

Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English

Sari Ylinen¹, Maria Uther², Antti Latvala¹, Sara Vepsäläinen¹, Paul Iverson³, Reiko Akahane-Yamada⁴, and Risto Näätänen¹

Abstract

Foreign-language learning is a prime example of a task that entails perceptual learning. The correct comprehension of foreign-language speech requires the correct recognition of speech sounds. The most difficult speech-sound contrasts for foreign-language learners often are the ones that have multiple phonetic cues, especially if the cues are weighted differently in the foreign and native languages. The present study aimed to determine whether non-native-like cue weighting could be changed by using phonetic training. Before the training, we compared the use of spectral and duration cues of English /i/ and /l/ vowels (e.g., beat vs. bit) between native Finnish and English speakers. In Finnish, duration is used phonologically to separate short and long phonemes, and therefore Finns were expected to weight duration cues more than native English speakers. The cross-

linguistic differences and training effects were investigated with behavioral and electrophysiological methods, in particular by measuring the MMN brain response that has been used to probe long-term memory representations for speech sounds. The behavioral results suggested that before the training, the Finns indeed relied more on duration in vowel recognition than the native English speakers did. After the training, however, the Finns were able to use the spectral cues of the vowels more reliably than before. Accordingly, the MMN brain responses revealed that the training had enhanced the Finns' ability to preattentively process the spectral cues of the English vowels. This suggests that as a result of training, plastic changes had occurred in the weighting of phonetic cues at early processing stages in the cortex. ■

INTRODUCTION

Second-language (L2) learning is a challenge for most adult learners. It is well established that one's native language exerts an influence on the perception and production of L2 speech sounds. Language learners' difficulties with L2 sounds are often due to dissimilarities in the use of phonetic cues in the L2 and native-language (L1) phonological systems. First, L2 and L1 may use the same cue in phonological distinction, but the critical values may be different (e.g., in the same formant space, there is one L1 vowel category and two L2 vowel categories; Flege, 1988). Second, L2 may have phonological distinctions that are cued by acoustic features that are not used phonologically in L1 (e.g., duration distinctions of quantity languages for native speakers of languages that do not use duration phonologically; McAllister, Flege, & Piske, 2002). Third, the recognition of many (if not most) L2 sounds requires the integration of multiple cues (e.g., Holt & Lotto, 2006; Flege & Hillenbrand, 1986). Because all cues are often not equally important for recognition, native speakers weight them differently, giving higher weight to the most reliable and informative cues (Holt & Lotto,

2006). If the cue weighting differs in L2 from that in L1, L2 users may ignore the critical cues and instead attempt to use cues that are not reliable in the identification of L2 sounds; a good example is the Japanese perception of English /r/ and /l/, which tends to focus on the irrelevant F2 formant rather than the relevant F3 (Iverson et al., 2003).

Regardless of the difficulties faced by L2 learners, with extensive exposure and practice, it is possible to approach a native-like level in the perception and production of L2 sounds (e.g., Bongaerts, 1999). To facilitate and to speed up this process, different training techniques have been developed (see, e.g., Iverson, Hazan, & Bannister, 2005). One technique that improved the identification of nonnative speech contrasts is high-variability phonetic training that uses multiple exemplars of different natural words produced by different speakers (Logan, Lively, & Pisoni, 1991; see also Lively, Logan, & Pisoni, 1993). By exposing learners to a wide range of speech stimuli, high-variability phonetic training aims at the reallocation of attention to the critical cues. It should enable one to learn which cues are, on one hand, critical and, on the other hand, irrelevant for recognition. Previous results have suggested that high-variability phonetic training improves the identification of L2 items and further that the improvement retains for a period of at least 3 months (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Lively,

¹University of Helsinki, ²Brunel University, UK, ³University College London, UK, ⁴ATR Cognitive Science Laboratories/ATR Learning Technology, Kyoto, Japan

Pisoni, Yamada, Tohkura, & Yamada, 1994). The improvement can also be transferred to the learners' speech production (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997).

The first aim of the present study was to compare L1 and L2 cue weighting, namely, the cue weighting of an English vowel pair between native speakers of English and Finnish. The second aim was to determine whether L2 cue weighting could be changed as a function of high-variability phonetic training to resemble more native-like weighting. In addition to the effect of phonetic training on overt, behavioral identification responses, we investigated its effect on brain responses to determine the automatic activation of phonetic representations without participants' conscious effort. In particular, the MMN component of the auditory ERP is a useful tool in this respect. The MMN (Näätänen, Gaillard, & Mäntysalo, 1978) reflects pre-attentive change detection in auditory stimulation, and it is elicited by any discriminable acoustic change in a regular stimulus stream (for reviews, see Näätänen, 2001; Picton, Alain, Otten, Ritter, & Achim, 2000). According to Winkler, Schröger, and Cowan (2001) and Winkler, Karmos, and Näätänen (1996), the MMN elicitation requires the detection of regularities in auditory input and a violation of extrapolations based on these regularities. The MMN elicited by speech stimuli does not, however, reflect change detection only but reveals the additional activation of long-term memory representations for phonetic categories (for a recent review, see Pulvermüller & Shtyrov, 2006). Thus, the MMN response for familiar or prototypical speech sounds is typically larger in amplitude than that to unfamiliar or non-prototypical sounds (e.g., Ylinen, Shestakova, Huotilainen, Alku, & Näätänen, 2006; Dehaene-Lambertz, Dupoux, & Gout, 2000; Winkler et al., 1999; Dehaene-Lambertz, 1997; Näätänen et al., 1997). In addition, sounds presented in a familiar context elicit a larger MMN than those presented in an unfamiliar context (Jacobsen, Schröger, Winkler, & Horvath, 2005; Jacobsen et al., 2004). This familiarity effect is likely due to more elaborate regularity representation for familiar sounds (Jacobsen et al., 2004, 2005). Thus, although the MMN elicitation does not necessarily require identification of sounds but rather discrimination, long-term memory representations facilitate the process, which is reflected as MMN enhancement.

Due to the participation of long-term memory on MMN elicitation, it has also been used to index the effects of auditory training in the brain. The effects of training on the MMN were first demonstrated with nonspeech sounds (Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen, 1993). Subsequent studies showed that speech-sound training also enhanced the area and the duration of the MMN (Tremblay, Kraus, & McGee, 1998; Tremblay, Kraus, Carrell, & McGee, 1997; Kraus et al., 1995). Studies measuring the magnetic counterpart of the MMN (MMNm) have demonstrated a stronger MMNm response to nonnative speech stimuli as a function of training (Menning, Imaizumi, Zwisserlood, & Pantev, 2002) and a laterality shift of the MMNm elicited by Morse-coded syllables from the opposite

hemisphere to the language-dominant hemisphere as a result of training (Kujala et al., 2003). Interestingly, Tremblay et al. (1998) found that neurophysiological changes indicated by the MMN that result from speech-sound training can precede changes in subjects' behavioral responses. Thus, in addition to being an objective tool for evaluating the discriminability of speech sounds in the brain (Näätänen, 2001), the MMN may also reveal more subtle training effects than traditional behavioral research methods.

In the present study, the English tense /i/ versus lax /I/ distinction was used to investigate the weighting of L2 phonetic cues in a group of native Finnish speakers who used English as their L2. Their responses were compared with those of native English speakers. Previous studies have suggested that although /i/ is typically longer in duration than /I/, native speakers of English base their identification primarily on spectral cues or on the integration of spectral and duration cues (Hillenbrand, Clark, & Houde, 2000; Flege, Bohn, & Jang, 1997; Bohn & Flege, 1990). In contrast, Finnish L2 users of English often seem to base their identification on vowel duration (Lintunen, 2004; Suomi, 1976; for the effect of other L2 backgrounds, see, however, Iverson & Evans, 2007). Apparently, this occurs because the Finnish phonological system uses duration phonologically in quantity distinctions separating short and long phonemes (e.g., /tili/ "account" vs. /ti:li/ "brick"; see Wiik, 1965). Flege (1988) suggested that quantity-language speakers reinterpret the tense versus lax distinction as a quantity distinction and consequently weight the duration cues. In line with this, Peltola et al. (2003) found that Finnish students of English did not show native-like MMN responses for the English tense versus lax vowel distinction, when isolated vowels of the same duration were used. To determine the roles of the spectral and the duration cues in the processing of English /i/ and /I/, the present study compared the responses to stimuli with normal duration and with modified duration with each other—in the latter case, the duration cues were ambiguous or equalized. We hypothesize that native English speakers process vowels with normal and modified duration similarly because they weight spectral cues. In their processing of /i/ and /I/, no differences are expected because both vowels represent L1 prototypes. By contrast, Finnish L2 users of English may have difficulties with modified durations because they likely weight duration cues in the vowel recognition. In addition, the Finns' responses may be modified by unbalanced vowel familiarity: the English /i/ closely resembles the Finnish /i/ and thus is more familiar to the Finns than the English /I/, which has no qualitative counterpart in Finnish (Wiik, 1965).

