International Journal of Psychophysiology 97 (2015) 23–29

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

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

Henna Tamminen [a,b,c,*], Maija S. Peltola [a,b,c], Teija Kujala [d,e], Risto Näätänen [e,f,g]

[a] Department of Phonetics, FIN-20014, University of Turku, Finland
[b] Centre for Cognitive Neuroscience, FIN-20014, University of Turku, Finland
[c] LAB-lab, Department of Phonetics, FIN-20014, University of Turku, Finland
[d] Cicero Learning Network, P.O. Box 9, FIN-00014, University of Helsinki, Finland
[e] Cognitive Brain Research Unit (CBRU), P.O. Box 9, Cognitive Science, Institute of Behavioural Sciences, FIN-00014, University of Helsinki, Finland
[f] Institute of Psychology, University of Tartu, Estonia
[g] Center of Functionally Integrative Neuroscience (CFIN), University of Århus, Denmark

ARTICLE INFO

Article history:
Received 15 October 2014
Received in revised form 23 April 2015
Accepted 24 April 2015
Available online 6 May 2015

Keywords:
Training
Speech perception
Phonological processing
Mismatch negativity (MMN)

ABSTRACT

Language-specific, automatically responding memory traces form the basis for speech sound perception and new neural representations can also evolve for non-native speech categories. The aim of this study was to find out how a three-day phonetic listen-and-repeat training affects speech perception, and whether it generates new memory traces. We used behavioural identification, goodness rating, discrimination, and reaction time tasks together with mismatch negativity (MMN) brain response registrations to determine the training effects on native Finnish speakers. We trained the subjects the voicing contrast in fricative sounds. Fricatives are not differentiated by voicing in Finnish, i.e., voiced fricatives do not belong to the Finnish phonological system. Therefore, they are extremely hard for Finns to learn. However, only after three days of training, the native Finnish subjects had learned to perceive the distinction. The results show striking changes in the MMN response; it was significantly larger on the second day after two training sessions. Also, the majority of the behavioural indicators showed improvement during training. Identification altered after four sessions of training and discrimination and reaction times improved throughout training. These results suggest remarkable language-learning effects both at the perceptual and pre-attentive neural level as a result of brief listen-and-repeat training in adult participants.

1. Introduction

Automatically responding language-specific memory traces (Näätänen et al., 1997) are the basis for speech sound perception. Categorical perception and the ability to discriminate between different speech sounds are also crucial for speech perception, discrimination being easier at the category boundary area and harder within the category (Liberman et al., 1957). Native language speech sound representations and categories evolve already in early childhood (Cheour et al., 1998), but new neural representations can also develop for non-native speech categories, e.g., in an authentic environment in the case of immigrants (Winkler et al., 1999), in classroom learning (Peltola and Aaltonen, 2005; Peltola et al., 2012), and in early immersion (Peltola et al., 2005). At least in the beginning of the learning process, second language is perceived through mother tongue categories, which makes foreign language learning particularly demanding. This difficulty may be explained through the Native Language Magnet effect (NLM) (Kuhl, 1991) which describes how the prototypical representatives of phoneme categories hinder discrimination near them and facilitate across category discrimination. Native speech sound categories and their prototypes may be located so that a non-native language category boundary is positioned at that same place as the prototype or in its immediate vicinity. A situation like this would cause problems for a language learner, since the native language prototype region would quite probably cause discrimination problems in perception and thus result also in production difficulties.

Neural mechanisms of auditory learning can be studied with the mismatch negativity (MMN) response, a component of the event-related potential. It is an excellent tool for investigating memory traces and speech perception at the pre-attentive level. MMN is automatically elicited by infrequent, or deviant, stimuli among frequent, or standard, stimuli. MMN reflects discrimination accuracy and its amplitude is

---

[*] Corresponding author at: Department of Phonetics, FIN-20014, University of Turku, Finland. Tel.: +358 2 333 5305.

E-mail addresses: henna.tamminen@utu.fi (H. Tamminen), maija.peltola@utu.fi (M.S. Peltola), teija.m.kujala@helsinki.fi (T. Kujala), risto.naatanen@helsinki.fi (R. Näätänen).

http://dx.doi.org/10.1016/j.ijpsycho.2015.04.020
0167-8760/

connected to discrimination results. It also reflects plastic changes caused by learning (for a thorough review on the MMN see e.g., Kujala and Näätänen, 2010). The MMN responses to native language contrasts are larger in amplitude and earlier in latency than those for foreign language contrasts, which are irrelevant in the mother tongue, as shown by Näätänen et al. (1997) in their research where MMN responses were elicited for Finns and Estonians for their mother tongue contrasts but not for Finns for the foreign Estonian contrast. Winkler et al. (1999) similarly showed MMN responses to a Finnish contrast in Finns but not in naïve Hungarians, but they also showed responses to the Finnish contrast in Hungarians fluent in Finnish. Speech discrimination studies often combine behavioural tests with psychophysiological measures and, for example, in Näätänen et al. (1997), the identification result nicely showed a lack of a category in the area of the Estonian vowel /õ/ for Finns while, of course, there was one for the Estonians. The study by Winkler et al. (1999) showed consistent results in the reaction times and the MMN. However, if the acoustic differences between the stimuli are vast, the influence of the native language disappears (Peltola et al., 2003). On the other hand, second language learning may lead to native-like responses (e.g., Peltola and Aaltonen, 2005; Peltola et al., 2012; Winkler et al., 1999).

