

Adult Learners and a One-day Production Training – Small Changes but the Native Language Sound System Prevails

Kimmo U. Peltola

Department of Phonetics, Learning, Age and Bilingualism Laboratory (LAB-lab), University of Turku, Finland

Tomi Rautaoja

Department of Phonetics, Learning, Age and Bilingualism Laboratory (LAB-lab), University of Turku, Finland

Paavo Alku

Department of Signal Processing and Acoustics, Aalto University, Finland

Maija S. Peltola

Department of Phonetics, Learning, Age and Bilingualism Laboratory (LAB-lab), University of Turku, Finland

Abstract—Theories and models of non-native phonetic acquisition emphasise the significance of the native sound system in the learning process. On the basis of extensive research, it is evident that the mother tongue phonology prevents the learner from perceiving redundant sound contrasts, which are of importance in the target language. These phonological difficulties are naturally reflected in the production patterns as well, and learners tend to articulate according to the native model. In this study the aim was to see, whether a short production training results in articulatory changes. For this purpose we trained native speakers of Finnish and American English to produce a non-native vowel contrast embedded in a pseudo-word context /ty:ti/ - /tʌ:ti/. We analysed the first two formants from both the baseline and after-training productions. The results showed significant group differences in the formant frequencies and in their standard deviations. These findings indicate that both groups produced the pseudo-words according to the mother tongue model and that a short training was not enough to overcome the strength of the native language sound system. However, training did reduce deviations of the formant frequencies, which implies that the learners found a more homogeneous production pattern with training.

Index Terms—native language, production training, adult learners, vowel production

I. INTRODUCTION

Speech sound production and perception form the basis for oral communication. Children are exposed to the native sounds from birth, or even earlier (Partanen et al. 2013) and as a result of this the native speech sound categories begin to develop (Kuhl, 1991; Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992). After the establishment of language-specific memory traces for mother tongue phonological categories (Näätänen et al. 1997), perception is altered in accordance with the ambient system and non-native speech sounds and their acoustic cues become redundant. Production patterns then start to develop and articulatory gestures are acquired so that the output of production matches the acoustic model provided by the mother tongue and as a result new production templates are formed (Perkell et al. 1997). These templates are constantly reinforced by matching the articulatory sensation with the acoustic output and also by comparing the output against the auditory input by other speakers. When non-native speech is encountered, the system needs to make adaptations and this may be the main challenge in acquiring non-native speech sounds.

According to the formulations of the traditional Contrastive Analysis Hypothesis (CAH), the phonetic foreign language learning difficulties result from the categorical differences between the native and the target speech sound systems. In particular, cases of under-differentiation were considered especially problematic, since in these situations the learner is unable to perceive – and thus produce – a contrast that is phonemic in the target language but phonetic, i.e. redundant, in the native language (Weinreich, 1953; Lado, 1957). Modern theories and models of non-native acquisition maintain these ideas. According to the Speech Learning Model (SLM, Flege, 1987), the second language (L2) speech sounds can have three types of relations with the native language (L1): Firstly, the sounds can be *Identical*, in which case there are no learning difficulties. Secondly, the L2 sound can be totally *New*, and then production difficulties are at maximum in the beginning stages of learning, but they can be overcome with effort. Thirdly, the relation can be of the type *Similar*, which implies persistent learning difficulties, since the target sounds are misinterpreted as being the same as in the mother tongue. This case bears a close resemblance to the situation of under-differentiation by earlier theories. Despite the more motoric or gestural background of the Perceptual Assimilation Model (PAM, Best & Strange, 1992),