METHODS

Participants

Twelve native speakers of Finnish (4 men, mean age = 26 years, range = 18–40 years) and 13 native speakers

of English (5 men, mean age = 24 years, range = 20–35 years) participated in the study. After preliminary analysis, two native English speakers were excluded from the analysis of the MMN data (one because of excessive eye movement and resultant poor signal-to-noise ratio and the other because of atypical response pattern that deviated from that of other native English speakers¹). All reported having normal hearing and having no speech- or language-related dysfunctions or problems, and three Finns and two native English speakers reported being left-handed. Six native English speakers were from the United States, two from Canada, and five from the United Kingdom (none of the participants were Scottish, who tend to use duration more systematically for /i/ vs. /I/ contrast than the other native English speakers), whereas none of the Finnish participants had lived in an English-speaking country or community. According to the native Finnish speakers' reports, 10 of them had studied English at school for 9–10 years and had started to learn English at the age of 9 years. One Finnish participant had studied English for 13–14 years starting from 9 years, and one Finn had studied English for 8 years starting from 11 years. They assessed being at an intermediate (6 participants) or advanced (6 participants) level of competence in English. The native English speakers, in turn, reported either no competence (6 participants) or a basic level of competence (6 participants) in Finnish, with the exception of one participant who was at advanced level. The native English speakers' skills in Finnish were, however, not expected to affect their responses to the native English language.

Ten of 12 native Finnish speakers (3 men, mean age = 28 years) returned to participate also in the training and the posttraining measurements. Only the data of these 10 participants were used in the comparison between the pretraining and the posttraining. The native English speakers did not participate in the training or posttraining measurements.

Behavioral Experiment

Stimuli and Procedure

Stimulus words used in the behavioral experiment were pronounced by a male native speaker of English from Southern England, representing typical British English pronunciation. He was asked to read a list of words several times (the words at the beginning and at the end of the list were excluded to avoid different list-reading properties). The utterances were recorded in an acoustically attenuated chamber with a digital recorder and further digitally processed. From the recordings, one exemplar of each word representing a typical English pronunciation was selected for the behavioral test. The chosen stimulus material consisted of 45 English minimal pairs with /I/ versus /i/ distinction. Nineteen of 45 minimal pairs used in the pretraining and the posttraining measurements were the same as in the training sessions, and 26 pairs were used only in the pretraining and the posttraining measurements (see Appendix A). Each stimulus had two versions, one with a normal vowel duration and another with a modified vowel duration. Stimuli with modified duration were used to determine whether vowels could be correctly identified although the duration cue was ambiguous. The vowel durations were modified as follows. The vowel durations were measured with Praat software (Boersma & Weenink, 2004). Further, the duration of the /i/ vowel in each minimal pair was manipulated so that it corresponded to the duration of the /I/ vowel in that minimal pair and vice versa (see Figure 1A). Because the duration of /i/ is typically longer than that of /I/, /i/ vowels were shortened and /I/ vowels lengthened. The manipulations were resynthesized using the pitch-synchronous overlap and add method (see Boersma & Weenink, 2004). Further, the stimuli were high-pass filtered at 80 Hz to eliminate any potential low-frequency noise in the recording, the intensity of the stimuli was normalized by scaling the peak to 90% maximum, and the first and the last 2 msec

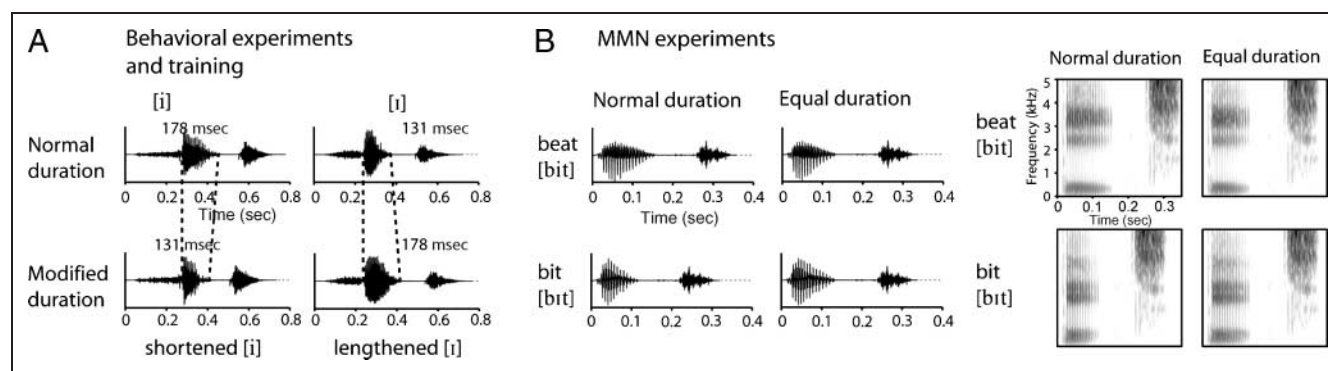


Figure 1. Stimuli. (A) An example of the stimuli used in the behavioral experiment and training. Top: The minimal pair *feet* versus *fit* with normal duration (i.e., as pronounced by a native English speaker). Bottom: The same word pair with modified vowel durations: [i] has been shortened to correspond to the original duration of [I], and [I] has been lengthened to correspond to the original duration of [i]. (B) The waveforms (left) and the spectrograms (right) of the stimuli used in the MMN experiment.

of the stimuli were ramped to eliminate possible click artifacts. The fundamental frequency (F0) of the stimuli varied within a natural range, the maximal values at the beginning of the vowel being 121–201 Hz (mean = 151 Hz) and the minimal values at the end being 102–168 Hz (mean = 121 Hz). Altogether, there were 180 stimuli in the behavioral experiment, each of which was presented twice during the test session, resulting in the total of 360 stimulus trials in the behavioral experiment.

The behavioral data were collected by using a minimal-pair perceptual identification task. The testing procedure was identical before and after the training. During the experiment, the participants sat in a chair in an acoustically attenuated chamber holding a response mouse in their hands. In front of them, there was a computer screen. The participants were presented with one auditory stimulus at a time via headphones, and simultaneously two response options were shown on the screen. If the auditorily presented word was, for example, *bit*, the options on the screen were “hit” and “heat” (the order of the options being random). The participants were instructed to press the left button of the mouse if they had heard the word appearing on the left-hand side of the screen or to press the right button of the mouse if they had heard the word appearing on the right-hand side of the screen. The press of a mouse button triggered the next trial after a 1000-msec delay.

The data were analyzed by calculating percent correct separately for /i/ and /I/ with and without duration modification. For statistical comparisons we used an ANOVA and Tukey HSD *post hoc* tests. For the cross-linguistic comparisons before and after the training, the between-subject factor of ANOVA was Language Group (Finnish vs. English) and the within-subject factors were Vowel (I vs. i) and Duration Type (normal duration vs. modified duration). For the training part, repeated measures ANOVA with within-subject factors of Test Session (pretraining vs. posttraining) \times Vowel \times Duration Type \times Word Type (trained vs. untrained words) was used.

MMN Experiment

Stimuli

The speech stimuli of the MMN experiment were synthesized words *beat* and *bit* (see Figure 1B) that were based on those used in a previous study (Iverson & Evans, 2007). They were synthesized using a Klatt synthesizer (Klatt & Klatt, 1990), and the acoustic parameters were selected based on best exemplars chosen by native British English speakers (Iverson & Evans, 2007). The formant values for F1–F3 were, respectively, 271, 2370, and 3220 Hz for *beat* and 436, 2051, and 2495 Hz for *bit*. For all stimuli, F0 was at the beginning of vowel 148 Hz and at the end 119 Hz. There were two pairs of stimuli: normal-duration stimuli with /i/ longer (130 msec) than /I/ (90 msec) and equal-duration stimuli, both with a

110-msec duration (i.e., the average of the two normal durations).² All other aspects of the stimuli were modeled to match a natural recording of a male native speaker of British English (see Iverson & Evans, 2007).