Neural plastic changes take place when we learn or train to perceive something new. Training induced MMN can be elicited for frequency differences in general (e.g., Näätänen et al., 1993; Menning et al., 2000; Atienza and Cantero, 2001; Kujala et al., 2001b). More importantly, training effects have been shown in studies using linguistic stimuli (e.g., Kraus et al., 1995; Tremblay et al., 1997, 1998; Menning et al., 2002; Tremblay and Kraus, 2002). Many of these studies have used some type of discrimination training (Näätänen et al., 1993; Kraus et al., 1995; Menning et al., 2000; Atienza and Cantero, 2001) or identification training (Tremblay et al., 1997, 1998). For example, one week discrimination training with synthetic speech stimuli varying in voice onset time (VOT) altered the neurophysiologic responses in the study by Kraus et al. (1995). In that study a same-different two-alternative forced-choice discrimination with visual feedback was used and there were six 1-hour training sessions between the pre- and post-training testing. Similarly, Tremblay et al. (1997) showed results in a nine day identification training study with synthesised speech stimuli varying in VOT. Visual feedback was used also in this study. The effects were seen in MMN duration and area increment as well as in discrimination and identification scores. Moreover, the training effects transferred to untrained stimuli with a different place of articulation, which was seen in the decreased onset latency of the MMN. The changes were observed after nine 20 min sessions of training in five days; pre- and post-training testing took two days each (Tremblay et al., 1997). In another study Tremblay et al. (1998) showed that changes resulting from VOT identification training with visual feedback can be seen in neural activity before behavioural learning. The training effects in a period of ten days were seen by the fourth day in the MMN, however, the occurrence of the behavioural changes was individual and the changes were seen either on the same day as in the MMN or during the following days (Tremblay et al., 1998). Also Menning et al. (2002) showed an increase in the behavioural performance together with an increase in the amplitude of the Mismatch Negativity Field (MMF) when the German participants were trained Japanese mora-timing. Forced-choice, two-alternative, self-adjusting staircase method discrimination training took place during ten consecutive workdays for approximately 1.5 h per day. Here, as well, visual feedback was given. Reaction times improved rapidly, already in the first session, whereas the MEG was recorded only before and after the training sessions and hence, the results were seen after 10 days. To sum up, these training studies using speech stimuli trained subjects with identification or discrimination methods. The training periods were 7–10 days consisting of several sessions (4–10) lasting approximately 20 min minimum to 1.5 h maximum per day (one study did not report the session time). Feedback was given in all studies. Training was executed either between pre- and post-testing or mixed with testing. Training effects were found in each study. All in all, it is fair to conclude that laboratory training can lead to learning, or more specifically, to "robust, linguistically-functional learning" (Bradlow, 2008, p. 299), even in difficult learning settings.

The goal of the present study was to determine how a listen-and-repeat training of foreign language words — or more precisely, a feature which is phonologically relevant in the foreign language but not in the mother tongue — affects neural and perceptual plasticity. In other words, the aim was to see how the listen-and-repeat training affects the formation of new memory traces and the perception of foreign language items. In order to study this, we measured the MMN response. We also used behavioural tests to determine whether the category boundary, goodness of the category exemplars, discrimination sensitivity, and reaction time are simultaneously affected by the same training.

Our training stimuli were two synthesised English words which pair up as a minimal pair — '*feel*' /fiːl/ and '*veal*' /viːl/ — differing only in VOT in the first labiodental fricative phoneme segment, the former being voiceless and the latter voiced. There is no such distinction in Finnish, and only the unvoiced /f/ phoneme is present in its sound system. In addition, Finnish uses no acoustic cue of voicing in any contrast, so the distinction is based on a totally new parameter. As both the English /f/ and /v/ assimilate to the Finnish /f/ — though unequally (/f/ better than /v/) — according to the Perceptual Assimilation Model (PAM) (Best and Strange, 1992), the discrimination of these sounds is expected to cause difficulties. Finns perceive the English /v/ as a poor representative of the Finnish /f/ and not as a representative of a different category. Also, according to the Speech Learning Model (SLM) (Flege, 1987), second language sounds which are similar, not identical or new, to native categories are to cause severe difficulties. Hence, the English /v/ is problematic for a Finnish learner who now has to perceive this voicing difference; the English /v/ is neither identical (in which case there would not be any problems) nor is it totally new (in which case there would be intermediate difficulties) to a Finnish language learner. What makes the situation even more difficult, is the fact that Finnish has a voiced labiodental approximant /ʋ/ which differs from the English /v/ by the manner of articulation. The Finnish orthography is transparent and nearly phonemic, so in writing, these two phonemes (Finnish /ʋ/ and English /v/) are represented by the same grapheme <v> and, because of this, Finnish learners of English quite often use the Finnish approximant sound instead of the correct fricative one. The transparent Finnish orthography links certain kinds of acoustics to a particular grapheme, but in this case the grapheme <v> contains acoustic properties completely alien to Finns (Peltola, 2004). On the other hand, seeing words like '*feel*' and '*veal*' written as two different words, may provide some help, for example, in a categorisation task. Nevertheless, the contrast is difficult for Finns (naïve and English learners) since when Finnish is considered, they are not required to perceive voicing concerning fricatives, whereas the opposite is true in the case of the English language (see Bradlow (2008) for a similar comparison of Japanese and English).