the same premises are valid in this formulation as well. The relation between the L1 and L2 sound system is in the core of the model, and the categories are described in terms of assimilatory patterns. Again as the *Identical* type in SLM, one pattern describes a setting, where there are no learning difficulties, i.e. when the phonemes contrasting in the L2 are assimilated into two separate L1 phonemes in the same manner. Also, intermediate difficulties are seen in cases where the L2 phonemes do not assimilate to any L1 categories and this is clearly closely related to the *New* category in SLM. Most importantly, there are two types of problematic assimilation patterns, firstly, a case where two L2 phonemes are assimilated unequally into one L1 category, and secondly, a pattern where these two L2 categories assimilate equally well into a single L1 phoneme. The last pattern is considered to result in major learning difficulties, since the learner does not even perceive category goodness difference between the two contrasting speech sounds. These two are linked with the type *Similar* in SLM and have their roots obviously in the concept on under-differentiation (Weinreich, 1953). Another approach to the role of the mother tongue in non-native speech learning is formulated by the Native Language Magnet Model (NLM, Kuhl, 1991). In the NLM framework, the emphasis is not on designing a typology for possible differences. Instead, the idea is that the perceptual magnets, i.e. category prototypes, developed in early infancy for the L1 categories, dominate perception and discrepancies between the prototypical exemplars in different languages result in misperceptions. In conclusion, irrespective of the theoretical formulation, it is evident that the role of the mother tongue phonology in non-native speech sound acquisition is significant and major difficulties arise when categorical distinctions are done in different ways.

Earlier research has shown that, despite the problems caused by the native system, learners are able to overcome the difficulties. Children have been shown to be plastic in perceptual reorganisation both when exposure is given in training (Giannakopoulou, Uther & Ylinen, 2013; Taimi, Alku, Kujala, Näätänen & Peltola, M. S., 2014a) or in early immersion programme (Cheour, Shestakova, Alku, Ceponiene & Näätänen, 2002; Peltola, M. S., Kuntola, Tamminen, Hämäläinen & Aaltonen, 2005). Child learners have also been shown to alter their production patterns with a *listen and repeat*-training (Taimi, Jähi, Alku & Peltola, M.S., 2014b), but senior learners are reported to also show production plasticity when trained with the same method (Jähi, Alku & Peltola, M. S., 2015). Adult learners have shown neural plasticity in cases of immigration (Winkler et al., 1999), classroom learning settings (Peltola, M. S., Tamminen, Toivonen, Kujala & Näätänen, 2012) and training in laboratory (Tamminen, Peltola, M. S., Kujala & Näätänen, 2015), and training has also resulted in improved behavioural perception scores (Iverson & Evans, 2009). Production learning has been reported in immigrants (Flege, Bohn, & Jang, 1997), classroom learners (Peltola, M. S., Lintunen & Tamminen, 2014) and training participants (Iverson, Pinet & Evans, 2012; Saloranta, Tamminen, Alku & Peltola, M. S., 2015; Peltola; K. U., Tamminen, Alku & Peltola, M. S., 2015). Within the training studies showing production learning in adult learners, several variations in the training methods emerge: Iverson, Pinet & Evans (2012) used a high variability training protocol for less and more experienced language learners and showed similar production plasticity for both groups. In contrast, Peltola, K. U., Tamminen, Alku & Peltola, M. S. (2015) utilised a *listen and repeat*-training system with an additional visual cue in the form of either transcription or orthography and only one acoustic exemplar per category. Their study showed that misleading orthography prevented production learning and that the visual channel seemed to dominate attention. Saloranta, Tamminen, Alku & Peltola, M. S. (2015) used the same stimulus setting as Peltola, K. U., Tamminen, Alku & Peltola, M. S. (2015), but in that study additional overt production instructions were provided and the results showed immediate production changes toward the acoustic target stimuli. It thus seems that, despite the difficulties imposed by the native system, learners may adopt new perceptual and production patterns in various learning settings.

The aim of this study was two-fold: Firstly, the objective was to see whether a very brief and simple *listen and repeat*-training protocol would enable learners to adopt new production patterns. Secondly and more importantly, the purpose was to investigate whether the mother tongue speech sound system would be reflected in the training outcomes. Earlier studies using the same stimulus set have utilised a two-day training protocol with twice the amount of exposure to the target stimuli (Peltola, K. U., Tamminen, Alku & Peltola, M. S., 2015; Saloranta, Tamminen, Alku & Peltola, M. S., 2015). These studies have shown clear learning effects, and therefore it is of interest to seek the minimum amount of training needed to induce production changes. In addition, these earlier studies have been conducted on native speakers of Finnish, and thus it is not possible to see the effects of the native system. In this study, the one-day protocol was identically presented not only to Finnish learners but also to native speakers of American English (AmE). The experiments were carried out in two countries to diminish the possibility of exposure to the target stimuli. In Finnish, the target stimulus /ʌ/ embedded in the pseudo-word context is non-phonemic and it assimilates either to the Finnish category /y/ or /u/. The other stimulus /y/ in the learning paradigm serves as a contrast vowel for Finns, since it is a native phoneme. For American English speakers, the setting is different: both stimuli are considered as actual targets, since the contrast is altogether non-phonemic in AmE. Thus the learning setting is significantly different depending on the mother tongue.

