Available online at www.sciencedirect.com

ScienceDirect

Lingua 302 (2024) 103697

Lingua

www.elsevier.com/locate/lingua

Training non-native speech sounds results in long-lasting plastic changes – Hard-wiring new memory traces takes time

Henna Tamminen [a,b,*], Teija Kujala [c], Maija S. Peltola [a,b]

[a] *Phonetics and Learning, Age & Bilingualism Laboratory (LAB-lab), Department of Computing, Faculty of Technology University of Turku, Turku, Finland*
[b] *Centre for Cognitive Neuroscience, University of Turku, Turku, Finland*
[c] *Cognitive Brain Research Unit, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland*

Received 21 June 2023; revised 8 February 2024; accepted in revised form 8 February 2024;

Abstract

Training of a foreign language speech sound pair may lead to the formation of new memory traces in young adults who are not very experienced in the target language, as well as to the strengthening of existing memory traces in advanced target language students. We used listen-and-repeat training to test whether previously formed memory traces exist approximately a year later. Further, we compared these learning results with those obtained from advanced target language students. Both groups participated in a recording that was identical to the one that the less experienced learners had already completed a year before, but only once and without training. The experiments included EEG recordings measuring mismatch negativity (MMN) and N1, as well as listening tests. The less experienced learners' MMN responses were similar to those of advanced students, and their N1 amplitude had increased close to that of the advanced students. These results suggest long-lasting speech memory traces resulting from a brief training and the development of increased sensitivity.

Keywords: Training; Speech perception; Phonological processing; Mismatch negativity (MMN); N1

---

1. INTRODUCTION

According to second language learning models, it is not the sounds that are completely new or the same that are challenging to learn, but the sounds that are similar sounding or almost the same. The Speech Learning Model

---

* Corresponding author at: Phonetics and Learning, Age & Bilingualism Laboratory (LAB-lab), Department of Computing, Koskenniemenkatu 4, FIN-20500 Turku, Finland.

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

https://doi.org/10.1016/j.lingua.2024.103697
0016-7037/

(SLM) (Flege, 1987) and its revised version, the SLM-r (Flege and Bohn, 2021) state that *similar* sounds that are neither *identical* to any of the native language sounds, nor completely *new*, cause learning problems. The Perceptual Assimilation Model (PAM) (Best and Strange, 1992) and its updated version, the Perceptual Assimilation Model of L2 Speech Learning (PAM-L2) (Best and Tyler, 2007), predict learning problems, for example, in a situation where two foreign language categories are unequally assimilated to a native language category. For example, the English sounds /f/ and /v/ can be described as follows in relation to the Finnish sound system: according to the SLM, the English /f/ is *identical* to the Finnish /f/ while the English /v/ is *similar* to the Finnish /f/, and according to the PAM, the English /f/ – /v/ pair is unequally assimilated to the Finnish /f/. At least at the initial stages of learning, a second language is perceived through native language speech categories, which hinders the learning process of difficult speech sounds. From another point of view, the native category prototypes make the nearby second language speech sounds difficult to discriminate, whereas sounds in the immediate vicinity of the category border are more easily discriminated, as described by the Native Language Magnet Model (NLM) (Kuhl, 1991).

Memory traces for the native language speech sound categories form in early childhood (Cheour et al., 1998; Dehaene-Lambertz and Baillet, 1998), which makes them strong, and hence they hinder later second language learning (e.g., Kuhl et al., 2008). These language-specific memory traces (Näätänen et al., 1997) are the foundation of speech perception, and they remain fairly stable even after long deafness (Salo et al., 2002). Also, immigrants preserve their native language perceptual skills even after a significantly long immigration period (Flege and MacKay, 2004) and the effect of the native language is persistently shown as a foreign accent in second language speech production (e.g., Flege et al., 1997; Moyer, 2013, p. 18). Second language learning and the consequent formation of memory traces for second language speech sounds is increasingly difficult after the age of about one year (Kuhl et al., 2008), but by no means impossible. An authentic learning environment is usually ideal for the learning of second language speech sounds, as shown by immigrant (Tsukada et al., 2005; Winkler et al., 1999) and immersion programme (Cheour et al., 2002; Peltola et al., 2005; Shestakova et al., 2003) studies showing native-like perception. However, classroom learning does not necessarily lead to native-like perception (and hence to the formation of new memory traces) in children (Jost et al., 2015) or in adults (Grimaldi et al., 2014; Peltola et al., 2003).

Different training methods have been proven effective for foreign speech sound perception and production learning. Many studies have used various behavioural perception training methods (Baese-Berk, 2019; Fouz-González and Mompean, 2021; Grenon et al., 2019; Kraus et al., 1995; Menning et al., 2002; Tremblay et al., 1998, 1997; Tremblay and Kraus, 2002). Others have used speech production training to measure the effects on production (e.g., Immonen and Peltola, 2018; Kartushina et al., 2016; Kartushina and Martin, 2019; Peltola et al., 2015) or on the combination of production and perception (Kartushina et al., 2015; Peltola et al., 2020; Saloranta et al., 2020). Our previous training studies have shown that listen-and-repeat training may result in the formation of new memory traces for foreign language speech sounds in less experienced young adult learners (Tamminen et al., 2015), or in the strengthening of existing memory traces in proficient target language students (Tamminen and Peltola, 2015), but no training effects were seen in elderly monolinguals (Tamminen et al., 2021).

