Listen-and-repeat training improves perception of second language vowel duration: evidence from mismatch negativity (MMN) and N1 responses and behavioral discrimination

Antti Saloranta [a,b] (corresponding author)
antti.saloranta@utu.fi

Paavo Alku [c]
paavo.alku@aalto.fi

Maija S. Peltola [a,b]
maija.peltola@utu.fi

[a]Department of Future Technologies
University of Turku
Koskenniemenkatu 4
20014 University of Turku
Finland

[b]Learning, Age & Bilingualism laboratory (LAB-lab)
University of Turku
Koskenniemenkatu 4
20014 University of Turku
Finland

[c]Department of Signal Processing and Acoustics
Aalto University
P.O. Box 12200, 00076 Aalto
Finland

Declarations of interest: none.

Abstract

The purpose of this study was to examine the efficacy of three days of listen-and-repeat training on the perception and production of vowel duration contrasts. Generalization to an untrained vowel and a non-linguistic sound was also examined. Twelve adults underwent four sessions of listen-and-repeat training over two days with the pseudoword contrast /tite/-/ti:te/. Generalization effects were examined with another vowel contrast, /tote/-/to:te/ and a sinusoidal tone pair as a non-linguistic stimulus. Learning effects were measured with psychophysiological (EEG) event-related potentials (mismatch negativity and N1), behavioral discrimination tasks and production tasks. The results showed clear improvement in all perception measurements for the trained stimuli. The effects also affected the untrained vowel by eliciting an N1 response, and affected the behavioral perception of the non-linguistic stimuli. The MMN response for the untrained linguistic stimuli, however, did not increase. These findings suggest that the training was able to increase the sensitivity of preattentive auditory duration discrimination, but that phoneme-specific spectral information may also be needed to shape the neural representation of phoneme categories.

Keywords: second language acquisition; production training; mismatch negativity; event-related potential; EEG

1 Introduction

Quantity languages, in which duration of individual speech segments differentiates between meanings of words and is therefore a phonological feature, can be problematic for learners, as phonological duration contrasts are somewhat uncommon in some of the most widely spoken languages of the world. Most difficulties in discriminating and producing unfamiliar second language segmentals and suprasegmentals stem from the influence of the native language, and are a result of the brain becoming desensitized to phonetic variation that is irrelevant for differentiating between native sounds (Iverson et al., 2003), such as duration in the case of non-quantity languages. Iverson et al. (2003) found that speakers of Japanese have a highly reduced sensitivity to sound frequency changes most relevant for the detection of the English /r/-/l/ contrast, explaining the major difficulties Japanese speakers have with this contrast. Similar results have been found for Finnish speakers, who, as speakers of a quantity language, show a higher sensitivity to phonetic duration differences than Germans (Kirmse et al., 2008); furthermore, Finns, unlike Russians, also seem to have developed phonemic categories for duration differences (Ylinen et al., 2005b). Popular models of second language acquisition, such as the Speech Learning Model (SLM) by Flege (e.g. Flege, 1987) and the Perceptual Assimilation Model (PAM) by Best and Strange (e.g. Best and Strange, 1992) posit that discrimination of second language phonemes is based on how they fit within the existing native language phoneme categories. According to both models, the most difficult learning situation arises when a second language phoneme shares some features with one or more native language categories, but still differs systematically from them. This would be the case, for example, for a native Spanish speaker hearing Finnish. Spanish has vowel phonemes that have a close spectral match with some Finnish vowels, such as /i/, but it does not have phonemic duration categories, which would likely mean that both short and long instances of

Finnish /i/ would be mapped into the same Spanish /i/ category, leaving the duration contrast overlooked.

The use of training to overcome the desensitization caused by the native language has been extensively studied. Perceptual methods are typically used for training, but learning outcomes may be measured in both perception and production. Strange and Dittmann (1984) used same-different discrimination training with immediate feedback with the English /r/-/l/ contrast and Japanese participants. They found improved identification and discrimination of a trained synthetic "rock" –"lock" stimulus series and an untrained "rake"-"lake" series with 14-18 training sessions over three weeks. The same /r/-/l/ contrast has also been trained with so called high variability phonetic training (e.g. Bradlow et al., 1999, 1997), where naturally produced minimal pairs from different speakers are used in perceptual identification training with feedback on correct or incorrect classifications. In identification training, participants are presented with stimuli and are asked to classify them according to predetermined categories. Bradlow et. al. (1997) found improved identification performance of the /r/-/l/ contrast with adult Japanese participants with 45 sessions of training over three weeks. The effects also generalized to new talkers and untrained stimuli, and the participants received better listener ratings from native evaluators on their productions of English words containing /r/ and /l/. Bradlow et. al. (1999) found that the learning effects achieved with high variability phonetic training were retained at least three months after the training has ended, again measured by identification performance for the perceptual domain and native evaluator rating in the production domain. Several earlier studies (e.g. Hirata, Whitehurst, & Cullings, 2007; Okuno, 2014; Tajima, Kato, Rothwell, Akahane-Yamada, & Munhall, 2008) suggest that perception of vowel and consonant duration can also be improved with perceptual (non-high variability) identification training, although the results are not entirely conclusive. All of

these three studies report improvement on the perception of duration, with Okuno (2014) also finding improvements in production accuracy. However, the training did not consistently result in generalization to new stimuli or talkers. It may be that duration contrasts are not acquired through training as easily as other features, such as the /r/-/l/ contrast or vowel quality, i.e. the differences in frequencies that are used differentiate between vowels, and that it requires a more multifaceted approach in training. Previous studies have indeed suggested that segment duration and quality are processed through separate neural mechanisms (e.g. Ylinen et al., 2005a) and this may be part of the reason they also behave differently in training studies.

A method that is notably absent from duration training studies is production training, despite several recent studies showing it to be an effective training tool in second language acquisition. Taimi et al. (2014) used listen-and-repeat training with young Finnish children in order to help them produce a Swedish vowel contrast not found in Finnish. In the study, 7-10-year-old children underwent four sessions of training over two days, consisting of a total of 120 repetitions of the novel contrast, and were able to achieve more native-like production of the new vowel already after three of the four short training sessions. This training also proved useful for individuals aged 62 to 73 years learning foreign language(s) by demonstrating improved production of a vowel contrast with four training sessions over two days when comparing them to individuals of similar age with no foreign language learning (Jähi et al., 2015). Saloranta et al. (2015) found learning effects in a study using the same contrast and a listen-and-repeat training procedure enhanced with instructions. The aim of the instructions was to make the participants explicitly aware of the feature being trained. With this training, 18-30-year-old Finnish adults were able to change their production of the novel Swedish vowel after just one out of four training sessions. Finally, Saloranta et al. (2017) reported

improved behavioral discrimination and production of vowel duration using listen-and-repeat as the training method. Participants underwent four sessions of listen-and-repeat training over two days, for a total of 150 repetitions of the trained stimulus pair, which improved the discrimination scores and changed the short/long syllable ratios in production of the trained stimuli. In addition to learning effects in behavioral discrimination and production of vowel quality and duration contrasts, listen-and-repeat has also been used to improve behavioral perception of consonant voicing contrasts. Two studies by Tamminen et al. (Tamminen et al., 2015; Tamminen and Peltola, 2015) found that listen-and-repeat training produced learning effects for behavioral discrimination and identification of voice onset time (VOT). In both studies, a group of 18-32-year-old native Finnish speakers was trained on an English VOT contrast (/fi:l – vi:l/) not found in their native language. The subjects underwent four sessions of listen-and-repeat training with no feedback over three days, with each session consisting of 30 repetitions of the voicing contrast for a total of 120 repetitions during training. Both studies reported increased sensitivity and decreased reaction times in discrimination tasks and changes in category boundary steepness in identification tasks, and additionally changes in stimulus goodness ratings in the former study and category boundary in the latter.

In the present study, psychophysiological event-related potentials are used to evaluate training effects, most importantly the mismatch negativity (MMN) and the N1 responses. Of these, the N1 is a preattentive negative polarity response that peaks at about 100 ms after stimulus onset or offset or after a change in stimulus energy (Näätänen and Picton, 1987). It is evoked by abrupt changes in energy levels of the stimuli, typically stimulus onset, i.e. the transition from silence to sound (Näätänen and Picton, 1987), and it is considered to have little linguistic significance (Kujala and Näätänen, 2010). However, N1 amplitudes may also be larger when a deviating stimulus is detected in a sequence of similar stimuli, such as a

1500 Hz tone in a sequence of 1000 Hz tones (Näätänen and Picton, 1987, p. 388), which may make it useful for examining auditory discrimination sensitivity. Some training effects on N1 have been demonstrated, for example, with discrimination training using sine tones (Brattico et al., 2003), where subjects were taught to discriminate a specific tone from a group of seven tones. N1 amplitudes diminished for all but the trained tone as a result of one hour of discrimination training. Tremblay et al. (2001) found an increase in N1-P2 peak-to-peak amplitude after English-speaking subjects underwent identification training with feedback in order to discriminate within-category VOT contrasts in CV syllables. These changes were coupled with behavioral improvement, suggesting that as the within-category VOT difference became meaningful, the neural response to it was enhanced.