II. MATERIAL AND METHODS

A. Subjects

Two groups of adult subjects participated in the production experiments and training. All subjects gave a written consent prior to testing. Group 1 (native Finns) consisted of 10 native speakers of Finnish (mean age 24.6 years, 5 females). None of the subjects studied foreign languages nor had lived in any other Nordic country except for Finland. They all were exposed to English via both the school system and the media and they also had minimal exposure to Swedish, since all Finns are required to study basic Swedish skills in secondary school. However, the subjects self-reported their proficiency level to be 1.7 on average (scale 0=non to 4=excellent), indicating a rather low proficiency. In contrast, Group 2 (native AmE) contained 9 (mean age 42.6 years, 2 females) native speakers of Western dialect American English, spoken in Sacramento, California. The subjects self-reported to have no prior exposure to Swedish and 5 indicated to have some basic knowledge of Spanish. All subjects in both groups reported to have a normal hearing and none of them suffered from any linguistic deficits. The study was approved by the Ethics Committee of the University of Turku, Finland.

B. Procedure

The subjects participated in the study during one day and the whole procedure lasted for approximately 15-20 minutes. Native Finns were tested in a sound attenuated laboratory room (Learning, Age and Bilingualism laboratory, LAB-lab), where the subject sat alone. The stimuli were presented via Sanako Headset SLH-07 and registered with Sanako Lab 100 –soft/hardware. The native AmE subjects were tested using Sanako Study Students 7.20 (7.20.030508.00) software attached to a laptop (Della Latitude E7240) and Beyerdynamic MMX 300 headset. The recordings were not performed in laboratory conditions, but due to the high quality microphone, the obtained data was of extremely high quality and comparable to that obtained in LAB-laboratory.

The experiment began with background scanning (age, foreign language skills obtained in school, linguistic knowledge etc.) followed by a baseline recording during which the subjects listened to word stimuli and repeated according to the provided model. After the baseline recording, the subjects performed the training block which was followed by the second recording. Altogether, the protocol consisted of two recordings (baseline and final) and one training block. The recordings contained the target/non-target word /ty:ti/ 10 times and the target word /tɜ:ti/ 10 times in turns, while the training consisted of 30 repetitions of each. The interstimulus interval (ISI) was 3 seconds, during which the subjects imitated the stimulus words according to the provided auditory model. Since the experiment was short in duration, the test subjects did not show any signs of fatigue.

C. Stimuli

The stimuli were created using a semi-synthetic method, which allows for complete controlling of the acoustic quality of the stimuli. The glottal pulse extracted from a real speaker ensures the natural human sounding element (Alku, Tiitinen & Näätänen, 1999). The relevant quality difference in the stimulus words was on the first syllable vowel, while the carrier pseudo-word was fixed. The stimulus /ty:ti/ contained the rounded close front vowel /y/ (mid-point F1: 269 Hz, F2: 1866 Hz, F3: 2518 Hz), which is a frequent speech sound in Finnish, but does not form a phoneme category in English. Thus this stimulus was a non-target word for Finns, but a target for native AmE speakers. The stimulus word /tɜ:ti/ was a target word for both groups, since this rounded close central vowel /ɜ/ (mid-point F1: 338 Hz, F2: 1258 Hz, F3: 2177 Hz) does not form a category in either language. The fundamental frequency (F0) was 126 Hz in the first syllable vowel and the duration of both stimuli was 624 ms. The same stimuli were used in earlier studies and a more detailed description is given in Taimi, Jähi, Alku & Peltola, M. S. (2014b).

D. Analysis

The acoustic analysis involved measuring formant frequency (F1, F2, F3) and fundamental frequency (F0) values of all recorded words by all subjects. Speech data consisted altogether of 760 productions. From this, 400 productions were recorded from Finnish subjects (20 words x 10 subjects x 2 measurements) and 360 from native speakers of English (20 words x 9 subjects x 2 measurements). The formant frequencies were extracted using Praat software (version 5.4.21) and the frequencies were measured from the estimated steady-state phase visible in the spectrogram. In addition, we calculated the standard deviation value for each formant. The F1 and F2 values as well as the standard deviations for these formants were subjected to Analysis of Variance (ANOVA) using SPSS (IBM SPSS Statistics, version 22). Further post hoc tests were performed when appropriate. Fundamental frequency values were not statistically analysed in this study.