The main difference in our study, compared to many training studies, is that our subjects trained with an articulatory listen-and-repeat training (a similar method is widely used in schools at foreign language classes, but it may also be combined with feedback), and this training took just a few minutes per day during only three consecutive days. In addition, in this study we provided no feedback, the amount of training was notably small and the participants were tested and trained every day. Perception of the unfamiliar voicing contrast is hypothesised to be challenging before training. In other words, /v/ is part of the category /f/ for Finns and thus the English /f/–/v/ category boundary is within the acoustic area of the Finnish /f/ prototype. It can be hypothesised that before training this is reflected in many ways: in the identification test the /v/ category should be smaller than /f/ and in the goodness rating it should be rarely rated as a good category representative; discrimination may be harder and slower, and the MMN response may be very small. However, it is expected that our training has some effects resulting in changes seen in the shifting of the category boundary, in the goodness ratings of the category boundary and /v/ category members in particular, and in discrimination sensitivity as well as in faster reaction times. Consistent with this, the MMN response should grow in amplitude during the training process, if new memory traces evolve during training.

2. Materials and methods

2.1. Subjects and stimuli

The participants were 12 native Finns (age range 18–32, mean 23.4 years, 7 females), who had not studied any languages after upper secondary school, and until which they had been studying English on average 8.8 years (range 5–11 years). English is usually taught at schools from the third grade (age 9) onwards in Finland. Despite this, none of the participants are considered bilinguals. All subjects were right-handed (tested with Edinburgh Handedness Inventory (Oldfield, 1971)), neurologically healthy, and had normal hearing (tested prior to experiments with an audiometer with perceptually relevant frequencies 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz). Each subject gave an informed consent prior to participating in the experiments. One subject had to be excluded from the identification and goodness rating results and another from the discrimination results due to unreadable data.

The stimuli consisted of 15 variants of the words /fiːl/ '*feel*' and /viːl/ '*veal*' (synthesised using HLSyn software 1.0 Sensimetrics Inc.). The stimuli varied only in the VOT of the first sound — from entirely voiceless to completely voiced fricative in 14 ms steps. The duration of each stimulus was 499 ms and from 197 ms onwards, they were identical, i.e., the vowel and lateral parts were identical in each stimulus. This stimulus continuum was used in the identification (ID) and goodness rating (GR) tests. The stimuli used in the discrimination and reaction time (RT) experiment, in the training and in the MMN registrations were two words from the stimulus continuum. Native English speakers had classified these two stimuli as belonging to different categories, the individual boundaries were approximately at the midpoint between the training stimuli. The stimuli were non-prototypical representations of the categories because the difference between prototypical stimuli would have been too obvious, leading to ceiling effects as in Peltola et al. (2003). The representative of the /fiːl/ category had a VOT of 113 ms and the representative of the /viːl/ category had a VOT of 71 ms.

2.2. Behavioural experiments

The ID and GR experiments were a forced choice task where the subjects heard (via headphones, Sennheiser HD25) each of the 15 stimuli 8 times in random order and were instructed to identify them as /fiːl/ or /viːl/ by pressing a button labelled "feel" or "veal". After each identification they were asked to evaluate the goodness of the word on a 1–7 scale where 1 was poor and 7 excellent. This behavioural task was self-paced with no feedback and there were two pauses during the test. All ID and GR tests started with a short familiarisation block during which the subject heard all 15 stimuli once in a random order and labelled and rated them; the familiarisation was not included in the actual data.

The subjects also participated in a discrimination and RT experiment where the stimuli (see Fig. 1) were presented in the oddball paradigm where /fiːl/ was the standard and /viːl/ the deviant stimulus. Interstimulus interval (ISI) was 1000 ms and deviant probability was 0.13 (130 standards and 20 deviants). The subjects were instructed to press a button as soon as they heard the deviating stimulus and there was no feedback in this test either. This experiment started with a short familiarisation, also not included in the actual data.

In the self-paced training experiment the stimuli were presented via headphones so that /fiːl/ and /viːl/ stimuli alternated. Altogether, there were 60 stimuli in one training session. The amount of training is smaller than in earlier studies where it has been between approximately 1000 and 2000 trials (Kraus et al., 1995; Menning et al., 2002; Tremblay et al., 1997, 1998). No feedback was given during any of the training sessions. The subjects were instructed to listen and repeat the stimuli very carefully. In other words, a stimulus (e.g., /fiːl/) was presented and the participant repeated it as carefully as possible, after which by pressing a button, the experiment continued and the next stimulus (in this case, /viːl) was presented. The productions in the training sessions were recorded with a microphone (AKG D660S) at approximately 20 cm from the subject. One training session lasted a few minutes, depending on the individual pace of the subjects.

2.3. ERP measurements

In the MMN registrations the stimuli were presented in the oddball paradigm where ISI was 650 ms and deviant probability was 0.13 (783 standards, 120 deviants). Standards following deviants where excluded from the analysis. The EEG was registered with 21 Sn electrodes (Electro-Cap International, Inc.) using Synamps amplifier (sampling rate 250 Hz; bandwidth 0.5–70 Hz). The electrodes used in the recording were Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, LM, and RM. Eye movements were monitored with electrodes attached below and near the outer canthus of the right eye, and impedance was kept under 5 kΩ. The subjects watched a silent non-subtitled movie while sitting in a comfortable armchair.

