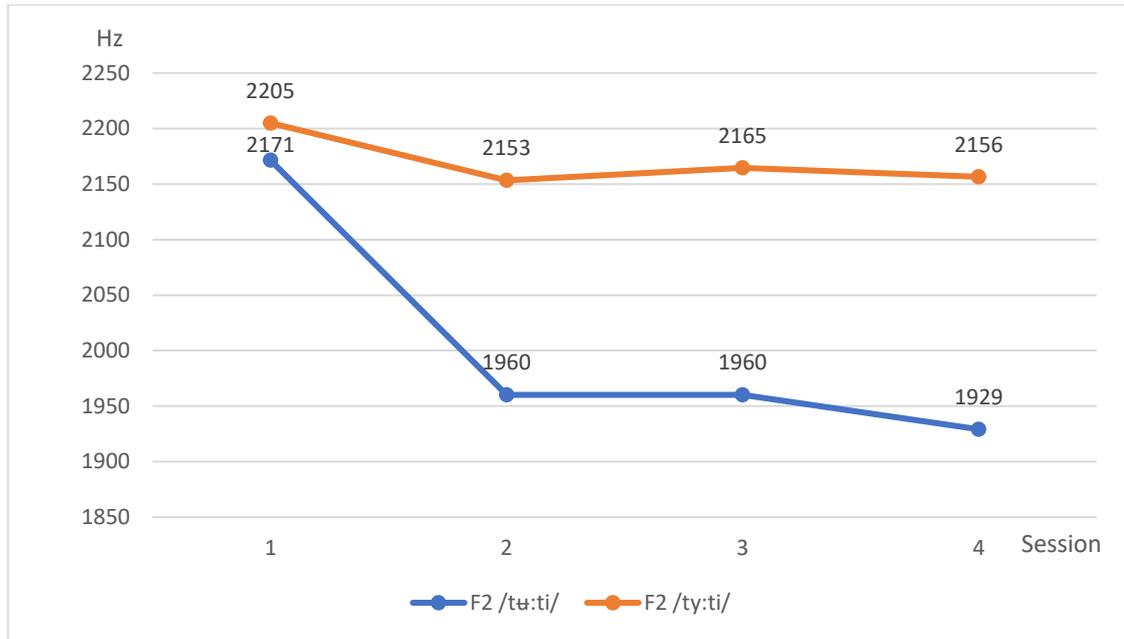
## 4.5 Results

The analysis began with measuring the average frequencies and standard deviations from the first two formants in the target vowel /ʉ/ and the non-target vowel /y/ that were embedded in the pseudowords /tʉ:ti/ and /ty:ti/, as was discussed in the previous section. The average frequencies of F1 and F2 for /tʉ:ti/ and /ty:ti/ in each session are presented in Table 1 below.

		Session 1	Session 2	Session 3	Session 4
/tʉ:ti/	F1	479	486	473	478
	F2	2171	1960	1960	1929
/ty:ti/	F1	478	484	466	474
	F2	2205	2153	2165	2156

**Table 1.** Average frequencies (Hz) of F1 and F2 in four sessions.

As can be seen from Table 1, the lowest average value of F1 in /tʉ:ti/ (473 Hz) was measured in the third session, whereas the highest value (486 Hz) was measured in the second session. The lowest average value of F2 in /tʉ:ti/ (1929 Hz), on the other hand, was measured in session four, whereas the highest value (2171 Hz) was measured in the first session. As for /ty:ti/, the highest average value of F1 (484 Hz) was measured in the second session and the lowest value (466 Hz) in the third session. The highest average value of F2 (2205 Hz) was measured in the first session and the lowest value (2153 Hz) in the second session. Overall, the average value of F1 seems to stay relatively stable, the variation between the values being only dozens of hertz. The F2 values, on the other hand, seem to vary more between the sessions. Figure 3 below compares the F2 values of /ʉ/ and /y/ in each session.



**Figure 3.** Average frequencies of F2 values (Hz) in four sessions.

As Figure 3 shows, the F2 values of /tʌ:ti/ seem to decrease after the first session, whereas the F2 values of /ty:ti/ seem to stay relatively stable. The results of the acoustical analysis and the decrease in the F2 values seen in Figure 3 lead to expect that there would be a statistically significant difference between the F2 values of /tʌ:ti/ and /ty:ti/ in session 2. This preliminary result was further investigated with statistical analyses which are reported further below.

		Session 1	Session 2	Session 3	Session 4
/tʌ:ti/	F1	47	55	70	70
	F2	211	323	290	264
/ty:ti/	F1	46	57	71	70
	F2	204	256	233	177

**Table 2.** Standard deviations of F1 and F2 in four sessions.

Standard deviations (SD) were calculated for each subject's productions. An average value of the participants' standard deviation values was then calculated for both words in each session. Table 2 above presents these standard deviations of F1 and F2 in both words in each session. As can be seen from the table, the values vary greatly. The lowest standard deviation of F1 in /tʌ:ti/ can be seen in session 1 (SD = 47), whereas the highest

value is the same in sessions 3 and 4 ( $SD = 70$ ). The lowest value of F2 can be seen in session 1 ( $SD = 211$ ) and the highest in session 2 ( $SD = 323$ ). As for /ty:ti/, the lowest standard deviation of F1 can be found in session 1 ( $SD = 46$ ) and the highest in session 3 ( $SD = 71$ ). The lowest value of F2, on the other hand, can be found in session 4 ( $SD = 177$ ), and the highest in session 2 ( $SD = 256$ ). After these acoustical analyses, the formant frequencies and standard deviations were analysed statistically.

The statistical analysis began with a Word (2) x Session (4) x Formant (2) Repeated Measures ANOVA for the formant frequencies in order to find out whether there were any differences in the productions of the target and the non-target vowels during the experiment, as was mentioned in 4.4. The cut-off value for statistical analysis was 0.05, only statistically significant results are reported here. A significant main effect of Word ( $F(1, 10) = 6.133, p = 0.033$ ) was found, meaning that the acoustical qualities of the words differed significantly. A significant main effect of Formant ( $F(1, 10) = 798.526, p < 0.001$ ) was also found, which indicates that the formants F1 and F2 were produced differently on the whole. Additionally, a significant Word x Formant interaction ( $F(1, 10) = 6.460, p = 0.029$ ) was found. This indicates that between the words, there was a difference in the formant values and thus the sessions were analysed separately to determine in which session the difference was significant. The effect for Session was not significant. Overall, the primary analysis revealed that the participants produced a difference between the target and the non-target vowels embedded in the pseudowords /tu:ti/ and /ty:ti/. Based on these results, each session was analysed separately with a Word (2) x Formant (2) Repeated Measures ANOVA in order to compare the formant values between the words in each session.

When inspecting each session individually, a significant main effect of Formant was found in all of the sessions (session 1:  $F(1, 10) = 738.270, p < 0.001$ ; session 2:  $F(1, 10) = 505.917, p < 0.001$ ; session 3:  $F(1, 10) = 600.426, p < 0.001$ ; session 4:  $F(1, 10) = 613.726, p < 0.001$ ), which indicates that the F1 and F2 values differed from each other, as was expected. Sessions 3 and 4 provided more meaningful results, as a significant Word x Formant interaction was found in sessions 3 ( $F(1, 10) = 5.133, p = 0.047$ ) and 4 ( $F(1, 10) = 9.520, p = 0.012$ ), meaning that in sessions 3 and 4 one or both formants differed significantly between the words. A significant main effect of Word ( $F(1, 10) = 10.048, p = 0.010$ ) was found in session 4, which means that the productions of the words

differed significantly from each other overall. Based on these results, the formant values were then analysed separately in each session with paired samples T-tests.

The T-tests analysed both formants individually. The F2 values of /tʌ:ti/ and /ty:ti/ were compared against each other in each session. F1 values were analysed in a similar manner. F1 values did not differ significantly between the words in any of the sessions, but F2 values differed significantly in session 4. As was discussed in section 4.2, it was expected that no difference would be shown in the F1 values, whereas the F2 values would differ if the participants had learned to perceive and produce the difference between the stimulus words. The F2 values between the words did not differ significantly in sessions 1, 2 or 3, but there was a significant difference in the fourth session ( $t(10) = -3.127, p = 0.011$ ). In session 3, however, the difference was not statistically significant but showed a tendency towards it ( $t(10) = -2.158, p = 0.056$ ).

