

Children Learning a Non-native Vowel – The Effect of a Two-day Production Training

Laura Taimi

Department of Phonetics, University of Turku, Turku, Finland

Katri Jähi

Department of Phonetics, University of Turku, Turku, Finland

Paavo Alku

Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland

Maija S. Peltola

Department of Phonetics, University of Turku, Turku, Finland

Abstract—The aim of this study was to investigate, how young children learn to produce a non-native vowel embedded in a pseudo-word context after short and intensive listen-and-repeat training sessions. The trained vowel contrast was chosen so that it would generate maximal learning difficulties according to models of second language learning. The group consisted of 13 7–10 year old girls. The child subjects participated in the study twice, on two consecutive days. Both days consisted of two training and two recording sessions. The model stimuli were pseudo-words /ty:ti/ and /tʌ:ti/, synthesized using a semi-synthetic method. The primary target word was /tʌ:ti/, which contained the vowel phoneme phonologically irrelevant in Finnish. The two lowest formants and the fundamental frequency were analysed from the recorded vowels. The results revealed that children learned to produce the new vowel after only three training sessions and that the effect was also seen after the fourth session. These results suggest that children learn to produce a new vowel after a short period of phonetic training.

Index Terms—children, language learning, vowel production

I. INTRODUCTION

The language that an infant hears in his/hers environment alters the perception of sounds and, by the age of six months, the perceptual vowel space has already changed so that the native language vowel prototypes start to act as magnets hindering the discrimination near a native language prototype but not near a non-prototype. The perception of native vowels by infants is also strongly in accordance with the perception of adult speakers of the language. Thus, the sound categories for the mother tongue evolve during the child's first year of life (Kuhl, 1991).

The sounds of a non-native language (L2) are filtered through the native language system. The native sound prototypes function as prototype magnets (Kuhl, 1991), which leads to difficulties in perceiving non-native sounds that are acoustically near the native language phonemes. According to the Speech Learning Model (SLM, Flege, 1987), the sounds of a non-native language can be divided into three types according to their correspondence with the native categories: “new”, “similar” and “identical”. The sounds that are classified as similar are the ones that are the most difficult to learn, since they resemble the native categories of the mother tongue but, however, differ from them significantly, e.g. regarding the exact formant values or phonological status. Sounds that belong to the group “new” are most often difficult at first, as they are something totally novel, but they do not result in major learning difficulties, since they are not confused with existing categories. “Identical” sounds are identical in both languages and thus their acquisition presents no problems. Another model of non-native sound acquisition, namely the Perceptual Assimilation Model (PAM, Best and Strange, 1992, Best 1995), approaches these difficulties by perceptual assimilation patterns where target language sound pairs are compared with the native language sound system. According to this model, the most difficult contrast to perceive is a case where two target language sounds are categorized into one native language category either equally well or equally poorly. This situation is called *Single Category (SC) assimilation* in the model and it is difficult because the sounds are not discriminated in the mother tongue, but they are treated as two equally good allophones of one category. The trained sound pair /y/ and /ʌ/ is considered to be *Category-Goodness Difference (CG Type)*, which is considered to produce considerable learning difficulties. In the present study, children's ability to learn a non-native vowel category was investigated. In particular, we were interested in studying whether an intensive two-day listen-and-repeat training, which is widely used in Finnish schools, could change the production of a non-native vowel. The trained category, /ʌ/, is an allophone of the Finnish vowel categories /y/ and /u/. This is considered to produce learning difficulties in all models of second language phonetic learning.

Earlier research by Iverson and Evans (2009) on intensive training effects with adults has shown that auditory training can be useful in learning to perceive foreign language sound categories. In their study, the identification of English vowels by German and Spanish speakers was investigated. The results showed that the improvement in identification also depended on the amount of vowel categories that the languages have: when there are fewer categories in the native language than in the target language (Spanish), the improvement took a longer time than when the vowel space was more crowded in the native language (German). In our study, the target language has three vowel categories in the same acoustic space where the native language has only two. This phenomenon also known as under-differentiation (Weinreich, 1953/1963) can lead to learning difficulties, since one has to learn a distinction non-existent in the native language.