The aim of the present study[1] was twofold: first, to investigate whether training induced memory traces still exist after one year, and second, how the memory traces of less experienced learners compare with those of advanced students, i.e., more experienced learners. In other words, does short training result in similar persistent plastic changes as advanced level studies?

To this end, we used electroencephalography (EEG) and perceptual tests. We recorded N1, which reflects sound encoding, and mismatch negativity (MMN), which reflects stimulus discrimination (Kujala and Näätänen, 2010). Both responses are indicators of learning, becoming larger after training (e.g., Kujala et al., 2007; Kujala and Näätänen, 2010; Pantev and Herholz, 2011). These responses are obtained by presenting stimulus sequences. Sound on– and offsets elicit N1, whereas sound changes elicit MMN, indicating that the auditory system has detected the difference between the sounds (Näätänen, 1992). With the perceptual tests, we measured identification (ID), goodness rating (GR), discrimination sensitivity (*d*′), and reaction time (RT).

Previously, both groups, the less experienced learners and the more experienced learners, were tested on three consecutive days as follows: 1) ID and GR, *d*′ and RT, EEG, and training, 2) training, ID and GR, *d*′ and RT, EEG, and training, and 3) training, ID and GR, *d*′ and RT, and EEG. In the present study, the less experienced learners participated in the same recordings (ID, GR, *d*′, RT, and EEG) once and without training approximately a year later. In order to find out whether the training induced memory traces were long-lasting, we compared the baseline or post training situation, from the previous study, to the follow-up year later. Learning effects were investigated by comparing the less experienced learners' follow-up and the more experienced learners' baseline.

---

[1] The study is part of the first author's PhD thesis.

The stimulus pair /fiːl/ – /viːl/ was selected as it represents an unequally assimilated English sound pair for Finns, since the English /f/ and /v/ assimilate to the Finnish /f/, /f/ better than /v/ (see e.g., Best and Tyler, 2007). The Finnish phonological system does not discriminate between the voiceless and voiced labiodental fricatives /f/ and /v/, and only the voiceless /f/ sound is part of the Finnish system (e.g., Suomi et al., 2008, p. 25). The labiodental approximant /ʋ/, however, exists in the Finnish system and, like the English /v/, is represented in writing with the grapheme <v>. Because of this, Finnish learners of English often substitute the English fricative /v/ with the Finnish /ʋ/. The English /v/ is difficult even for students majoring in English (e.g., Lintunen, 2004, pp. 174–178). The less experienced learners, native Finnish speakers, did not have a memory representation for the English /v/, but a memory trace evolved in only three days (Tamminen et al., 2015). However, the advanced target language learners did show a memory trace for the English /v/ already in the baseline measurements and it was further strengthened during the three-day training (Tamminen and Peltola, 2015).

Our hypothesis was that the training induced memory trace might be detectable in the follow-up in the same manner as native language memory traces after long deafness (Salo et al., 2002). On the other hand, the memory traces formed through laboratory training a year ago are probably not as strong compared to those formed during long-lasting learning in a classroom, as even the native language seems to require constant use in order to remain intact, as suggested by Flege and MacKay (2004). The question remains, if the memory traces are still intact, whether the amount of English language in everyday Finnish context is sufficient to either sustain the required memory trace or even strengthen it.

2. METHODS

The participants, stimuli and procedure are the same as described in detail in Tamminen et al. (2015) and Tamminen and Peltola M. S (2015) and are summarised here. The study was approved by the Ethics Committee of the University of Turku, Finland.

2.1. Participants

Twelve Finns, who were less experienced learners of English, participated in the first three sessions conducted a year before the current follow-up experiment. Eight (age range 19–27 years, mean 22,6 years, 5 females) of them agreed to participate in the follow-up. One participant was excluded from the ID and GR results in the follow-up due to technical errors, but data for discrimination and EEG measurements are from 8 participants. Until the end of upper secondary school, the participants had studied English for 8.8 years on average (range 5–11 years), after which they had not studied any languages. Even though the participants had studied English in primary and lower secondary education and upper secondary school, they were not considered to be sensitised to the phonetic details between Finnish and English. These details as well as contrastive analyses of the two sound systems are not on the national core curriculum, and thus are not at the centre of attention in language teaching. The participants reported that their English proficiency in speaking was excellent (2), good (2), satisfactory (2), or bad (1) and their proficiency in understanding was excellent (3), or good (4). They also reported hearing English daily a lot (2), or quite a lot (6). The follow-up was conducted one year and two months from the first session on average (range 10–17 months). None of the participants reported any changes in their neurological condition and none had resided abroad for longer periods during this time.