The standard deviations were analysed similarly to the average frequencies in order to find more about the consistency of the participants' productions. The analysis began with a Word (2) x Session (4) x Formant (2) Repeated Measures ANOVA for standard deviations. A significant main effect of Formant was found ( $F(1, 10) = 59.853, p < 0.001$ ) which indicates that the standard deviations of F1 and F2 differed from each other, which was to be expected. A significant Word x Formant interaction ( $F(1, 10) = 5.089, p = 0.048$ ) was also found, which indicates that the standard deviations differ from each other at least in one session. To find out in which session the formant values differed from each other, all the sessions were inspected separately with a Word (2) x Formant (2) Repeated Measures ANOVA. The results show that no significant effect was found in session 1, 3 or 4. Session 2, however, revealed a significant main effect of Word ( $F(1, 10) = 7.166, p = 0.023$ ) and Word x Formant interaction ( $F(1, 10) = 11.562, p = 0.007$ ). Thus, the words differ from each other with respect to the standard deviations of F1 and F2 in session 2. To find out which formant's standard deviation differed significantly between the words, the formants were compared in the second session with paired samples T-tests. The difference between the words resulted from a difference in F2 values ( $t(10) = 3.035, p = 0.013$ ), which was expected as the second formant was the main differentiating factor between the target vowel /ʌ/ and the non-target vowel /y/, as discussed earlier.

Overall, the productions of the target and the non-target vowels were not significantly different in the first three sessions, meaning that the formant values were very similar in

terms of their frequencies. In the fourth session, however, the words were produced differently enough for the distinction to be statistically significant, the difference being in the F2 values. The analysis of the standard deviations revealed that there was a significant difference between the standard deviations of /u/ and /y/ in the second session, indicating that the participants have been unsure in their articulation, which is reflected in the vowel's backness.

This chapter has discussed the practical dimensions of the study, that is, the participants and methodology, as well as presented the results of the study. The results of the study are next discussed in a broader view and in relation to other studies in Chapter 5.

## 5. Discussion

The present study investigated children's ability to learn to produce the Swedish vowel contrast /y/ – /u/. The statistical analysis presented in 4.5 revealed that the participants, 12 preschool girls, learned to produce the foreign vowel /u/ in two days. Additionally, the statistical analysis of standard deviations revealed that the listen-and-repeat training had an effect on the participants' productions in a way that the standard deviation grew between sessions 1 and 2. This chapter reflects on the results, discussing possible reasons behind them, as well as their implications on early language training.

As discussed in Chapter 4, the present study aimed to find out how listen-and-repeat training affects children's ability to perceive and produce a non-native vowel in a linguistically challenging situation. Based on the research and theories introduced in Chapters 2 and 3, it was hypothesised that listen-and-repeat training would affect the children's ability to produce the vowel contrast /y/ – /u/. As was hypothesised, the results revealed that children learned to produce the target and the non-target vowels differently. For the children to learn to produce the vowels differently, they would first have to learn to perceive the difference between the sounds, as was discussed in Chapter 1. The results indicate that the participants had indeed perceived the distinction and learned to produce the non-native vowel.

According to the SLA models discussed in Chapter 2, the target vowel /u/ and the vowel contrast /y/ – /u/ represent the most difficult learning situation for Finnish language

learners. However, the vowel contrast could be any other one that would provide an equally difficult learning situation. To interpret the findings in a broader perspective, the vowel contrast in the present study should be considered to be a representative of a linguistically difficult learning situation in general, rather than a single vowel pair. Additionally, the participants were faced with another challenging linguistic situation as the stimuli were based on a male speaker's vocal tract model.

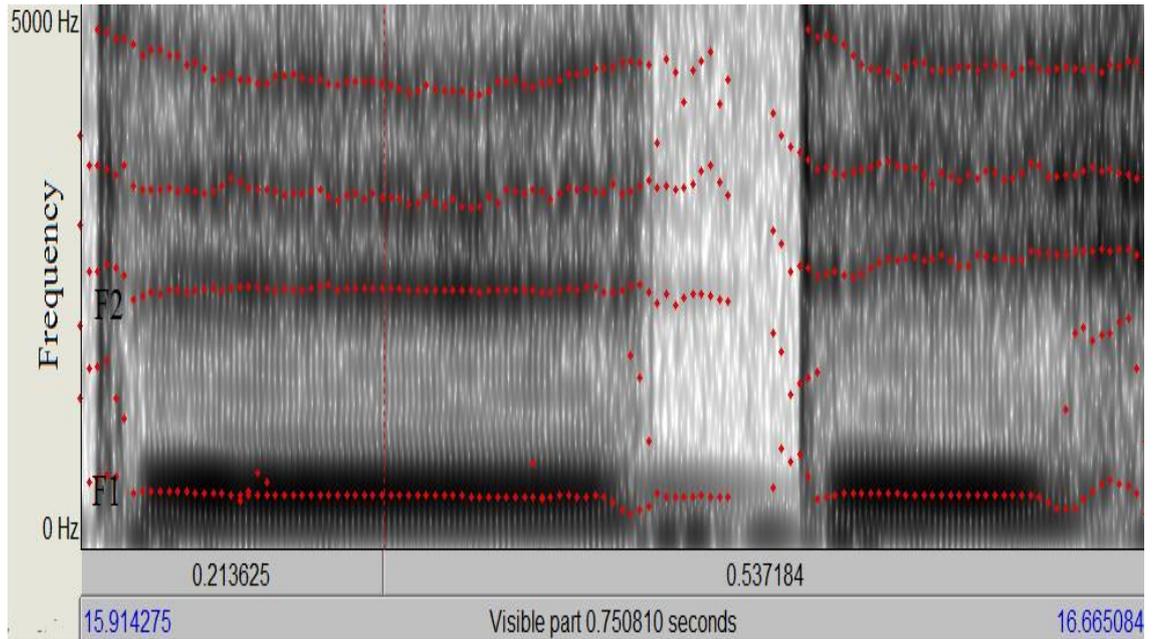
Male and female formant frequencies differ from each other due to anatomical constraints of the vocal tract, the male voice being typically much lower in nature than the female voice due to descending larynx in males during puberty (Rosner and Pickering 1994, 66–67). In the present study, the formant frequencies of the semisynthetic stimuli were typical for an adult male speaker (F1 = 338 Hz, F2 = 1258 Hz for /ʌ/ and F1 = 269 Hz, F2 = 1866 Hz for /y/), as was mentioned in 4.2. The fundamental frequency for both of the stimuli was 126 Hz. The fundamental frequency affects the proportional frequency of the higher formants, due to which children are not generally able to produce formants as low as adult males, as their F0 is typically much higher. The average F0 of the participants was 250 Hz, whereas the average F1 and F2 values were 477 Hz and 2087 Hz, respectively. Although the model the children heard differed from their own production, the children were able to learn to produce the non-native vowel /ʌ/ according to the model. It is common for the formant frequencies to vary across speakers, yet listeners are able to recognise spoken words despite this variation. This *speaker normalisation* is essential in speech perception. Listeners seem to compensate effortlessly and automatically for variation in the productions of different speakers already in early infancy (Rosner and Pickering 1994, 218).

The first recording of the first day was used as a baseline recording, which the later recordings could be compared to. The acoustical analysis showed that the participants produced the target and the non-target vowels very similarly in the first session (see Table 1), the difference between /ʌ/ and /y/ being only 34 Hz in the F2 values (/tʌ:ti/ F2 = 2171, /ty:ti/ F2 = 2205). As was discussed in 4.5, the initial acoustical analysis led to expect a statistically significant difference already in session 2, where the F2 values of /ʌ/ and /y/ differed by nearly 200 Hz (/tʌ:ti/ F2 = 1960, /ty:ti/ F2 = 2153, see Table 1 and Figure 3). This hypothesis, however, was not met as the statistical analysis revealed that the productions of the target vowel /ʌ/ and the non-target vowel /y/ differed significantly in