The *beat* and *bit* stimuli were presented in a passive oddball paradigm that contained a repetitive standard stimulus and occasional deviant stimuli. They were divided to 12 stimulus blocks, each including 434 stimuli. In three blocks, *beat* with normal vowel duration was presented as the repetitive, standard, stimulus ($p = .85$) and *bit* with normal duration as a deviant ($p = .15$). In another three blocks, the roles of the standard and deviant were reversed. The stimuli with equal durations were presented in a similar manner in three blocks and another three blocks with reversed roles of the standard and deviant. The ISI (from offset to onset) in these blocks varied randomly between 400 and 600 msec.

In addition to speech stimuli, one oddball block of 1200 nonspeech control stimuli was included in the experiment to allow the assessment of the stability of MMN across the language groups and test sessions. The nonspeech stimuli were complex tones. The repetitive standard stimulus ($p = .80$) had a fundamental frequency of 500 Hz (harmonic partials 1000 and 1500 Hz) and a duration of 75 msec. There were two deviant stimuli ($p = .10$ each), one deviating from the standard in frequency and the other in duration. The frequency deviant had the same duration as the standard, but its fundamental frequency was 550 Hz (harmonic partials 1100 and 1650 Hz). The duration deviant, in turn, had the same frequency as the standard, but its duration was 25 msec. With nonspeech stimuli, the ISI randomly varied between 250 and 350 msec. All stimuli in the MMN experiment were presented in a pseudorandom order where each deviant was preceded by at least two standard stimuli.

Data Acquisition and Analysis

The data acquisition and analysis were identical before and after the training. During the experiment, the participants sat in an electrically and acoustically attenuated room and watched a self-selected muted video film with subtitles in their native language whereas auditory stimuli were presented via headphones at the intensity of 50 dB above sensation level. The participants were instructed to ignore stimuli and to concentrate on watching the film. The nose-referenced EEG was recorded at a 250-Hz sampling rate with NeuroScan system and Synamps amplifier using a 32-channel electrode cap with Ag/AgCl electrodes. Eye movements were monitored with a bipolar EOG recorded from the canthi of both eyes and below and above the right eye. After the recording, the data were off-line filtered with a 1.5- to 20-Hz band-pass filter (24 dB/octave), artifacts exceeding $\pm 75 \mu\text{V}$ at any channel were rejected, the epochs of -100 to 900 msec from stimulus onset (-100 to 400 msec for nonspeech controls) were averaged separately for each stimulus type, and the baseline

was corrected to a 50-msec time window before change onset. The data were rereferenced to the average of the mastoids to improve the signal-to-noise ratio of MMN because MMN typically reverts its polarity at the mastoid sites. For responses to speech, the difference waves were created by subtracting the ERP response to a standard stimulus from that to the same stimulus as a deviant. For nonspeech controls, the same standard response was subtracted from the duration and the frequency deviants.

Similarly to some previous MMN training studies (Tremblay et al., 1997, 1998; Kraus et al., 1995), the pretraining and the posttraining comparison was based on the MMN area because the difference between the sessions does not necessarily reach its maximum at the MMN peak. The MMN area was measured in a time window between the latencies where MMN amplitude was 40% of the grand-average peak amplitude at the Fz scalp site. The MMN areas at F3, Fz, F4; C3, Cz, C4; and P3, Pz, P4 scalp sites were submitted to one-tailed *t* tests to assess that a statistically significant MMN response was elicited and were analyzed further using ANOVA. The Finnish L2 users' data were compared with that of the native English speakers in separate ANOVAs before and after the training. The ANOVA for these cross-linguistic comparisons had between-subject factor of Language Group (Finnish vs. English) and within-subject factors of Vowel in standard stimulus (/I/, unfamiliar to Finns vs. /i/, familiar to Finns), Duration Type (normal duration vs. equal duration), Coronal Scalp Site (frontal vs. central vs. parietal), and Sagittal Scalp Site (left vs. midline vs. right). The repeated measures ANOVA for the training part had within-subject factors of Test Session (pretraining vs. posttraining), Familiarity of Standard Stimulus (unfamiliar /I/ vs. familiar /i/), Duration Type, Coronal Scalp Site, and Sagittal Scalp Site. For nonspeech controls, the factors were Group or Test Session for cross-linguistic or training parts, respectively, Coronal Scalp Site, and Sagittal Scalp Site.

To test the differences in early exogenous components and to determine whether the standard responses account for possible changes in MMN, we measured and analyzed the peak amplitude of N1 and P2 of standard responses with corresponding ANOVAs as the MMN data.³ For obligatory responses, scalp distribution differences are reported only if they interacted with factors Group or Test Session because they were not our main interest.

Training

Stimuli and Procedure

Holt and Lotto (2006) suggested that increasing variance on the cue that the learner is trained to ignore may effectively change cue weighting. Therefore, in the present study, stimuli with modified duration were also used in the training to make duration cues unreliable, to guide Finnish listeners not to assimilate the stimuli to the Finn-

ish length categories, to draw their attention to spectral cues, and thus to guide them to use the cues that native English speakers tend to use in vowel recognition.⁴ Stimulus words used in the training were pronounced by two male and two female native speakers of English from Southern England (different than the speaker of experimental stimuli). They were considered to represent typical English pronunciation. The recordings, the selection of the stimuli, and the modification of their duration corresponded to those described in the context of the behavioral experiment, with the exception that the training session included only 19 minimal pairs. Because the modification of all 19 pairs pronounced by all four speakers was not successful, some of the pairs were excluded (e.g., utterances with salient intonation changes or creaky voicing were not included), resulting in the inclusion of 13–19 pairs from each speaker. Altogether, the stimulus material of the training sessions consisted of 244 stimuli, including 61 minimal pairs and their duration-modified variants.

Training was conducted by using a minimal-pair perceptual identification task. The participants had 10 training sessions that took place within three weeks period between the pretraining and the posttraining measurements. Each training session lasted about 20–25 min and was divided to three blocks. Participants were allowed a break between the blocks. During the training sessions, the participants sat in front of a computer screen in a quiet room. As in the behavioral tests, they were presented with one auditory stimulus at the time via headphones, and simultaneously two response options showed up on the screen (one on the left, another on the right). If the auditorily presented word was, for example, *bit*, the options on the screen were “hit” and “heat” (the order of options was random). The participants were instructed to click with a computer mouse the word they had just heard. A correct response triggered a positive feedback signal (a happy animated bunny) and, further, the next trial. An incorrect response resulted in a negative feedback signal (a sad animated bunny) and the repetition of the trial. The order of response options on the screen could be randomly switched in the repetition trial, so that the correct answer was not necessarily obtained by just changing the side of the response.

RESULTS

Behavioral Results

Figure 2A illustrates the percent correct of vowel recognition in the two language groups before the training. The main result revealed by ANOVA was a significant Language Group \times Vowel \times Duration Type interaction, $F(1,23) = 9.98$, $p < .01$. According to a *post hoc* test, this was due to lower percent correct in Finns than in the native English speakers for the duration-modified vowels ($p < .001$) but not for the vowels with normal duration. In the Finns, the