The other group consisted of eleven (age range 20–28, mean age 23.6, 6 females) more experienced learners, advanced Finnish university students of English. The more experienced learners' three-day training results were reported earlier in Tamminen and Peltola M. S. (2015) and only the baseline results of this group are reported in the current study. They had majored in English for 3.5 years (range 1–7 years) and had studied English in primary and secondary schools for 10 years before that. Contrary to primary and lower secondary education and upper secondary school, phonetic details between Finnish and English, and contrastive analyses of the two sound systems, are part of the English curriculum at the university. The participants reported that their English proficiency in speaking was excellent (7), good (3), or satisfactory (1) and their proficiency in understanding was excellent (10), or good (1). They also reported hearing English daily a lot (7), or quite a lot (4). All participants were right-handed (Oldfield, 1971), neurologically healthy, and of normal hearing. An informed consent was obtained from each participant.

2.2. Stimuli

A continuum of fifteen variants of the words /fiːl/ ‘*feel*’ and /viːl/ ‘*veal*’ were synthesised (HLSyn software) so that the voicing of the first sound varied in 14 ms steps, from totally voiceless to entirely voiced. The continuum was used in the ID and GR tests, while two stimuli (representing different categories classified by native English speakers) were chosen for the oddball discrimination and EEG recordings. The voicing started at 113 ms in the representative of /fiːl/ and at

71 ms in /vi:l/ (detailed description in Tamminen et al., 2015). Hence, the difference between the two stimuli started at 71 ms.

2.3. Experiments and procedure

During the self-paced, forced choice ID and GR experiment, the 15 stimuli were presented 8 times in a random order. The task was, first, to identify the stimuli as either ‘feel’ or ‘veal’, and then rate the goodness of the identified stimuli on a scale of 1 to 7, where 1 was poor and 7 excellent. In the oddball discrimination task /fi:l/, with the familiar sound, was the standard stimulus and /vi:l/, with the unfamiliar, difficult sound, the deviant. The instruction in the discrimination task was to press a given button as soon as the deviating stimulus was detected. There were 130 standards and 20 deviants in the discrimination test, deviant probability being 0.13. In the EEG recording, the participants were instructed to ignore the stimuli while they were watching a silent movie. The EEG was registered with 21 electrodes using Synamps amplifier (sampling rate 250 Hz, bandwidth 0.5–70 Hz). There were 783 standards and 120 deviants in the EEG recording, deviant probability being 0.13. The stimuli were the same as in the discrimination task. The training task was also self-paced and the participants were instructed to listen to the stimulus and repeat it (this was recorded, but the results are not discussed here). No feedback was given in order not to provide any cues of the stimuli and to discover only the possible training effects. The same stimuli were used in the training as in the discrimination and EEG recordings. The two stimuli were presented in an alternating manner, 30 times each.

During the initial three testing sessions, the ID and GR task was performed first, after which the oddball discrimination and the EEG measurements were conducted in a counterbalanced order. Training was carried out last during the first day, before and after testing during the second day, and before testing on day three. In the follow-up, the participants performed the same tasks and measurements in the same manner, but only once and without training. This was the overall set-up for the less experienced learners. The data from the less experienced learners reported here are the baseline, the third and the follow-up session measures. The more experienced learners were overall tested otherwise similarly but without the follow-up. However, the results of the more experienced learners reported here are from the baseline measurements only (see Fig. 1).

2.4. Data analyses

The ID data was subjected to a logit transformation analysis to obtain category boundary location and boundary steepness. The mean GR values were calculated from the ten repetitions for each stimulus. The RT was measured from the onset of the stimulus and responses within ± 3 standard deviations were included. The *d*’ was calculated using the hits, misses, false alarms, and correct rejections.

The event-related potential (ERP) epochs (550 ms window including a 50 ms pre-stimulus period) were off-line filtered with a 1–30 Hz bandpass filter and the artefact criterion was ± 100 µV. Baseline correction period was from 50 ms before stimulus onset to stimulus onset. By subtracting the standard response from the deviant response, the difference waveforms were created for each participant. The time window for quantifying the MMN response was 300–340 ms for both groups. The time window for the N1 response was 205–225 ms for the less experienced learners and 190–210 ms for the more experienced learners. The time windows were selected around the maximum amplitudes,

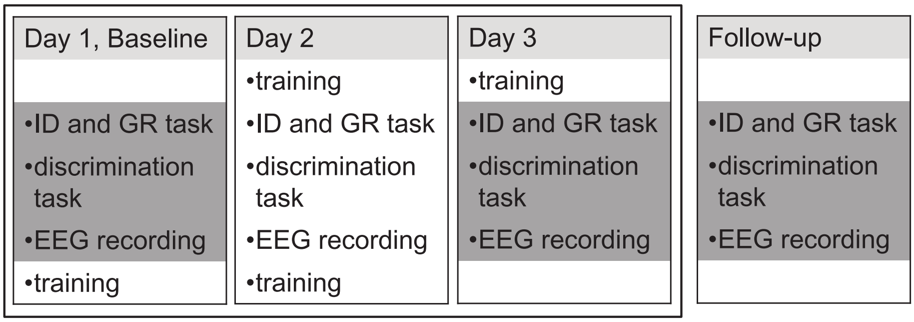

Fig. 1. The outline of the study. The first testing and training took place on three consecutive days for the less experienced learners (Tamminen et al., 2015) and for the more experienced learners (Tamminen and Peltola, 2015). The follow-up for the less experienced learners was conducted a year later. Data reported here are the baseline for both groups as well as the third day and the follow-up for the less experienced learners (shaded area).

and therefore the N1 time windows are different for the two groups. The difference between the standard and deviant stimuli started at 71 ms resulting in seemingly greater response latencies (both MMN and N1), and the time windows were selected accordingly. Fricatives also increase the latency of the response (e.g., Pereira et al., 2018).

