

# Motoric training alters speech sound perception and production – active listening training does not lead into learning outcomes

Kimmo U. Peltola

Phonetics and Learning, Age & Bilingualism laboratory (LAB-lab), Department of Future Technologies, University of Turku, FINLAND

Henna Tamminen

Phonetics and Learning, Age & Bilingualism laboratory (LAB-lab), Department of Future Technologies, University of Turku, FINLAND

Paavo Alku

Department of Signal Processing and Acoustics, Aalto University, FINLAND

Teija Kujala

Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, FINLAND

Maija S. Peltola

Phonetics and Learning, Age & Bilingualism laboratory (LAB-lab), Department of Future Technologies, University of Turku, FINLAND

**Abstract** - Accurate speech-sound perception and production are prerequisites for communication in a non-native language. Earlier research has shown that new categorization and pronunciation patterns evolve in different learning settings and that these skills can be trained with various methods. We tested the effects of two types of training protocols on the production and identification of L2 vowels embedded in a pseudo-word context. Group 1 (Producers) participated in a listen and repeat training, where they produced the target stimulus /tɛ:ti/ and the non-target stimulus /ty:ti/ after the example in a pseudo-randomized order; Group 2 (Listeners) was instructed to count the number of targets /tɛ:ti/ in the same stimulus train without any motoric production movements. The results showed clearly that listen and repeat training led to plastic changes both in production and in identification, while no learning effects were obtained with the listening paradigm. This suggests a significant role of motoric experience in the acquisition of speech.

**Index terms** - non-native vowel production, vowel identification, active auditory and motoric training.

## I. INTRODUCTION

The acquisition of a non-native language poses demands for the perception and production of speech sounds. Fundamentally, the most severe problems arise from the confusing similarities and dissimilarities between the native and the target language. This is based on the language-specificity of the neural processing of the speech signal, which has led the brain to categorize input according to the L1 phonological system (Näätänen et al., 1997). Models of L2 phonetic/phonological learning explain that the most problematic cases are those that are confused with the native categories. According to Speech Learning Model (Flege, 1987), these difficult speech sounds are termed as *similar*, while the L2LP model developed by Escudero (2005) calls them *new*. In addition, they may also be considered as cases of an equal two-to-one assimilation pattern as in the framework of the Perceptual Assimilation Model, PAM (Best & Strange, 1992).

Despite difficulties caused by the native sound system, learning is possible not only in children, whose brain is malleable in language learning (Peltola, M. S., Kuntola, Tamminen, Hämäläinen & Aaltonen, 2005; Cheour, Shestakova, Alku, Ceponiene & Näätänen, 2002) but also in adults (Flege & MacKay, 2004; Tamminen, Peltola, M. S., Kujala & Näätänen, 2015) and even seniors (Jähi, Alku & Peltola, M. S., 2015), and even with maximally difficult targets. Various types of trainings have proven powerful. Jamieson & Morosan (1986) showed that synthetic speech stimuli can be used in training the perception of an L2 dental fricative contrast, and Lively, Pisoni, Yamada, Tohkura & Yamada (1994) showed that training effects induced by a High Variability procedure were also of long-lasting. High variability training has proven valid in several studies (Iverson, Pinet & Evans, 2012; Lambacher, Martens, Kakehi, Marasinghe & Molholt, 2005). A low variability training was also found functional in immigrants training to produce extremely difficult vowel sounds embedded in pseudo-word contexts (Savo & Peltola, M. S., 2019). Even simpler training protocols seem to lead into plastic changes, since a listen and repeat procedure with only one exemplar per speech category resulted in changes in vowel production (Peltola, K. U., Alku & Peltola, M. S., 2017a). However, reduced amount of training had different degree of learning effects depending on the mother tongue which implies the significant interaction between the native

language and the amount of training needed (Peltola, K. U., Rautaoja, Alku & Peltola, M. S., 2017b), which implies that the amount of exposure has a critical threshold. Specific pronunciation instructions appear to lead into extremely rapid production changes (Saloranta, Tamminen, Alku & Peltola, M. S., 2015), which may be taken as evidence for the positive role of overt knowledge in pronunciation learning.