The MMN is a later preattentive response, occurring 150-250 ms after a deviation from a pre-established pattern, such as a string of identical auditory stimuli followed by an occasional deviant (Kujala et al., 2007), differing, for example, in amplitude or frequency. MMN is elicited even when subjects are not attending to the stimuli (Kujala et al., 2007) and it is sensitive to language (Näätänen et al., 1997), meaning that the elicitation of the response to the same group of stimuli depends on the listener's native language. Language-specific MMN responses have been demonstrated for, for example, vowel quality (e.g. Näätänen et al., 1997; Winkler et al., 1999), lexical tone contours (Chandrasekaran et al., 2009) and vowel duration (e.g. Ylinen et al., 2006). It has also been demonstrated that memory traces induced by learning can be examined with MMN, and that MMN responses can change during second language acquisition, for example as a result of language exposure after immigrating to a new country (e.g. Winkler et al., 1999), as a result of training (e.g. Tamminen et al., 2015; Tremblay et al., 1998) or during language immersion (Peltola et al., 2005). Training effects for mismatch negativity have been observed for several linguistic structures. Tremblay et al.

(1997) trained English speaking adults to discriminate voice onset time contrasts in CV syllables starting with a stop consonant. Subjects participated in 9 sessions of identification training with feedback, where they had to identify whether the stimulus they heard was prevoiced, voiced, or voiceless. Learning outcomes were measured with behavioral identification and discrimination tests as well as MMN, and they showed that behavioral performance improved in both the discrimination and identification tests related to the baseline, and that duration and area of the MMN also increased. Furthermore, the effects generalized to a set of VOT stimuli. Menning et al. (2002) trained adult German subjects to discriminate Japanese mora structures and analyzed the results with behavioral discrimination testing as well as with mismatch negativity. Subjects underwent 10 days of adaptive discrimination training that got more difficult as discrimination performance increased. Clear learning effects were observed in both behavioral discrimination and the MMN: reaction times decreased and discrimination accuracy increased and MMN amplitudes were higher after the completion of the training. Of particular interest for the current study are the results from Tamminen et al. (2015) and Tamminen & Peltola (2015), who used MMN to study second language voicing contrast acquisition using listen-and-repeat training. Both studies showed statistically significant increases in MMN amplitudes; Tamminen et al. (2015) reported that a three-day listen-and-repeat training paradigm was able to elicit an MMN response to a non-native voicing contrast in Finnish young adult learners of English, while Tamminen & Peltola (2015) showed that the amplitude of an existing MMN response to the same voicing contrast could be further increased with listen-and-repeat training.

The purpose of the current study is to answer three main questions: first, can vowel duration perception or production be improved with the same amount of listen-and-repeat training as perception and production of vowel quality or consonant voicing contrasts? Previous studies

have shown that listen-and-repeat can be an effective method for the latter two, but it has been suggested that the phonological system may be divided into separate levels: phoneme duration, for example, is processed independently from phoneme quality (Ylinen et al., 2005a). Second, if vowel duration processing can be trained, are the effects generalized to other, untrained vowels? Finally, if generalization occurs, is it limited to linguistic sounds or is the processing of non-linguistic sounds affected as well? Liégeois-Chauvel et al. (1999) found that temporal processing of speech and non-speech utilizes the same neural mechanism, and a training scheme that affects duration processing in speech could therefore also induce learning effects in non-speech sounds.

2 Materials and methods

2.1 Stimuli

The linguistic stimuli used in the experiment were disyllabic, semisynthetic Finnish pseudoword pairs. Semisynthetic stimulation refers to the sound generation approach where the excitation of the natural human speech production mechanism is combined with a digital model of the vocal tract in order to produce vowel stimuli; more information on the method can be found in Alku et al. (1999). The method uses an extracted glottal excitation waveform from a real speaker, producing natural sounding stimuli with phonetic features that can be carefully controlled. The pairs used were /tite/ – /ti:te/ and /tote/ – /to:te/. These words were chosen due to due to their highly different places of articulation, and due to phonemes similar to /i/ and /o/ being among the 10 most common vowel segments in the languages of the world (Maddieson and Disner, 1984, p. 125). A stimulus pair consisting of sinusoidal tones and noise was also created to serve as a non-linguistic stimulus. All of the stimuli had an

identical CVCV structure, and the linguistic stimuli were identical in aspects other than the target vowel. The mean fundamental frequency of all vowels was 110 Hz. The first (F1) and second formant (F2) for the target vowel was 330 Hz and 2129 Hz respectively in /i/, and 452 Hz and 805 in /o/. The formants for the final /e/ in both words were 477 Hz for F1 and 1692 Hz for F2. The duration of occlusion of the middle consonant was 58 ms in all stimuli. In the non-linguistic stimuli, the explosion phases of the consonants consist of white noise, and the parts representing the vowels consist of sinusoidal tones. The frequency of the lowest tone was adjusted to be equal to the mean of the F1 and F2 frequencies in the corresponding vowel of the linguistic stimuli. In addition to the lowest tone, the non-linguistic stimuli consisted of one sinusoidal per every 1 kHz. In this article, /tite/ – /ti:te/ and /tote/ – /to:te/ will from here on be referred to as the trained linguistic and untrained linguistic stimuli, respectively. All short stimuli were 392 ms long (first syllable 154 ms) and the long ones 428 ms (first syllable 194 ms). This difference between the stimuli was confirmed by several native phoneticians to be difficult but discriminable. The difference between the members of the stimulus pairs begins at 120 ms. More information on the creation and features of these stimuli can be found in Saloranta et al. (2017).

2.2 Participants

A total of 12 participants (4 men) took part in the experiment. Subjects were healthy 19-29-year-olds who were recruited among exchange students entering the University of Turku. All subjects volunteered to take part in the project and gave their written consent. The study was approved by The Ethics Committee of the University of Turku. The subjects' linguistic background was carefully examined in order to exclude participants with any phonological quantity contrasts in their native language or languages they were highly proficient in. The

native languages of the participants were French (4), Spanish (2), English (2), Russian (1), Lithuanian (1), Mandarin (1) and Nepalese (1). All of these languages contain vowel phonemes similar to the ones in the linguistic stimuli (Augustaitis, 1964, p. 12; Fougeron and Smith, 1993; Khatiwada, 2009; Lee and Zee, 2003; Martnez-Celdrn et al., 2003; Yanushevskaya and Bunčić, 2015). The subject's handedness was assessed by self-evaluation using the Edinburgh Handedness Inventory (Oldfield, 1971); all subjects were right-handed. The subjects' hearing on the 100-4000 Hz range at 5-25 dB was tested using a Grason-Stadler GSI 18 audiometer. All subjects had normal hearing in this range. Subjects were also asked to self-evaluate their Finnish skills on a scale of 0-4, (0 = no skills, 1= basic, 2 = satisfactory, 3= manages in everyday situations, 4 = excellent). The mean level was 0.67 (stdev 0.47), with no participants rating themselves higher than 1. At the end of the experiment, each subject was asked to give a brief, oral self-evaluation of their performance and to offer their own suggestions as to what they thought the research was about.