The statistical analyses were conducted so that the less experienced learners' baseline or third session was compared to the follow-up session, and in addition, the less experienced learners' baseline and follow-up were compared to the more experienced learners' baseline sessions. The within-group comparisons were conducted with repeated measures Analysis of Variance (ANOVA) and the between group analyses were carried out by Multivariate Analysis of Variance (MANOVA). Pairwise comparisons were performed when necessary. The ERP data were analysed separately for the MMN and N1 with a Group (2) × Electrode (6) analysis when the groups were compared, or Session (2) × Electrode (6) analysis when the less experienced learners' sessions were compared. The MMN is most prominent at fronto-central areas (e.g., Kujala et al., 2007) and hence electrodes Fz, Cz, F3, F4, C3, and C4 were selected for the analysis. The behavioural data were analysed separately for the category boundary location, category boundary steepness, discrimination sensitivity, and RTs, all with a Group (2) or Session (2) analysis. The GR data were analysed with a Group (2) × Stimulus (5) analysis, when groups were compared, or Session (2) × Stimulus (5) analysis, when the less experienced learners' sessions were compared. The goodness ratings were analysed for the prototypes (less experienced learners: 3 and 13; more experienced learners: 2 and 13) of the categories, the trained stimuli (7 and 10) and the border (9). The data were normally distributed and the p-values varied as follows: ID 0.085–0.187; GR 0.054–0.951; RT 0.082–0.877; *d*' 0.050–0.569; MMN 0.060–0.991; and finally, N1 0.075–0.953.

3. RESULTS

3.1. ERP results

The statistical analyses showed no significant effects when the MMN responses of the less experienced learners in the last session during the initial training and the follow-up session were compared. In other words, the established training effect was seen also in the follow-up. Further, the analyses did not reveal any difference in the MMN response between the less experienced learners' follow-up session and the more experienced learners' baseline session since no main effect of Group was found. The analysis comparing the baselines showed a tendency for enhanced amplitudes in the more experienced learners compared to the less experienced learners (*F*(1,17) = 3.837, p = 0.067). See Fig. 2 and Table 1 for the MMN results.

The N1 responses in the first three sessions did not differ from each other as shown by the pairwise comparison of the sessions. However, in the less experienced learners, the N1 response was larger in the follow-up than in the baseline session as shown by the main effect of Session (*F*(1,7) = 6.095, p = 0.043). Moreover, no main effect of Group was found when the less experienced learners' follow-up and the more experienced learners' baseline N1 responses were compared, suggesting similar N1 responses in these groups. However, there was a significant main effect of Group (*F*(1,17) = 6.164, p = 0.024) when the baselines were compared, indicating a larger N1 in the more experienced learners than in the less experienced learners. See Fig. 2 and Table 1 for the N1 results.

3.2. Behavioural results

All the behavioural results are shown in Table 2. Statistical analyses revealed no significant changes in the category boundary when comparing the baseline and the follow-up sessions of the less experienced learners. However, there was a significant change in category steepness (*F*(1,6) = 6.008, p = 0.050) between the less experienced learners' baseline and follow-up. There were no differences between groups when the less experienced learners' follow-up and the more experienced learners' baseline sessions were compared. However, the category steepness was significantly different, i.e. more systematic in the more experienced learners, (*F*(1,17) = 6.397, p = 0.022) when the baseline sessions of the two groups were compared, but the category boundary was the same.

The comparison of the GR data between the baseline and follow-up sessions in the less experienced learners revealed the main effect of Stimulus *F*(2,24) = 5.449, p = 0.003) indicating that the stimuli were rated differently. An interaction between Stimulus and Session (*F*(4,24) = 3.136, p = 0.033) was also found; i.e., the stimuli got different ratings between the sessions. When the two sessions were analysed per stimulus, only stimulus 3, the prototype in the /fiːl/ category, was rated statistically differently between the sessions; it was rated as a better representative of the category in the follow-up than in the baseline session. There were no differences between the groups when the follow-up of the less experienced learners and the baseline of the more experienced learners were compared. However, the comparison of the groups' baselines showed an interaction between Stimulus and Group (*F*(4,64) = 3.673, *p* = 0.009), indicating that the groups rated the stimuli differently.