The link between production and perception is one of the fundamental questions of speech sciences (Liberman & Mattingly, 1985; Perkell, 2012). In a study using a High Variability method with identification training, the results showed changes not only in the identification (ID) performance, but also in production ratings (Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997). The effect seems valid also into the opposite directions, since motoric training changed identification scores (Lametti, Rochet-Capellan, Neufeld, Shiller & Ostry, 2014). In addition, production training with visual feedback resulted in discrimination and production changes, though not in synchrony (Kartushina, Hervais-Adelman, Frauenfelder & Golestani, 2015). Also, Peltola, K. U., Alku & Peltola, M. S. (2017a) showed that mere exposure to non-native vowels changed production performance in a similar manner as a listen and repeat training.

In the light of earlier findings, it seems that even the most difficult non-native sounds can be trained and that changes in the perception may induce changes in production, or vice versa. The aim of the present study is to determine, whether a listen and repeat training and an active listening training result in similar learning effects in production and identification.

## II. METHODS

### A. Participants

Altogether 22 subjects participated in the experimental procedure. All subjects were native speakers of Finnish and none of them had lived abroad. They all had a normal hearing and they were all right handed (tested using the Edinburgh Handedness Inventory (Oldfield, 1971)). All subjects reported to be healthy and no speech deficits or hearing disorders of any kind were reported in the background questionnaires. The subject pool was divided into two groups on the basis of the type of training that they were to be exposed: Group 1 (Producers) consisted of 11 subjects (6 females, mean age 24.2 years, range 18–38), while Group 2 (Listeners) contained 11 subjects (5 females, mean age 27.9 years, range 23–38).

### B. Stimuli

The stimuli were created using semisynthetic speech generation (SSG), a method that separates speech signals into the glottal excitation generated by the vocal folds and the vocal tract (Alku, Tiitinen & Näätänen, 1999). By using an adaptive two-stage procedure, SSG computes the glottal excitation in form of a digital time-domain signal and the adjustable vocal tract in form of a digital all-pole filter. SSG has been previously used in several studies on speech perception (Ceponiene et al., 2013).

The synthesis began by recording (sampling frequency 22050 Hz, 16 bits) natural speech utterances by a bilingual (Finnish/Swedish) male (age 24 years), with no history of any speech or hearing disorder. He pronounced the target word /tʌ:ti/ and the non-target word /ty:ti/ in an anechoic chamber five times altogether and the middle one was chosen for further processing (high-quality condenser microphone (Bruel & Kjaer 4188)). The signal waveform of both words was cut into three sections: the beginning (/t/), the middle (/ʌ:/ and /y:/) and the end (/ti/). SSG was first used in order to estimate the glottal excitation, denoted by  $g(n)$ , of the middle section. Second, vocal tract all-pole filters were computed over the middle sections of both /tʌ:ti/ and /ty:ti/. Both of these filters were next transformed to line spectral pairs (LSPs) (Soong & Juang, 1984). The LSPs of the vocal tract filters were converted to Mel frequency domain (Stevens & Volkman, 1940). A continuum was formed in the LSP space between the point of the vocal tract filter of /ʌ:/ and that of /y:/. By stepping this continuum using a constant step, nine new artificial vocal tracts were created with an equal Euclidean distance between consecutive vocal tracts. We also created one vocal tract by stepping the continuum one step beyond the original end points. By filtering  $g(n)$  through the obtained vocal tract models, semi-synthetic vowels were created that had their vocal tracts set according to the described stepping procedure. Next, word-level stimuli were generated by concatenating the beginning section cut from the original waveform of /tʌ:ti/, the semi-synthetic vowel section corresponding to each point in the continuum, and the end section cut from the original waveform of /tʌ:ti/. Finally, all the generated waveforms were normalized in intensity.

Due to the use of the SSG method, it is worth noting that the generated 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 (/ʌ/ F1 338 Hz, F2 1258 Hz; /y/ F1 269 Hz, F2 1866 Hz). This was achieved without compromising the auditory quality of the stimuli due to the use of a glottal excitation extracted from a natural utterance.