III. RESULTS

The formant frequency data was first analysed using a Group (2) x Session (2) x Word (2) x Measure (2) Repeated measures ANOVA, which merely suggested a tendency for interaction between Word and Group ($f(1,17)=4.220$, $p=0.056$). In order to examine this further, One-way ANOVAs were performed. These analyses showed a significant main effect of Group in the F1 values of the vowel /ɜ/ both in the baseline session ($f(1,18)=5.805$, $p=0.028$) and after training ($f(1,18)=8.700$, $p=0.009$) indicating that two groups produced the target vowel /ɜ/ with different F1 values to begin with and that the difference was evident also post-training. The Groups differed also in their production of the /y/ F1 value in the baseline measurement ($f(18)=5.191$, $p=0.036$), but this difference disappeared as a result of training. A closer examination of the data was executed by running t-tests on the groups separately. The analysis showed that the

native Finnish Group was able to produce the two words with different F1 values in the baseline measurement ($t(9)=2.309$, $p=0.046$) as well as after training ($t(9)=4.208$, $p=0.002$) and the same was valid for the F2 values in baseline ($t(9)=-5.925$, $p<0.001$) and post-training ($t(9)=-5.930$, $p<0.001$). The Group of American English speakers also produced the two target words with different acoustic patterns and the analysis for F1 reached significance both prior to ($t(8)=2.496$, $p=0.037$) and after ($t(8)=3.612$, $p=0.007$) training. This was the case also with the F2 values (baseline $t(8)=-4.599$, $p=0.002$) and post training $t(8)=-3.777$, $p=0.005$). These findings clearly showed that both groups were able to produce the stimuli with different acoustic qualities already before training. However, the groups differed in the manner in which they produced the target vowel / ʌ / since native Finns produced higher F1 values. In addition, it seems evident that the groups changed their productions slightly, since the pre-training difference in / y / F1 values disappeared as a function of training. The average formant values for both groups and sessions are shown in Table 1.

TABLE 1.

	Finnish group N=10				English group N=9			
	Min	Max	Mean	SD	Min	Max	Mean	SD
/ ʌ / S1F1	312	463	405.5	44.7	318	412	361.3	33.7
/ ʌ / S1F2	1027	1913	1486	264.8	1330	1892	1611.7	162.8
/ ʌ / S2F1	360	464	408.5	34.5	316	404	363.2	32.2
/ ʌ / S2F2	999	1944	1467.3	334.6	1473	1769	1621.9	89
/ y / S1F1	321	451	390	38.9	289	402	350.3	36.7
/ y / S1F2	1581	2113	1844.7	197.6	1630	2198	1841.2	176.9
/ y / S2F1	324	458	386.5	42.1	290	388	349.7	35.5
/ y / S2F2	1572	2162	1861.3	208.3	1669	2092	1847.4	147.8

Descriptive statistics of the formant values (Hz) of the both groups. S1 and S2 refer to the first and second recording session, respectively. F1 and F2 refer to the first and second formant, respectively.