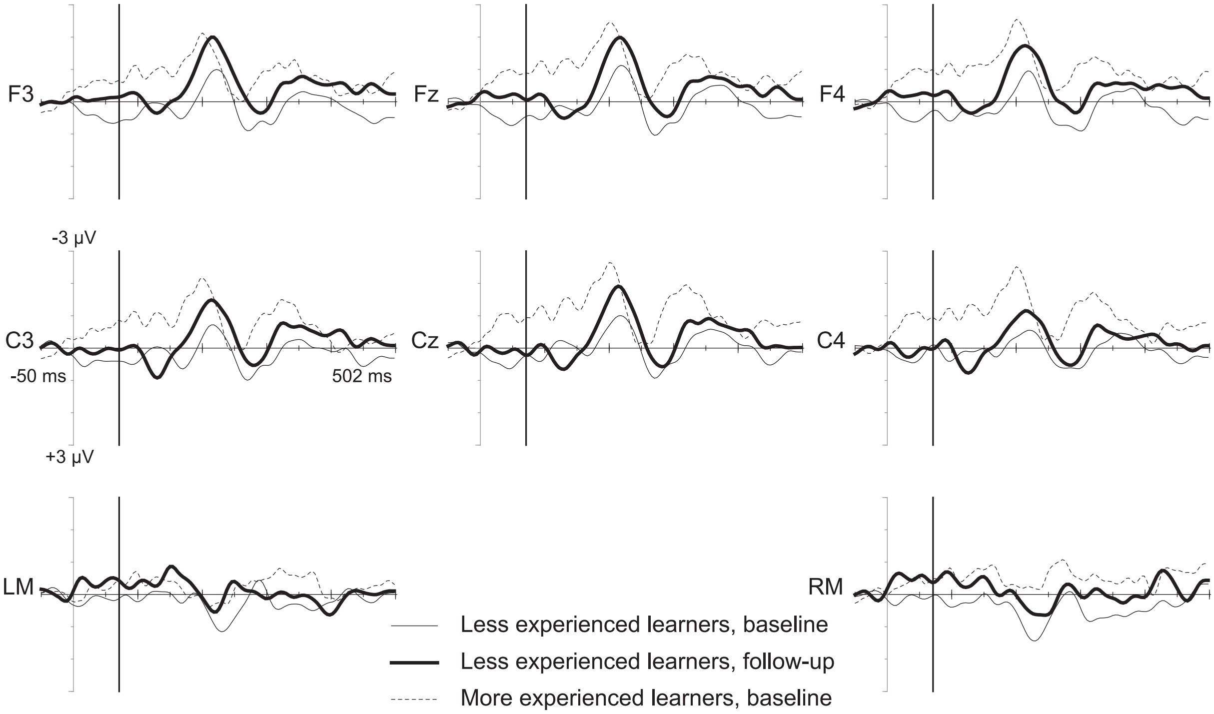

Fig. 2. The grand average difference waveforms for Less experienced learners and More experienced learners. The thin line represents the baseline measurement and the thick line the follow-up session of the Less experienced learners and the dotted line represents the baseline of the More experienced learners. The difference between the stimuli begins at 71 ms shown by the vertical black line at each electrode.

Table 1
The mean N1 amplitudes from the 205–225 ms and 190–210 ms time windows for the Less experienced learners and the More experienced learners, respectively, and the mean MMN amplitudes from the 300–340 ms time window for both groups. Standard deviations are in brackets.

	Fz	Cz	F3	F4	C3	C4
N1						
Less experienced learners	−0.692	−0.590	−0.529	−0.479	−0.370	−0.237
baseline	(2.129)	(2.393)	(2.044)	(2.083)	(2.066)	(2.206)
Less experienced learners	−1.911	−1.813	−1.907	−1.681	−1.415	−1.100
year after	(1.307)	(1.323)	(1.250)	(1.297)	(1.293)	(1.281)
More experienced	−2.317	−2.479	−1.952	−2.357	−2.020	−2.336
learners baseline	(1.295)	(0.752)	(1.306)	(1.184)	(0.923)	(0.912)
MMN						
Less experienced learners	0.053	−0.284	0.247	0.219	−0.126	0.059
baseline	(1.275)	(1.363)	(1.311)	(1.300)	(1.254)	(1.363)
Less experienced learners	−0.382	−0.615	−0.369	−0.355	−0.537	−0.429
year after	(0.644)	(0.750)	(0.611)	(0.849)	(0.786)	(0.902)
More experienced	−1.345	−1.778	−1.088	−1.269	−1.300	−1.529
learners baseline	(1.964)	(1.927)	(2.026)	(1.790)	(1.973)	(1.484)

Reaction times were different ($F(1,7) = 11.299$, $p = 0.012$) between the baseline and follow-up sessions in the less experienced learners, i.e. the follow-up RT was faster. However, the discrimination sensitivity was the same. The reaction times or the discrimination sensitivities did not differ between the less experienced learners' follow-up and the more experienced learners' baseline sessions, or between the two groups' baseline sessions.

Table 2
The mean ID (category boundary location and steepness), GR (prototypes, training stimuli and border), and discrimination (*d'* and RT) values. Standard deviations are in brackets.