### C. Procedure

The experiments were conducted in a sound attenuated laboratory, where the subjects were constantly monitored via a video camera. Prior to the actual testing, the subjects filled in a detailed questionnaire, where their linguistic and other background factors were checked carefully. In addition, hearing was evaluated with an audiometer, where the subjects were presented with frequencies relevant for the accurate perception of vowel sounds.

The procedure was conducted on two consecutive days, so that Day 1 began with the baseline tests (identification and production) followed by the training paradigm and Day 2 started with the training and ended in the final tests in the

reversed order. In other words, the protocol design was for Day 1: baseline test ID; baseline test production; production or listening training. And for Day 2: production or listening training; final test ID; final test production. The training sets consisted of four blocks on both days, 30 stimuli per block, where the order and probability of the target stimulus varied across the blocks (60 targets and 60 non-targets in total per day). Group 1 (Producers) trained by a listen and repeat exercise, where the subjects were instructed to repeat after the model immediately after the stimulus had been presented (ISI 3 seconds), while Group 2 (Listeners) counted the number of the target stimuli ( $t\epsilon:ti$ ) within the block and reported these calculations to the researcher. The order of the training blocks was randomised between Day 1 and Day 2. The stimuli in the production experiment were presented (via SLH-07 head set) in turns using Sanako Lab 100 language lab system with identical 3 second inter-stimulus intervals. In the production tests the subjects in both groups were instructed to repeat immediately after the model. Both the target and the non-target stimuli were repeated 10 times. In the identification experiment, the stimuli from the acoustic continuum with thirteen steps from /y/ to /u/ were delivered using the similar sound system. The subjects were asked to identify the words by pressing a response button. The identification used a forced-choice method and the presentation of the stimuli was self-paced. The continuum consisted of 13 words, each containing a first-syllable vowel with slightly different quality. The participants were presented with 8 repetitions of each token per test and they were instructed to respond immediately without hesitation, but no reaction times were measured. Presentation software (version 16.3) was used for both presenting the stimuli and for registering the responses.

#### D. Analysis

The production data were first analysed with Praat, version 5.3.01 (Boersma, 2001). Using the Linear Predictive Coding (LPC) Burgh algorithm, we determined the values for the two lowest formant frequencies, namely F1 and F2 (Hz), of the first syllable vowel from the steady-state phase in the acoustic signal. We also calculated the standard deviation Hz values of the F1 and F2 from each subject. A Repeated Measures Analysis of Variance (ANOVA, IBM SPSS, version 23) was then performed with both the formant data and the standard deviation values. The data were subjected to further post hoc tests when required.

For the Identification data, we calculated two values for the category boundary, namely the location of the boundary and the steepness value of that boundary. We used the Logit transformation analysis to obtain these values and then subjected these data separately to a Repeated Measures Analysis of Variance (IBM SPSS, version 23). Again, further post hoc tests were performed when applicable.

### III. RESULTS

To test the effects of the training protocols we performed a Group (2) x Session (2) x Word (2) x Measure (2) Repeated Measures ANOVA for the formant values and the standard deviations. The formant data (Table 1.) showed the main effect of Measure ( $F(1,20) = 2328.435, P < 0.001$ ) and a Word x Measure interaction ( $F(1,20) = 123.510, p < 0.001$ ) indicating that the F1 and F2 values were systematically different in the target and non-target words. More importantly, we discovered the main effect of Word ( $F(1,20) = 101.623, p < 0.001$ ) showing that the two words were produced consistently differently. The most important discovery was the main effect of Session ( $F(1,20) = 4.569, p = 0.045$ ), which showed that the training had an effect on vowel production. Further analyses revealed that the training effects were valid only in the target word F2 values (main effect of Session ( $F(1,29) = 4.471, p = 0.047$ )). Interestingly, the data from Group 1 (Producers) showed the main effect of Session for the F2 values ( $F(1,190) = 5.257, p = 0.045$ ), while no such effect was found from Group 2 (Listeners), indicating thus changes in the Producers, but not in Listeners.