Bradlow et al. (1997) demonstrated that perceptual training of the /r/ -/l/ contrast in adult Japanese learners of English transfers also into motor skills by improving the productions of the liquids. The subjects belonging to the experimental group participated in a high-variability identification training and their identification skills as well as productions of the two consonants were measured both pretest and post-test. The identification skills improved significantly from the pretest to post-test. The more interesting finding was that the perceptual training facilitates also the production since native English listeners identified the sounds produced by Japanese subjects significantly more correct in the post-test than in the pretest. This improvement generalized also to novel items that the subjects had not heard in the high-variability training or the previous tests. Similar results were also found in a distinct study where children with phonological impairment were investigated (Rvachew, 1994). In that study it was found that the productions of /ʃ/ improved more in groups of children who, in addition to traditional speech therapy, had additional perception training of /ʃ/. Cheour et al. (2002) described the neural plasticity of young (3—6-year-old) Finnish children, who learned to discriminate preattentively a non-native vowel distinction in only two months after joining a French school or day care center. Also Peltola et al. (2005) reported similar learning effects in children, who had participated in an immersion program at least for two years. These results suggest that children are able to learn to discriminate a difficult sound contrast fast and the learning effect can be seen in production as well as in the preattentive perceptual tasks. A more recent study by Giannakopoulou et al. (2013) also gives strong evidence on children's motor plasticity compared to adults. In their study Greek participants trained on perceptual identification and discrimination of the English tense /ɹ :/ and lax /ɹ / using high-variability training. All subjects participated in pretest, training and post-test. Pretest and post-test consisted of Perceptual Identification and Auditory Discrimination. The high-variability training between the tests consisted of 10 sessions over a two week period. Both groups could benefit from the training but the 7—8 year old group improved their performance significantly more than adults and the authors propose that this supports the notion of enhanced plasticity for spoken language in children.

In the present investigation, monolingual Finnish children's ability to produce a vowel that is not phonologically relevant in their native language was studied. Vowels are composed of acoustic resonances, formants, and they are identified by their formant structure. The formant frequencies measured for the vowels depend on the size and shape of the supraglottal cavities (Fant, 1970; Gilbert et al., 1997). Several studies demonstrate that children have higher formant frequencies than adults and male speakers lower frequencies than female speakers, which can be explained with the size differences and with the influence of fundamental frequency on the formant frequencies (Hillenbrand et al., 1995; Huber et al., 1999; Lee et al., 1999; Peterson and Barney, 1952). In this study the child participants were all girls, since some studies have shown that even young children have different formant values depending on their gender (Huber et al., 1999).

Finland is a bilingual country with Finnish and Swedish as the official languages. In 2012 approximately five percent (5,4%) of the population had Swedish as their mother tongue. Although other languages have substantially as big representation, since 4,9 percent of the population have some other language than Finnish or Swedish as their mother tongue (Official Statistics of Finland (OSF): Population structure [e-publication]). According to *Statistics Finland*, there are 5,4% Swedish mother tongue speakers and 8,7% with some other language than Finnish or Swedish in Turku (StatFin database: Population according to age (1-year), sex, marital status and language by area 1990 - 2012 [Internet]). However, the Finnish speaking majority do not hear Swedish in abundance in everyday life, because most of the Swedish mother tongue speakers are in fact bilinguals and use Swedish mostly among the minority. It is though possible to choose only one language in official forms so there are no statistics of the amount of bilingual Finnish-Swedish speakers. Since Swedish is an official language it is quaranteed by law that e.g. Swedish education for the minority is offered and because of that the two language groups attend different schools, where only Finnish or Swedish is used. Consequently, the child subjects do not have any superior knowledge of Swedish so that it would impact the results.