2.4. Procedure and analysis

The experiments were conducted on three consecutive days as follows:

Day 1) ID and GR, discrimination and RT, MMN registration, and training

Day 2) training, ID and GR, discrimination and RT, MMN registration, and training

Day 3) training, ID and GR, discrimination and RT, and MMN registration.

The order of the discrimination and RT experiment and the MMN recording was counterbalanced between the subjects. During these three days, the subjects trained each stimulus 120 times altogether. Training was performed twice on day 2 to enhance the amount of input of the stimuli. The training and testing session time per day was maximally 2 h. The experiments were conducted according to the guidelines defined by the Ethics Committee of the University of Turku, Finland.

The identification data were subjected to logit transformation analysis (SPSS) which automatically provides both the category boundary locations and steepness values. The category boundary is the cross-over point in the continuum where the distribution of answers is 50%; the steepness value is the calculated tilting of the response curve from the beginning point to the end point in the continuum. The steepness value for the boundary indicates the consistency of the subject's answers. These data were statistically analysed with a Repeated measures analysis of variance (ANOVA). The goodness rating results were statistically analysed for five stimuli (category boundary, training stimuli, and prototypes of both categories) by a Session (3) × Stimulus (5) ANOVA. All ID data were checked prior to statistical analyses in order to rule out potential outliers. All presented data are within the pre-experiment set criteria according to which subjects not showing a category boundary would have been excluded.

The discrimination data consisted of hits, misses, false alarms, and correct rejections according to which d′ values were calculated as follows $d' = z(H) - z(F)$ (H = hits, F = false alarms) (Macmillan and Creelman, 1991). The RTs were measured from the onset of the deviant stimulus, and button presses within ±3 standard deviation were

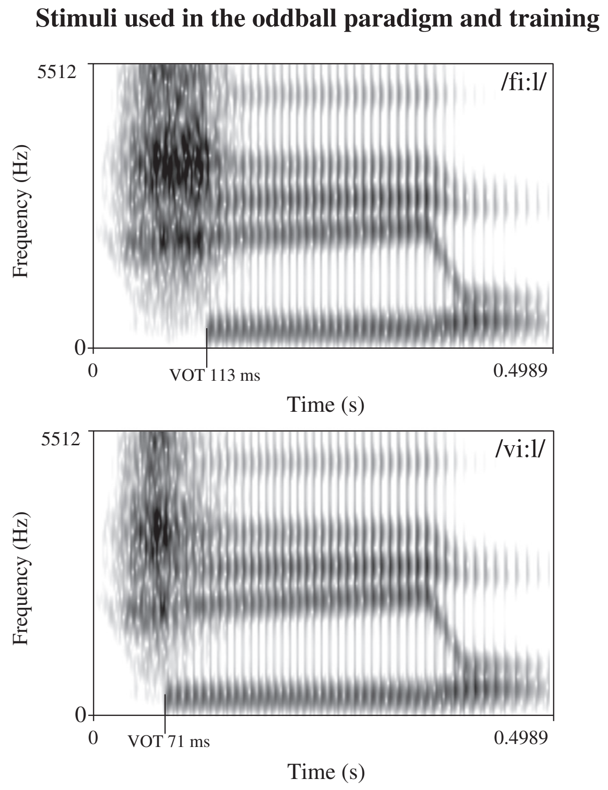


Fig. 1. The spectrograms of the two stimuli used in the discrimination task, in training and in the MMN recordings. Voicing starts at 113 ms in /fiːl/ (top) and 71 ms in /viːl/ (bottom).

included in the analysis. These d′ and RT values were separately subjected to a Repeated measures analysis of variance.

The EEG was filtered off-line with a 1–30 Hz bandpass filter with an artefact rejection criterion set at ± 100 μV and divided into 550 ms epochs including a 50 ms pre-stimulus period during which baseline was corrected. The average number of accepted trials in the analysis was 107 out of 120 (range 85–120). Epochs were separately averaged for the standard and deviant stimuli for each subject and difference waveforms were created by subtracting the response elicited by the standard stimulus from that by the deviant stimulus. Two consecutive time windows were selected around the maximum amplitudes in the grand average difference waveforms in Fz and Cz (300–340 ms and 340–380 ms). Since the MMN is most prominent at the fronto-central scalp areas (e.g., Kujala et al., 2007) we selected Fz, Cz, F3, F4, C3, and C4 electrodes for the statistical analyses. One sample t-test was carried out to determine whether the MMN significantly differed from zero. The MMN data were then statistically analysed using a Session (3) × Time Window (2) × Electrode (6) ANOVA (sphericity assumed). Post hoc tests were performed where appropriate.

3. Results

3.1. Behavioural results

The results of the identification and goodness rating tests are shown in Fig. 2. The statistical analysis verified that the category boundary location changed during the three-day training period as there was the main effect of session ($f(2,20) = 5.561$, $p = 0.012$). Paired sample test showed that the category boundary location was significantly different between sessions 1 and 3 ($t(10) = 3.265$, $p = 0.009$). The category boundary location was near the /viːl/ training stimulus (stimulus number 10) in the baseline measurements, but shifted towards the /fiːl/ category approximately between the two training stimuli (stimuli 7 and 10). The steepness of the boundary also changed during training since there was the main effect of session ($f(2,20) = 6.42$, $p = 0.007$). Further, paired sample t-test showed that the steepness was significantly different between sessions 1 and 2 ($t(10) = 2.421$, $p = 0.036$) and between sessions 1 and 3 ($t(10) = 2.88$, $p = 0.016$). There were no systematic differences in the GR test. The mean category boundary and steepness values of each session are shown in Table 1.