The standard deviation data (Table 2.) revealed also the main effects of Measure ( $F(1,20) = 104.725, p < 0.001$ ) and Session ( $F(1,20) = 15.213, p = 0.001$ ) as well as the Measure x Word interaction ( $F(1,20) = 23.595, p < 0.001$ ), but also important interactions of Session x Measure ( $F(1,20) = 12.470, p = 0.002$ ) and Session x Group ( $F(1,20) = 7.090, p = 0.015$ ). Most significantly, we found Group x Session x Measure ( $F(1,20) = 10.102, p = 0.002$ ) and Group x Session x Word x Measure ( $F(1,20) = 4.475, p = 0.047$ ) interactions revealing differences between the groups as a function of training. Further analyses showed that there were no effects in the non-target word /ty:ti/, but several in the target word /tu:ti/ (Measure ( $F(1,20) = 71.691, p < 0.001$ ); Session ( $F(1,20) = 9.126, p = 0.007$ ); Session x Measure ( $F(1,20) = 8.161, p = 0.010$ )). The most important discoveries were the Group x Session ( $F(1,20) = 7.995, p = 0.010$ ) and the Group x Session x Measure ( $F(1,20) = 11.171, p = 0.003$ ) interactions. These findings showed that the groups performed differently as a function of training. Further analysis revealed this to be due to difference in the F2 values ( $F(1,20) = 9.813, p = 0.005$ ). Interestingly, only Group 1 (Producers) showed the main effect of Session in F2 ( $F(1,19) = 14.737, p = 0.003$ ) indicating that the standard deviations of target word F2 changed in the Producers but not in the Listeners.

TABLE 1.

The average F1 and F2 (Hertz) values for Group 1 and Group 2 in two sessions.

	Group 1		Group 2	
	Session 1	Session 2	Session 1	Session 2
/tu:ti/ F1	421	431	412	412
/tu:ti/ F2	1276	1383	1285	1303
/ty:ti/ F1	392	394	367	366
/ty:ti/ F2	1897	1909	1821	1835

TABLE 2.

The average F1 and F2 standard deviation values for Group 1 and Group 2 in two sessions.

	Group 1		Group 2	
	Session 1	Session 2	Session 1	Session 2
/tu:ti/ F1	29	23	24	18
/tu:ti/ F2	217	119	146	148
/ty:ti/ F1	21	20	28	24
/ty:ti/ F2	88	59	80	62

We analysed the boundary location and steepness data (Table 3.) from ID experiment separately with a Group (2) x Session (2) Repeated Measures ANOVA to test the potential group differences and training effects. The boundary location analysis showed only the significant main effect of Group ( $F(1,20) = 5926, P=0.024$ ) indicating an overall difference between the groups. The steepness value data revealed the main effect of Session ( $F(1,20) = 5.389, P=0.031$ ), which suggested that training altered the consistency of the boundary. Further analysis showed that only Group 1 (Producers) exhibited changes as a function of training (main effect of Session ( $F(1,19) = 7.235, P=0.023$ )), suggesting that production training resulted in a steeper category boundary (Table 3., Figure 1. and Figure 2.).

TABLE 3.

The boundary locations and steepness values from Group 1 and Group 2 on two sessions.

	Group 1		Group 2	
	Session 1	Session 2	Session 1	Session 2
Boundary	7,74	7,40	6,95	6,75
Steepness	1,36	1,78	1,67	1,74

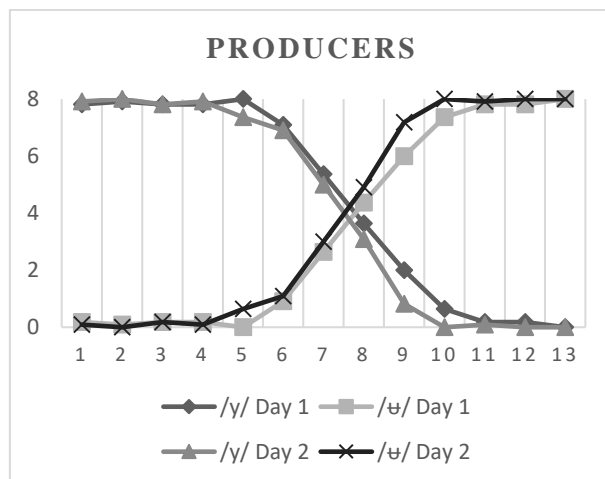


Figure 1. The ID data from Group 1 (Producers) showing the boundary location and its steepness.