The standard deviation analysis revealed some significant differences both between the sessions and groups. The Group (2) x Session (2) x Word (2) x Measure (2) Omnibus ANOVA revealed a Word x Measure interaction ($f(1,17)=12.651$, $p=0.002$) indicating that the standard deviation values in the target vowels were differently distributed over the two measured formants. More importantly, the main effect of Session was found ($f(1,17)=4.704$, $p=0.045$), which showed that the standard deviation values significantly changed as a result of training. The most interesting finding was the triple interaction between Group, Word and Measure, signaling a difference in the deviation values of the groups in the two words. In addition, a Word x Group interaction ($f(1,17)=9.168$, $p=0.008$) strengthened this existence of a group difference. A further Group (2) x Session (2) x Measure (2) analysis was performed on the two words separately. The analysis on the word / ty:ti / revealed a Group x Measure interaction ($f(1,17)=6.786$, $p=0.018$) and yet a further One-Way ANOVA analysis showed this to be due to a Group difference in the standard deviation values for F2 in the baseline ($f(1,18)=5.252$, $p=0.035$) and post-training ($f(1,18)=5.237$, $p=0.035$) measurements. This analysis showed no statistically significant differences in the standard deviation values of the target word / tʌ:ti /. Separate paired-samples t-tests for both Groups were performed in order to examine the findings in more detail. These analyses yielded no differences in the deviation within the English speaking Group. However, the Finnish Group showed a significant difference in the standard deviation of the F2 values for / ty:ti / and / tʌ:ti / both in the baseline ($t(9)=2.939$, $p=0.017$) and post-training ($t(9)=3.342$, $p=0.009$). It seems evident that the groups differed in the level of hesitation and this difference was connected with the F2 standard deviation values. Native Finns showed minor deviation values for the vowel / y / in comparison with the American English speakers, but the deviation values were similar in the target vowel / ʌ /. The average standard deviations are shown in Table 2.

TABLE 2.

	Finnish group N=10			English group N=9		
	Min	Max	Mean	Min	Max	Mean
/ ʌ / S1F1	10	56	22.4	10	26	15.8
/ ʌ / S1F2	42	524	195.2	43	166	119.7
/ ʌ / S2F1	11	22	15.6	7	22	15.1
/ ʌ / S2F2	59	264	125.2	52	219	94.1
/ y / S1F1	7	26	16.5	9	30	19.7
/ y / S1F2	28	90	53	28	282	111
/ y / S2F1	9	29	17	11	38	18
/ y / S2F2	21	103	57.5	61	142	81.7

Descriptive statistics of the standard deviation values (Hz) of the both groups.

S1 and S2 refer to the first and second recording session, respectively. F1 and F2 refer to the first and second formant, respectively.

IV. DISCUSSION

The present study was conducted in order to see, how native speakers of two non-related and phonetically distant languages learn a non-native speech sound contrast. The study utilised a well-documented learning protocol (e.g. Taimi,

Jähi, Alku & Peltola, M. S., 2014b; Peltola, K. U., Tamminen, Alku & Peltola, M. S., 2015; Saloranta, Tamminen, Alku & Peltola, M. S., 2015), where learners were trained to listen and repeat vowels embedded in pseudo-word contexts. In earlier studies trainings have been longer (two days and four training blocks) and the subjects have all been native speakers of Finnish. Thus the present study tested the potential effects of extremely short exposure of merely one training session and, more importantly, the role of different native languages. The hypothesis was two-fold: Firstly, it is possible that even a short training session may alter productions, which would indicate strong and quick plasticity. Secondly, it may be that the native language prevails and that more training is needed to overcome the mother tongue strength. The results of the analysis seemed to indicate both possibilities. The finding that the groups differed in their vowel /y/ F1 values in the baseline session but not after the training, may indicate that some learning has occurred. Also, the overall reduction of standard deviation in both groups suggests that the productions became less hesitant and the learners started to produce the vowels more homogeneously. In contrast, the general result that the groups produced different kinds of formant values altogether gives strong support on the power of the mother tongue. In addition, the fact that the standard deviation values were different in the vowel /y/ in the two groups gives further evidence of the native language strength. This finding shows clearly that the mother tongue sound system affects productions even after training, since Finns familiar with the vowel /y/ showed smaller deviations for this sound in comparison with the non-native /ʌ/. In line with this, the deviations were equally extensive for both target vowels in the American English speakers for whom both stimuli contained non-native vowels. Altogether, the results showed that the articulatory patterns responsible for the native speech sound production prevail and that a short *listen and repeat* training does not provide sufficient amount of practise for learning new motoric gestures. However, it seems evident on the basis of the standard deviation analysis that the subjects in both groups started to adjust their productions towards the model, thus indicating a beginning phase in motoric learning. This may suggest that, with more training, articulation could change. Most importantly, our results clearly show how significantly the native speech sound system affects the production of foreign speech. This finding has implications also to the way in which non-native speech production may be taught to native speakers of various languages. In addition, the results give rise to speculations concerning the way in which theories of non-native speech sounds acquisition could be modulated.