Discrimination sensitivity also changed significantly during training since there was the main effect of session ($f(2,20) = 9.962$, $p = 0.001$). Paired sample test showed that the discrimination sensitivity differed between sessions 1 and 2 ($t(10) = -2.944$, $p = 0.015$), between sessions 2 and 3 ($t(10) = -2.426$, $p = 0.036$), and most importantly, between sessions 1 and 3 ($t(10) = -3.382$, $p = 0.007$). The RTs, too, changed (the main effect of session ($f(2,20) = 15.796$, $p < 0.001$)). Furthermore, the RTs differed between sessions 1 and 2 ($t(10) = 2.749$, $p = 0.021$), between sessions 2 and 3 ($t(10) = 3.963$, $p = 0.003$), and most significantly, between sessions 1 and 3 ($t(10) = 4.726$, $p = 0.001$). The mean values of the discrimination sensitivity and the RTs of each session are presented in Table 1.

3.2. ERP results

The MMN results are shown in Fig. 3 and Table 2. One-sample t-tests were carried out first to determine whether the MMN response at Fz and Cz significantly differed from zero. The analyses showed that the MMN was not statistically significant in the first session in either time window. However, in the second session, it reached significance in

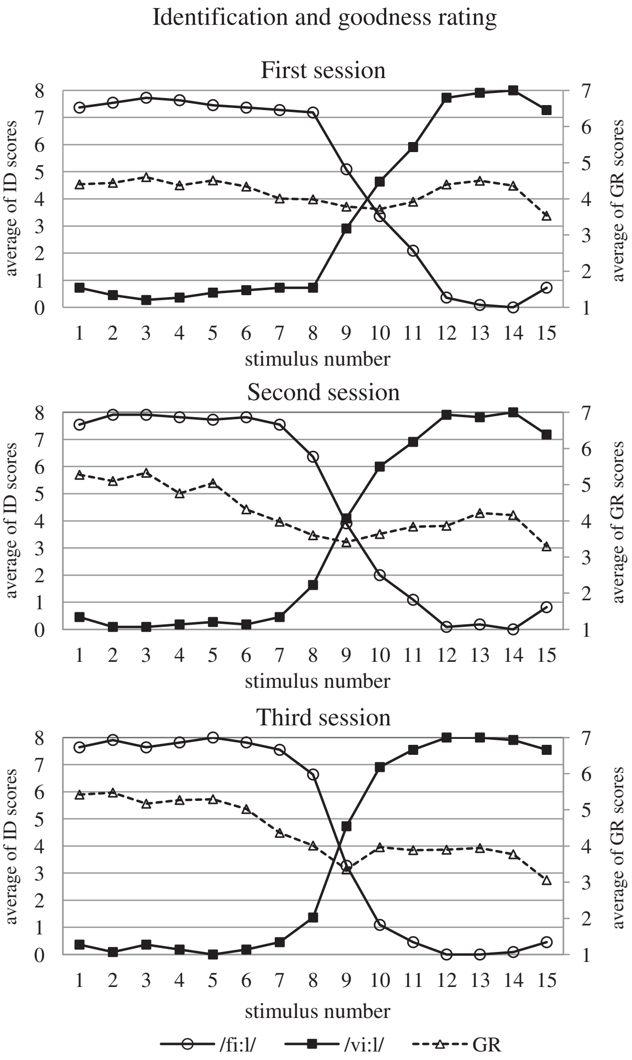


Fig. 2. The identification and goodness rating scores. The *X*-axis shows the stimulus continuum with 15 exemplars, where number one was the voiceless one and fifteen was the totally voiced one; the left *Y*-axis shows the number of the occasions the stimulus was identified as a member of each category (max 8 times); and the right *Y*-axis represents the goodness ratings on scale 1–7 (1 = poor, 7 = excellent). The change in the location and steepness of the category boundary is seen most clearly between the first and the third sessions.

both time windows. In the third session the MMN was significant in the second time window.

When both time windows were included in an ANOVA there was the main effect of session (f(2,22) = 3.633, p = 0.043), suggesting an amplitude increase, visible also in Fig. 3 and Table 2, as a function of training. The main effect of Time Window (f(1,11) = 16.583, p = 0.002) suggested that the MMN amplitude was different in the two time windows. These findings are in line with the t-tests showing that the MMN was non-existent in session 1, but prominent in sessions 2 and 3. There was also the main effect of Electrode (f(5,55) = 2.621, p 0 0.034) suggesting MMN amplitude differences between the electrode sites. There were no interactions between any of the factors.

Table 1
The mean category boundary location, boundary steepness, discrimination sensitivity (d′), and reaction time (RT) values. Standard deviations are in brackets.