2.3 Test structure

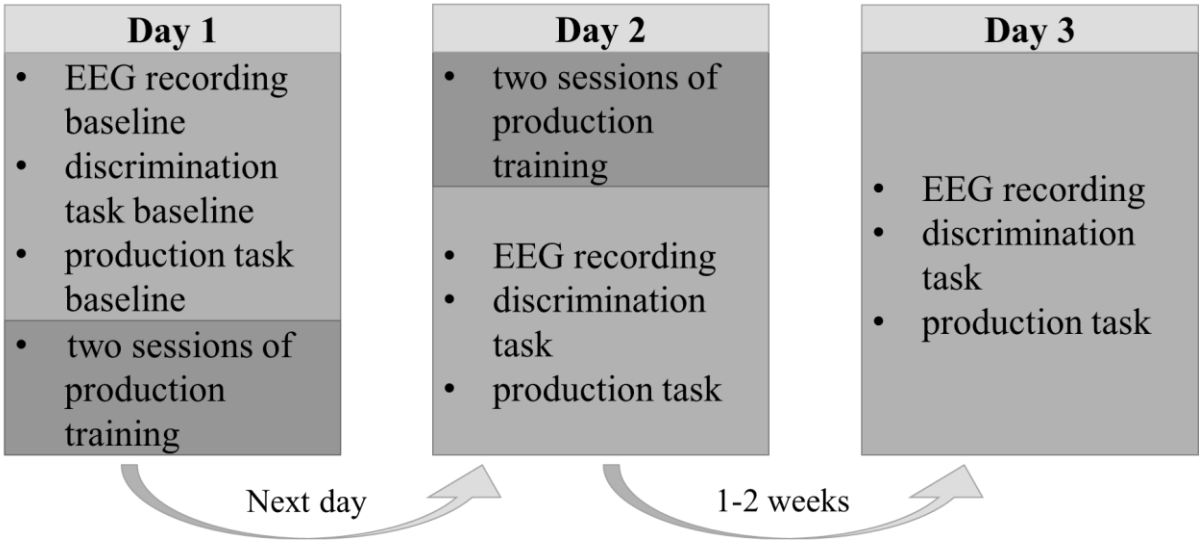


Figure 1. (adapted from Saloranta et al. 2017) Structure of the experiment. On Day 1 and Day 3, all three stimulus pairs were used in the baseline/progress measurements, apart from the production task baseline in

which the untrained linguistic or the non-linguistic pairs were not used. All four training sessions and the Day 2 progress measurements were conducted with only the trained linguistic stimuli. Furthermore, the untrained linguistic stimuli were also used in the production task on Day 3.

The three-day test structure (Figure 1) was the same as the one employed by Saloranta et al. (2017), with the first day containing all baseline measurements for all stimuli and the first two blocks of training, the second day containing the final two blocks of training and measurements for only the trained linguistic stimuli, and the third day containing the final measurements for all stimuli.

2.4 Production task and training

The same basic test structure was used during both the production recordings and the production training. The stimuli were presented to the subjects diotically (i.e the same mono sound arriving at both ears) in an alternating long-short pattern with an interstimulus interval (ISI) of three seconds, using the Sanako SLH-07 headset and Sanako Lab 100 language lab software and hardware. Subjects were instructed to listen to each token carefully, then repeat it clearly and calmly in a normal voice as accurately as they could. During the production tasks, the stimulus pairs were presented 10 times, and during the training 30 times, for a total of 150 repetitions of the trained linguistic pair (120 during training and 30 during recordings) and 10 repetitions of the untrained linguistic pair throughout the experiment. In the training block the subjects were instructed to repeat each word and that they could use this block as practice. No feedback was given in any of the production tasks or during training. The production task was performed with only the trained linguistic stimulus pair on the first two days, and with both linguistic pairs on the third day. The untrained linguistic pair was only recorded on the third day in order to minimize practice effect, and because it was assumed the

performance between the two linguistic pairs would be initially similar. This would allow for straightforward comparison of performance despite there not being a baseline for the untrained pair. The recorded production tasks were acoustically analyzed in Praat 6.0.0.5 (Boersma and van Heuven, 2001) for total production length and first syllable vowel length for every token. Only the first syllables were analyzed from the productions as this was where the contrast occurred in the stimuli and it was the most likely part to show any changes. In order to minimize differences in individual speaking rates, this data was normalized by dividing the first syllable vowel durations of the repetitions of the long members of the stimulus pairs by the first syllable vowel durations of the repetitions of the short members. The ratios acquired in this way for each participant were used for the final statistical analyses. For comparison, the same ratio of both linguistic stimulus pairs was 1.23.

2.5 EEG recordings

EEG recordings were performed with a Brain Products ActiCHAmp system and Brain Products Recorder software, version 1.20.0801. The setup used 32 active electrodes to record the EEG. Vertical eye movement was monitored with two electrodes placed above and below the left eye, and horizontal movement with frontal electrodes F7 and F8 at the sides of the head. The impedance of the electrodes was kept under 10 kΩ. Subjects were instructed to sit still and watch a silent film on a TV screen while the stimuli were presented diotically with a PC running Presentation version 16.3 by NeuroBehavioral Systems through Sennheiser HD 25-1 II headphones. The stimuli were presented in an oddball paradigm (deviant probability 0.13) with the short members of each stimulus pair acting as the standard (874 repetitions) and the long one as the deviant (140 repetitions). Short members of the stimulus pairs were

used as the standards, as they were thought to be more phonologically familiar to the participants than the long ones, and therefore more likely to be perceived correctly. The interstimulus interval was 650 ms.

The EEG was offline referenced to the average left and right mastoids and filtered with a 1-30 Hz bandpass filter. Artifact rejection was set at ± 100 μV. Epochs for analysis started at 100 ms before stimuli onset and ended 500 ms after it. The 100 ms prestimulus period was used for baseline correction. The first and second standard stimuli after a deviant were rejected from the analysis, as they would likely display non-standard-like responses due to the change from the deviant to the standard stimulus, thus distorting the average standard responses. Averages were calculated separately for all valid standard and deviant epochs, and difference waveforms were then created by subtracting responses elicited by the standard waveforms from the responses to the deviants. Next, 30 ms time windows were chosen for each stimulus type, centered around the peak amplitudes for each response in the difference waveforms. Time windows did not differ between sessions, but different stimuli had different windows. For the trained linguistic stimuli, four time windows were chosen: N1 windows were set at 195-225 ms and 225-255 ms, and MMN windows at 310-340 ms and 340-370 ms. Two windows were used for each response as they both clearly had two amplitude peaks at different times. For the untrained linguistic stimuli, two time windows were used: the N1 window was set at 220-250 ms and the MMN 330-360 ms. Single windows were used as no double amplitude peaks could be observed. Finally, for the non-linguistic stimuli only the MMN window was set at 350-380 ms, as there was no discernible N1 response in the difference wave. Mean amplitudes were analyzed for each window and used in statistical analysis. Electrodes C3, C4, Cz, F3, F4 and Fz were used in the analyses.

2.6 Discrimination task

The discrimination task employed an oddball paradigm, with the short members of each stimulus pair as the standard and the long member as the deviant. The ISI was 1000 ms, and the deviant probability was 0.13 (130 standard, 20 deviant). The stimuli were presented with a PC running Presentation (version 16.3) by NeuroBehavioral Systems and delivered diotically with Sennheiser HD 25-1 II headphones. Subjects were told to push a response button as fast as possible when they noticed a change in the stimulus stream. Nothing specific about the stimuli was said beforehand, except that they would be either words or sounds. No feedback was given. Detection rates of and reaction times to the deviants were measured from each block, and the former were used to calculate the discrimination sensitivity scores (d'). If a participant had not responded to any deviant in some block, the stimulus onset asynchrony value (1428 ms) was used as the average reaction time. Ceiling level for discrimination accuracy is 4.62; a participant who did not respond once would score 0.7. The average reaction times and the d' scores were subjected to statistical analysis. All statistical analyses were performed with IBM SPSS Statistics 22. P-values are Bonferroni adjusted to account for multiple comparisons.

3 Results

In their self-evaluations, all subjects correctly identified vowel duration as the linguistic feature being studied. Most of them felt that all the tasks had become easier as they progressed through the experiment, though not all stimuli were considered equally difficult: the trained linguistic stimuli were considered to be the easiest overall, while the non-linguistic ones were thought to be somewhat difficult throughout the experiment.

One-sample t-tests were performed at Fz and Cz electrodes in order to determine whether responses significantly differed from zero at the different time windows. N1 was not statistically significant at either electrode site in the first time window in the first and second sessions, but it was significant in both time windows in the third session. N1 was also not statistically significant at either electrode site for the untrained linguistic stimuli in the first session, but it was significant at both sites in the third session. All other analyzed responses significantly differed from zero.