Theories of non-native speech sound acquisition (Flege, 1987; Best & Strange, 1992; Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992) argue that the manner in which the non-native categories are perceived on the basis of the mother tongue has a decisive role in the amount of learning difficulties encountered. This is certainly valid and several studies have indicated towards this interpretation in perceptual (e.g. Tyler, Best, Faber & Levitt, 2014) and production experiments (e.g. Peltola, M. S., Lintunen & Tamminen, 2014) and many studies have discussed both aspects of speech communication (Flege, Bohn & Jang, 1997; Ingram & Park, 1997). This is based on the idea that perception is difficult in cases of under-differentiation (Weinreich 1953, Lado 1957). In more modern terms, difficulties arise when the target speech sounds are of the type Similar according to the Speech Learning Model (Flege, 1987) or if the assimilation pattern is of the type Single Category assimilation or Category Goodness in Perceptual Assimilation Model (Best, 1995). However, when two languages such as Finnish and American English are contrasted, the setting may be less simplistic and the interpretation may be different depending on whether looking at perception or production. For Finns there are two alternatives on how the target vowel /ʌ/ is perceived: in terms of SLM, it may be Similar to /y/ or /u/, in the latter case it would be easy to discriminate from the contrasting stimulus /y/, but in the former case it would be extremely difficult to discriminate from /y/. From the point of view of PAM, the /y/ - /ʌ/ contrast may show a Single Category assimilation so that both are perceived as exemplars of /y/, but with different goodness ratings, or it could be of the type Two-Category Assimilation with /ʌ/ representing a poor version of /u/. In the latter case, no difficulties should arise, while in the former, extreme problems are anticipated. For native speakers of American English, the situation is different. Both vowels are of the type Similar in SLM terms and with PAM it can be analysed as representing a case of either Single Category assimilation or Both Uncategorizable. In the latter case intermediate difficulties are expected, while in the former severe problems should persist. From the point of view of the present study, it seems that the analyses need to be of the type where extreme difficulties should arise, since the new learning outcomes were almost non-existent. On the other hand – and most importantly – is it actually of significance to know which pattern of assimilation is correct, since the learners need to acquire production patterns in any case? It certainly is difficult, if the subjects do not even hear a difference in the two stimuli that need to be learned, but apart from this initial problem, it may well be that the main issue is the creation of a new motoric command pattern. In other words, whether the target vowel /ʌ/ is Similar, Single Category, Category Goodness or Uncategorizable, in order to learn to produce it according to the model, articulation needs to change. Thus it may be that the predictions of the models of non-native speech acquisition do not function as accurately in production learning as they do in perceptual learning.

Contrastive studies have the potential of providing further knowledge on how the native language may play a role in non-native speech learning. It may well be that different kinds of trainings may function better in speakers of different mother tongues and further study is needed to shed light into this question. However, it becomes clear from our study that the role of the mother tongue is strong, it governs the production of target speech sounds and short training cannot overcome the life-long exposure and established production patterns. However, training seems to start affecting productions slightly and thus even short exposure may be beneficial for students of foreign languages. Theoretically, it may be that while the connection of production and perception is evident, in non-native speech learning models the two

sides of the coin could in fact work differently: perceptual patterns explain persistent perceptual problems, but motor patterns need to be learned for all types of non-native sounds.

ACKNOWLEDGEMENT

The authors wish to thank all the subjects for participating in this study. We also want to express our gratitude to Center for Mind and Brain, UC Davis – and professor Steven Luck and professor Ron Mangun, in particular – for hosting a visit to their laboratory. We also wish to thank Sanako Corp., for sponsoring the LAB-laboratory. The first author is financed by doctoral program Utuling of University of Turku.