	ID		Discrimination	
	Boundary	Steepness	d′	RT
First session	9.4 (0.78)	1.02 (0.55)	3.49 (0.88)	749 (117)
Second session	9.1 (1.12)	1.31 (0.61)	4.18 (0.37)	677 (97)
Third session	8.8 (0.92)	1.53 (0.65)	4.42 (0.19)	610 (90)

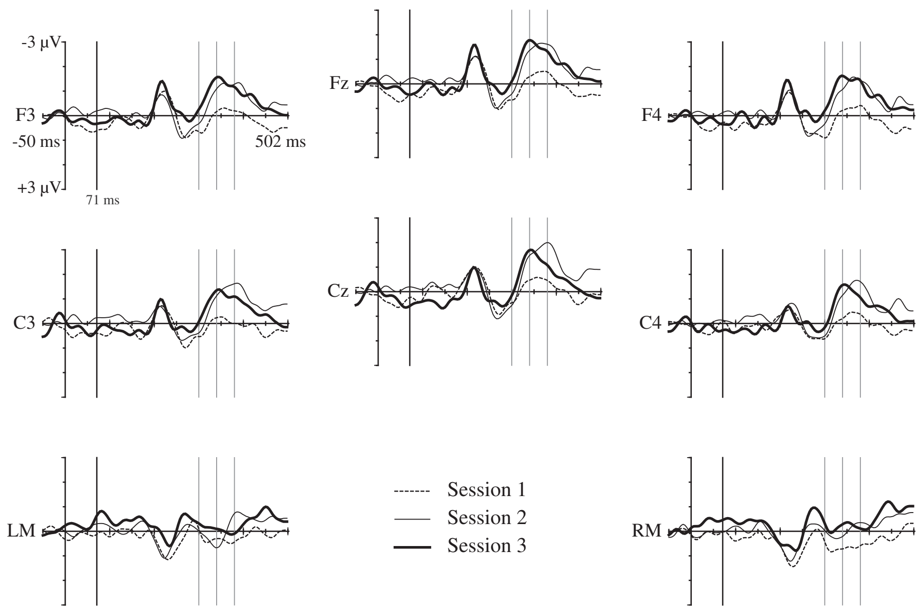
4. Discussion

The aim of this study was to find out how a three-day listen-and-repeat training with linguistic stimuli affects the perception of foreign phonemes containing phonological features not found in the mother tongue. Our results showed how the three-day phonetic training, with only 240 repetitions of the trained words, resulted in changes in both behavioural and pre-attentive level. This was seen in the category boundary location shift towards the /fiːl/ category enlarging the /v/ category and the boundary consistency improvement by training. It was further seen in the improvement of the discrimination sensitivity and in the RTs which became faster as the subjects learned the trained contrast. Finally, the training effects were seen in the enhancement of the MMN amplitude. Thus, the adult participants learned to perceive the difficult voicing contrast, not belonging to their native language, during only four short sessions of training on three days, and this was accomplished without any feedback. The fact that our training with the 240 repeated words resulted in behavioural and neural changes indicates that with this amount of training it is possible to see the development of a new neural memory trace and the change in behaviour.

Category boundary location shifted after four training sessions. This boundary transition is a minute one, but it is clearly a systematic change. Phonologically it is of significance, since the boundary tends to be at around the centre between the trained stimuli (7 and 10). The category boundary in the baseline session is phonemically significantly in a less optimal position than in the final session. Category boundary steepness, d′ and RTs, on the other hand, changed already after two training sessions; the change in steepness reached the maximum effect at this point, while d′ and RT continued to improve as there were significant differences between sessions 2 and 3. Despite the fact that there were no statistically significant changes in the goodness ratings, because training did not alter the within category hierarchy, it is possible to state that behaviour changed during the training process. In fact, the result that the goodness ratings of the /f/ category members did not change was an expected finding, since they are probably identical with the Finnish ones, even though the context word is not familiar. The finding that the goodness ratings of the /v/ category did not change was not as expected, but it was reasonable since learning of category hierarchy probably takes more time and needs potentially more variance in the presented stimuli. Training induced changes in goodness rating may demand a more hierarchically structured category which may develop slowly while discrimination sensitivity may arise prior to that. This has been suggested for memory trace formation as well (Peltola et al., 2007). Some sort of “high variability” training approach (see e.g., Bradlow, 2008), where the participants are exposed to a large set of varying stimuli from both categories, could be a better training method for creating a native-like hierarchical category.

In addition to this behavioural plasticity, the effects of the training were evident in the growth of the MMN amplitude, suggesting the generation of memory traces for the trained stimuli. The baseline MMN response was non-existent. After only two training sessions, a response evolved and it was also prominent in the last session (shown by the fact that MMN differed from zero in sessions 2 and 3). The enhancement of the MMN amplitude during the training suggests a formation of a memory trace for the trained sounds, which leads to an enhanced accuracy in discriminating these sounds from one another (Kujala and Näätänen, 2010). The memory traces are the basis for the perceptual changes shown by the behavioural tests. Changes were seen simultaneously in the category boundary steepness, discrimination sensitivity, RTs and MMN amplitude already in the second session after two training sessions. The category boundary location changed the latest, in the