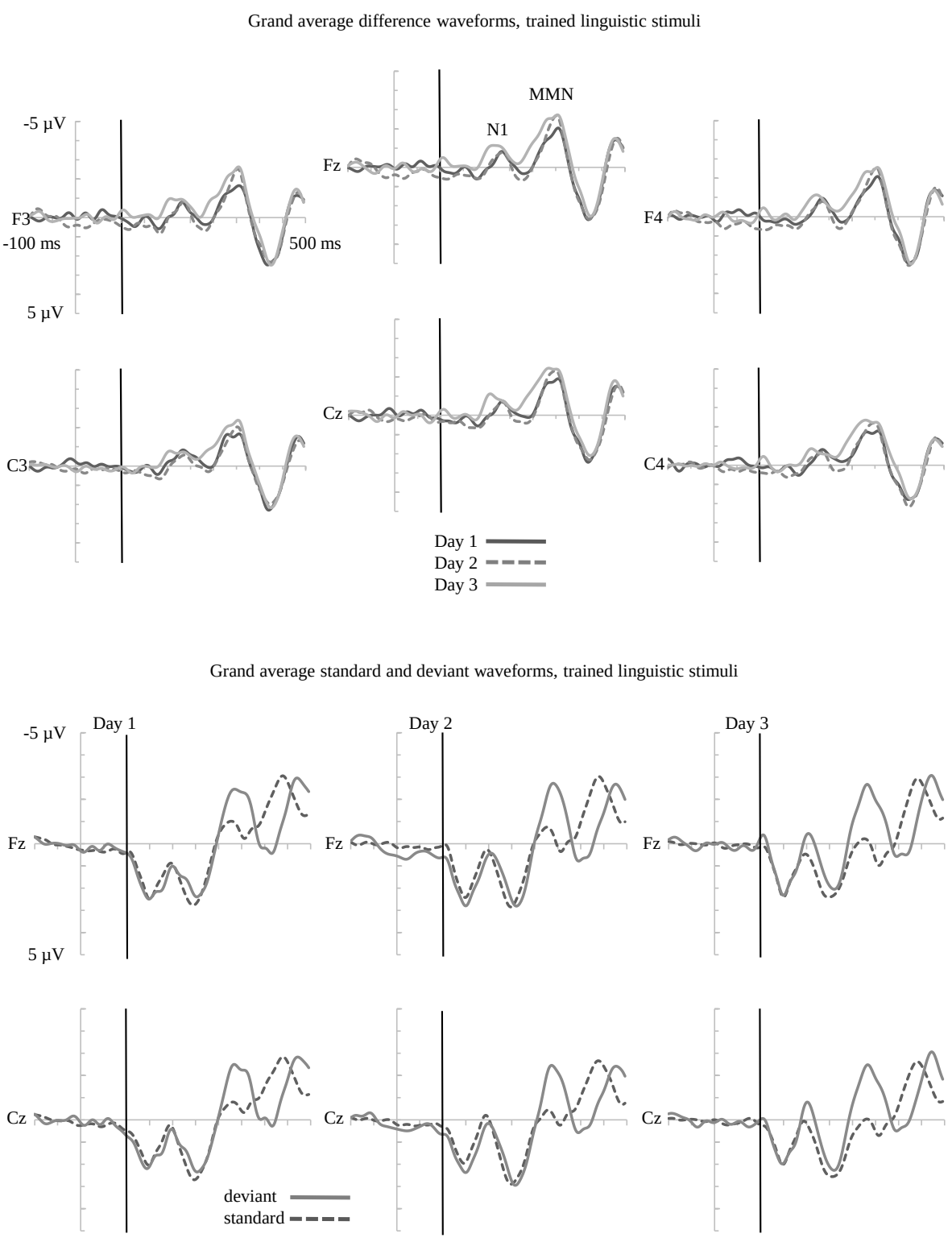
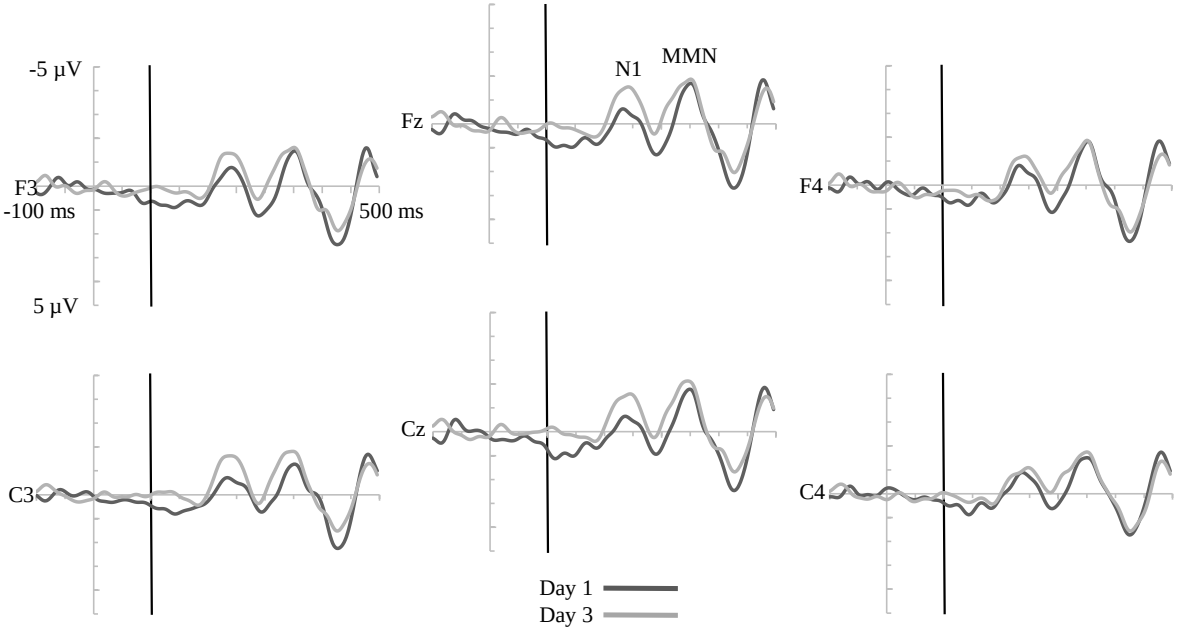


Figure 2. The grand average difference waveforms for the C3, C4, Cz, F3, F4 and Fz electrodes and grand average standard and deviant waveforms for the trained linguistic stimuli. The vertical line in the waveforms indicates where the difference between the stimuli begins.

Grand average difference waveforms, untrained linguistic stimuli



Grand average standard and deviant waveforms, untrained linguistic stimuli

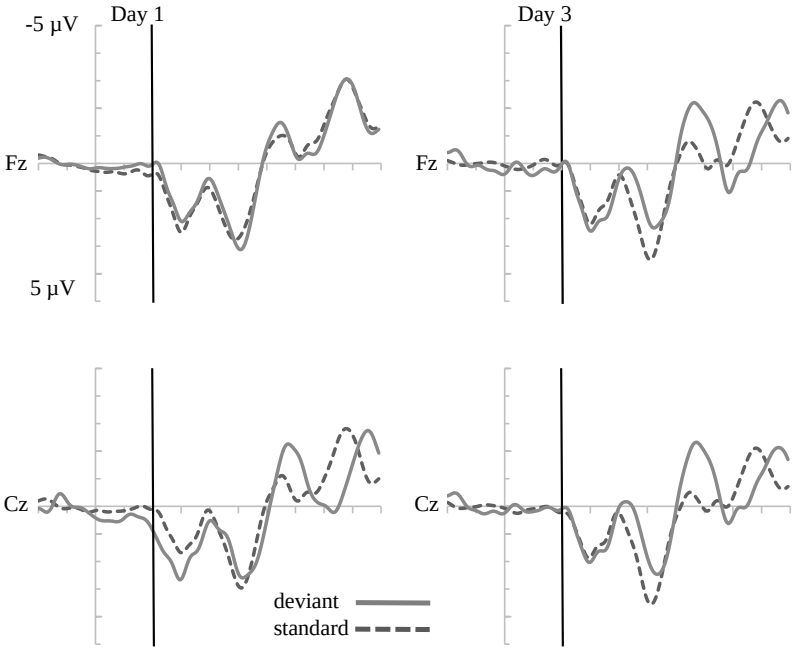
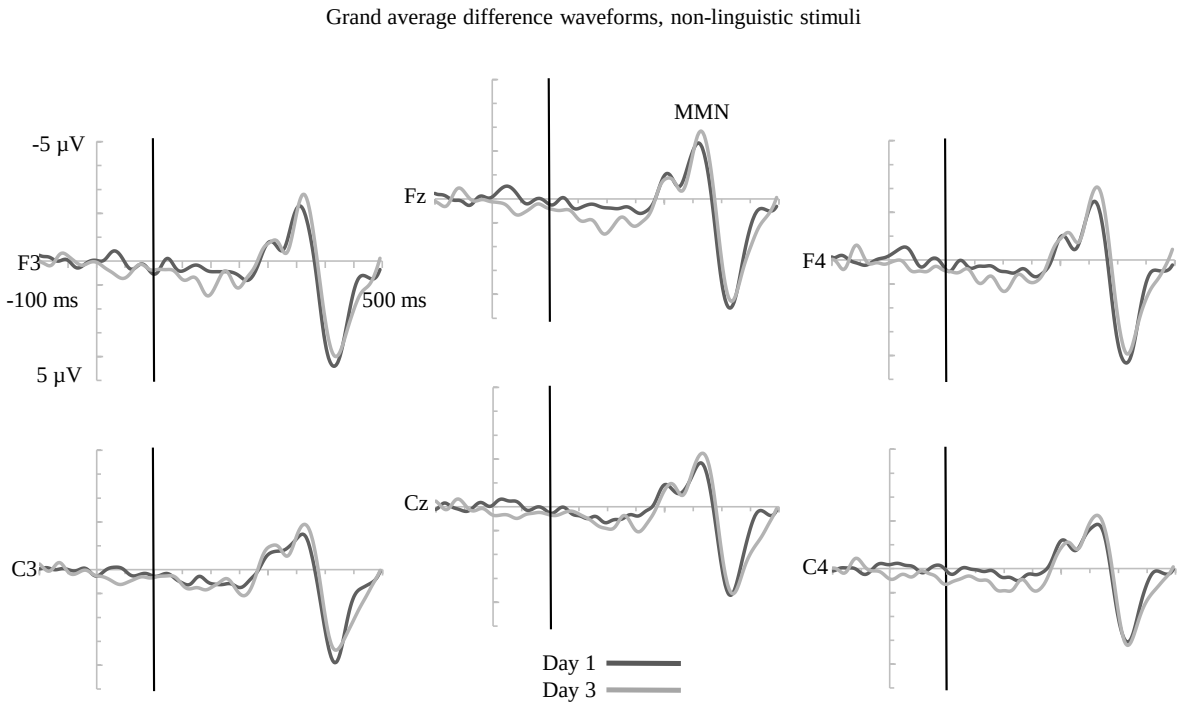