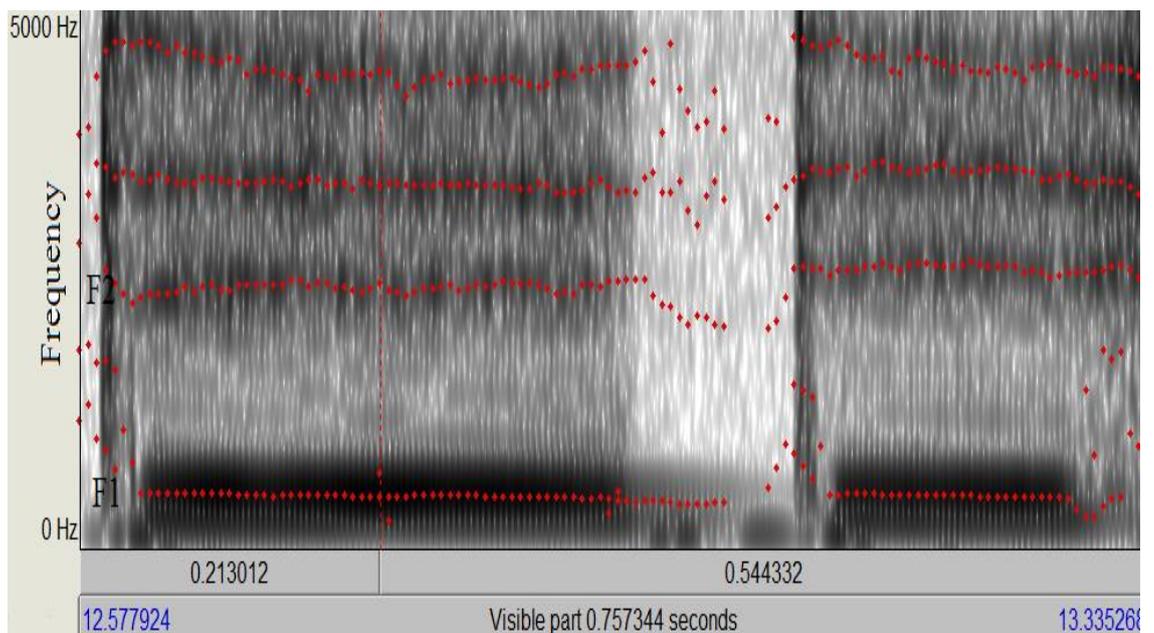
terms of F2 values only in session 4, where the difference between the F2 values was 227 Hz (/tʌ:ti/ F2 = 1929, /ty:ti/ F2 = 2156).

The inspection of individual standard deviations, in turn, revealed that there was a significant difference between the standard deviations of F2 values in session 2, suggesting that the participants became unsure of their productions and thus the mean SD grew in session 2. This would indicate that individual participants would have become more unsystematic in their own productions in session 2, producing sometimes vowels resembling /ʌ/ and sometimes /y/. No statistical significance was found in sessions 3 or 4, which indicates that the productions became more systematic at individual level. In a group this small, the standard deviation is a great factor affecting the results.

Recent research has emphasized individual differences in speech sound perception studies. For instance, Iverson et al. (2008) studied adults with different language backgrounds and their individual differences in learning a non-native sound distinction. Golestani and Zatorre (2009) also aimed to characterise individual differences of learners in a study of adult English speakers who trained to distinguish the Hindi dental–retroflex contrast. Individual differences in English vowel learning have also been studied with regard to the subjects' native vowel processing ability (Lengeris and Hazan 2010). Although inspecting individual participants does not provide generalisable information on L2 learning by children, it can give an insight into the results and help to understand them. Figures 4, 5 and 6 below present one participant's productions of /tʌ:ti/ and /ty:ti/ from two different sessions.

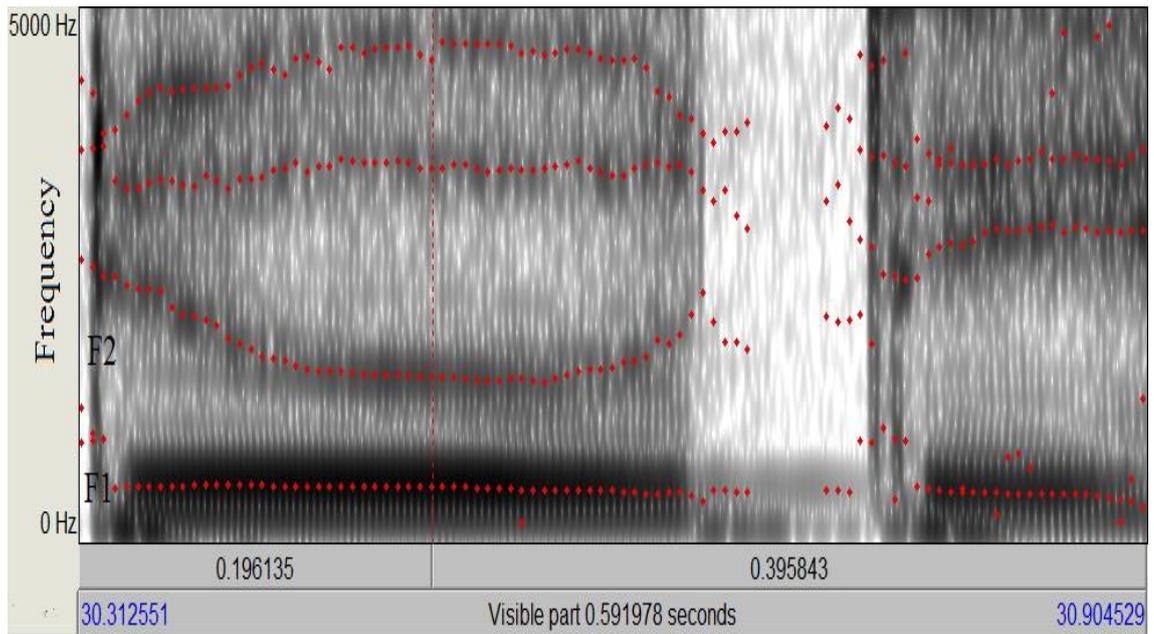


**Figure 4.** The third pseudoword /tʌ:ti/ from session 1 produced by a participant.



**Figure 5.** The second pseudoword /ty:ti/ from session 1 produced by the same participant.

Figure 4 is a spectrogram of the third pseudoword /tʌ:ti/ from the first session, whereas Figure 5 is a spectrogram of the second /ty:ti/ from the same session. As can be seen from Figures 4 and 5, the participant has produced the pseudoword /tʌ:ti/ in session 1 according to their L1. The production resembles very much the word /ty:ti/, as the F2 value is relatively high (2371 Hz in Figure 4 and 2369 Hz in Figure 5).



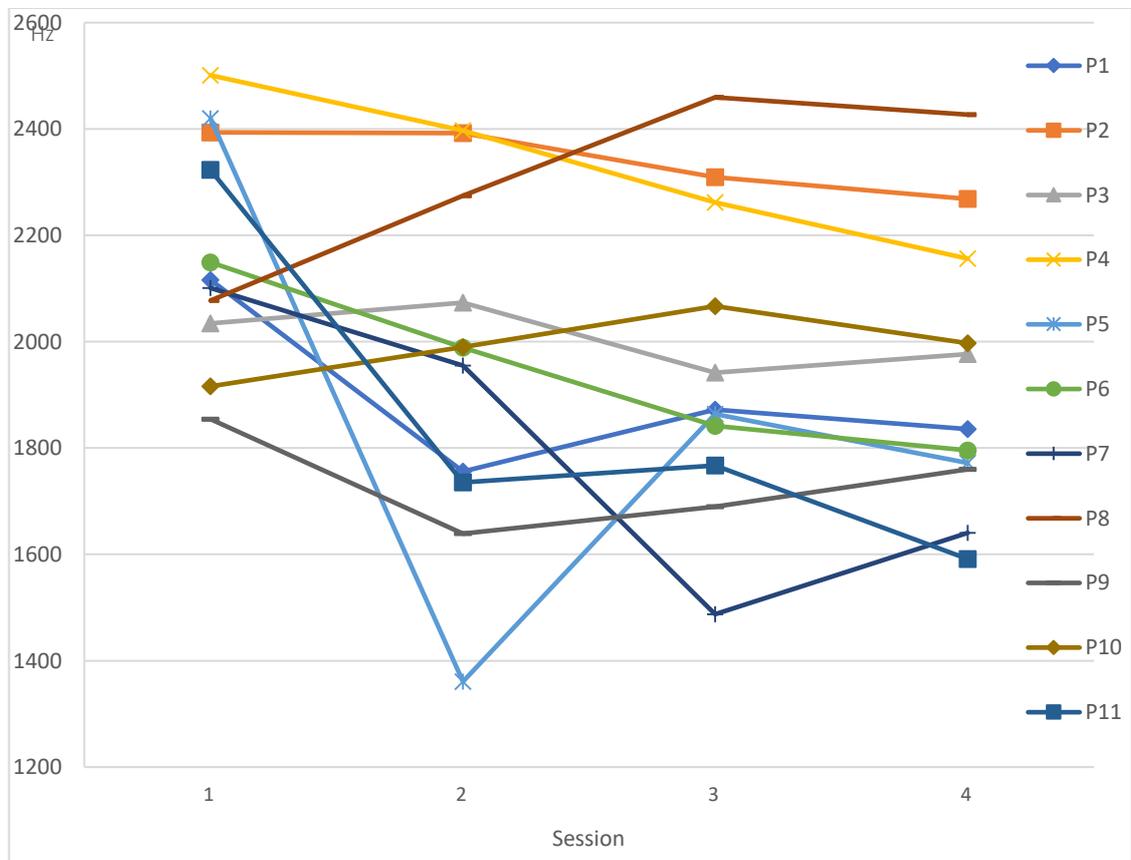
**Figure 6.** The fifth pseudoword /tʌ:ti/ from session 3 produced by the same participant.