The aim of our study was to find out the effect of four short training sessions on the production of a vowel category that is not phonologically relevant in the subject's native language, but which is close to two native language categories. Our hypothesis was that children would be able to alter their productions of the target vowel according to the given model because of high behavioural and neural plasticity, but that the productions of the native language vowel in the nontarget word would not change.

II. MATERIAL AND METHODS

A. Subjects

A total of 13 girls volunteered to participate in the study. The mean age was 9;1 years (range 7;1—10;9). None of the subjects reported of any hearing or language deficits. All of them had Finnish as their mother tongue and their families did not use any other languages at home. All of the subjects had lived their whole lives in Finland. Furthermore, they all had very limited exposure to other languages: eight of the subjects had some minimal passive exposure to another language (e.g. Italian, French). All subjects were enrolled in the Finnish elementary school system. Compulsory education in Finland starts in the year when the child turns seven years. Subjects reported various hobbies, e.g. listening to music and different sports. Altogether the group represented a sample of monolingual native Finnish-speaking children with no linguistic deficits. All subjects and their parents gave an informed written consent. The study was carried out with the permission of the Ethics Committee of the University of Turku.

B. Procedure

The subjects participated in the study twice, on two consecutive days and on both days there were recording and training blocks in turns. The subjects were told that they would hear Swedish words and they would have to repeat them after the model. They were also instructed that they could take breaks between the sessions, if they wanted to, or suspend the experiment upon their wish. The experiment was done in a sound attenuated room where only the experimenter and the subject were present. The stimuli were presented binaurally with Sanako Headset SLH-07 and registered with Sanako Lab 100 -software.

The first day began with a recording block (baseline) and after that there was a training session followed by a new recording block and training session. On the second day there were also four blocks starting with the training session and finishing with the last recording block. As a result, the experiment consisted of four recording and training sessions. Prior to the first session on both days, the subjects could set the volume to a comfortable level during a familiarization block, which consisted of six words so that both stimuli were repeated three times one after another. In the recording and training blocks the subjects repeated the target word and nontarget word stimuli in turns after the model. In the training block there were 30 repetitions of the target word /tʌ :ti/ and 30 of the nontarget word /ty:ti/, altogether 60 words. The recording block included 10 target words and 10 nontarget words, thus it contained 20 words altogether. In total, there were 320 words to be repeated by the subject. The whole experiment lasted for under half an hour per day per subject. Due to technical problems seven words produced by one subject and three by another had to be excluded from the analysis.

C. Stimuli

The rounded closed vowel space is divided into three categories in Swedish but only two in Finnish. The words used in this experiment were chosen so that they would result in maximal learning difficulties, as the vowel /ʌ/ is not phonological in Finnish but it is acoustically located near the categories /y/ and /u/. The distinction between the vowels is mostly made on the basis of the second formant values. The stimuli were semi-synthetic pseudowords /ty:ti/ and /tʌ :ti/. The stimuli of the study were created using the Semi-synthetic Speech Generation method (SSG), a method that models the production of natural speech with digital signal processing methodologies (Alku et al., 1999). Due to the use of the SSG method, it is worth noting that the generated two word stimuli are of equal acoustic and prosodic features (e.g. duration, fundamental frequency, intonation, voice quality) except during the long vowel in the middle of the words when the stimuli differ *only* in terms of their formant frequencies. This was achieved, importantly, without compromising the auditory quality of the stimuli: due to the use of a glottal excitation extracted from a natural utterance, both word stimuli were of naturalness equal to that of real speech.