Figure 3. The grand average difference waveforms for the C3, C4, Cz, F3, F4 and Fz electrodes and grand average standard and deviant waveforms for the untrained linguistic stimuli. The vertical line in the waveforms indicates where the difference between the stimuli begins.



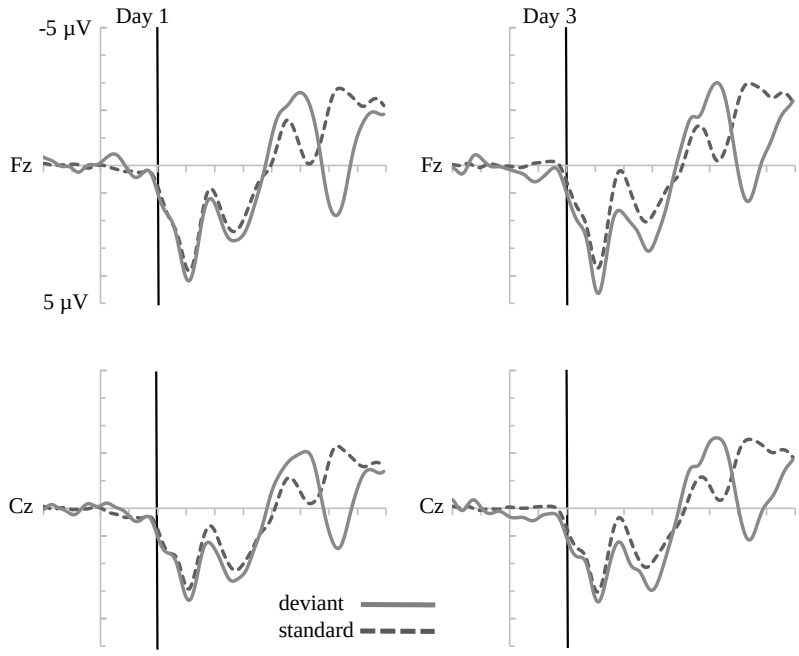


Figure 4. The grand average difference waveforms for the C3, C4, Cz, F3, F4 and Fz electrodes and grand average standard and deviant waveforms for the non-linguistic stimuli. The vertical line in the waveforms indicates where the difference between the stimuli begins.

	trained				untrained		non-linguistic
	N1-1 195-225	**N1-2 225-255**	**MMN1 310-340**	**MMN2 340-370**	**N1 220-250**	**MMN 330-360**	**MMN 350-380**
Fz							
Session 1	-0.13 (0.9)	-0.61* (0.84)	-1.23* (1.68)	-1.78* (1.26)	-0.47 (0.88)	-1.44* (1.94)	-1.79* (1.14)
Session 2	-0.23 (1.11)	-0.60* (0.79)	-1.40* (1.35)	-2.28* (1.16)	-	-	-
Session 3	-1.03* (0.74)	-0.86* (0.78)	-2.21* (1.73)	2.49* (1.38)	-1.40* (0.65)	-1.72* (1.65)	-2.36* (1.89)
Cz							
Session 1	-0.1 (0.71)	-0.57* (0.59)	-1.37* (1.4)	-1.68* (1.16)	-0.49 (0.82)	-1.54* (1.96)	-1.49* (1.09)
Session 2	-0.11 (0.97)	-0.46* (0.66)	-1.51* (1.07)	-1.94* (1.11)	-	-	-
Session 3	-0.96* (0.8)	-0.66* (0.63)	-2.23* (1.52)	-2.20* (1.32)	-1.46* (0.81)	-1.99* (1.21)	-1.97* (1.74)
C3							
Session 1	-0.34 (0.62)	-0.71 (0.61)	-1.35 (1.06)	-1.49 (1.15)	-0.6 (0.44)	-1.09 (1.69)	-1.1 (1)
Session 2	-0.09 (0.94)	-0.45 (0.6)	-1.14 (0.95)	-1.72 (1.1)	-	-	-
Session 3	-0.7 (0.64)	-0.64 (0.54)	-1.88 (1.53)	-2.1 (1.22)	-1.55 (0.86)	-1.73 (1.1)	-1.62 (1.46)
C4							
Session 1	-0.13 (0.58)	-0.67 (0.64)	-1.52 (1.22)	-1.62 (1.03)	-0.73 (0.75)	-1.4 (1.56)	-1.6 (1.01)
Session 2	-0.28 (0.9)	-0.48 (0.64)	-1.64 (0.9)	-1.85 (0.97)	-	-	-
Session 3	-0.71 (0.48)	-0.58 (0.74)	-2.25 (1.3)	-2.08 (1.25)	-0.92 (0.83)	-1.61 (1.19)	-1.96 (1.35)
F3							
Session 1	-0.16 (0.7)	-0.56 (0.73)	-1.12 (1.46)	-1.51 (1.1)	-0.59 (0.84)	-1.19 (1.84)	-1.6 (0.99)
Session 2	-0.15 (1.08)	-0.57 (0.74)	-1.12 (1.27)	-2.12 (1.19)	-	-	-
Session 3	-0.9 (0.8)	-0.79 (0.7)	-1.85 (1.73)	-2.34 (1.38)	-1.31 (0.64)	-1.49 (1.53)	-2.25 (1.71)
F4							
Session 1	-0.17 (0.92)	-0.77 (0.73)	-1.24 (1.52)	-1.83 (1.14)	-0.63 (0.97)	-1.48 (1.68)	-1.84 (1.25)
Session 2	-0.21 (0.99)	-0.57 (0.74)	-1.38 (1.32)	-2.18 (0.95)	-	-	-
Session 3	-0.89 (0.63)	-0.83 (0.62)	-2.11 (1.58)	-2.36 (1.09)	-1.1 (0.76)	-1.61 (1.5)	-2.62 (1.93)

Table 1. Time windows (ms), mean amplitudes (µV) and standard deviations (in brackets, µV) for the psychophysiological measurements for each stimulus in each session for each electrode. - = no recordings were made on the second day for the untrained and non-linguistic stimuli. * = responses that statistically differ from zero (only Fz and Cz).

A Session(2) X Time window(2) X Electrode(6) repeated measures Analysis of Variance (ANOVA) was performed for the N1 response for the trained linguistic stimuli between the baseline and final sessions, resulting in a significant Session X Time window interaction ($F(1,11) = 6.855$; $p = 0.024$; $\eta_p^2 = 0.384$). This suggests that the latency of the N1 response decreased as a result of the training in the trained stimuli. For the untrained linguistic stimuli, a Session(2) X Electrode(6) repeated measures ANOVA for the N1 response between the baseline and final sessions resulted in a significant main effect of Session ($F(1,11) = 7.889$; $p = 0.017$; $\eta_p^2 = 0.418$), indicating that the N1 was elicited in the untrained linguistic stimuli as a result of training. Finally, the mean N1 amplitudes for the trained linguistic stimuli immediately after the training were also analyzed with a Session(2) X Time window(2) X Electrode(6) repeated measures ANOVA between both the first and second session and second and third session. No significant effects or interactions emerged.