As can be seen in Figure 6, the production of /tʌ:ti/ by the same participant has changed in the third session, as the second formant has lowered significantly (1550 Hz). When compared against the model word /tʌ:ti/ that was used as stimulus (see Figure 1), the production presented in Figure 6 closely resembles the model. The three spectrograms presented above illustrate the learning process at individual level. The participant in question seems to have learned to perceive the difference between /ʌ/ and /y/ as well as to realize it in their own production already in session 3. The changes in one individual's production are not significant in terms of group results but provide some insight into the possible reasons behind the results. The productions of individual participants are further inspected below with the help of Table 3 that shows the statistical information of the F2 values at group level and Figure 7 that presents the average F2 values of /ʌ/ in /tʌ:ti/ by every participant (P1–11) in each session (1–4).

	Minimum (Hz)	Maximum (Hz)	Mean	SD
Session 1	1854	2501	2171	211
Session 2	1361	2397	1960	323
Session 3	1488	2460	1960	290
Session 4	1591	2427	1929	264

**Table 3.** Minimum, maximum, mean and standard deviation of F2 (/ʌ/) at group level.

Table 3 above presents the minimum, maximum, mean and standard deviation values at group level in each session for the target vowel /ʌ/. Table 3 shows that the SD value increases in session 2. This could mean that, at group level, the participants did not perceive the difference between the stimuli words /tʌ:ti/ and /ty:ti/ in the first session, or at least the difference did not show in their productions. In the second session however, the productions of the participants become more unsystematic as a whole, which can be seen as an increase in the SD value. This could indicate that some individuals may have produced a difference between the stimuli words, while others may have not, making the productions at group level very unsystematic. Some participants learned to produce the difference much sooner than others, but the group result may be greatly affected by some participants whose productions were very unsystematic and thus raised the standard deviation. Table 3 also shows that the maximum values seem to stay stable, but the minimum values vary between the sessions. This is further demonstrated in Figure 7 which shows the minimum and maximum values in each session by indicating all of the average F2 values of /ʌ/ by each participant in each session.



**Figure 7.** Average F2 values of /ʌ/ in each session by each participant.

As can be seen from Figure 7, the productions of the participants are very varied. Some individuals seem to have learned to differentiate /y/ and /ʉ/ in their productions, based on the gradually lowering values of F2 in /tʉ:ti/. For example, the F2 values of participants 4 (S1 = 2501 Hz), 7 (S1 = 2100 Hz) and 11 (S1 = 2322 Hz) have each lowered several hundred hertz by the final session (P4 = 2156 Hz, P7 = 1640 Hz, P11 = 1591 Hz). The productions of some participants, on the other hand, seem to remain relatively stable, which suggests that they have not learned to discriminate /ʉ/ from /y/, or at least it does not show in their productions. For instance, the difference between F2 values in session 1 and session 4 for participants 3 and 10 is 58 Hz and 81 Hz, respectively. The F2 value of participant 10 is in fact higher in session 4 than session 1, as is the F2 value of participant 8, whose F2 value has risen for over 150 Hz by the final session.

Figure 7 reveals that at individual level, it is evident that some individuals perform better than others. Due to the small sample size, individual differences become important in determining the overall group result. In the present study, the participant group learned to produce a distinction between /tʉ:ti/ and /ty:ti/ in the fourth session, even though at individual level some participants did not seem to learn the difference at all and others seemed to have learned it already in session 3, as Figures 4–7 suggest.

Earlier studies have shown that perceptual training can facilitate the learning of a new sound category, as discussed in 3.3. The results of the present study are in line with those of previous studies. In a similar research setting to the present study, Taimi et al. (2014) also found out that children can learn to produce a non-native vowel in two days. Their study was identical to the present study in terms of the stimuli, /tʉ:ti/ – /ty:ti/, and the two-day testing procedure involving training and recording sessions. In contrast to the present study, however, Taimi et al. (2014) studied slightly older children as the participant group consisted of 7–10-year-old girls (n=13), as opposed to the 6–7-year-old girls (n=12) in the present study. The results of Taimi et al. (2014) revealed that the participants learned to produce the new non-native vowel /ʉ/ after only three sessions, whereas the results of the present study show that the younger participants learned to produce the new sound after four sessions. Possible explanations for and implications of this difference are considered below.

The results of the present study, as well as the results of Taimi et al. (2014), support the idea of children's plasticity and fast ability to acquire languages. However, considering

that the only difference between these studies is the age of participants, it could be concluded that age in particular is the factor affecting the results. The idea that earlier is better (see discussion in Chapter 3) is here contradicted by the results that 7–10-year-old children learned the non-native sound faster than younger, 6–7-year-old children. Children’s brains are plastic and capable of acquiring new language-related information quickly. Nevertheless, the possible benefits of younger age in terms of brain plasticity and adaptability may be overridden by the still developing cognitive abilities of younger children.

Even though preschool aged children would be more plastic, their cognitive abilities may not be developed enough so that language learning in this research setting would be more efficient. The inability to understand and concentrate on the task at hand may hinder the learning process, giving advantage to older children with more developed cognitive abilities. Additionally, individual differences, as was discussed earlier, may be a factor affecting the results. The individual differences of young, 6–7-year-old children may be very notable, and the results of the whole group may be affected by the results of those individuals who are too tired or otherwise unable to focus on the learning situation. Perhaps children 7–10 years old in the study conducted by Taimi et al. (2014) were as a group better able to concentrate on the task or to understand what was required of them.

The results of Taimi et al. (2014) and the present study are much related to the introduction of earlier language teaching in the Finnish education system. Language teaching in Finnish education usually begins in the third grade when students start to learn English. In addition to English, Finnish students start to learn the second national language Swedish (or Finnish for Swedish-speaking Finns) typically in the sixth grade. According to the Finnish National Agency for Education (*Opetushallitus*), these three languages comprise 80 % of Finnish pupils’ language studies (Opetushallitus 2019). To tackle the dominance of English and to provide an earlier start in language learning, the Finnish government formulated its Government Program for the years 2016–2019 in which the Government Key Project for Languages focuses on “integrating early language learning into Finnish education with a much wider scope than before, providing pupils with a wider language repertoire, and creating a welcoming and encouraging attitude towards language learning” (Finnish National Agency for Education 2017). In addition to starting language teaching in the first grade, the project aimed at introducing language

learning in day-care, even before starting the actual language syllabus (Finnish National Agency for Education” 2018b).

Based on the results of the study by Taimi et al. (2014) and the present study, learning new sounds with listen-and-repeat training seems to be easier for older children. This result provides preliminary evidence for the question whether language learning should start already in day-care. When judged by short-term learning results, introducing language learning in day-care does not seem to be more advantageous compared to introducing it in primary school. The Finnish National Agency for Education argues for the early introduction of language learning with brain plasticity and sensitivity periods (see also Chapter 3 of this study) and states that “these periods present *the most opportune* ages for specific linguistic features to be learned ... [and] are also strongly related to brain plasticity” (Finnish National Agency for Education 2017, emphasis in original). The Finnish National Agency for Education mentions “early childhood” as the best age to start language learning and, based on the above-mentioned Key Project, this early childhood takes place in the day-care years.

## 6. Conclusion

The purpose of this study was two-fold: first, to examine children as language learners in the light of theories and studies concerning second language learning and second, to determine how production training affects children’s ability to perceive and produce a non-native sound which was theoretically as challenging as possible.

In the light of SLA theories examined in Chapter 2, the difficulties learners face when learning an L2 are the same for all learners, as difficulties rise from the similarities and differences between the learner’s L1 and L2. However, age is an important factor in L2 learning, as discussed in Chapter 3. The SLM and PAM are widely used theories that do not take age much into account but focus instead on the differences between the learner’s L1 and L2 and the learning difficulties caused by them (see sections 2.1 and 2.2). The NLM takes age into account in stating that difficulties rise from the L1 speech sound prototypes that have formed in early infancy (see section 2.3). Due to this the NLM, however, cannot predict differences in the learning outcome of adults and children who

have already acquired their native language, as the difficulties that rise from the differences in L1 prototypes and L2 sounds are the same.

Nevertheless, children differ from adults as second language learners. This is often explained with a sensitive period for language learning during which the brain's plasticity is in its peak and the brain is more sensitive for new language information, as discussed in 3.1. The sensitive period for language learning has widely been studied but no consensus on its existence, or starting and ending ages, has been reached. The ultimate attainment of an L2 is by no means impossible after childhood and the sensitive period. L2 learning is affected by multiple factors other than age, such as learning environment, motivation and other personal factors, as was explained in 3.2. In general, however, it can be concluded that children are more successful in L2 learning than adults, especially in a natural environment and in terms of pronunciation.