SSG is based on the source filter theory of speech production according to which a voiced speech sound can be separated into two main parts: the excitation airflow generated by the fluctuating vocal folds, the glottal waveform, and the vocal tract (Fant, 1970). By using a natural voiced speech signal recorded by a microphone as input, SSG utilises an automatic adaptive two-stage procedure resulting in two outputs: the estimated glottal flow (in form of a digital time-domain signal) and the vocal tract filter (in form of a digital all-pole filter). The estimated glottal flow can then be used as an excitation to digital vocal tract models to synthesise speech stimuli. Since the excitation waveform is extracted from a natural utterance, the method is named *semi-synthetic*. Using glottal waveforms extracted from natural speech renders a realistic jitter in the synthesized stimuli, which is characteristic to natural utterances. SSG has been previously used successfully in several studies on speech perception (e.g. Ceponiene et al., 2003) but also recently as a vocoding technique in a statistical parametric speech synthesizer (Raitio et al., 2011).

In order to obtain raw material for the SSG synthesis, words produced by a natural talker were recorded in an anechoic chamber using a high-quality condenser microphone (Bruel&Kjaer 4188). The speaker was a bilingual (Finnish/Swedish) male student of the Aalto University, aged 24 years, with no history of any speech or hearing disorder. He pronounced the target word /tʌ :ti/ and the nontarget word /ty:ti/ by repeating the same word five times. All sounds were recorded with a sampling frequency of 22050 Hz and a resolution of 16 bits. From the recorded five repetitions, the middle one was chosen for further processing both for /tʌ :ti/ and /ty:ti/. The signal waveform of both words was cut into three parts corresponding to the beginning section (i.e. waveform corresponding to /t/ in both words), middle section (i.e. waveform corresponding to /ʌ :/ in /tʌ :ti/ and waveform corresponding to /y:/ in /ty:ti/) and end section (i.e. waveform corresponding to /i/ in both words). This segmentation was straightforward to conduct because

the long vowel in the middle of the both words was preceded and followed by the unvoiced plosive /t/. From the segmented utterances, SSG was first used in order to estimate the glottal waveform, denoted by $g(n)$, of the middle section cut from /tʌ :ti/. Second, the vocal tract filter was computed with SSG over the middle section of word /ty:ti/. By filtering $g(n)$ through the obtained vocal tract model, a semi-synthetic vowel /y:/ was created. Third, two signal waveforms were created as follows. The beginning, middle and end sections cut from /tʌ :ti/ were concatenated hence resulting in a waveform of stimulus /tʌ :ti/ (i.e. the procedure corresponded copying the waveform of the natural /tʌ :ti/ word). To obtain the stimulus waveform of /ty:ti/, the beginning section of /tʌ :ti/, the semi-synthetic /y:/ vowel, and the end section of /tʌ :ti/ were concatenated. The intensities of the generated two stimulus waveform were finally normalised by adjusting the square sums of the digital signals to the same value. The formant values of the target word /tʌ :ti/ were F1=338 Hz, F2=1258 Hz, F3=2177 Hz and for the nontarget word /ty:ti/ F1=269 Hz, F2=1866 Hz, F3=2518 Hz as measured from the midpoint of the vowel (190 ms after stimulus onset). The fundamental frequency was 126 Hz in the long vowels of both the target and nontarget words. The duration of the stimuli was 624 milliseconds and the inter-stimulus interval (ISI) was set to three seconds.

D. Analysis

Speech signals produced in the recording sessions by the test subjects were analysed by Praat software, version 5.3.01 (Boersma & Weenink, 2013). The following measurements were made for the first vowel in each recorded word: frequency of the first formant (F1), frequency of the second formant (F2) and the fundamental frequency (F0). The formants were measured from a steady-state phase of the vowel using the Linear Predictive Coding (LPC) Burg algorithm. In the statistical analysis we focused on the F1 and F2 because it is known that the two first formants are enough for distinguishing between vowel categories (Dew and Jensen, 1977). In addition, special attention was directed to the F2, since the main difference between /y/ and /ʌ / is on that acoustic cue. F0 values were measured and analysed in order to secure that there were no outliers that would distort the formant results. No considerable deviants were found.