In order to analyze the final effects of the training, a Session(2) X Time window(2) X Electrode(6) repeated measures ANOVA was run for the mean amplitude of the MMN response for the trained linguistic stimuli between the baseline and final sessions. This resulted in a main effect of Session ($F(2,11) = 5.794$; $p = 0.035$; $\eta_p^2 = 0.345$), indicating that the mean MMN amplitude increased significantly from the baseline but its latency did not change since there was no effect of time window. No other effects or interactions reached significance. Analysis was continued with the data for the untrained linguistic stimuli between the baseline and final sessions, with no significant effects or interactions for the MMN response. No significant MMN effects or interactions emerged for the non-linguistic stimuli. Next, in order to gauge immediate effects of the training on the trained linguistic stimuli, a Session(2) X Time window(2) X Electrode(6) repeated measures ANOVA for the mean amplitude of the MMN response for the trained linguistic stimuli was run between the

first and second sessions, resulting in a Session X Time Window X Electrode interaction ($F(5,55) = 2,593$; $p = 0.035$; $\eta_p^2 = 0.191$) and a Time Window X Electrode interaction ($F(5,55) = 6,548$; $p < 0.001$; $\eta_p^2 = 0.373$). No other significant effects or interactions emerged, indicating that while training effects were visible on the third day, they did not appear immediately after the training. Another Session(2) X Time window(2) X Electrode(6) repeated measures ANOVA was then run with the second and third session, resulting in a main effect of Session ($F(1,11) = 5.361$; $p = 0.041$; $\eta_p^2 = 0.328$). This shows that the increase seen in the mean MMN amplitude developed between the end of the training and the final measurements. No other main effects or interactions reached significance.

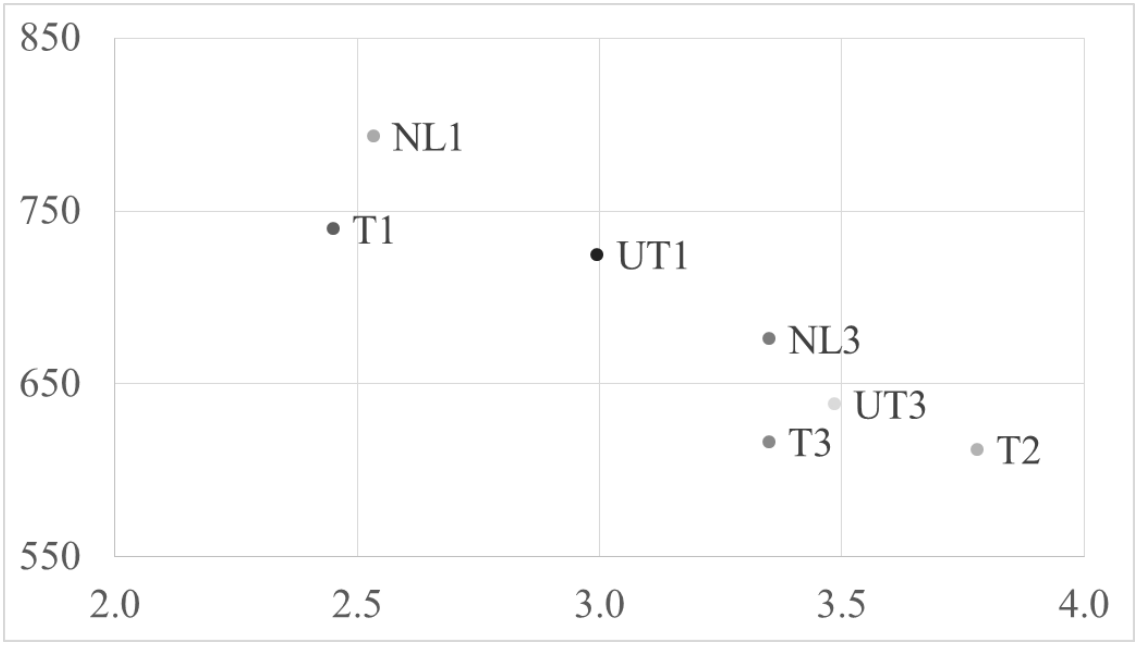


Figure 5. Behavioral discrimination reaction times (vertical axis, ms) and sensitivity scores (horizontal axis) for each stimulus pair. T = trained, UT = untrained, NL = non-linguistic. Stimuli marked with "1" are from Day 1, "2" from Day 2 and "3" from Day 3. Proximity to bottom right corner indicates improved performance, i.e. lower reaction times and higher discrimination sensitivity.

	trained linguistic		untrained linguistic		non-linguistic	
	d'	**RT**	**d'**	**RT**	**d'**	**RT**
Day 1	2.45 (1.53)	792 (233)	3 (1.5)	791 (241)	2.53 (1.04)	874 (214)
Day 2	3.78 (0.82)	617 (105)	-	-	-	-
Day 3	3.35 (1.18)	612 (81)	3.48 (1.1)	648 (132)	3 (1.2)	686 (98)

Table 2. Mean discrimination reaction times (RT) in milliseconds and discrimination accuracy scores (d') and their standard deviations (in brackets) for all stimuli. - = no measurements were made on the second day for the untrained and non-linguistic stimuli.

To analyze discrimination reaction times, a Session(2) X Stimulus(3) repeated measures ANOVA was run with all three stimuli between the baseline and final sessions, which again resulted in a main effect of Session ($F(1,11) = 6.545$; $p = 0.027$; $\eta_p^2 = 0.373$), suggesting that the overall reaction times were faster between the first and last sessions, as suggested by Figure 5. Further analysis was carried out with a Session(2) repeated measures ANOVA for each stimulus between Days 1 and 3, resulting in a main effect of Session for both the trained linguistic stimuli ($F(1,11) = 5.168$; $p = 0.044$; $\eta_p^2 = 0.320$) and the non-linguistic stimuli ($F(1,11) = 6.633$; $p = 0.026$; $\eta_p^2 = 0.376$), suggesting that the subjects were able to respond faster to both the trained linguistic stimuli and the non-linguistic ones by the end of the experiment. No significant effects emerged for the untrained linguistic stimuli. Further analysis was carried out with paired samples t-tests comparing the reaction times for the trained linguistic and non-linguistic stimuli within the same sessions. No significant difference were found in the baseline measurements, but in the final session the difference was significant ($t(11) = -2.468$; $p = 0.031$; $d = 0.656$), suggesting that while reaction times decreased with both stimuli during training, the final reaction times for the trained linguistic stimuli were significantly faster than the ones for the non-linguistic ones. As with the psychophysiological data, analyses were next carried out to find any immediate effects the training may have had on the trained stimuli by running a Session(2) repeated measures

ANOVA with the reaction times for Day 1 and Day 2, which resulted in a significant main effect of Session ($F(1,11) = 5.986$; $p = 0.032$; $\eta_p^2 = 0.352$). Next, the same analysis was conducted between Days 2 and 3, which did not yield significant results. These results suggest that the observed decrease in reaction times for the trained linguistic stimuli was there immediately after the training ended, and the reaction times did not decrease further between the second and third sessions. This finding is demonstrated by Figure 5.

Analysis of discrimination sensitivity began with a Session(2) X Stimulus(3) repeated measures ANOVA of the discrimination sensitivity scores with all three stimuli and the baseline and final sessions. This resulted in a main effect of Session ($F(1,11) = 6.030$; $p = 0.032$; $\eta_p^2 = 0.354$), indicating that overall, the discrimination sensitivity scores were different between the first and last days of the experiment; the values seen in Figure 5 show that all stimuli had higher scores on Day 3 than Day 1. No other effects or interactions reached significance. Further analysis was carried out with a Session(2) repeated measures ANOVA for each stimulus, resulting in a main effect of Session ($F(1,11) = 11.842$; $p = 0.006$; $\eta_p^2 = 0.518$) for the trained linguistic stimuli, indicating that behavioral discrimination sensitivity increased as a result of training. No significant effects or interactions emerged for the untrained linguistic or non-linguistic stimuli. Finally, the discrimination sensitivity values for the trained stimuli immediately after the training were analyzed with a Session(2) repeated measures ANOVA between Day 1 and 2, resulting in a main effect of Session ($F(1,11) = 21.157$; $p = 0.001$; $\eta_p^2 = 0,658$), suggesting increased discrimination sensitivity between Days 1 and 2. No significant effects were found between Day 2 and 3, suggesting that, similarly to the reaction times, the observed improvements occurred right after the training and did not change further between the second and last sessions.