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. “doi: [http://dx.doi.org/10.1016/S1388-2457\(99\)00088-7](http://dx.doi.org/10.1016/S1388-2457(99)00088-7)”.
- [2] Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (eds.) *Speech perception and linguistic experience*, Baltimore: York press, 171–206.
- [3] Best, C. T. & W. Strange. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20, 305–330.
- [4] 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.
- [5] 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.
- [6] Flege, J.E., O.-S. Bohn & S. Jang. (1997). Effects of experience on non-native speakers’ production and perception of English vowels. *Journal of Phonetics* 25, 437–470.
- [7] 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.
- [8] Ingram, J. C. L. & S. Park. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics* 25, 343–370.
- [9] Iverson, P., M. Pinet & B. G. Evans. (2012). Auditory training for experienced and inexperienced second language learners: Native French speakers learning English vowels. *Appl. Psycholinguist.* 33, 145–160. “doi:10.1017/S0142716411000300”.
- [10] 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. “doi:10.1121/1.3148196”.
- [11] Jähi, K., P. Alku & M. S. Peltola. (2015). Does interest in language learning affect the non-native phoneme production in elderly learners? In The Scottish Consortium for ICPhS 2015 (eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 234.
- [12] 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.
- [13] Kuhl, P. K., K. A. Williams, F. Lacerda, K. N. Stevens & B. Lindblom. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606–608.
- [14] Lado, R. (1957). *Linguistics across cultures*. Michigan: University of Michigan Press.
- [15] Näätänen, R., A. Lehtokoski, M. Lennes, M. Cheour, M. Huotilainen, A. Iivonen, M. Vainio, P. Alku, R. J. Ilmoniemi, A. Luuk, J. Allik, J. Sinkkonen, & K. Alho. (1997). Language specific phoneme representations revealed by electric and magnetic responses. *Nature* 385, 432–434.
- [16] Partanen, E., T. Kujala, R. Näätänen, A. Liitola, A. Sambeth & M. Huotilainen. (2013). Learning-induced neural plasticity of speech processing before birth. *Proc Natl Acad Sci U S A.* 2013 Sep 10, 110(37): 15145–15150. Published online 2013 Aug 26. doi: 10.1073/pnas.1302159110. Psychological and Cognitive Sciences.
- [17] Peltola, K. U., H. Tamminen, P. Alku & M. S. Peltola. (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 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 236.
- [18] 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.
- [19] Peltola, M. S., H. Tamminen, H. Toivonen, T. Kujala & R. Näätänen. (2012). Different kinds of bilinguals – Different kinds of brains: The neural organisation of two languages in one brain. *Brain Lang.* 121, 261–266. “doi: 10.1016/j.bandl.2012.03.007”.
- [20] Peltola, M. S., P. Lintunen & H. Tamminen. (2014). Advanced English learners benefit from explicit pronunciation teaching: an experiment with vowel duration and quality. In: P. Lintunen, M.S. Peltola & M.-L. Varila (Eds.) *AFinLA-e Soveltavan kielitieteen tutkimuksia*, 6, 86–98.
- [21] Perkell, J. S., M. Matthies, H. Lane, F. Guenter, R. Wilhelms-Tricarico, J. Wozniak, & P. Guiod. (1997). Speech motor control: acoustic goals, saturation effects, auditory feedback and internal models. *Speech communication* 22, 227–250.
- [22] Saloranta A., H. Tamminen, P. Alku & M. S. Peltola. (2015). Learning of a non-native vowel through instructed production training. In The Scottish Consortium for ICPhS 2015 (eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 235.
- [23] Taimi, L., P. Alku, T. Kujala, R. Näätänen & M. S. Peltola. (2014a). The effect of production training on non-native speech sound perception and discrimination in school-aged children: An MMN and behavioural study. *Linguistica Lettica* 22, 114–129.
- [24] Taimi, L., K. Jähi, P. Alku & M. S. Peltola. (2014b). Children Learning a Non-native Vowel – The Effect of a Two-day Production Training. *Journal of Language Teaching and Research* 5, 1229–1235.

- [25] Tamminen, H., M. S. Peltola, T. Kujala & R. Näätänen. (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(2015), 23–29.
- [26] Tyler, M. D., C. T. Best, A. Faber & A. G. Levitt. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71, 4–21.
- [27] Weinreich, U. (1953/1963). Languages in contact. Mouton, London.
- [28] Winkler, I., T. Kujala, H. Tiitinen, P. Sivonen, P. Alku, A. Lehtokoski, I. Czigler, V. Csépe & R. Näätänen. (1999). Brain responses reveal the learning and of foreign language phonemes. *Psychophysiology* 26, 638–642. “doi: 10.1159/000356237”.

Kimmo U. Peltola, MA, is a PhD candidate 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.

Tomi Rautaoja, MA, is a PhD candidate from the Department of Phonetics and Learning, Age and Bilingualism Laboratory (LAB-lab), University of Turku, Finland. His research is focused on the role of the mother tongue in speech sound perception and production.

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.