	ID		GR					Discrimination	
	boundary	steepness	proto	training	border	training	proto	*d'*	RT
Less experienced learners baseline	9.3 (0.42)	0.95 (0.55)	4.6 (1.41)	3.9 (1.08)	3.7 (1.39)	4.0 (1.17)	4.7 (0.92)	3.45 (0.89)	763 (118)
Less experienced learners year after	9.1 (0.89)	1.44 (0.44)	5.4 (0.91)	3.9 (0.93)	2.9 (0.71)	3.6 (1.18)	4.5 (1.00)	3.87 (0.39)	640 (107)
More experienced learners baseline	8.9 (0.65)	1.58 (0.49)	5.8 (0.96)	4.2 (1.35)	2.9 (0.66)	3.8 (0.84)	4.4 (0.91)	3.64 (0.85)	674 (98)

4. DISCUSSION

The purpose of our study was to find out whether listen-and-repeat training induces long-lasting foreign language sound representations for less experienced learners, and whether training induced learning results in similar processing as a prolonged classroom learning. Foreign, or second, language perception in these two groups was measured using psychophysiological measures (MMN and N1) and behavioural listening tasks.

The plastic changes of neural processing in the less experienced learners were persistent since the MMN response, which originally became enhanced by training, did not differ between the follow-up session a year later and in the last session of the original training sessions. The training induced response, hence, did not diminish significantly during the year. In addition, the less experienced learners' follow-up MMN response was not different from the baseline MMN of the more experienced learners. Further, the comparison of the baseline MMN responses showed larger MMNs in the more experienced learners than in the less experienced learners (Table 1 and Fig. 2), the effect remaining a tendency probably because of the unfortunately small group of less experienced learners. These results indicate that the three-day listen-and-repeat training forms long-term memory traces, which are comparable to those of highly advanced target language students. In behavioural performance, training effects have been shown to last for a month (Fouz-González and Mompean, 2021; Kraus et al., 1995) and even for a year (Escudero and Williams, 2014).

Our behavioural results corroborate the MMN results. None of the behavioural measures differed between the less experienced learners' follow-up and the more experienced learners' baseline. However, the baseline comparisons revealed between-group differences in some of the behavioural measures. The difference in category boundary steepness indicates that the less experienced learners were not as systematic in placing the category boundary before training as the more experienced learners were. The two groups also rated the goodness of the stimuli differently in the baseline sessions. The native language of the less experienced learners most probably hindered systematic identification of the foreign sounds before training. Comparisons within the less experienced learners, however, revealed some changes: the systematicity in placing the category boundary was better, the prototype of the /f/ category was rated higher and the RTs were faster in the follow-up compared to the baseline. The native language /f/ is identical to that of the target language which might explain the higher rating of the prototypical member of the category (Kuhl, 1991), even though the fricative was presented in a foreign language context (/fiːl/). Reaction times tend to become faster in the course of a training study as participants become familiarised with the task and the stimuli. However, in this study, the decrease in the RTs was found in the follow-up a year later, which is why the familiarisation effect might not be as strong. All in all, production training has been shown to improve perception whether the training is instructed (e.g., Kartushina et al., 2015; Lee et al., 2020; Linebaugh and Roche, 2015) or not (e.g., Saloranta et al., 2022, 2020; Tamminen et al., 2015). However, these studies did not measure training effects in long-term follow-up sessions. To summarise the behavioural results, training was both effective and long-lasting, and there were no group differences in any of the behavioural measures between the less experienced learners' follow-up and the more experienced learners' baseline. In other words, the plastic changes achieved in behavioural perception by the less experienced learners were at the same level with the more experienced learners. However, the unfortunately small number of participants in the follow-up session and the possible implications it may have had on the results need to be acknowledged.

The most surprising finding was the change in the N1 response. During the original three-day training sessions of the less experienced learners, the N1 remained stable. However, after one year, it was significantly larger. Further, the less experienced learners' follow-up N1 was similar to the baseline N1 of the more experienced learners, whereas the baseline N1 responses were different between the two groups. The increased N1 response, elicited without further training, may be connected with the mechanism behind the effortless acquisition of speech. In other words, some kind of increase in sensitivity had occurred in the less experienced learners during the year between the initial training and the follow-up. The amplitude of this sensitivity indicator in the less experienced learners in the follow-up was at the same level as in the more experienced learners with extensive studies and with excellent language proficiency. This increased sensitivity might refer to a mechanism that enables the speech processing apparatus to be open for further learning, and to deep-rooted learning, which takes place automatically by itself. Finns are exposed to English through, for example, TV, which may have facilitated the consolidation of the speech memory traces after the initial training sessions. Consistent with our results, Saloranta et al. (2020) showed in their speech sound duration training study, with listen-and-repeat training, that an N1 response was elicited by an untrained linguistic contrast a week from the last training session. They suggested that this reflects the brain becoming more sensitive to the contrast. Since the N1 generates to attention-catching properties of an event and may reflect the sensitivity growth to the previously learned (Kujala and Näätänen, 2010; Näätänen et al., 2011), its enhancement by training might reflect an increased sensitivity of the speech processing apparatus, making it more alert to detect the different auditory features (ibid.), i.e. the voiced and voiceless fricatives.