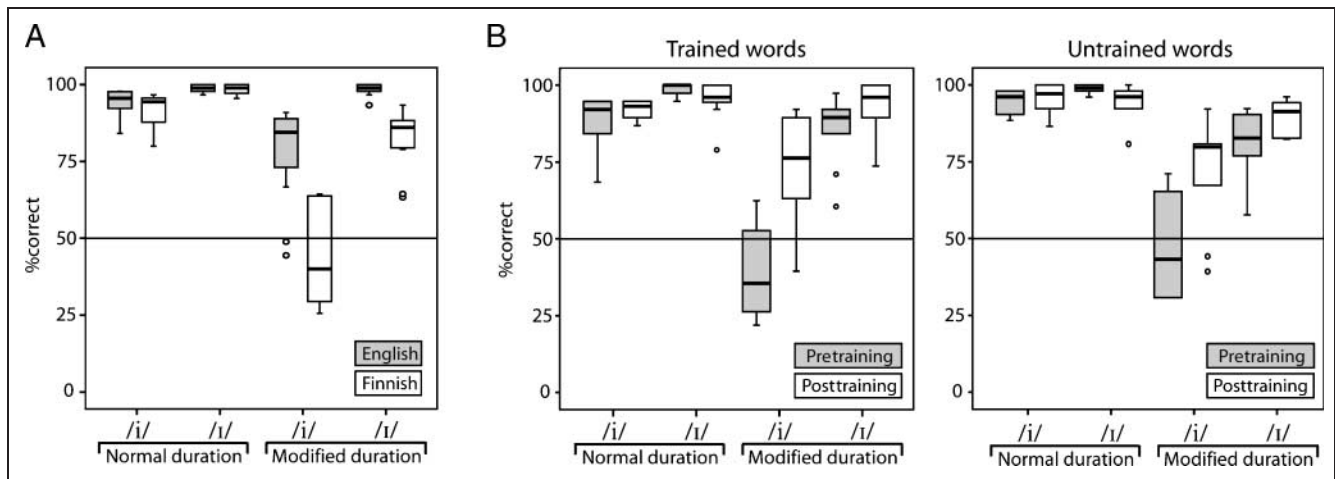


Figure 2. (A) Identification accuracy in the native speakers of Finnish and English in the behavioral cross-linguistic comparison. (B) Identification accuracy for trained (left) and untrained words (right) in behavioral pretraining and posttraining measurements in the Finns. The whiskers show smallest and largest observation, the box shows lower and upper quartile, and the line within the box shows the median. Outliers are marked with small circles.

duration-modified vowels always had a lower percent correct than the vowels with normal duration ($p < .001$), whereas in the native English speakers, a difference between the two duration types was observed for /i/ only ($p < .001$). ANOVA indicated also significant main effects of Language Group, $F(1,23) = 33.75$, $p < .001$, Vowel, $F(1,23) = 78.23$, $p < .001$, and Duration Type, $F(1,23) = 163.50$, $p < .001$, as well as Language Group \times Vowel interaction, $F(1,23) = 6.47$, $p < .05$, Language Group \times Duration Type interaction, $F(1,23) = 53.68$, $p < .001$, and Vowel \times Duration Type interaction, $F(1,23) = 88.97$, $p < .001$, but these were in more detail explained by the interaction of all three factors.

The identification accuracy in the pretraining and the posttraining behavioral experiments is illustrated in Figure 2B. The main finding with regard to training effects was a significant Test Session \times Vowel \times Duration Type interaction, $F(1,9) = 6.18$, $p < .05$. According to a *post hoc* test, before the training the percent correct for duration-modified /i/ was significantly lower ($p < .001$ for all comparisons) than that to any other stimulus type in either test session. After the training, the percent correct for the duration-modified /i/ was significantly improved as compared with the pretraining level ($p < .001$), although percent correct was still lower than that to other stimulus types in the same session ($p < .01$). In addition to /i/, also duration-modified /I/ was more difficult to identify before the training than its counterpart with normal duration ($p < .01$). With the duration-modified /I/, the improvement observed between the pretraining and the posttraining measurements did not reach significance, however. The interaction of all three factors explained in more detail three significant main effects: Test Session, $F(1,9) = 71.84$, $p < .001$; Duration Type, $F(1,9) = 67.65$, $p < .001$; and Vowel, $F(1,9) = 77.24$, $p < .001$. In addition, the main effect of Word Type, $F(1,9) = 5.64$, $p < .05$, indicated that the words used in the training were more difficult than the ones that were used in the pretraining and the post-

training measurements only. The comparison between the native English speakers' data and the Finns' posttraining data indicated no significant differences between the groups after the training.

MMN Results

Significant MMN responses (see Figures 3 and 5) were elicited in both language groups for all stimulus types at the frontal and central scalp sites ($p < .05$), except for nonspeech control frequency change at C3 in the training group of Finns before the training (when all 12 Finns and not only the training group were included, the MMN was significant at all fronto-central sites). For the cross-linguistic MMN data elicited by speech stimuli, the main finding was a significant Language Group \times Vowel in standard stimulus \times Coronal Scalp Site interaction, $F(2,42) = 5.07$, $p < .05$. A *post hoc* test indicated that when the deviant had been in the context of the standard stimulus /I/ that is unfamiliar to Finns, the Finns' MMN was significantly ($p < .001$) smaller at fronto-central sites than the native English speakers' MMN. No significant difference was found between the groups' MMNs elicited in the context of /i/ that is familiar to Finns. In addition, at frontal sites, the Finns' MMN elicited in the context of familiar /i/ was larger than that elicited in the context of unfamiliar /I/ ($p < .001$), whereas the native English speakers' MMN was larger in the context of /I/ than /i/ at fronto-central sites ($p < .05$). ANOVA revealed also a significant main effect of Duration Type, $F(1,21) = 7.92$, $p < .05$, that was due to larger MMN area for normal-duration stimuli than for equal-duration stimuli, and a significant main effect of Coronal Scalp Site, $F(2,42) = 48.67$, $p < .001$, that was caused by larger MMN amplitudes at frontal and central than at parietal scalp sites. The MMN responses elicited by nonspeech stimuli did not differ significantly between the two language groups, Duration Change, $F(1,21) = 3.21$, *ns*, and Frequency Change, $F(1,21) =$