Grand average difference waveforms



Grand average standard and deviant waveforms

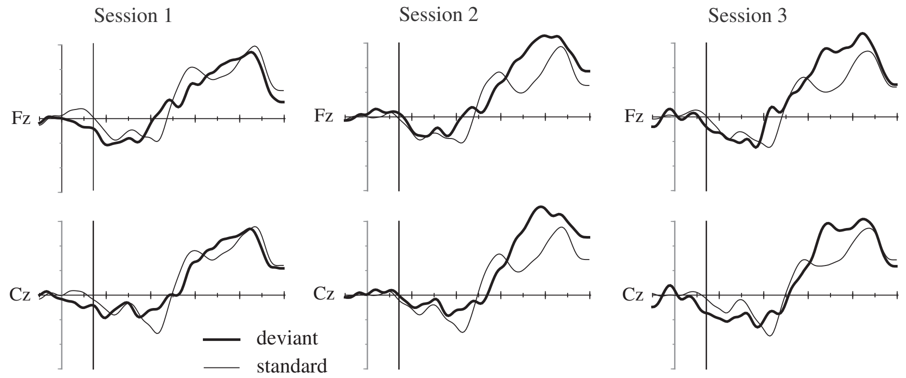


Fig. 3. The grand average difference waveforms from the three sessions. The difference of the stimuli starts at 71 ms which is shown by the vertical line at each electrode site. The time windows marked in the figure are at 300–340 ms and 340–380 ms. The figure shows the non-existent response in the first and the prominent responses in the last two sessions at the frontal and central electrode sites Fz, F3, F4, Cz, C3, and C4. The reversed polarity is seen stronger in the left than in the right mastoid.

third session after four training sessions. Altogether, it is striking how, despite the highly difficult nature of the contrast to be trained and learned, the MMN evolved and the behavioural indicators changed at such a speed during our training sessions and even without any feedback. The behavioural discrimination ability and the MMN response developed in synchrony, which supports the suggestion that the MMN reflects perceptual accuracy and neural plasticity (Kujala and Näätänen, 2010), and is in accordance with a large number of studies (e.g., Amenedo and Escera, 2000; Kujala et al., 2001a). It is fairly reasonable to hypothesise that the production would not change significantly during this training time since perception learning usually precedes production learning (see e.g., Flege, 1993).

Certainly, in addition to the listen-and-repeat training per se, the subjects were exposed to the stimuli during the testing in the protocol

Table 2
The mean MMN amplitudes for two time windows (300–340 ms, 340–380 ms) for all sessions. Standard deviations are in brackets. One sample t-test results at Fz and Cz.

	Fz		Cz		F3		F4		C3		C4	
	300–340	340–380	300–340	340–380	300–340	340–380	300–340	340–380	300–340	340–380	300–340	340–380
First session	0.145	−0.452	−0.079	−0.525	0.285	−0.247	0.268	−0.323	0.108	−0.140	0.114	−0.387
	(1.047)	(1.256)	(1.203)	(1.176)	(1.062)	(1.314)	(1.071)	(1.331)	(1.073)	(1.425)	(1.102)	(1.079)
Second session	−0.600*	−1.580**	−0.561*	−1.777**	−0.451	−1.258	−0.528	−1.417	−0.440	−1.456	−0.413	−1.565
	(0.777)	(1.407)	(0.847)	(1.741)	(0.816)	(1.076)	(0.703)	(1.254)	(0.868)	(1.364)	(0.779)	(1.478)***
Third session	−1.110	−1.517**	−0.953	−1.376*	−1.019	−1.340	−1.060	−1.484	−0.846	−1.187	−0.844	−1.341
	(2.002)	(1.581)	(1.905)	(1.691)	(1.982)	(1.521)	(1.882)	(1.548)	(1.925)	(1.675)	(1.730)	(1.684)

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

as well. They heard the whole stimulus continuum in the ID and GR test and the training stimuli during the discrimination test and the EEG registration, and this may have an impact on the results. Hence, the whole research protocol (all the tests and the training) provided the sufficient amount and type of training to result in behavioural and psychophysical changes. Perhaps it could even be argued that the mere exposure to the stimuli during the testing could have been enough to elicit the effect. Nevertheless, the small amount of exposure was sufficient. A follow-up study, however, would show the effectiveness of the training protocol in the long run. It would be of great relevance to show whether training induced changes, behavioural and psychophysical, could be long lasting or permanent even. Then, the effectiveness of this type of training method often used at schools would have stronger support. In fact, training effects have been shown to be maintained after one month in behavioural performance (Kraus et al., 1995) and after three weeks in the MMF (Menning et al., 2000).

One of the main questions remaining is whether this type of learning effect would be generalizable into other contrasts. As mentioned in the Introduction, training effects have been shown to transfer to stimuli with, for example, different place of articulation than the trained stimuli (Tremblay et al., 1997). For example, the English /s/–/z/ contrast for a Finn would be equal to the /f/–/v/ contrast, and therefore it can be argued that the training could generalise to that contrast as well. On the basis of our results, it is apparent that the simple articulatory training method widely used in schools seems to work; new memory traces for non-native speech contrasts, which are phonemically challenging, evolve in just three days during a listen-and-repeat training. To conclude, our training resulted in changes as did the training in, e.g., Kraus et al. (1995), Menning et al. (2002), Tremblay et al. (1997) and Tremblay et al. (1998). However, the training effects reported here were shown with listen-and-repeat training rather than with identification or discrimination training, on three consecutive days with four few minute training sessions rather than during a longer period with more training. In addition, the changes were obtained without any feedback whatsoever. Hence, the similar kind of training used in schools with a small amount of repetition without feedback resulted in behavioural improvement and in the formation of memory traces.