The SLA theories provided a basis for the present study in the selection of stimuli, as these theories help in determining the most challenging sound contrasts to learn. When the foreign sound or sound contrast is maximally difficult to learn, a study utilising this sound contrast provides important information on how factors such as age affect the second language learning process. The stimuli used in the present study were two semi-synthetic pseudo-words /tæ:ti/ and /ty:ti/. The target-vowel /æ/ is phonemically irrelevant in Finnish but relevant in Swedish, which creates a difficult learning situation. The participant group consisted of 12 Finnish preschool girls who participated in the study on two consecutive days. Each day consisted of two training and two recording sessions. The recordings were acoustically analysed to find out the first two formant frequencies of the target and the non-target vowel. These formant frequencies were then statistically analysed, as well as their standard deviations.

The results showed that the children learned to produce the difficult sound contrast on the second day, after four training and recording sessions. This indicates that the participants learned to differentiate between the vowels /æ/ and /y/ and, after perceiving the difference, also to produce the stimuli words differently. The results support previous findings on children's fast ability to learn foreign sounds. The children also showcased their ability to learn language by listening and repeating only. The participants received no explicit information on how to produce the new sound but had to rely on their hearing and perception of the sounds in order to discriminate between the two stimulus words. If the

participants were told that they would be hearing two different words, they may have learned to produce the words differently earlier. This would require a further study with two participant groups that would receive different instructions prior to the test. To form a more comprehensive view on preschool-aged children's ability to perceive and produce a non-native vowel, further research could also include preschool-aged boys as well as girls.

## References

- Alku, P., Tiitinen, H., & Näätänen, R. 1999. "A Method for Generating Natural-Sounding Speech Stimuli for Cognitive Brain Research." *Clinical Neurophysiology* 110 (8): 1329–1333. [https://doi.org/10.1016/S1388-2457\(99\)00088-7](https://doi.org/10.1016/S1388-2457(99)00088-7).
- Best, C., McRoberts, G., and Sithole, N. 1988. "Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants." *Journal of Experimental Psychology: Human Perception and Performance* 14 (3): 345–60. <https://doi.org/10.1037/0096-1523.14.3.345>.
- Best, C., and Tyler, M. 2007. "Nonnative and Second-Language Speech Perception," In M. Munro & O.-S. Bohn (Eds.) *Language experience in second language speech learning* 13–34. Amsterdam: John Benjamins.
- Best, C. 1991. "The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model: A Window on Early Phonological Development." *Haskins Laboratories Status Report on Speech Research*, 1–30.
- . 1995. "A Direct Realist View of Cross-Language Speech Perception." In W. Strange (Ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* 171–204.
- Boersma, P & Weenink, D. 2018. Praat: doing phonetics by computer. Version 6.0.22, retrieved 20 November 2018 from <http://www.praat.org>.
- Borden, G. J., Harris, K. S. & Raphael, L. J. 2003 (4<sup>th</sup> ed.) *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*. Baltimore: Lippincott Williams & Wilkins.
- Bundgaard-Nielsen, R., Best, C., & Tyler, M. 2011. "Vocabulary Size Is Associated with Second-Language Vowel Perception Performance in Adult Learners." *Studies in Second Language Acquisition* 33 (3): 433–61. <https://doi.org/10.1017/S0272263111000040>.
- Chambers, G. 1998. "Pupils' Perceptions of the Foreign Language Learning Experience." *Language Teaching Research* 2 (3): 231–59. <https://doi.org/10.1177/136216889800200304>.
- Csizer, K., & Dörnyei, Z. 2005. "The Internal Structure of Language Learning Motivation and Its Relationship with Language Choice and Learning Effort." *The Modern Language Journal* 89 (1): 19–36. [www.jstor.org/stable/3588549](http://www.jstor.org/stable/3588549).
- Curtiss, S. 1977. *Genie : A Psycholinguistic Study of a Modern-Day "Wild Child."* New York: Academic Press, Inc.
- Dewaele, J.-M., & Furnham, A. 1999. "Extraversion: The Unloved Variable in Applied

- Linguistic Research.” *Language Learning* 49 (3): 509–44.  
<https://doi.org/10.1111/0023-8333.00098>.
- Dörnyei, Z. 1994. “Motivation and Motivating in the Foreign Language Classroom.” *The Modern Language Journal* 78 (3): 273–84. <https://www.jstor.org/stable/330107>.
- Dörnyei, Z., & Clément, R. 2001. “Motivational Characteristics of Learning Different Target Languages: Results of a Nationwide Survey.” In Z. Dörnyei R. Schmidt (Eds.) *Motivation and Second Language* 399–432. Honolulu: University of Hawaii.
- Finnish National Agency for Education. 2017. “Finland Invests in Early Language Learning.” Retrieved from [https://www.oph.fi/download/187642\\_Finland\\_invests\\_in\\_early\\_language\\_learning.pdf](https://www.oph.fi/download/187642_Finland_invests_in_early_language_learning.pdf).
- . 2018a. “Early childhood education and care.” Retrieved from [https://www.oph.fi/english/education\\_system/early\\_childhood\\_education](https://www.oph.fi/english/education_system/early_childhood_education).
- . 2018b. “Introducing an Earlier Start in Language Teaching : Language Learning to Start as Early as in Kindergarten.” Retrieved from [https://www.oph.fi/download/193013\\_Introducing\\_an\\_earlier\\_start\\_in\\_language\\_learning.pdf](https://www.oph.fi/download/193013_Introducing_an_earlier_start_in_language_learning.pdf)
- Flege, J. 1987. “The Production of " New " and " Similar " Phones in a Foreign Language : Evidence for the Effect of Equivalence Classification.” *Journal of Phonetics* 15 (1): 47–65.
- . 1988. “The Production and Perception of Foreign Language Speech Sounds.” In H. Winitz (Ed.) *Human Communication and Its Disorders: A Review* 224–401. Norwood: Ablex.
- . 1995. “Second-Language Speech Learning: Theory, Findings, and Problems.” In W. Strange (Ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 233–277. Timonium: York Press.
- . 1997. “English Vowel Production by Dutch Talkers: More Evidence for the "Similar" vs. "New" Distinction.” In A. James & J. Leather (Eds.) *Second-Language Speech, Structure and Process*, 11–52. Berlin: De Gruyter.
- . 2005. “Evidence for Plasticity in Studies Examining Second Language Speech Acquisition.” ISCA Workshop on Plasticity in Speech Perception 15–17 June 2005, London. Retrieved from [http://jimflege.com/files/London\\_June\\_2005.pdf](http://jimflege.com/files/London_June_2005.pdf).
- Flege, J, Frieda, E., & Nozawa, T. 1997. “Amount of native-language (L1) use affects the pronunciation of an L2.” *Journal of Phonetics* 25: 169–186.
- Fowler. C. 1986. “An Event Approach to the Study of Speech Perception from a Direct-Realist Perspective” *Journal of Phonetics* 14 (1): 3–28.
- Fowler, C, & Rosenblum, L. 1991. “The Perception of Phonetic Gestures.” *Haskins Laboratories Status Report on Speech Research* 99/100: 102–17.