To conclude, the listen-and-repeat training of the difficult foreign sound contrast induced long-term memory traces, reflected by MMN and behavioural tests, in less experienced learners, and the memory traces were detectable one year from the initial training. Further, these memory traces were comparable to those of more experienced learners. The most striking finding was the less experienced learners' increased N1 response similar to that of the more experienced learners, which may indicate sensitisation towards the learned difference and the openness of the learning process. These results suggest that a small amount of repetition training without feedback, similar to that used in schools, is effective and leads to perception that is comparable to that achieved with prolonged classroom learning.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Henna Tamminen: Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Teija Kujala:** Writing – review & editing, Data curation. **Maija S. Peltola:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Data availability

The data that has been used is confidential.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGEMENTS

The authors thank all the subjects who participated and MA Jaana Hirvelä and MA Eetu Laine for their highly appreciated assistance in data collection. We also thank the anonymous reviewers and proof reader. Funding: This work was supported by the Emil Aaltonen Foundation and the Research Council of Finland [grant number 316970].

References

Baese-Berk, M.M., 2019. Interactions between speech perception and production during learning of novel phonemic categories. Atten. Percept. Psychophys. 81, 981–1005. https://doi.org/10.3758/s13414-019-01725-4.

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

Best, C.T., Tyler, M.D., 2007. Non-native and second-language speech perception: commonalities and complementarities. In: Bohn, O.-S., Munro, M.J. (Eds.), Language Experience in Second Language Speech Learning: In Honor of James Emil Flege, Language Learning & Language Teaching. John Benjamins Publishing Company, pp. 13–45.

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. https://doi.org/10.1038/1561.

Cheour, M., Shestakova, A., Alku, P., Ceponiene, R., Näätänen, R., 2002. Mismatch negativity shows that 3–6-year-old children can learn discriminate non-native speech sounds within two months. Neurosci. Lett. 325, 187–190. https://doi.org/10.1016/S0304-3940(02)00269-0.

Dehaene-Lambertz, G., Baillet, S., 1998. A phonological representation in the infant brain. Neuroreport 9, 1885–1888. https://doi.org/10.1097/00001756-199806010-00040.

Escudero, P., Williams, D., 2014. Distributional learning has immediate and long-lasting effects. Cognition 133, 408–413. https://doi.org/10.1016/j.cognition.2014.07.002.

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., Bohn, O.-S., 2021. The Revised Speech Learning Model (SLM-r). In: Wayland, R. (Ed.), Second Language Speech Learning: Theoretical and Empirical Progress. Cambridge University Press, pp. 3–83.

Flege, J.E., Frieda, E.M., Nozawa, T., 1997. Amount of native-language (L1) use affects the pronunciation of an L2. J. Phon. 25, 169–186. https://doi.org/10.1006/jpho.1996.0040.

Flege, J.E., MacKay, I.R.A., 2004. Perceiving Vowels in a Second Language. Stud. Second. Lang. Acquis. 26, 1–34. https://doi.org/10.1017/S0272263104261010.

Fouz-González, J., Mompean, J.A., 2021. Exploring the potential of phonetic symbols and keywords as labels for perceptualtraining. Stud. Second. Lang. Acquis. 43, 297–328. https://doi.org/10.1017/S0272263120000455.

Grenon, I., Kubota, M., Sheppard, C., 2019. The creation of a new vowel category by adult learners after adaptive phonetic training. J. Phon. 72, 17–34. https://doi.org/10.1016/j.wocn.2018.10.005.

Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P., Brattico, E., 2014. Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: a behavioral and ERP study. Front. Hum. Neurosci. 8, 1–14. https://doi.org/10.3389/fnhum.2014.00279.

Immonen, K., Peltola, M.S., 2018. Finnish children producing English vowels — studying in an English immersion class affects vowel production. J. Language Teach. Res. 9, 27–33.

Jost, L.B., Eberhard-Moscicka, A.K., Pleisch, G., Heusser, V., Brandeis, D., Zevin, J.D., Maurer, U., 2015. Native and non-native speech sound processing and the neural mismatch responses: a longitudinal study on classroom-based foreign language learning. Neuropsychologia 72, 94–104. https://doi.org/10.1016/j.neuropsychologia.2015.04.029.

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. J. Acoust. Soc. Am. 138, 817–832. https://doi.org/10.1121/1.4926561.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U.H., Golestani, N., 2016. Mutual influences between native and non-native vowels in production: evidence from short-term visual articulatory feedback training. J. Phon. 57, 21–39. https://doi.org/10.1016/J.WOCN.2016.05.001.

Kartushina, N., Martin, C.D., 2019. Talker and acoustic variability in learning to produce nonnative sounds: evidence from articulatory training. Lang. Learn. 69, 71–105. https://doi.org/10.1111/lang.12315.

Kraus, N., McGee, T., Carrell, T.D., King, C., Tremblay, K.L., Nicol, T., 1995. Central auditory system plasticity associated with speech discrimination training. J. Cogn. Neurosci. 7, 25–32. https://doi.org/10.1162/jocn.1995.7.1.25.