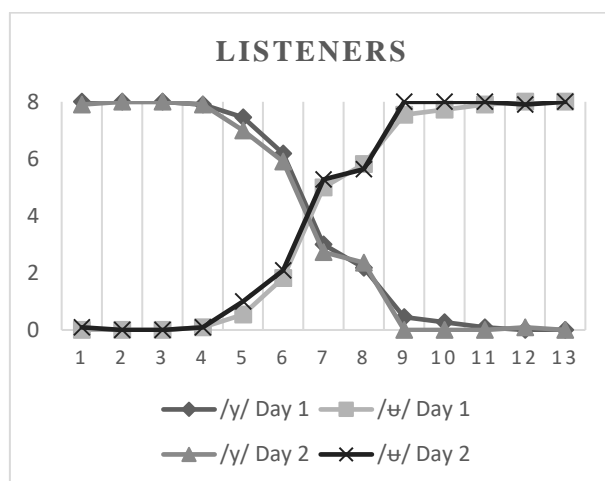


Figure 2. The ID data from Group 2 (Listeners) showing the boundary location and its steepness.

In summary, the statistical analyses revealed that changes both in production and in identification were exclusively present in the subjects, who trained with the listen and repeat paradigm, while no changes were present in the actively listening subjects. We were thus able to show that the motoric practice is required for production improvement and that the acquisition of new motoric patterns is transferred to perceptual categorization accuracy.

#### IV. CONCLUSIONS

Various types of training protocols have been found beneficial for changing either the perception or production of non-native speech sounds (e.g. Lively, Pisoni, Yamada, Tohkura & Yamada, 1994; Tamminen, Peltola, M. S., Kujala, & Näätänen, 2015). The interplay between speech perception and production is of significance in non-native speech learning and earlier studies have shown that motoric training may affect speech perception (Lametti, Rochet-Capellan, Neufeld, Shiller & Ostry, 2014), and that production may be changed with perceptual training (Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997). Our present findings showed that learners who participated in a training protocol involving motoric activities, acquired new production patterns for the non-native vowel very quickly, but active listening did not result in such learning effects. More interestingly, the motoric exercises affected the identification performance so that the categorisation systematicity increased, while mere listening did not have such transferrable training effects.

In the light of earlier studies and the present ones, it seems that the learning of new speech sounds involves operations from both domains, the motoric and the auditory. In addition, since no learning effects were observed when motoric practice was absent, it appears that motor activations are – if not prerequisites – then at least fundamental factors in the acquisition of non-native speech sound categories.

## ACKNOWLEDGEMENTS

We wish to thank Sanako Corp., for sponsoring the LAB-lab training software and all our volunteer subjects for their willingness to participate. Also, our gratitude goes to active students in Phonetics, whose assistance in data collection was highly important.