- Giannakopoulou, A., Uther, M., & Ylinen, S. 2013. "Enhanced plasticity in spoken language acquisition for child learners: Evidence from phonetic training studies in child and adult learners of English." *Child Language Teaching and Therapy* 29 (2): 201–218. <https://doi.org/10.1177/0265659012467473>.
- Golestani, N., & Zatorre R. 2009. "Individual Differences in the Acquisition of Second Language Phonology." *Brain and Language* 109 (2–3): 55–67. <https://doi.org/10.1016/j.bandl.2008.01.005>.
- Guion, S., Flege, J., Akahane-Yamada, R., & Pruitt, J. 2000. "An Investigation of Current Models of Second Language Speech Perception: The Case of Japanese Adults' Perception of English Consonants." *The Journal of the Acoustical Society of America* 107 (5): 2711–24. <https://doi.org/10.1121/1.428657>.
- Horwitz, E., Horwitz, M., & Cope, J. 1986. "Foreign Language Classroom Anxiety." *The Modern Language Journal* 70 (2): 125–132. <https://www.jstor.org/stable/327317>.
- Hurford, R. 1991. "The Evolution of the Critical Period for Language-Acquisition." *Cognition* 40 (3): 159–201. [https://doi.org/10.1016/0010-0277\(91\)90024-X](https://doi.org/10.1016/0010-0277(91)90024-X).
- Iverson, P., & Ekanayake, D. 2008. "Category and Perceptual Interference in Second-Language Phoneme Learning: An Examination of English / w / - / v / Learning by Sinhala, German, and Dutch Speakers" *Journal of Experimental Psychology: Human Perception and Performance* 34 (5): 1305–16. <https://doi.org/10.1037/0096-1523.34.5.1305>.
- Jamieson, D., & Morosan, D. 1986. "Training Non-Native Speech Contrasts in Adults: Acquisition of the English /ð/-/θ/ Contrast by Francophones." *Perception & Psychophysics* 40 (4): 205–15. <https://doi.org/10.3758/BF03211500>.
- Johnson, J., & Newport E. 1989. "Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language." *Cognitive Psychology* 21: 60–99. [https://doi.org/10.1016/0010-0285\(89\)90003-0](https://doi.org/10.1016/0010-0285(89)90003-0).
- Krashen, S. 1982. *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press Inc.
- Kuhl, P. 1991. "Human Adults and Human Infants Show a 'Perceptual Magnet Effect' for the Prototypes of Speech Categories, Monkeys Do Not." *Perception & Psychophysics* 50 (2): 93–107. <https://doi.org/10.3758/BF03212211>.
- . 1994. "Learning and Representation in Speech and Language." *Current Opinion in Neurobiology* 4 (6): 812–22. [https://doi.org/10.1016/0959-4388\(94\)90128-7](https://doi.org/10.1016/0959-4388(94)90128-7).
- Kuhl, P., Coffey-Corina, S., Padden, D., & Dawson, G. 2005. "Links between Social and Linguistic Processing of Speech in Preschool Children with Autism: Behavioral and Electrophysiological Measures." *Developmental Science* 8 (1): F1–12. <https://doi.org/10.1111/j.1467-7687.2004.00384.x>.
- Kuhl, P., Conboy, B., 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*. <https://doi.org/10.1098/rstb.2007.2154>.
- Kuhl, P., Tsao, F.-M., & Liu, H.-M. 2003. "Foreign-Language Experience in Infancy: Effects of Short-Term Exposure and Social Interaction on Phonetic Learning." *Proceedings of the National Academy of Sciences* 100 (15): 9096–9101. <https://doi.org/10.1073/pnas.1532872100>.
- Lengeris, A., & Hazan, V. 2010. "The Effect of Native Vowel Processing Ability and Frequency Discrimination Acuity on the Phonetic Training of English Vowels for Native Speakers of Greek" *The Journal of Acoustical Society of America* 128 (6). <https://doi.org/10.1121/1.3506351>.
- Lenneberg, E. 1967. *Biological Foundations of Language*. New York: John Wiley & Sons.
- Liberman, A., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. 1967. "Perception of the Speech Code." *Psychological Review* 74 (6): 431–61. <https://doi.org/10.1037/h0020279>.
- Logan, J, Lively, S., & Pisoni, D. 1991. "Training Japanese Listeners to Identify English /r/ and /l/: A First Report" *The Journal of Acoustical Society of America* 89 (2): 874–86. <https://doi.org/10.1121/1.1894649>.
- MacKay, I., Flege, J., Piske, T., & Schirru, C. 2001. "Category Restructuring during Second-Language Speech Acquisition." *The Journal of the Acoustical Society of America* 110 (1): 516–28. <https://doi.org/10.1121/1.1377287>.
- Munro, M., Flege, J., & Mackay, I. 1996. "The Effects of Age of Second Language Learning on the Production of English Vowels." *Applied Psycholinguistics* 17 (03): 313–34. <https://doi.org/10.1017/s0142716400007967>.
- Newport, L. 1990. "Maturational Constraints on Language Learning" *Cognitive Science* 14 (1): 11–28. [https://doi.org/10.1207/s15516709cog1401\\_2](https://doi.org/10.1207/s15516709cog1401_2).
- Opetushallitus. 2019. "Mitä Kieliä Perusopetuksessa Opiskellaan?" *Faktaa Express* 1A: 1–6. Retrieved from [https://www.oph.fi/download/195501\\_Faktaa\\_Express\\_1A\\_2019.pdf](https://www.oph.fi/download/195501_Faktaa_Express_1A_2019.pdf).
- Paganus, A, Mikkonen, V.-M., Mäntylä, T., Nuutila, S., Isoaho, J., Aaltonen, O., & Salakoski, T. 2006. "The Vowel Game: Continuous Real-Time Visualization for Pronunciation Learning with Vowel Charts." In T. Salakoski, F. Ginter, S. Pyysalo, T. Pahikkala (Eds.) *Advances in Natural Language Processing*. Berlin: Springer.
- Paradis, M. 2004. *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins Publishing Company.
- Patkowski, M. 1980. "The Sensitive Period for the Acquisition of Syntax in a Second Language." *Language Learning* 30 (2): 449–68. <https://doi.org/10.1111/j.1467-1770.1980.tb00328.x>.

- Peltola, K. U., Tamminen, H., Alku, P., & Peltola, M. S. 2015. "Non-native production training with an acoustic model and orthographic or transcription cues." In The Scottish Consortium for ICPhS 2015 (eds.) *Proceedings of the 18<sup>th</sup> International Congress of Phonetic Sciences*. Glasgow: The University of Glasgow.
- Peltola, M. S. 2003. *The Attentive and Preattentive Perception of Native and Non-Native Vowels: The Effect of Second Language Learning*. Turku: Painosalama.
- Peltola, M. S., Kuntola, M, Tamminen, H., Hämäläinen, H., & Näätänen, R. 2005. "Early exposure to non-native language alters preattentive vowel discrimination." *Neuroscience Letters* 388, 121–125.
- Penfield, W, & Roberts, L. 1981. *Speech and Brain Mechanisms*. Princeton: Princeton University Press.
- Perkell, J. 2012. "Movement goals and feedback and feedforward control mechanisms in speech production." *Journal of Neurolinguistics* 25: 382–407.
- Robertson, P. 2002. "The Critical Age Hypothesis. A Critique of Research Methodology." *Asian EFL Journal* 4 (1): 4–25.
- Rosner, B. S., & Pickering, J. B. 1994. *Vowel Perception and Production*. Oxford University Press.
- Ruben, R. 1999. "A Time Frame of Critical/Sensitive Periods of Language Development." *IJO & HNS* 51 (3): 85–89.
- Rvachew, S. 1994. "Speech Perception Training Can Facilitate Sound Production Learning." *Journal of Speech Language and Hearing Research* 37 (2): 347. <https://doi.org/10.1044/jshr.3702.347>.
- Ryan, R., & Deci, E. 2000. "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions." *Contemporary Educational Psychology* 25: 54–67. <https://doi.org/10.1006/ceps.1999.1020>.
- Saloranta, A., Tamminen, H., Alku, P., & Peltola, M. S. 2015. Learning of a non-native vowel through instructed production training. In The Scottish Consortium for ICPhS 2015 (eds.) *Proceedings of the 18<sup>th</sup> International Congress of Phonetic Sciences*. Glasgow: The University of Glasgow.
- Singleton, D. M., & Ryan, L. 2004. *Language Acquisition : The Age Factor*. Clevedon: Multilingual Matters.
- Snow, C., & Hoefnagel-Höhle M. 1978. "The Critical Period for Language Acquisition : Evidence from Second Language Learning." *Child Development* 49 (4): 1114–1128. <https://www.jstor.org/stable/1128751>.
- Taimi, L., Jähi, K., Alku, P., & Peltola, M. S. 2014. "Children Learning a Non-Native Vowel-The Effect of a Two-Day Production Training." *Journal of Language Teaching and Research* 5 (6): 1229–1235. <https://doi.org/10.4304/jltr.5.6.1229-1235>.

- Tamminen, H., Peltola, M. S., Kujala, T., & Näätänen, R. 2015. "Phonetic Training and Non-Native Speech Perception — New Memory Traces Evolve in Just Three Days as Indexed by the Mismatch Negativity ( MMN ) and Behavioural Measures." *International Journal of Psychophysiology* 97 (1): 23–29. <https://doi.org/10.1016/j.ijpsycho.2015.04.020>.
- Twyford, C. W. 1987. "Age-Related Factors in Second Language Acquisition." *New Focus*, no. 02: 1–8.
- Wang, Y., & Kuhl, P. K. 2003. "Evaluating the 'Critical Period' Hypothesis: Perceptual Learning of Mandarin Tones in American Adults and American Children at 6, 10 and 14 Years of Age." In M. J. Solé, D. Recasens, J. Romero (Eds.) *Proceedings of 15th ICPHS*: 1537–40. Barcelona : Universitat Autònoma de Barcelona.
- Wang, Y., Spence, M., Jongman, A., & Sereno, J. 1999. "Training American Listeners to Perceive Mandarin Tones." *The Journal of the Acoustical Society of America* 106 (6): 3649–58. <https://doi.org/10.1121/1.428217>.
- Williams, M., & Burden, R. 1999. "Students' Developing Conceptions of Themselves as Language Learners." *The Modern Language Journal* 83 (2): 193–201. <https://doi.org/10.1111/0026-7902.00015>.
- Yeni-Komshian, G., Flege, J., & Liu, S. 2000. "Pronunciation Proficiency in the First and Second Languages of Korean–English Bilinguals." *Bilingualism: Language and Cognition* 3 (2): 131–49. <https://doi.org/10.1017/s1366728900000225>.