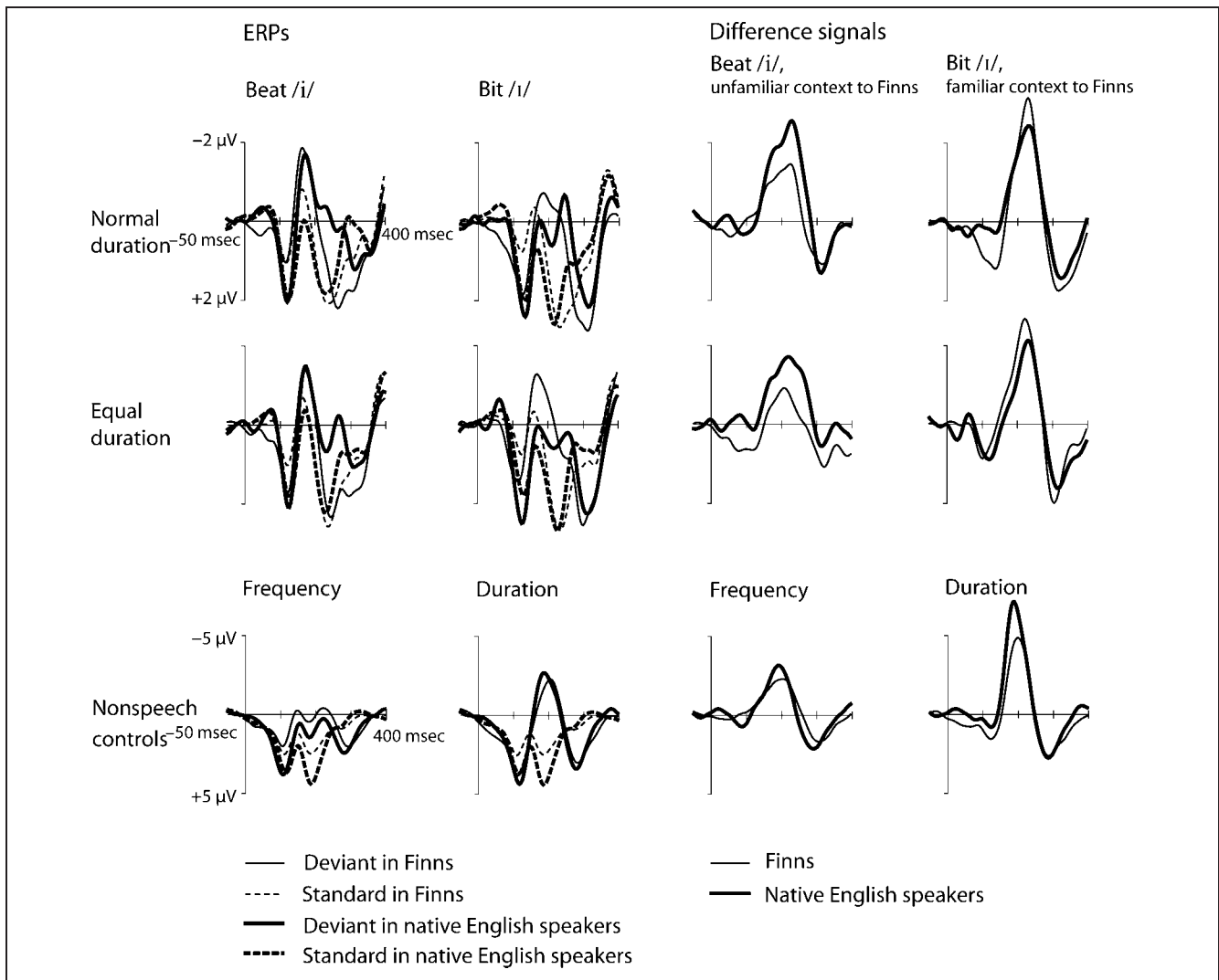


Figure 3. The MMN responses at Fz in the native speakers of English (thick line) and Finnish (thin line) before training as illustrated by ERPs (left) and deviant-minus-standard difference signals (right). Top rows: The MMN responses to speech stimuli *beat* and *bit* with normal and equal duration. Bottom row: The MMN responses to frequency and duration changes in the nonspeech control condition. Note the different scales for speech and nonspeech.

0.27, *ns*, although the native English speakers tended to have larger MMN responses, as illustrated by Figure 3.

The effect of training on the MMN responses in the pretraining and the posttraining measurements are illustrated in Figures 4 and 5. An ANOVA of the MMN area for speech stimuli revealed a significant Test Session \times Sagittal Scalp Site interaction, $F(2,18) = 4.56, p < .05$. According to a *post hoc* test, it was due to a significantly larger MMN at left and midline sites after than before the training ($p < .001$). In addition, whereas before the training the MMNs at the three sagittal scalp sites did not differ from each other, after the training the MMNs at the left and midline sites were significantly larger than those at the right hemisphere sites ($p < .05$). This interaction explained the main effects of Test Session, $F(1,9) = 9.04, p < .05$, and Sagittal Scalp Site, $F(2,18) = 6.27, p < .01$. ANOVA revealed also a significant Familiarity of Standard Stimulus \times Coronal Scalp Site interaction, $F(2,18) = 4.18, p < .05$, that explains the main effect of Coronal Scalp

Site, $F(2,18) = 17.36, p < .001$. According to the *post hoc* test of the interaction, at fronto-central sites the MMN elicited in the context of familiar standard /i/ was larger than that elicited in the context of unfamiliar standard /I/ ($p < .05$). The responses to the nonspeech control stimuli that were not trained did not differ significantly between pretraining and posttraining sessions: Duration Change, $F(1,9) = 1.50, ns$; and Frequency Change, $F(1,9) = 0.05, ns$.

The comparison between the native English speakers and the Finns after the training indicated a significant Group \times Sagittal Scalp Site interaction, $F(2,38) = 8.67, p < .001$, explaining the main effect of Sagittal Scalp Site, $F(2,38) = 6.00, p < .01$. A *post hoc* test for this interaction indicated that the native English speakers' MMNs were larger than the Finns' MMNs at midline and right sites ($p < .05$), whereas no difference was found between the groups over the left hemisphere. The Finns' MMNs were larger on the left than on the right ($p < .01$), whereas the native English speakers' MMNs were larger over the

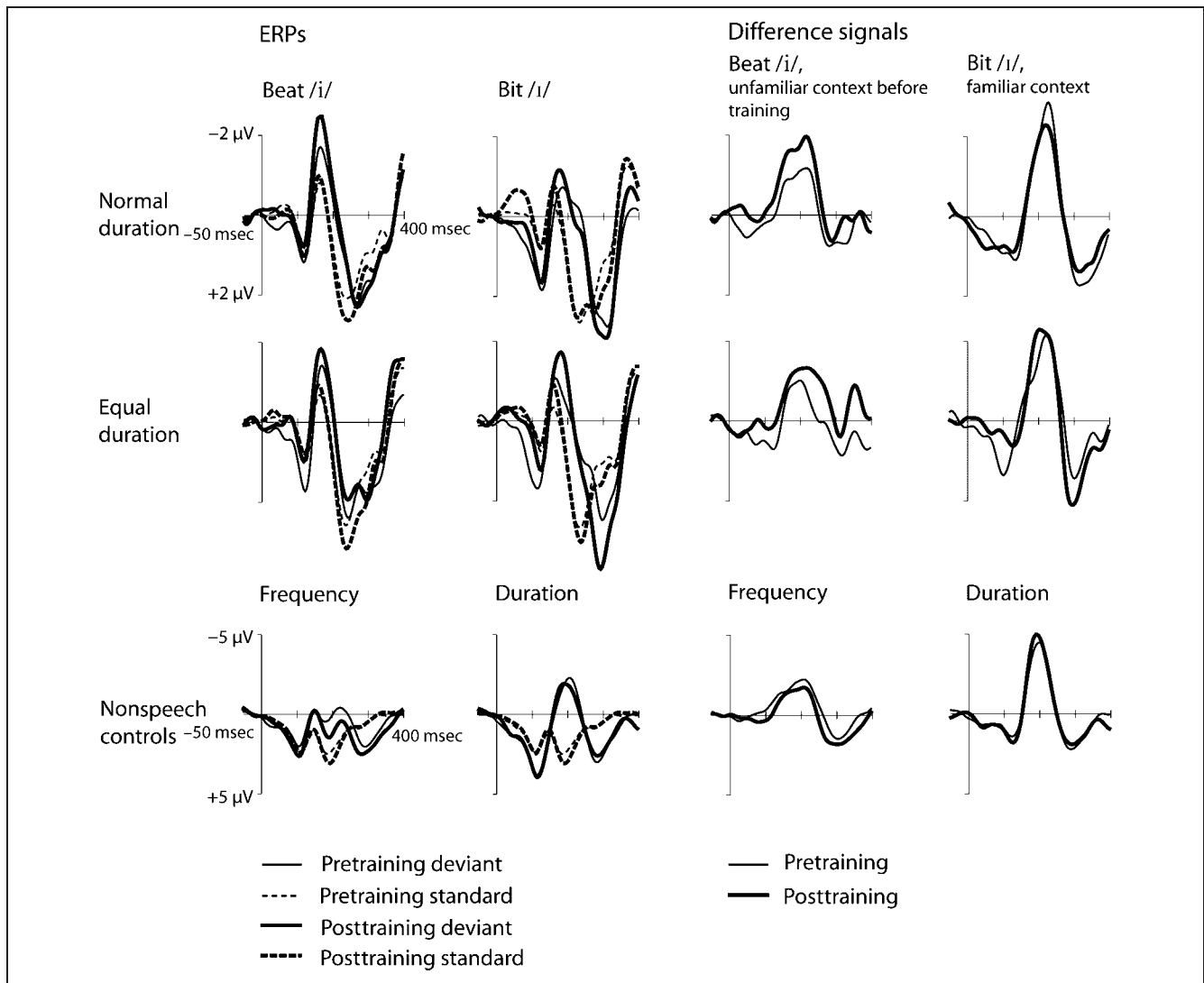


Figure 4. The MMN responses at Fz in the pretraining (thin line) and the posttraining (thick line) as illustrated by ERPs (left) and deviant-minus-standard difference signals (right). Top rows: The MMN responses to speech stimuli *beat* and *bit* with normal and equal duration. Bottom row: The MMN responses to frequency and duration changes in the nonspeech control condition. Note the different scales for speech and nonspeech.

midline than on the left ($p < .05$). Also a significant main effect of Coronal Scalp Site, $F(2,38) = 39.09$, $p < .001$, was found. It was due to larger MMN at frontal and central than at parietal scalp sites ($p < .001$). No significant differences were found between the groups in nonspeech sounds, Duration Change, $F(1,19) = 1.04$, *ns*, and Frequency Change, $F(1,19) = 0.41$, *ns*.