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. https://doi.org/10.3758/BF03212211.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T., 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). Philos. Trans. R. Soc., B 363, 979–1000. https://doi.org/10.1098/rstb.2007.2154.

Kujala, T., Näätänen, R., 2010. The adaptive brain: a neurophysiological perspective. Prog. Neurobiol. 91, 55–67. https://doi.org/10.1016/j.pneurobio.2010.01.006.

Kujala, T., Tervaniemi, M., Schröger, E., 2007. The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. Biol. Psychol. 74, 1–19. https://doi.org/10.1016/j.biopsycho.2006.06.001.

Lee, B., Plonsky, L., Saito, K., 2020. The effects of perception- vs. production-based pronunciation instruction. System 88,. https://doi.org/10.1016/j.system.2019.102185 102185.

Linebaugh, G., Roche, T.B., 2015. Evidence that L2 production training can enhance perception. J. Acad. Language Learn. 9, A1–A17.

Lintunen, P., 2004. Pronunciation and phonemic transcription: a study of advanced Finnish learners of English. Anglicana Turkuensia, 24. University of Turku, Turku.

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. https://doi.org/10.1101/lm.49402.

Moyer, A., 2013. Foreign Accent: The Phenomenon of Non-native Speech. Cambridge University Press, Cambridge hhttps://doi.org/10.1017/CBO9780511794407.

Näätänen, R., 1992. Attention and Brain Function. Lawrence Erlbaum Associates, Hillsdale N.J..

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. https://doi.org/10.1038/385432a0.

Näätänen, R., Kujala, T., Winkler, I., 2011. Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. Psychophysiology 48, 4–22. https://doi.org/10.1111/j.1469-8986.2010.01114.x.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4.

Pantev, C., Herholz, S.C., 2011. Plasticity of the human auditory cortex related to musical training. Neurosci. Biobehav. Rev. 35, 2140–2154. https://doi.org/10.1016/j.neubiorev.2011.06.010.

Peltola, K.U., Tamminen, H., Alku, P., Peltola, M.S., 2015. Non-native production training with an acoustic model and orthographic or transcription cues. In: Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow.

Peltola, K.U., Tamminen, H., Alku, P., Kujala, T., Peltola, M.S., 2020. Motoric training alters speech sound perception and production — active listening training does not lead into learning outcomes. J. Language Teach. Res. 11, 10–16.

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. https://doi.org/10.1016/j.neulet.2003.08.013.

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. https://doi.org/10.1016/j.neulet.2005.06.037.

Pereira, O., Gao, Y.A., Toscano, J.C., 2018. Perceptual encoding of natural speech sounds revealed by the N1 event-related potential response. Auditory Percept. Cognit. 1, 112–130. https://doi.org/10.1080/25742442.2018.1545106.

Salo, S., Peltola, M.S., Aaltonen, O., Johansson, R., Lang, A.H., Laurikainen, E., 2002. Stability of memory traces for speech sounds in cochlear implant patients. Logoped. Phoniatr. Vocol. 27, 132–138.

Saloranta, A., Alku, P., Peltola, M.S., 2020. Listen-and-repeat training improves perception of second language vowel duration: Evidence from mismatch negativity (MMN) and N1 responses and behavioral discrimination. Int. J. Psychophysiol. 147, 72–82. https://doi.org/10.1016/j.ijpsycho.2019.11.005.

Saloranta, A., Heikkola, L.M., Peltola, M.S., 2022. Listen-and-repeat training in the learning of non-native consonant duration contrasts: influence of consonant type as reflected by MMN and behavioral methods. J. Psycholinguist. Res.. https://doi.org/10.1007/s10936-022-09868-6.

Shestakova, A., Huotilainen, M., Ceponiene, R., Cheour, M., 2003. Event-related potentials associated with second language learning in children. Clin. Neurophysiol. 114, 1507–1512. https://doi.org/10.1016/S1388-2457(03)00134-2.

Suomi, K., Toivanen, J., Ylitalo, R., 2008. Finnish sound structure: Phonetics, phonology, phonotactics and prosody. University of Oulu, Oulu.

Tamminen, H., Peltola, M.S., 2015. Non-native memory traces can be further strengthened by short term phonetic training. In: Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow.

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. Int. J. Psychophysiol. 97, 23–29. https://doi.org/10.1016/j.ijpsycho.2015.04.020.

Tamminen, H., Kujala, T., Näätänen, R., Peltola, M.S., 2021. Aging and non-native speech perception: a phonetic training study. Neurosci. Lett. 740,. https://doi.org/10.1016/j.neulet.2020.135430 135430.

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.L., 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. https://doi.org/10.1121/1.420139.

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

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., Flege, J.E., 2005. A developmental study of English vowel production and perception by native Korean adults and children. J. Phon. 33, 263–290. https://doi.org/10.1016/j.wocn.2004.10.002.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I.i., Csépe, V., Ilmoniemi, R.J., Näätänen, R., 1999. Brain responses reveal the learning of foreign language phonemes. Psychophysiology 36, 638–642. https://doi.org/10.1111/1469-8986.3650638.