## Appendix 1. Questionnaire

Pvm: \_\_\_\_\_ Kh nro: \_\_\_\_\_ (Tutkija täyttää)

### TAUSTAKYSELYLOMAKE

Lapsen nimi: \_\_\_\_\_

Ikä ja syntymäaika: \_\_\_\_\_

Puhelinnumero & s-posti: \_\_\_\_\_

### KIELI

1. Onko lapsella todettu tai tiedossa kielenkehityksen häiriötä?

\_\_\_ Kyllä on

\_\_\_ Ei ole

Mitä häiriötä kielenkehityksessä on todettu?

\_\_\_\_\_  
\_\_\_\_\_

2. Onko lapsella todettu kuulon heikentymää?

\_\_\_ Kyllä on

\_\_\_ Ei ole

3. Mitä kieliä lapsi osaa? Arvioi samalla, kuinka hyvin kyseistä kieltä hän osaa (äidinkieli, erinomaisesti, hyvin, tyydyttävästi, huonosti).

suomi: \_\_\_\_\_

ruotsi: \_\_\_\_\_





## **Appendix 2: Finnish Summary**

### **Preschool Children Learning a Foreign Vowel Through a Two-Day Listen-And-Repeat Training**

Tämä tutkielma käsittelee lasten vieraan kielen oppimista ja siihen vaikuttavia tekijöitä. Tutkielmassa tarkasteltiin lasten vieraan kielen oppimista erilaisten vieraan kielen oppimisen teorioiden ja mallien valossa ja selvitettiin kuinka tuottoharjoittelu vaikuttaa lasten kykyyn havaita ja tuottaa vaikea vieraskielinen äänne.

#### **Vieraan kielen oppimisen teoriat**

Tutkimuksen teoreettisessa osassa tarkasteltiin kolmea vieraan kielen oppimisen teoriaa, jotka olivat Speech Learning Model (SLM), Perceptual Assimilation Model (PAM) ja Native Language Magnet (NLM) teoria. Nämä teoriat käsittelevät kielenoppijan kohtaamia haasteita uuden kielen äännteitä opetellessa. Kaikki teoriat ovat kontrastiivisia ja jaottelevat vieraan kielen äännteitä suhteessa oppijan äidinkieleen.

SLM:n mukaan vieraan kielen äännteet voidaan jaotella kolmeen luokkaan: äännteet voivat olla identtisiä, uusia tai samankaltaisia suhteessa äidinkielen äännteisiin. Identtiset äännteet ovat akustisesti samanlaisia, eikä niiden oppimisessa pitäisi olla vaikeuksia. Uudet äännteet voivat tuottaa suuriakin oppimisvaikeuksia, sillä vieraan kielen äänne ei kuulu äidinkielen äännekategorioiden ja sille tulee luoda oma kategoriansa. Samankaltaisilla äännteillä on jokin tunnistettava vastine äidinkielessä, vaikka äännteet eroavatkin akustisesti toisistaan. Samankaltaiset äännteet ovat SLM:n mukaan hankalimpia, sillä ne voivat olla vaikeita erottaa äidinkielen äännteistä. SLM ottaa kantaa myös iän vaikutukseen vieraan kielen ääntämyksen oppimisessa, sillä sen mukaan oppimistulokseen vaikuttavia tekijöitä ovat äidinkielen ja vieraan kielen käyttömäärä ja vieraan kielen oppimisen aloitusikä.

PAM:n mukaan vieraan kielen äännteet havainnoidaan äidinkielen äännekategorioiden kautta. Vieraan kielen äännteet voidaan assimiloida äidinkielen äännekategorioiden kuudella tavalla. Kaksi vieraan kielen äännettä voi assimiloitua kahteen äidinkielen

äännekategoriaan, ja näiden äänteiden erottamistarkkuus on hyvä. Kaksi vieraan kielen äännettä voi myös assimiloidua yhteen äidinkielen kategoriaan, jolloin toinen äänneistä havaitaan äidinkielen kategorian hyvänä edustajana mutta toinen ei. Tällöin eron havaitseminen voi olla kohtalaista tai helppoa, riippuen äänteiden välisistä eroista. Kaksi vieraan kielen äännettä voi myös assimiloidua yhteen äidinkielen äännekategoriaan niin, että ne havaitaan kumpikin yhtä hyvinä tai huonoina äidinkielen äännekategorian edustajina. Näiden äänteiden välisten erojen havaitseminen on silloin PAM:n mukaan hyvin vaikeaa. Kaksi vieraan kielen äännettä voidaan myös havaita niin, että toinen assimiloiduu äidinkielen äännekategoriaan, mutta toinen ei. Tällöin näiden äänteiden erottelukyky on hyvä. Kumpikin vieraan kielen äänne voidaan myös havaita puheena, mutta kumpikaan ei assimiloitu äidinkielen kategorioihin. Tällaisessa tilanteessa erottelu on vaikeaa tai helppoa, riippuen äänteiden samankaltaisuudesta suhteessa äidinkielen äänneisiin. Kaksi vieraan kielen äännettä voivat myös olla niin kaukana äidinkielen äänneistä, ettei niitä edes havaita puheeksi. Näiden erottelukyvyn oletetaan olevan hyvä.

NLM:n mukaan vastasyntyneen aivot mukautuvat äidinkielen äänneisiin ensimmäisen puolen vuoden aikana, mikä vaikuttaa lapsen havainnointikykyyn. Sensitiivisyys äidinkielen äänneille muokkaa havainnointikykyä niin, että äidinkielen äänneille alkaa muodostua prototyyppisiä. Prototyypit edustavat lapsen eniten kuulemia äännevariantteja, joista muodostuu foneettisten kategorioiden keskiöt. Nämä prototyypit toimivat eräänlaisina magneetteina, jotka vetävät kategorian rajalla olevia äänneitä puoleensa. Magneettiefektin vuoksi äännekategorioiden rajalla äänteiden erottaminen toisistaan on erityisen vaikeaa. Vieraan kielen oppimisessa magneettiefekti vaikeuttaa uusien äänteiden oppimista, sillä vieraan kielen äänneet havaitaan äidinkielen kautta. Jos vieraskielinen äänne on lähellä äidinkielen äännettä, äänneitä on vaikea havaita erilleen.

### **Lapset vieraan kielen oppijoina**

Lapsia pidetään yleisesti parempina kielen oppijoina kuin aikuisia. Yleisesti hyväksytyyn näkemykseen mukaan vieraan kielen oppiminen on helpompaa nuoremmalla iällä, ja erityisesti natiivinkaltaisen ääntämyksen saavuttaminen on vaikeampaa aikuisena. Aikaa, jolloin kielten oppimista pidetään helpompana, kutsutaan kriittiseksi iäksi. Kriittinen ikä -hypoteesin mukaan äidinkielen omaksumisen on tapahduttava varhaislapsuudessa, sillä murrosikään mennessä aivojen plastisuus vähenee ja kielen

omaksuminen on vaikeampaa. Hypoteesin mukaan äidinkieli on omaksuttava kielellisesti rikkaassa ympäristössä varhaislapsuuden ja murrosiän välillä, jotta kielellinen ilmaisu ja ymmärrys kehittyvät normaalisti ja tasapainoisesti. Todisteena kriittinen ikä -hypoteesin puolesta pidetään yleensä tapauksia, joissa lapsi on jäänyt vaille edellä mainitun kaltaista ympäristöä, minkä seurauksena lapsen äidinkieli ei ole kehittynyt ikään nähden normaalille tasolle. Kriittinen ikä -hypoteesia on laajennettu käsittämään myös vieraan kielen oppiminen. Kriittisestä iästä, sen alkamisesta ja loppumisesta, ei ole yksimielisyyttä tutkijoiden keskuudessa. Tutkittaessa kielen eri osa-alueita, kuten fonologista tai syntaktista kehitystä, kriittisen iän ajoituksen on todettu osuvan eri ikäkausille kielen osa-alueesta riippuen.