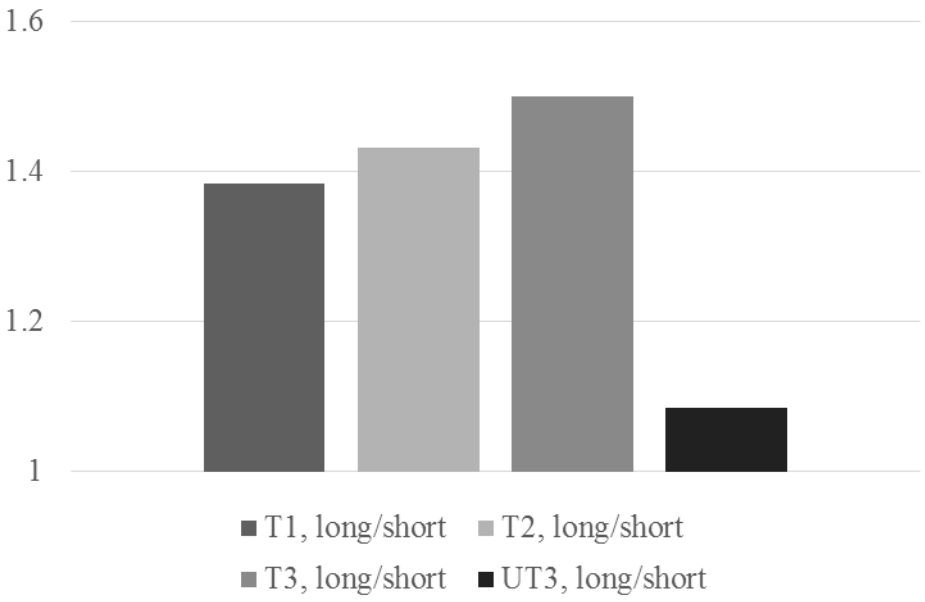


Figure 6. Average long/short production ratios of the first syllables of both stimulus pairs, calculated by dividing the vowel durations of the repetitions of the long members of the pairs by the duration of the short ones. Values above 1 indicate that repetitions of the long vowels were longer than the short ones.

	trained linguistic long/short	**untrained linguistic long/short**
Day 1	1.38 (0.23)	-
Day 2	1.43 (0.24)	-
Day 3	1.5 (0.31)	1.08 (0.05)

Table 3. Mean first syllable long/short production ratios and their standard deviations (in brackets) for the linguistic stimuli. The untrained linguistic stimuli were only measured on the final day.

No significant effects emerged from between-session analysis of the production ratios (Figure 6), suggesting that production of the trained linguistic stimuli remained unaffected by the training. Comparison of the ratios between the trained linguistic stimuli and the only measurement for the untrained linguistic stimuli, using paired samples t-tests, showed that the trained linguistic stimuli were produced with significantly higher long/short ratios than the untrained ones in all sessions: Day 1 = $t(11) = 4.567$; $p = 0.001$; $d = 1.802$; Day 2 = $t(11) = 4.604$; $p = 0.001$; $d = 1.458$; Day 3 = $t(11) = 4.871$; $p < 0.001$; $d = 1.354$.

4 Discussion

The purpose of the study was to evaluate the efficacy of a listen-and-repeat training paradigm on improving perception and production of vowel duration contrasts, and whether potentially improved perception of vowel length is generalized to untrained vowels or to non-linguistic sounds. The effects were studied using psychophysiological measures of discrimination sensitivity (MMN and N1 responses), behavioral discrimination tasks and production tasks. The results show that the training had clear, statistically significant learning effects on the perception of the trained linguistic stimuli, specifically the increase of the mean amplitude of the MMN response, the reduced latency of the N1 response, the increased behavioral discrimination scores and reduced discrimination reaction times. All of these are indicators of training effects. This is in agreement with previous studies showing that perception of non-native duration contrasts can be improved with training (e.g. Hirata et al., 2007; Okuno, 2014; Tajima et al., 2008). The results are also largely in line with previous production training studies: Saloranta et al. (2017) found behavioral learning effects for trained stimuli after the same amount of training that was used in this study while examining learning of vowel duration, and both studies of second language VOT acquisition by Tamminen et al. (Tamminen et al., 2015; Tamminen and Peltola, 2015) found improvements in behavioral identification ability and increases in MMN amplitudes after only three days of listen-and-repeat training. It seems, therefore, that this type of training can be effective for learning not only spectral, but durational features in second language vowels. The fact that the significant amplitude change in the MMN response only emerged between the second and third day, at least a week after the training had stopped, may suggest consolidation effects: Atienza et al. (2004) found that after discrimination training of complex auditory stimuli, the amplitudes of the MMN responses induced by the training continued to increase without further training for

up to 72 hours. It should also be noted that although the difference was not significant when compared to the third day, the behavioral discrimination scores for the trained stimuli were at their highest by the end of the second day, right after training had ended. This is likely a combination of true learning effects and laboratory training effects, resulting from a very high exposure to the stimuli over the previous two days. The scores on the third day are therefore to be a more likely to be an accurate representation of the participants' performance level.

Regarding generalization effects, subjects reported that the tasks involving the untrained linguistic stimuli had become easier, but no significant improvement in the MMN response, behavioral discrimination or production tasks was seen. It therefore seems that a generalization effect to the untrained vowel could not be shown, as none of the linguistically significant measures exhibited improvement. Similar results were observed by Tajima et al. (2008), who used a short but intensive identification training paradigm, 5 days and 3600 trials, and failed to find generalization effects to untrained contrasts, although identification performance improved with the trained stimuli. These findings differ from many other training studies, who have reported generalized learning effects either with spectral contrasts (e.g. Bradlow et al., 1999, 1997) or duration contrasts (e.g. Hirata et al., 2007; Okuno, 2014). One key difference between studies that find generalization and those that do not could be the amount of training, as the studies that reported generalization had participants training for up to several weeks, in contrast with the five days reported by Tajima or the two days of training in the current study. More studies using listen-and-repeat training that also test generalization effects are needed in order to determine whether generalization is dependent on time, stimulus type or some other factor.

While no linguistic generalization took place, changes did occur in the preattentive perception of the untrained linguistic contrast, as a significant increase in the N1 response was elicited in the EEG recordings as a result of training. These results are compatible with earlier studies suggesting that the N1 complex can be affected with training (Brattico et al., 2003; Tremblay et al., 2001). N1 is not considered to be a linguistic component, but rather a response to observed changes in the physical features of the stimuli. Given the somewhat separated processing of phoneme quality and quantity (Ylinen et al., 2005a) and the general processing mechanism for duration regardless of linguistic significance, (Liégeois-Chauvel et al., 1999), it is possible that duration was correctly detected as the relevant cue from the training stimuli and the brain became more sensitive to it, and this was reflected by the increased N1 amplitude. Näätänen suggests that N1 amplitudes may be selectively enhanced for "relevant stimuli" at least partially due to "a general increase in sensory sensitivity" (Näätänen, 1992, p. 132). This would suggest that the training was able to access and shape general neural duration processing, lowering the detection threshold and allowing for improved detection and discrimination of the physical duration difference. However, this increased sensitivity did not result in an MMN amplitude increase for the untrained vowel, as the memory trace that evokes the MMN for the phoneme is acoustically complex in nature, consisting of both a temporal and a spectral component. With limited spectral information to link the duration cue to the specific memory trace, no increase in the MMN resulted in the untrained vowel. This suggests that while the training was able to engage the general duration processing system, improved processing of vowel duration may not be easily generalized to other vowels due to the complex nature of the memory trace involved. Spectral information may be needed for modification of the memory trace. Increased general sensitivity to duration could also be behind the decrease in N1 response latency for the trained stimuli, and the significant decrease in reaction times for the non-linguistic stimuli in the discrimination test.

It seems, therefore, that processing of duration may have been generally affected by the training, though the change is not reflected in all measurements.

Learning may also be further hindered if a similar vowel exists in the learners' native language. Nenonen et al. (2005) found that MMN responses to second language duration contrasts were lower, if the phoneme in question could be mapped through the native phonological system where no duration contrast exists. This is also compatible with models of second language acquisition, such as PAM (Best and Strange, 1992) and SLM (Flege, 1987), which both suggest that if a second language phoneme is similar to a native one but somehow systematically different, it is likely mapped to one of the existing native categories. Phonemes highly similar to both of the fairly common vowels in the stimuli can be found in all the participants' native languages, so it may be that there was enough training to overcome this effect and increase the amplitude of the MMN response for the trained linguistic stimuli, but not the untrained one, despite the overall enhancement of the duration processing system.

While perception measurements showed some kinds of changes in all stimuli, results from production analyses exhibited no learning effects. This is at odds with earlier studies using listen-and-repeat training (e.g. Jähi et al., 2015; Saloranta et al., 2015; Taimi et al., 2014) that all showed learning effects in the production of a non-native vowel quality contrast. This may, however, be simply explained by the fact that the long/short ratios for the productions of the trained stimuli were already at a high level in the baseline measurements, resulting in a ceiling effect and leaving little room for improvement. While the ratios for the trained stimuli were already high at baseline, the same ratio for the untrained pair was low. This could suggest that the subjects were able to produce the duration difference in the trained stimulus pair, but not the untrained one. This interpretation is supported by the fact that the observed

ratios for the productions of the trained pair are notably higher in all sessions than the ratio of 1.23 in the stimuli themselves (Figure 6), which was judged to be discriminable but difficult to native listeners. The reasons for this are not entirely clear, as discrimination performance and MMN responses on the first day were comparable between the linguistic pairs, with no statistically significant differences emerging. This may be another indication of the disconnect between temporal and spectral features with the untrained stimuli that was suggested by the psychophysiological measurements.