## REFERENCES

- [1] Alku, P., Tiitinen, H. & Näätänen, R. (1999). A method for generating natural-sounding speech stimuli for cognitive brain research. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology* 110(8): 1329-1333.
- [2] Best, C. T. & Strange, W. (1992) Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20: 305-330.
- [3] Boersma, P. (2001) Praat, a system for doing phonetics by computer. *Glott International* 5(9/10): 341-345.
- [4] Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R. & Tohkura, Y. (1997) Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101(4): 2299-310.
- [5] Ceponiene, R., Lepistö, T., Shestakova, A., Vanhala, R., Alku, P., Näätänen, R. & Yaguchi, K. (2003) Speech-sound selective auditory impairment in autism: can perceive but do not attend. *Proceedings of the National Academy of Sciences*: 100(9): 5567-5572.
- [6] Cheour, M., Shestakova, A., Alku, P., Ceponiene, R. & Näätänen, R. (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.
- [7] Escudero, P. (2005) *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Utrecht: Netherlands Graduate School of Linguistics.
- [8] 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.
- [9] Flege, J. E. & MacKay, I. R. (2004) Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26(01): 1-34.
- [10] Iverson, P., Pinet, M. & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics* 33(1): 145-160.
- [11] Jamieson, D. G. & Morosan, D. E. (1986) Training non-native contrasts in adults: Acquisition of English /ð/ - /θ/ contrast by francophones. *Perception & Psychophysics*, 40(4): 205-215.
- [12] Jähi, K., Alku, P. & Peltola, M. S. (2015) Does interest in language learning affect the non-native phoneme production in elderly learners. *Proceeding of the 18th International Congress of Phonetic Sciences*, Glasgow, ISBN 978-0-85261-941-4. Paper number 234.
- [13] Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H. & Golestani, N. (2015) The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *Journal of the Acoustical Society of America*, 138(2): 817-832.
- [14] Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M. & Ostry, D. J. 2014. Plasticity in the human speech motor system drives changes in speech perception. *Journal of Neuroscience*, 34(31): 10339-46.
- [15] Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. & Yamada, T. (1994) Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *J. Acoust. Soc. Am.* 96(4), 2076-2087.
- [16] Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A. & Molholt, G. (2005) The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26(2): 227-247.
- [17] Liberman, A. M. & Mattingly, I. G. (1985) The motor theory of speech perception revisited. *Cognition*, 21: 1-36.
- [18] Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J. & Alho, K. (1997) Language specific phoneme representations revealed by electric and magnetic responses. *Nature*, 385: 432-434.
- [19] Oldfield, R. C. (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9: 97-113.
- [20] Peltola, K. U., Alku, P. & Peltola, M. S. (2017a). Non-native speech sound production changes even with passive listening training. *Linguistica Lettica*, 25: 158-172.
- [21] Peltola, K. U., Rautaoja, T., Alku, P. & Peltola, M. S. (2017b) Adult learners and a one-day production training - small changes but the native language sound system prevails. *Journal of Language Teaching and Research* 8(1): 1-7.
- [22] Peltola, M. S., Kuntola, M., Tamminen, H., Hämäläinen, H. & Aaltonen, O. (2005) Early exposure to non-native language alters preattentive vowel discrimination. *Neuroscience Letters*, 388(3): 121-125.
- [23] Perkell, J. S. (2012) Movement goals and feedforward mechanisms in speech production. *Journal Neurolinguistics*, 25(5): 382-407.
- [24] Savo, S. & Peltola, M. S. (2019) Arabic-speakers learning Finnish vowels: short-term phonetic training supports second language vowel production. *Journal of Language Teaching and Research*, 10(1): 45-50.

- [25] Saloranta A., Tamminen, H., Alku, P. & Peltola, M. S. (2015) Learning of a non-native vowel through instructed production training. *Proceeding of the 18th International Congress of Phonetic Sciences*, Glasgow, ISBN 978-0-85261-941-4. Paper number 235.
- [26] Soong, F. K. & Juang, B-H. (1984) Line spectrum pair (LSP) and speech data compression. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 1.10.1-1.10.4. San Diego, USA.
- [27] Stevens, S. & Volkmann, J. (1940) The relation of pitch of frequency: A revised scale. *American Journal of Psychology*, 53: 329-353.
- [28] 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: 23-29.

**Kimmo U. Peltola**, MA, is a University Teacher from the Department of Phonetics and Learning, Age and Bilingualism Laboratory (LAB-lab), University of Turku, Finland. He is interested in the role of the mother tongue in speech sound perception and production. In addition, his interests include second language training research, bilingual perception and production as well as psychophysiology of speech processing.

**Henna Tamminen**, MA, is a University Teacher from the Department of Phonetics and Learning, Age and Bilingualism Laboratory (LAB-lab), University of Turku, Finland. She is interested in foreign language speech perception learning, bilingual perception, second language training research, psychophysiology of speech processing, and production learning of non-native speech.

**Teija Kujala**, PhD, is a Professor of Psychology at the University of Helsinki, Finland. Professor Kujala is Director of Cognitive Brain Research Unit (CBRU) and the Chair of the Research Infrastructure Platform Cognitive Brain Laboratory. Her current interests involve the neural basis of language, plasticity, and language-learning deficits and their amelioration.

**Paavo Alku**, Dr. Tech, is an Academy Professor from Department of Signal Processing and Acoustics, Aalto University, Finland. His research interests include a wide range of topics related to the speech communication technology, e.g. analysis and parameterization of voice production and HMM-based speech synthesis.

**Maija S. Peltola**, PhD, is an Associate Professor and 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.