Obligatory Responses

In the training part, ANOVAs for N1 and P2 to speech sounds indicated that N1 and P2 were larger after than before the training: Test Session main effects for N1 and P2, respectively, $F(1,9) = 10.72$, $p < .01$ and $F(1,9) = 6.75$, $p < .05$. For P2 to nonspeech standards, the Test Session \times Sagittal Scalp Site interaction was significant, $F(2,18) = 3.61$, $p < .05$, and a *post hoc* test indicated larger P2s for all scalp-site levels after than before the training ($p < .01$). The cross-linguistic ANOVAs for N1 and P2 to speech in-

dicated only vowel effects: N1 was larger for /i/ than /I/, Vowel main effect, $F(1,21) = 17.21$, $p < .001$, and P2 was larger for /I/ than /i/, Vowel main effect, $F(1,21) = 13.41$, $p < .01$. In ANOVA for nonspeech standards, a significant Group \times Coronal Scalp Site interaction, $F(2,42) = 3.56$, $p < .05$, was found. According to a *post hoc* test, the native English speakers' P2 to these sounds was larger than that of the Finns at fronto-central electrodes ($p < .01$).

DISCUSSION

The present study aimed to train the brain to weight L2 phonetic cues differently, as reflected by behavioral and brain responses. First, a cross-linguistic comparison was conducted to determine whether behavioral and brain measures reflect differences in cue weighting of English /I/ and /i/ vowels between native English speakers and Finnish L2 users of English.

Behavioral Performance

Before the training, the behavioral identification accuracy for the duration-modified vowels was significantly lower in the Finns than in the native English speakers. The modification of the duration of /i/ dropped its correct recognition in the Finns to chance level, and it had a significant distorting effect on the recognition of /I/ as well. In the native English speakers, the distorting effect of duration modification was smaller, although for /i/ it was significant. The Finns' difficulty with duration-modified /i/ is likely due to its assimilation to Finnish length categories. Because the English /i/ is qualitatively very similar to the Finnish /i/ (Wiik, 1965), the Finns may have assimilated the English /i/ with normal duration to the Finnish long /i:/ category and the English /i/ with shortened duration to the Finnish short /i/ category. /I/, in turn, may have been less readily assimilated with the Finnish length categories because it is qualitatively nonprototypical in Finnish (Wiik, 1965). The stronger effect of duration modification on vowel recognition in the Finns as compared with the native English speakers suggests that Finns indeed seem to weight duration cue more than native speakers of English. However, because no difference between the groups was found in the normal-duration vowels, the Finnish L2 users appeared to be able to identify the vowels correctly based on the duration or the combination of duration and quality.

For the training part, the behavioral results suggested that after the training the Finns' identification accuracy

was overall higher, and no further significant differences were found in the performance of the two language groups. The improvement in the percent correct after the training was mainly due to the improvement with duration-modified /i/, for which the percent correct was the lowest before the training. As discussed above, the most probable reason for the Finns' poor pretraining performance is that they weighted the duration cue more than the native English speakers did and, possibly, mapped the /i/ vowels onto the Finnish length categories. After the training, however, the identification performance of the duration-modified /i/ had significantly improved, indicating that as a result of training, the Finns were more reliably able to hear the difference between the two vowels based on their spectral quality. This result suggests a change in the cue weighting as a function of training. Importantly, the behavioral training effects did not apply to trained words only, but they were generalized to untrained words: A similar improvement in the identification accuracy was observed for both word types.

Obligatory ERP Responses

In the Finns, N1 was found to be larger after than before the training. Although this result would be line with the findings of Ylinen and Huotilainen (2007) suggesting that N1 to standard stimuli may reflect vowel familiarity, it seems controversial that the native English speakers' N1 tended to be, in fact, less negative than that of the

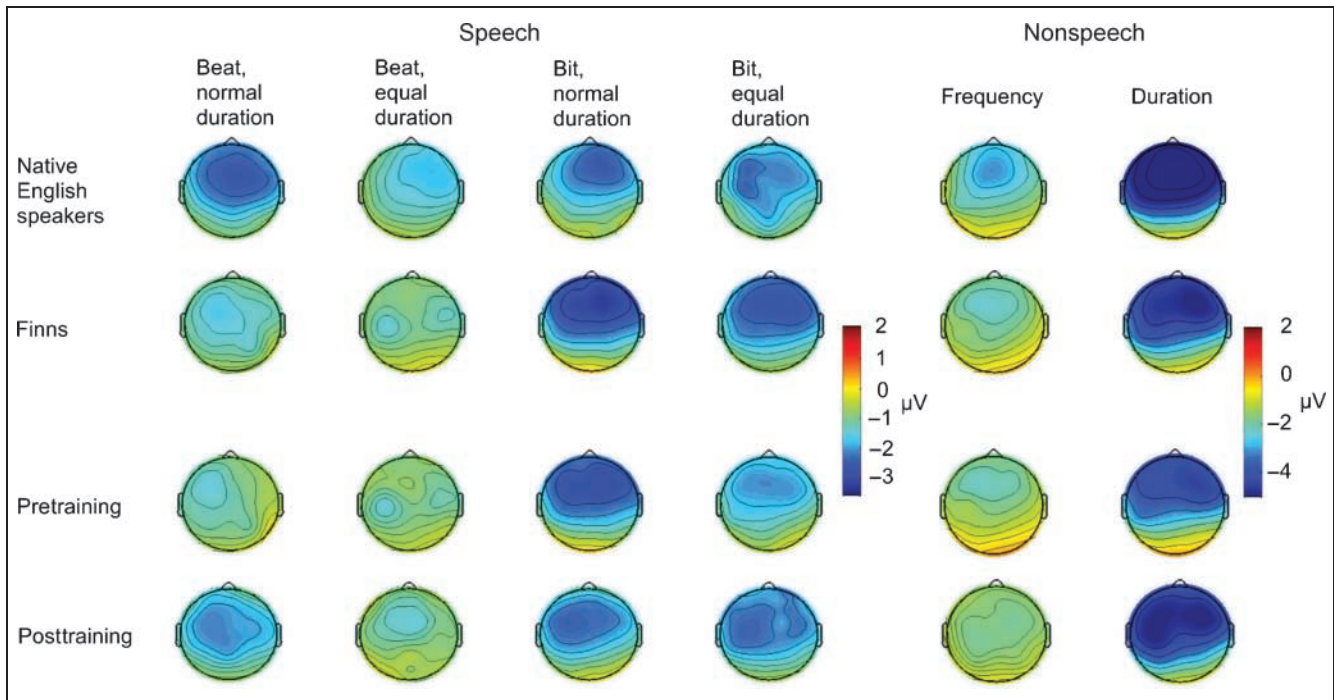


Figure 5. Topographical maps for the MMN (deviant-minus-standard difference signals). Top rows: Cross-linguistic comparison between the native English speakers and the Finns. Bottom rows: Pretraining and posttraining topography in the Finns. Note the different scales for speech (left) and nonspeech (right). The differences between the Finns' maps for the cross-linguistic comparison and pretraining are due to unequal number of participants ($n = 12$ vs. 10, respectively).

Finns. In addition to N1, in the Finns also P2 was found to be larger after than before the training, which is compatible with some previous studies on P2 enhancement after speech–sound training (Reinke, He, Wang, & Alain, 2003; Tremblay, Kraus, McGee, Ponton, & Otis, 2001). Recently, however, Sheehan, McArthur, and Bishop (2005) found that P2 was enhanced between two test sessions in both training and nontraining groups (see also Reinke et al., 2003), suggesting that exposure to stimuli during the first test session may be sufficient to enhance P2 in the second session and that P2 is not associated with perceptual accuracy. In our data, this view was supported by the fact that P2 was enhanced also for nonspeech control sounds that were not trained between the sessions. In addition, regardless of the training effect, no difference was found between the language groups in P2 to speech sounds. It is also noteworthy that the Finns' P2 enhancement between the sessions is not a likely account for their MMN enhancement because in nonspeech controls, P2 was enhanced but MMN was not. It is thus plausible that P2 was similarly enhanced in deviants and standards, although in deviants this was not detectable due to MMN overlap. The difference found between the two language groups for P2 to nonspeech control standards is likely related to a more positive displacement of the whole P1–N1–P2 complex in the native English speakers as compared with the Finns (see Figure 3). To conclude the discussion on obligatory responses, training appeared to have effects on them as well, but none of these account for the MMN enhancement that was of primary interest in the present study.