One overall trend that can be observed from the data is that it lends support to the Desensitization Hypothesis by Bohn (1995), as existing ability to discriminate and produce the novel duration contrasts was visible in the data. MMN responses were elicited for all stimuli on the first day, average discrimination accuracy scores were well over 2, when 0,7 indicates no responses to the deviants at all, and the trained stimulus pairs were repeated with long/short ratios even longer than the ones for the stimuli themselves. It is, in fact, likely due to this pre-existing ability that no changes were observed in the production of the linguistic stimuli: though no specific ceiling level exists for these types of tasks, it is unlikely that the subjects would have changed their production very much when they could already clearly produce an audible difference between the long and short vowels. The training was, however, able to affect the participants beyond that existing level when it came to duration discrimination. These results, again, closely mirror Tamminen and Peltola's (Tamminen and Peltola, 2015) results for VOT: they demonstrated both behavioral and psychophysiological learning for advanced learners of English, who showed baseline behavioral discrimination scores of 3.6 (ceiling level 4.61, same as the current study) and had MMN responses to the English voicing contrast already at the beginning of the experiment.

In summary, the findings from this study suggest two main results: listen-and-repeat training can be used to train and improve perception of duration contrasts in vowels, and the training also generally affects the general duration processing system, making it more sensitive to duration differences. However, it seems that while the brain is able to extract the duration signal from the stimuli used in the training and use it to enhance general duration processing, this enhancement is not transferred to the existing complex memory traces that produce the MMN response, and spectral information may be needed in conjunction with the duration information to modify these memory traces. Furthermore, generalization may have been impacted by the native system's resistance to change when the stimuli could be mapped through the native system. Taken together with existing literature, it seems that listen-and-repeat can be a useful tool in the acquisition of second language phonemes.

5 Acknowledgments

We offer our sincere thanks to the Centre for Language and Communication Studies at the University of Turku for their help in the recruitment of participants for the experiment.

Funding: The research was supported by Alfred Kordelin Foundation (grant numbers 150420 and 170356), Turku University Foundation (grant number 12314), Utuling doctoral program and the University of Turku. The Lab 100 language lab system was provided by Sanako Corporation.

6 References

Alku, P., Tiitinen, H., Näätänen, R., 1999. A method for generating natural-sounding speech stimuli for cognitive brain research. Clin. Neurophysiol. 110, 1329–1333. https://doi.org/10.1016/S1388-2457(99)00088-7

Atienza, M., Cantero, J.L., Stickgold, R., 2004. Posttraining Sleep Enhances Automaticity in Perceptual Discrimination. J. Cogn. Neurosci. 16, 53–64. https://doi.org/10.1162/089892904322755557

Augustaitis, D., 1964. Das litauische Phonationssystem. Sagner, Munich.
Best, C.T., Strange, W., 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. J. Phon. 20, 305–330.
Boersma, P., van Heuven, V., 2001. Praat, a system for doing phonetics by computer. Glot Int. 5, 341–347.
Bohn, O.-S., 1995. Cross-language speech perception in adults: first language transfer doesn't tell it all, in: Speech Perception and Linguistic Experience: Issues in Cross-Language Research. York Press, Baltimore, pp. 279–304.
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. Percept. Psychophys. 61, 977–985. https://doi.org/10.3758/BF03206911
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. J. Acoust. Soc. Am. 101, 2299–2310.
Brattico, E., Tervaniemi, M., Picton, T.W., 2003. Effects of brief discrimination-training on the auditory N1 wave. Neuroreport 14, 1–4. https://doi.org/10.1097/01.wnr.0000098748.87269.a1
Chandrasekaran, B., Krishnan, A., Gandour, J.T., 2009. Sensory processing of linguistic pitch as reflected by the mismatch negativity. Ear Hear. 30, 552–558.
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.
Fougeron, C., Smith, C.L., 1993. French. J. Int. Phon. Assoc. 23, 73. https://doi.org/10.1017/S0025100300004874
Hirata, Y., Whitehurst, E., Cullings, E., 2007. Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates. J. Acoust. Soc. Am. 121, 3837–3845. https://doi.org/10.1121/1.2734401
Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C., 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. Cognition 87, B47–B57. https://doi.org/10.1016/S0010-0277(02)00198-1
Jähi, K., Peltola, M.S., Alku, P., 2015. Does interest in language learning affect the non-native phoneme production in elderly learners?, in: Proceedings of the 18th International Congress of Phonetic Sciences.
Khatiwada, R., 2009. Nepali. J. Int. Phon. Assoc. 39, 373–380. https://doi.org/10.1017/S0025100309990181
Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schröger, E., Jacobsen, T., 2008. Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. Int. J. Psychophysiol. 67, 131–143. https://doi.org/10.1016/j.ijpsycho.2007.10.012
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/S0301-0511(06)00140-2 [pii]
Lee, W.-S., Zee, E., 2003. Standard Chinese (Beijing). J. Int. Phon. Assoc. 33, S0025100303001208. https://doi.org/10.1017/S0025100303001208
Liégeois-Chauvel, C., De Graaf, J.B., Laguitton, V., Chauvel, P., 1999. Specialization of left auditory cortex for speech perception in man depends on temporal coding. Cereb. Cortex 9, 484–496. https://doi.org/10.1093/cercor/9.5.484
Maddieson, I., Disner, S.F., 1984. Patterns of sounds. Cambridge University Press, London.
Martnez-Celdrn, E., Fernndez-Planas, A.M., Carrera-Sabat, J., 2003. Castilian Spanish. J. Int. Phon. Assoc. 33, 255–259. https://doi.org/10.1017/S0025100303001373
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
Näätänen, R., 1992. Attention and brain function. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.
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., Picton, T.W., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of component structure. Psychol. Sci. 24, 375–425.
Nenonen, S., Shestakova, A., Huotilainen, M., Näätänen, R., 2005. Speech-sound duration processing in a second language is specific to phonetic categories. Brain Lang. 92, 26–32. https://doi.org/10.1016/j.bandl.2004.05.005

Okuno, T., 2014. Acquisition of L2 Vowel Duration in Japanese by Native English Speakers. Diss. Abstr. Int. Michigan State U.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
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
Saloranta, A., Alku, P., Peltola, M.S., 2017. Learning and generalization of vowel duration with production training: behavioral results. Linguist. Lettica 25, 67–87.
Saloranta, A., Tamminen, H., Alku, P., Peltola, M.S., 2015. Learning of a non-native vowel through instructed production training, in: Proceedings of the 18th International Congress of Phonetic Sciences. University of Glasgow, Glasgow.
Strange, W., Dittmann, S., 1984. Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. Percept. Psychophys. 36, 131–145. https://doi.org/10.3758/BF03202673
Taimi, L., Jähi, K., Alku, P., Peltola, M.S., 2014. Children Learning a Non-native Vowel – The Effect of a Two-day Production Training. J. Lang. Teach. Res. 5, 1229–1235. https://doi.org/10.4304/jltr.5.6.1229-1235
Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., Munhall, K.G., 2008. Training English listeners to perceive phonemic length contrasts in Japanese. J. Acoust. Soc. Am. 123, 397–413. https://doi.org/10.1121/1.2804942
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.
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/http://dx.doi.org/10.1016/j.ijpsycho.2015.04.020
Tremblay, K., Kraus, N., Carrell, T.D., McGee, T., 1997. Central auditory system plasticity: Generalization to novel stimuli following listening training. J. Acoust. Soc. Am. 102, 3762–3773. https://doi.org/10.1121/1.420139
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 Hear. 22, 79–90. https://doi.org/10.1097/00003446-200104000-00001
Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csepe, V., Ilmoniemi, R.J., Näätänen, R., 1999. Brain responses reveal the learning of foreign language phonemes. Psychophysiology 36, 638–642.
Yanushevskaya, I., Bunčić, D., 2015. Russian. J. Int. Phon. Assoc. 45, 221–228. https://doi.org/10.1017/S0025100314000395
Ylinen, S., Huotilainen, M., Näätänen, R., 2005a. Phoneme quality and quantity are processed independently in the human brain. Neuroreport 16, 1857–1860.
Ylinen, S., Shestakova, A., Alku, P., Huotilainen, M., 2005b. The Perception of Phonological Quantity Based on Durational Cues by Native Speakers, Second-Language Users and Nonspeakers of Finnish. Lang. Speech 48, 313–338. https://doi.org/10.1177/00238309050480030401
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 Res. 1072, 175–185. https://doi.org/http://dx.doi.org/10.1016/j.brainres.2005.12.004