The statistical analysis was performed using IBM SPSS software, version 19. The analysis consisted of the formant frequencies for the formants. In order to find out if the two words were produced differently and whether the productions change as a function of training, we analysed statistically the results of the F1 and F2 values using a Word (2) X Session (4) X Formant (2) Repeated Measures analysis of variance (ANOVA). Each formant was analysed separately using (Word (2) X Session (4) ANOVA) and the sessions were compared with paired samples t-tests. One-tailed p -values are reported of the t-tests because we were interested to see *decrease* in the formant values and the p -values are Bonferroni corrected due to multiple simultaneous comparisons. An alpha level of .05 was used in all tests and only the statistically significant results are reported here.

III. RESULTS

The omnibus ANOVA analysis of the formant values revealed the significant main effect of word ($F(1,12)=6.261$, $p=0.028$). There was also a significant interaction between Word and Formant ($F(1,12)=9.188$, $p=0.010$), signalling that the difference found in the two words was concentrated more on one of the measured formants. We also found an interaction between Session and Formant ($F(3,10)=6.332$, $p=0.011$), implying that one of the formants changed as a result of training, while the other remained unaffected. Further analysis of the relevant F2 value revealed the significant main effect of word ($F(1,12)=7.856$, $p=0.016$), indicating that the F2 values were different in the native and the target word. More importantly, the analysis on F2 revealed the main effect of the Session ($F(3,10)=4.440$, $p=0.031$), suggesting that the change was valid in the values of the F2 between sessions. Most importantly, the paired samples t-tests conducted to compare the sessions were statistically significant between first and third session ($t(12)=2.621$, $p=0.033$) and between first and fourth session ($t(12)=3.047$, $p=0.015$) on the target word. The analysis, thus, revealed that the F2 value of the target word changed by the third session and that the change prevailed also at the fourth session. These findings can also be observed in the formant values shown in Table 1 and Fig. 1. Altogether, the F2 value of the target word's first vowel increases by the third session and remains low in the fourth session.

TABLE 1.
FORMANT VALUES (HZ) AND THEIR STANDARD DEVIATION VALUES IN PARENTHESES FOR THE TWO FIRST FORMANTS FOR THE TARGET AND NONTARGET WORDS IN EACH SESSION.

	F1				F2			
	Session 1	Session 2	Session 3	Session 4	Session 1	Session 2	Session 3	Session 4
Target word	513 (38)	532 (61)	539 (52)	533 (50)	2020 (178)	1976 (222)	1873 (336)	1872 (307)
Nontarget word	506 (39)	508 (39)	516 (47)	511 (50)	2144 (119)	2109 (110)	2103 (135)	2109 (115)

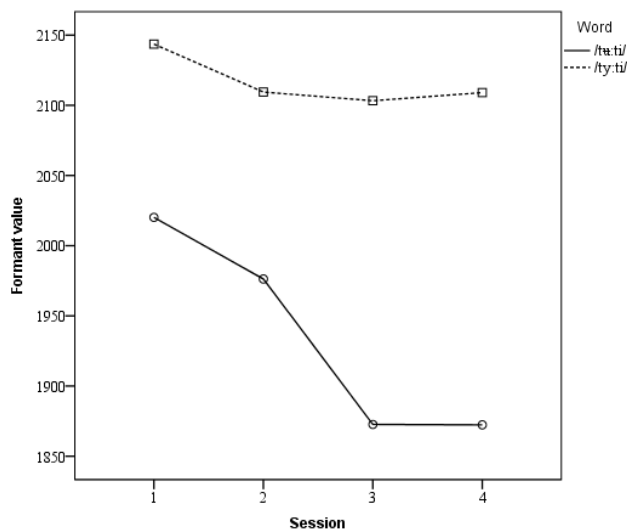


Figure 1. Second formant values by session in both words.