MMN Responses

In the MMN experiment, the main difference between the language groups occurred in the processing of the two vowels, whereas no significant differences were found in nonspeech control sounds (the native English speakers' nonsignificant tendency to larger grand-average responses to nonspeech control sounds as compared those of the Finns is likely due to individual variation). In the Finns, the MMN was smaller for familiar /i/ presented in the context of unfamiliar /I/ than that for unfamiliar /I/ presented in the context of familiar /i/, whereas in the native English speakers an opposite pattern was observed (for corresponding findings on asymmetrical vowel perception in native English speakers, see Polka & Bohn, 2003). Some previous MMN studies suggesting an enhanced MMN for familiar speech deviants as compared with unfamiliar deviants (e.g., Pulvermüller et al., 2001; Dehaene-Lambertz, 1997; Näätänen et al., 1997) would have predicted an opposite pattern of results also in the Finns. However, the result is compatible with previous reports, suggesting that the typicality of the standard stimulus that serves as a context for deviants enhances the MMN amplitude because the long-term memory representations for familiar stimuli enable the formation of more elaborate regularity repre-

sentations required for MMN elicitation (Jacobsen et al., 2004, 2005). Thus, the present results reflecting asymmetry in vowel processing in the Finns are suggested to be due to the fact that long-term memory facilitates the formation of elaborate regularity representation for familiar /i/ but not for the unfamiliar /I/. Therefore, larger MMNs may have been elicited for [I] deviants presented among [i] standards than vice versa. This interpretation is also in line with recent findings of Amitay, Irwin, and Moore (2006), who demonstrated that training with a set of identical stimuli resulted in improved discrimination accuracy, suggesting that the training had tuned the brain representations that were used as the basis of discrimination. It is likely that in the present study, similarly to the study of Jacobsen et al. (2005), the enhancing effect of prototypical standard stimuli on MMN became detectable because we had a fully crossed design and used ERP responses to physically identical stimuli to create difference signals.

The MMN experiment indicated also training effects in the Finnish L2 users: the MMN was significantly larger after than before the training over the left hemisphere and midline sites. Because no change occurred between the sessions for nonspeech control sounds, the MMN enhancement in the speech sounds must have been induced by the phonetic training rather than, for example, by familiarization with the stimuli during the two test sessions. Although the Test Session \times Familiarity of Standard Stimulus interaction was not significant, the posttraining MMN enhancement tended to be more pronounced for familiar /i/ presented in the context of previously unfamiliar /I/ than for unfamiliar /I/ presented in the context of familiar /i/ (see Figure 4). As a result of repeated perceptual exposure and active efforts to map the stimuli correctly onto English words in the training, the Finns have become more familiar with the spectral quality of previously nonprototypical /I/. This is likely to enable the formation of more accurate representations for standard /I/ stimuli, resulting in the enhancement of the MMN elicited by /i/ deviants. Directing the L2 users' attention to the spectral cues by reducing the reliability of duration cue in the categorization may also have reduced the assimilation of the English vowels with the long-term memory representations for the Finnish vowel length. This could, in turn, enable the development of a new long-term memory representation for /I/ that is independent of the Finnish phonological system.

The main difference between the behavioral and the MMN results was that in the behavioral results, the superior performance of the native speakers and the posttraining change in the cue weighting were reflected especially in the duration-modified vowels, whereas in the MMN experiments, these effects were observed in both duration types. The duration cue has probably played a less important role in the MMN elicitation than in the behavioral identification for the reason as follows. According to Winkler, Czigler, Jaramillo, Paavilainen, and Näätänen (1998), two successive changes in some auditory regularity

occurring within the temporal window of integration (~200 msec) may elicit only one MMN triggered by the first change because the two deviations are integrated into one unit. Accordingly, in the present study, the stimuli with normal duration appeared to elicit single MMNs regardless of the successive changes in the spectral quality and duration. Although normal-duration stimuli elicited slightly larger MMNs than equal-duration stimuli, duration probably played only a secondary role in the MMN elicitation, and the change in vowel quality occurring ~115 msec before the duration change was the primary determinant of MMN. As a consequence, the lack of interactions with the duration type in the MMN experiments is not likely to be due to, for example, smaller duration changes in the MMN stimuli as compared with the stimuli used in the behavioral experiments. The spectral account for the MMN elicitation does not, however, imply that no change in cue weighting was reflected in the MMN results. On the contrary, although the MMN was likely to be primarily determined by the spectral cues, the enhancement of the MMN as a result of training indeed suggests that the Finns used the spectral cues more efficiently even at the preattentive level after than before the training. Because it is implausible that the duration processing of vowels was enhanced as well, the enhanced ability to use spectral cues can be interpreted as reflecting a change in cue weighting in the MMN experiment. Thus, although both behavioral and brain measures suggested a change in L2 cue weighting as a function of training, the MMN results demonstrate that the training effects are reflected in brain responses as early as 200 msec after stimulus onset.

The present findings are in accordance with previous evidence (e.g., Iverson et al., 2003) that cue weighting is language specific and thus involves phonetic long-term memory representations. Accordingly, the enhanced post-training MMN responses indicate that the training resulted in plastic changes in the cortex that may be related to the development of a new long-term memory representation for the vowel /I/. This notion was supported by the fact that the MMN enhancement was stronger in the left hemisphere where long-term memory representations of speech sounds have previously been localized (Shtyrov, Pihko, & Pulvermüller, 2005; Näätänen et al., 1997). In addition, the comparison of the MMN responses of the native English speakers and the posttraining responses of the Finns suggests that after the training, the Finns' brain responses to the English vowels were similar to those of the native English speakers at the left hemisphere scalp sites. Still, as compared with the Finns, the native English speakers' responses were larger over the midline and at the right hemisphere sites, which may reflect a stronger bilateral activation in the native English speakers as opposed to the left-hemispheric lateralization in the Finns or, at least, a somewhat different orientation of sources in the two groups. However, unlike the Finns' topography, that of the native English speakers is inconsistent with the previous findings showing that the MMN elic-

ited by L1 speech sounds is lateralized to the left especially in word contexts (Shtyrov et al., 2005). Therefore, conclusions on laterality differences between the groups should be made with caution. Due to the relatively high test-retest stability of the MMN (Tervaniemi et al., 1999), the differences or changes in topography may be more reliably captured in repeated measures design used within the training group than in between-groups design used in the cross-linguistic comparison (cf. between-groups and repeated measures responses to nonspeech control sounds in Figures 3 and 4, respectively).

Assessment of the Effectiveness of Training

After training, no significant differences were observed in the two language groups' behavioral performance and also the Finns' MMN responses to L2 vowels were enhanced. Still, the posttraining differences between the MMN responses of the Finns and the native English speakers suggest that in addition to a short training period, more authentic exposure and practice with English is certainly needed before a truly native-like level in the processing of vowels is reached. To become permanent, the new phoneme representations would probably need repeated reinforcement in the form of attending to the critical L2 features when listening to L2 speech and efforts to pronounce L2 sounds similarly to native speakers. Nevertheless, the changes we observed in preattentive, automatic processing are likely to benefit the L2 users' speech perception in real-life communication. Instead of the Finnish length representations, more accurate L2 long-term memory representations may be automatically activated by L2 sounds, which allows immediate mapping onto correct word representations and reserves capacity to other processes, such as semantics and grammar. Although Finns speaking English (even at advanced level) typically produce no qualitative difference between /i/ and /I/ (Lintunen, 2004), more native-like perceptual representations are likely to improve also speech production.

Due to the significant improvement in behavioral performance and the enhancement of MMN brain responses, our training method resembling the high-variability phonetic training by Logan et al. (1991) can be considered effective. However, Iverson et al. (2005) recently found some other training techniques as effective as the high-variability phonetic training in teaching the English /r/-/l/ distinction to Japanese listeners. A successful training method helps learners to focus attention on relevant features. This was recently demonstrated by Polley, Steinberg, and Merzenich (2006) who trained two groups of rats with the same stimulus set to attend either sound frequency or intensity. The rats' identification performance improved, and changes occurred in their brain representations for the attended feature only, suggesting that the interaction of the bottom-up and top-down inputs is essential for perceptual learning. Correspondingly, Alain, Snyder, He, and Reinke (2007) observed rapid perceptual

learning and the enhancement of its neural correlates only when vowel stimuli had been attended. Thus, it seems that training with different techniques can be successful as long as they draw learners' attention to relevant cues and help them to learn to ignore the irrelevant ones, regardless of whether the cues occur in natural or manipulated speech.