Ikä on merkittävä vieraan kielen oppimiseen vaikuttava tekijä. Kuitenkin myös oppimisympäristöllä, oppijan motivaatiolla ja henkilökohtaisilla ominaisuuksilla on merkitystä vieraan kielen oppimisessa. Oppimisympäristö eroaa usein lapsilla ja aikuisilla. Lapset oppivat vieraan kielen luonnollisessa ympäristössä käydessään vieraskielisissä päiväkodeissa ja kouluissa. Tällaisessa ympäristössä lapset altistuvat vieraalle kielelle samalla tavoin kuin olisivat altistuneet äidinkielelleenkin. Aikuiset toisaalta opiskelevat vierasta kieltä luokkahuoneympäristössä. Aikuisetkin voivat omaksua kieltä luonnollisessa ympäristössä. Aikuisilla, toisin kuin pienillä lapsilla, on kuitenkin mahdollisuus turvautua kommunikoinnissa välikieleen, esimerkiksi englantiin, mikä vaikuttaa vieraan kielen oppimiseen.

Motivaatio on myös merkittävä oppimiseen vaikuttava tekijä. Motivaatio on voima, joka ajaa ihmisen tekemään jotakin ja myöhemmin myös pitää yllä tätä tekemistä. Motivaatio voidaan jakaa sisäiseen ja ulkoiseen. Sisäinen motivaatio tarkoittaa jonkin tekemistä sen aiheuttaman mielihyvän ja kiinnostuksen vuoksi. Ulkoinen motivaatio taas viittaa asian tekemiseen palkkion toivossa tai rangaistuksen pelossa. On havaittu, että vanhempia kielen oppijoita ohjaa enemmän ulkoinen motivaatio, kuten akateeminen tai taloudellinen menestys, kun taas nuoremmat oppijat ovat sisäisesti motivoituneempia. Oppimisympäristön ja motivaation lisäksi myös asenteilla vierasta kieltä ja sen oppimista kohtaan on merkitystä oppimistuloksien kannalta. Myös itseluottamus, ahdistuneisuus ja muut tunteisiin liittyvät tekijät vaikuttavat vieraan kielen oppimiseen ja oppimistuloksiin. Huono itseluottamus ja ahdistuneisuus ovat tutkimuksissa olleet yhteydessä huonoihin oppimistuloksiin.

## Tutkittavat ilmiöt

Tutkimuksessa tarkasteltiin kaksipäiväisen kuuntele-ja-toista -harjoituksen vaikutusta esikouluikäisten tyttöjen kykyyn havaita ja tuottaa vieraskielinen äänne, joka oli teorioiden valossa mahdollisimman vaikea. Tutkimuksen tavoitteena oli selvittää oppivatko lapset tuottamaan eron kahden äänteen välille ja kuinka nopeasti oppiminen tapahtuu valitulla harjoittelumenetelmällä. Menetelmä mukailee koulussa yleisesti käytössä olevaa kuuntele-ja-toista -menetelmää, jossa oppijan on havaittava äänteiden väliset erot vain akustisen informaation perusteella.

## Koehenkilöt, ärsykkeet ja tutkimusmenetelmät

Koehenkilöinä toimi 12 6–7 -vuotiasta esikouluikäistä tyttöä. Kaikki koehenkilöt olivat yksikielisiä ja äidinkieleltään suomenkielisiä, eikä heillä ollut merkittäviä kielellisiä tai terveydellisiä ongelmia. Koehenkilöiden kielitaito, kuulon tila ja mahdolliset kielelliset häiriöt kartoitettiin taustatietolomakkeella ennen tutkimusta.

Tutkimuksen ärsykeinä käytettiin kahta semisynteettistä epäsanaa /tu:ti/ ja /ty:ti/, joista ensimmäinen sisältää tutkimuksessa tarkastellun kohdevokaalin /u/. Vokaalit /u/ ja /y/ valittiin kohdesanoihin sillä perusteella, että niiden erottaminen on suomenkieliselle oppijalle mahdollisimman vaikea kielellinen oppimistehtävä. Suomessa suppeita pyöreitä vokaaleja on kaksi, /y/ ja /u/, kun taas ruotsin kielessä vokaaliavaruus jakautuu näiden kohdalta kolmeen kategoriaan /y/, /u/ ja /u/.

Tutkimus suoritettiin kahtena peräkkäisenä päivänä. Molempina päivinä koehenkilöt osallistuivat kahteen harjoitusosioon ja kahteen nauhoitusosioon, eli yhteensä neljään harjoitus- ja nauhoitusosioon. Kummassakin osiossa ärsykesanat /tu:ti/ ja /ty:ti/ vuorottelivat. Harjoitusosiossa kumpikin ärsykesana esiintyi 30 kertaa, kun taas nauhoitusosiossa kumpikin ärsykesana esiintyi 10 kertaa. Koehenkilöt kuulivat ärsykesanat kuulokkeista ja heidän tuli toistaa aina kummassakin osiossa kuulemansa sana.

Nauhoituksista analysoitiin akustisesti tutkimuksen kohteena olevien vokaalien /u/ ja /y/ formanttiarvot ja yksilökohtaiset keskihajonnat Praat-ohjelmistolla. Ensimmäinen ja toinen formantti ovat tärkeitä vokaalien erottamisessa toisistaan. Vokaalien /u/ ja /y/

erottamiseksi toisistaan on tarkasteltava erityisesti toisen formantin arvoja, sillä sen korkeus on yhteydessä vokaalin takaisuuteen. Akustisesta analyysistä saadut ensimmäisen ja toisen formantin arvot sekä yksilökohtaiset keskihajonnat analysoitiin sen jälkeen tilastollisesti IBM SPSS Statistics -ohjelmistolla, jotta voitiin selvittää, erosivatko vokaalien /u/ ja /y/ formanttiarvot toisistaan tutkimuksen edetessä.

## **Tulokset ja pohdintaa**

Tutkimuksen tuloksista kävi ilmi, että koehenkilöt muuttivat tuottoaan kohdevokaalin suuntaan toisena koepäivänä, neljänteen nauhoitusosioon mennessä. Tulokset osoittavat, että lapset oppivat havaitsemaan vokaalien /u/ ja /y/ välisen eron ärsykesanoissa ja tuottamaan jonkinlaisen eron ärsykesanojen välille. Tutkimuksen tulokset tukevat aiempia tutkimustuloksia lasten nopeasta vieraan kielen äänneiden omaksumiskyvystä. Lapset eivät saaneet koetilanteessa ohjeistusta kuulemiensa sanojen tuottamisesta, vaan heidän tuli toistaa kuulemansa. Tutkimuksen tulokset kertovat lasten kyvystä oppia kieltä vain kuuntelemalla ja toistamalla.

Akustisen analyysin perusteella lapset tuottivat tarkasteltavat vokaalit /u/ ja /y/ hyvin samankaltaisesti ensimmäisessä nauhoituksessa. Akustisen analyysin perusteella tilastollisesti merkitsevän eron oletettiin näkyvän jo toisessa nauhoituksessa, mutta vokaalit tuotettiin tilastollisesti merkitsevillä F2 arvon eroilla neljänteen nauhoitukseen mennessä. Koehenkilökohtaisten keskihajontojen tarkastelu paljasti, että toisessa nauhoituksessa F2 arvojen keskihajonta oli tilastollisesti merkitsevä. Toisessa nauhoituksessa lapset olivat mahdollisesti epävarmempia omasta tuotostaan ja tuottivat sanat yksilötasolla epäsystemaattisesti. Yksilötasolla oli suurta vaihtelua sanojen tuotossa. Toiset oppivat tuottamaan ärsykesanat erilleen nopeammin kuin toiset. Näin pienessä koehenkilöryhmässä yksilölliset erot nousevat merkittäviksi ja voivat vaikuttaa kokonaistulokseen.

Tuloksia verrattiin aiempaan lasten plastisuutta ja äänneiden oppimiskykyä tarkastelleeseen tutkimukseen, joka oli toteutukseltaan samanlainen kuin tässä tutkielmassa esitelty tutkimus, vain koehenkilöt olivat vanhempia. Aiemman tutkimuksen 7–10 -vuotiaat lapset oppivat tuottamaan ärsykeäänteet erilleen jo kolmanteen nauhoitukseen mennessä. Vaikka nuorempien lasten aivojen plastisuus on suurempaa, voi oppimistulokseen vaikuttaa vanhempien lasten kehittyneemmät kognitiiviset kyvyt.

Tulosten pohjalta näyttää siltä, että vieraan kielen opettaminen esikouluikäisille lapsille kuuntele-ja-toista menetelmällä ei tuota parempia oppimistuloksia alakouluikäisiin lapsiin verrattuna.