IV. DISCUSSION

In this study, typically developing children's ability to learn to produce a non-native vowel was investigated. The children were trained on the target vowel and the nontarget vowel by imitating a given model in four short sessions, and their productions were recorded in four recording sessions. The target vowel was considered to be difficult for Finnish speaking learners according to second language learning models (Best 1995; Flege 1987). The results showed that children can learn to produce the non-native target vowel in two days after only three short training sessions. The F2 values for the target word's first vowel in the first session were significantly higher than in the third and fourth session, which indicates that the vowel produced by the subjects had changed from resembling the vowel /y/ to resembling more the vowel /u/. Our results are in line with our hypothesis, as only the target vowel changed, and with previous studies indicating fast learning effects and motor plasticity in children. Earlier research has shown that perceptual training can facilitate the learning of a new sound category in children with difficulties in consonant productions (Rvachew, 1994). A similar effect has also been demonstrated in a distinct study with adults learning a sound category that does not belong to their native system, and that the effect is noticeable even three months after training (Bradlow et al., 1999). Furthermore, also in pre-attentive level in studies using MMN, it has been shown that new memory traces can evolve for second language sounds even in a couple of months in immersion programs or in French school for Finnish children (Cheour et al., 2002). Altogether, studies using different methods have shown that children are fast learners, and that the learning effect can be seen both pre-attentively and on attention-demanding tasks. Our results are consistent with these previous findings. The novelty element is that the teaching method often used in Finnish schools, namely the listen and repeat training, significantly alters the production and therefore facilitates learning.

The stimuli were produced using a male speaker's vocal tract model as a basis, resulting in formants that were low in nature. Even though this was a challenging linguistic situation for the child subjects, as the model they heard for the words differed from their own production, they still learned to produce the non-native vowel according to the given model. We suggest that this would be evidence of children's high motor plasticity, as children seem to update their production patterns when a new kind of a model is provided. In line with this is the clinical study by Bates et al. (2001) which offers strong evidence on children's behavioural and brain plasticity after early brain injury. In their study children with left hemisphere injury were in the normal range for their age in several different tasks and differed significantly from the control group only on a couple of measures. In contrast, adults with similar injuries scored significantly lower than normal adults. When these children were compared with adults using age-related Z-scores, children had a significant advantage. Similar results were shown in a behavioural study by Giannakapoulou et al. (2013) where children outperformed adults on perceptual identification and discrimination. These studies imply that children's brain and motor plasticity is good and significantly higher compared to adults.

In conclusion, we show that a listen-and-repeat training can facilitate the production of a non-native vowel in young children. There were only four training sessions for each subject and the change in the production of the non-native vowel /u/ was seen already after three sessions and the change remained after the fourth and last session, too. This illustrates that a training, which is also commonly used in schools, is an effective tool for training the production of non-native sounds. Altogether, these results indicate that children's motoric plasticity is high and new articulations are acquired rapidly.

ACKNOWLEDGMENT

The authors thank all the children and their parents, who participated in this study. We thank Piia Savilepo, M.A., and Katja Immonen, M.A., for their assistance in data collection. This study was financially supported by TOP-Foundation to the first author. We warmly thank Sanako for sponsoring the LAB-laboratory.