Conclusions

Both behavioral and electrophysiological data suggested significant training effects on the processing of English /i/ and /I/ vowels and, in particular, on the cue weighting in native speakers of Finnish. After the training, the Finns were able to use spectral cues more reliably and depend less on duration cues in the recognition of words with /i/ and /I/ vowels, as reflected by their behavioral responses. In addition, the enhanced MMN responses observed after the training suggested that training had improved the Finns' ability to process the spectral cues of the English vowels at the preattentive level. Taken together, the results imply reorganization in the weighting of phonetic cues in the cortex as a result of training.

APPENDIX A

Words used in behavioral experiment and training:

bean—bin	neat—knit
beat—bit	peach—pitch
cheap—chip	peak—pick
deep—dip	seat—sit
feet—fit	seek—sick
gene—gin	seen—sin
heat—hit	sleet—slit
lead—lid	team—Tim
least—list	teen—tin
mead—mid	

Words used in behavioral experiment only:

bead—bid	keen—kin
beaker—bicker	Keith—kith
beater—bitter	leafed—lift
beef—biff	leak—lick
bees—biz	litter—litter
bleep—blip	neap—nip
dean—din	Pete—pit
deed—did	seed—Sid
each—itch	sheep—ship
ease—is	skeet—skit
feast—fist	sleep—slip
fleet—flit	sneak—snick
heed—hid	speak—spick

Acknowledgments

The research presented in this article was supported by the Academy of Finland (project numbers 212819, 77322, 79820, and 79821) and the Engineering and Physical Sciences Research

Council UK (EPSRC; a grant GR/S55095/02 awarded to Dr. Uther). The authors thank Miika Järvenpää for technical assistance.

Reprint requests should be sent to Sari Ylinen, Cognitive Brain Research Unit, Department of Psychology, P.O. Box 9, FIN-00014 University of Helsinki, Finland, or via e-mail: sari.ylinen@helsinki.fi.

Notes

1. Deviation was three interquartile ranges above the upper quartile of the native English speakers' MMN amplitudes.
2. It is noteworthy that two successive changes within the temporal window of integration (~200 msec) may elicit only one MMN triggered by the first change (Winkler et al., 1998). Thus, there was a possibility that in the present experiment, the MMN to normal-duration stimuli would be elicited by spectral features only because they were detectable first. Still, both normal- and equal-duration stimuli were used to determine possible typicality effects and at the same time to control the contribution of duration and spectral cues to the MMN. In the behavioral experiments, fully crossed duration modification was used because we wanted to test the training effects with similar stimulus material as what was used in training. For the purpose of training, then, the ~20- to 30-msec difference between the normal and the equal average durations was thought not to be salient enough to catch the learners' attention and to help them to give up using the duration cue; more extreme durations within the natural duration range of tense and lax vowels was considered beneficial for training (see Holt & Lotto, 2006).
3. A preliminary analysis was conducted also for P1 to test the significance of small differences observed in ERPs at this latency, but no significant differences were found between the groups or between the training sessions.
4. The manipulation of stimuli differentiates the training used in the present study from the high-variability phonetic training of Logan et al. (1991).

REFERENCES

- Alain, C., Snyder, J. S., He, Y., & Reinke, K. S. (2007). Changes in auditory cortex parallel rapid perceptual learning. *Cerebral Cortex*, *17*, 1074–1084.
- Amitay, S., Irwin, A., & Moore, D. R. (2006). Discrimination learning induced by training with identical stimuli. *Nature Neuroscience*, *9*, 1446–1448.
- Boersma, P., & Weenink, D. (2004). Praat: Doing phonetics by computer (Version 4.2.17) [Computer program]. Retrieved October 1, 2004 from <http://www.praat.org/>.
- Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*, 303–328.
- Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133–160). Mahwah, NJ: Lawrence Erlbaum.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*, 977–985.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299–2310.
- Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, *8*, 919–924.

- Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*, 635–647.
- Flege, J. E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders, a review 1988* (pp. 224–401). Norwood, NJ: Ablex Publishing Corporation.
- Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.
- Flege, J. E., & Hillenbrand, J. (1986). Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *Journal of the Acoustical Society of America*, *79*, 508–517.
- Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America*, *108*, 3013–3022.
- Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, *119*, 3059–3071.
- Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *Journal of the Acoustical Society of America*, *122*, 2842–2854.
- Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America*, *118*, 3267–3278.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47–B57.
- Jacobsen, T., Horvath, J., Schröger, E., Lattner, S., Widmann, A., & Winkler, I. (2004). Pre-attentive auditory processing of lexicality. *Brain and Language*, *88*, 54–67.
- Jacobsen, T., Schröger, E., Winkler, I., & Horvath, J. (2005). Familiarity affects the processing of task-irrelevant auditory deviance. *Journal of Cognitive Neuroscience*, *17*, 1704–1713.
- Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, *87*, 820–857.
- Kraus, N., McGee, T., Carrell, T. D., King, C., Tremblay, K., & Nicol, T. (1995). Central auditory system plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience*, *7*, 25–32.
- Kujala, A., Huotilainen, M., Uther, M., Shtyrov, Y., Monto, S., Ilmoniemi, R. J., et al. (2003). Plastic cortical changes induced by learning to communicate with non-speech sounds. *NeuroReport*, *14*, 1683–1687.
- Lintunen, P. (2004). *Pronunciation and phonemic transcription: A study of advanced Finnish learners of English* (Anglicana Turkuensia No. 24). Turku, Finland: University of Turku.
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*, 1242–1255.
- Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *96*, 2076–2087.
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874–886.
- McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*, 229–258.
- 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. *Learning & Memory*, *9*, 253–267.
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *38*, 1–21.
- Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*, 313–329.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434.
- 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.
- 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. *Neuroscience Letters*, *352*, 25–28.
- Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology & Neuro-Otology*, *5*, 111–139.
- Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*, 221–231.
- Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *Journal of Neuroscience*, *26*, 4970–4982.
- Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., et al. (2001). Memory traces for words as revealed by the mismatch negativity. *Neuroimage*, *14*, 607–616.
- Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, *79*, 49–71.
- Reinke, K. S., He, Y., Wang, C., & Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Research*, *17*, 781–791.
- Sheehan, K. A., McArthur, G. M., & Bishop, D. V. (2005). Is discrimination training necessary to cause changes in the P2 auditory event-related brain potential to speech sounds? *Cognitive Brain Research*, *25*, 547–553.
- Shtyrov, Y., Pihko, E., & Pulvermüller, F. (2005). Determinants of dominance: Is language laterality explained by physical or linguistic features of speech? *Neuroimage*, *27*, 37–47.
- Suomi, K. (1976). *English voiceless and voiced stops as produced by native and Finnish speakers* (Jyväskylä Contrastive Studies, No. 2). Jyväskylä, Finland: Department of English, University of Jyväskylä.
- Tervaniemi, M., Lehtokoski, A., Sinkkonen, J., Virtanen, J., Ilmoniemi, R. J., & Näätänen, R. (1999). Test-retest reliability of mismatch negativity for duration, frequency and intensity changes. *Clinical Neurophysiology*, *110*, 1388–1393.
- Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: Generalization to novel

- stimuli following listening training. *Journal of the Acoustical Society of America*, *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.
- Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1–P2 complex after speech-sound training. *Ear and Hearing*, *22*, 79–90.
- Wiik, K. (1965). *Finnish and English vowels* (Annales Universitatis Turkuensis, Series B, Vol. 94). Turku, Finland: University of Turku.
- Winkler, I., Czigler, I., Jaramillo, M., Paavilainen, P., & Näätänen, R. (1998). Temporal constraints of auditory event synthesis: Evidence from ERPs. *NeuroReport*, *9*, 495–499.
- Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, *742*, 239–252.
- Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csepe, V., et al. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research*, *7*, 357–369.
- Winkler, I., Schröger, E., & Cowan, N. (2001). The role of large-scale memory organization in the mismatch negativity event-related brain potential. *Journal of Cognitive Neuroscience*, *13*, 59–71.
- Ylinen, S., & Huotilainen, M. (2007). Is there a direct neural correlate for memory-trace formation in audition? *NeuroReport*, *18*, 1281–1284.
- Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., & Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. *Brain Research*, *1072*, 175–185.