Acknowledgements

The study was financially supported by Emil Aaltonen Foundation, the Academy of Finland (No. 206352), and the Institutional Research Grant IUT02-13 to Juri Allik from the Estonian Ministry of Education and Research. The authors thank all the subjects who participated, and Jaana Hirvelä, M.A. for her assistance with data collection.

References

Amenedo, E., Escera, C., 2000. The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. Eur. J. Neurosci. 12, 2570–2574.

Atienza, M., Cantero, J.L., 2001. Complex sound processing during human REM sleep by recovering information from long-term memory as revealed by the mismatch negativity (MMN). Brain Res. 901, 151–160.

Best, C.T., Strange, W., 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. J. Phon. 20, 305–330.

Bradlow, A.R., 2008. Training non-native language sound patterns. In: Hansen Edwards, J.G., Zampini, M.L. (Eds.), Phonology and Second Language Acquisition. John Benjamins Publishing Company, Amsterdam, pp. 287–308.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., Näätänen, R., 1998. Development of language-specific phoneme representations in the infant brain. Nat. Neurosci. 1, 351–353.

Flege, J.E., 1987. The production of "new" and "similar" phones in a foreign language evidence for the effect of equivalence classification. J. Phon. 15, 47–65.

Flege, J.E., 1993. Production and perception of a novel, second-language phonetic contrast. J. Acoust. Soc. Am. 93, 1589–1608.

Kraus, N., Mc Gee, T., Carrell, T.D., King, C., Tremblay, K., Nicol, T., 1995. Central auditory system plasticity associated with speech discrimination training. J. Cogn. Neurosci. 7, 25–32.

Kuhl, P.K., 1991. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. Percept. Psychophys. 50, 93–107.

Kujala, T., Näätänen, R., 2010. The adaptive brain: a neurophysiological perspective. Prog. Neurobiol. 91, 55–67.

Kujala, T., Kallio, J., Tervaniemi, M., Näätänen, R., 2001a. The mismatch negativity as an index of temporal processing in audition. Clin. Neurophysiol. 112, 1712–1719.

Kujala, T., Karma, K., Ceponiene, R., Belitz, S., Turkkila, P., Tervaniemi, M., Näätänen, R., 2001b. Plastic neural changes and reading improvement caused by audiovisual training in reading-impaired children. Proc. Natl. Acad. Sci. 98, 10509–10514.

Kujala, T., Tervaniemi, M., Schröger, E., 2007. The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. Biol. Psychol. 74, 1–19.

Liberman, A.M., Harris, K.S., Hoffman, H.S., Griffith, B.C., 1957. The discrimination of speech sounds within and across phoneme boundaries. J. Exp. Psychol. 54, 358–368.

Macmillan, N.A., Creelman, C.D., 1991. Detection Theory: A User's Guide. Cambridge University Press, Cambridge (407 pp.).

Menning, H., Roberts, L.E., Pantev, C., 2000. Plastic changes in the auditory cortex induced by intensive frequency discrimination training. NeuroReport 11, 817–822.

Menning, H., Imaizumi, S., Zwitserlood, P., Pantev, C., 2002. Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. Learn. Mem. 9, 253–267.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., Paavilainen, P., 1993. Development of a memory trace for a complex sound in the human brain. NeuroReport 4, 503–506.

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 brain responses. Nature 385, 432–434.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh Inventory. Neuropsychologia 9, 97–113.

Peltola, M.S., 2004. Extensive language learning affects the perception of vowels. In: Peltola, M.S., Tuomainen, J. (Eds.), Studies in Speech Communication. Digipaino, Turku, pp. 43–60.

Peltola, M.S., Aaltonen, O., 2005. Long-term memory trace activation for vowels depends on the mother tongue and the linguistic context. J. Psychophysiol. 19, 159–164.

Peltola, M.S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., Näätänen, R., 2003. Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. Neurosci. Lett. 352, 25–28.

Peltola, M.S., Kuntola, M., Tamminen, H., Hämäläinen, H., Aaltonen, O., 2005. Early exposure to non-native language alters preattentive vowel discrimination. Neurosci. Lett. 388, 121–125.

Peltola, M.S., Tuomainen, O., Koskinen, M., Aaltonen, O., 2007. The effect of language immersion education on the preattentive perception of native and non-native vowel contrasts. J. Psycholinguist. Res. 36, 15–23.

Peltola, M.S., Tamminen, H., Toivonen, H., Kujala, T., Näätänen, R., 2012. Different kinds of bilinguals — different kinds of brains: the neural organisation of two languages in one brain. Brain Lang. 121, 261–266.

Tremblay, K.L., Kraus, N., 2002. Auditory training induces asymmetrical changes in cortical neural activity. J. Speech Lang. Hear. Res. 45, 564–572.

Tremblay, K., Kraus, N., Carrell, T.D., McGee, T., 1997. Central auditory system plasticity: generalization to novel stimuli following listening training. J. Acoust. Soc. Am. 102, 3762–3773.

Tremblay, K., Kraus, N., McGee, T., 1998. The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. NeuroReport 9, 3557–3560.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi, R.J., Näätänen, R., 1999. Brain responses reveal the learning of foreign language phonemes. Psychophysiology 36, 638–642.