REFERENCES

- [1] Alku, P., H. Tiitinen & R. Näätänen. (1999). A method for generating natural-sounding speech stimuli for cognitive brain research. *Clinical Neurophysiology* 110, 1329—1333.
- [2] Bates et al. (2001). Differential effects of unilateral lesions on language production in children and adults. *Brain and Language* 79, 223—265.
- [3] Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (ed.) *Speech perception and linguistic experience*, Baltimore: York press, 171—206.
- [4] Best, C. T. & W. Strange. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20, 305—330.
- [5] Boersma, P. & D. Weenink. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.01, available from <<http://www.praat.org/>>. (accessed 4 February 2013).
- [6] Bradlow, A. R., D. B. Pisoni, R. Akahane-Yamada, & Y. Tohkura. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am.* 101.4, 2299—2310.
- [7] Bradlow, A. R., R. Akahane-Yamada, D. B. Pisoni, & Y. Tohkura. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics* 61.5, 977—985.
- [8] Ceponiene et al. (2003). Speech-sound selective auditory impairment in autism: can perceive but do not attend. *PNAS*, 100.9, 5567-5572.
- [9] Cheour, M., A. Shestakova, P. Alku, R. Ceponiene & R. Näätänen. (2002). Mismatch negativity shows that 3—6-year-old children can learn to discriminate non-native speech sounds within two months. *Neuroscience Letters* 325, 187—190.
- [10] Dew, D. & P. D. Jensen. (1977). *Phonetic Processing: The Dynamics of Speech*. Bell & Howell Company, Columbus.
- [11] Fant, G. (1970). *Acoustic theory of speech production: with calculations based on X-ray studies of Russian*. Mouton, The Hague.
- [12] Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47—65.
- [13] Giannakopoulou, A., M. Uther & S. Ylinen. (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.
- [14] Gilbert, H. R., M. P. Robb & Y. Chen. (1997). Formant frequency development: 15 to 36 months. *Journal of Voice* 11.3, 260—266.
- [15] Hillenbrand, J., L. A. Getty, M. J. Clark & K. Wheeler. (1995). Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97.5, 3099—3111.
- [16] Huber J. E., E. T. Stathopoulos, G. M. Curione, T. A. Ash & K. Johnson. (1999). Formants of children, women and men: The effects of vocal intensity variation. *J. Acoust. Soc. Am.* 106.3, 1532—1542.
- [17] Iverson, P. & B. G. Evans. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *J. Acoust. Soc. Am.* 126.2, 866—877.
- [18] 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.
- [19] Lee, S., A. Potamianos & S. Narayanan. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *J. Acoust. Soc. Am.* 105.3, 1455—1468.
- [20] Official Statistics of Finland (OSF): Population structure [e-publication]. ISSN=1797-5395. 2012, Appendix table 2. Population according to language 1980—2012. Helsinki: Statistics Finland. Available from: <http://www.tilastokeskus.fi/til/vaerak/2012/vaerak_2012_2013-03-22_tau_002_en.html>. (accessed 22/1/ 2014).
- [21] Peltola, M. S., M. Kuntola, H. Tamminen, H. Hämäläinen & O. Aaltonen. (2005). Early exposure to non-native language alters preattentive vowel discrimination. *Neuroscience Letters* 388, 121—125.
- [22] Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24.2, 175—184.
- [23] Raitio et al. (2011). HMM-based speech synthesis utilizing glottal inverse filtering. *IEEE Transactions on Audio, Speech, and Language Processing*, 19.11, 153-165.
- [24] Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech and Hearing Research* 37.2, 347—357.
- [25] StatFin database: Population according to age (1-year), sex, marital status and language by area 1990 - 2012 [Internet]. Helsinki: Statistics Finland. Available from: <http://pxweb2.stat.fi/database/StatFin/databasetree_en.asp>. (accessed 22/1/2014).
- [26] Weinreich, U. (1953/1963). *Languages in contact*. Mouton, London.

Laura Taimi, M.A., is a PhD student from the Department of Phonetics University of Turku, Finland. Her research is focused on how school-aged children learn to produce and perceive non-native sounds.

Katri Jähi, M.A., PhD student from the Department of Phonetics, University of Turku, Finland. Topic of her PhD thesis is aging and second language learning and she focuses on speech production and perception.

Paavo Alku, Dr. Tech., is working as a Professor of Speech communication technology at the Department of Signal Processing and Acoustics, Aalto University, Finland.

Maija S. Peltola, PhD, Adj.Prof. is the head of the Department of Phonetics and Learning, Age and Bilingualism -laboratory at the University of Turku, Finland. Her research interests include a wide range of topics related to the perceptual and productional acquisition of non-native speech.