A neurophysiological investigation of processing phoneme substitutions in L2

Merel van Rees Vellinga[1], *Adriana Hanulíková*[1], *Andrea Weber*[1] *and Pienie Zwitserlood*[2]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Westfälische Wilhelms-Universität Münster, Münster, Germany

{merel.vanreesvellinga; adriana.hanulikova; andrea.weber}@mpi.nl, zwitser@uni-muenster.de

ABSTRACT

The present electrophysiological study examined whether linguistic experience with pronunciation variants in non-native speech (i.e., th-substitutions in English) influences processing at the prelexical level. In a mismatch negativity (MMN) study, the English pseudoword /θond/ was presented along with the deviants /tond/ and /sond/, while ERP-data were collected from Dutch listeners. If experience influences the MMN, smaller amplitude deviances should be observed for /tond/, the deviant typical for Dutch learners of English, than for the less typical deviant /sond/. If, on the other hand, the MMN mainly reflects perceptual similarity, smaller amplitude deviances should be found for the perceptually similar /sond/ deviant. The results of Dutch listeners revealed a significant deviance interaction, with /sond/ eliciting a smaller MMN and longer latencies than /tond/. This suggests that perceptual similarity rather than linguistic experience influences prelexical processing of th-substitutions for Dutch L2 listeners.

Keywords: MMN, th-substitutions, Dutch, L2.

1. INTRODUCTION AND BACKGROUND

An experience-based account of processing pronunciation variants assumes that a frequent variant will be processed differently from an infrequent one. The present study takes variant forms that vary in production frequency in foreign-accented speech as the starting-point for investigating the prelexical processing of such L2 mispronunciations. An answer is sought to the question of whether preferred variants of Dutch learners of English lead to different neurophysiological responses in Dutch listeners than disfavoured variants. The case is made with the English interdental fricative [θ], a sound that poses great difficulties for many learners of English. In particular, Dutch speakers often substitute [θ] with [t] (e.g., they pronounce *theft* as *teft*), and to a lesser extent also with [s] (Wester et al. 2007; Hanulíková & Weber 2010).

Recently, Hanulíková and Weber (in prep.) showed in an eye-tracking study that spoken-word recognition by L2 listeners is influenced by how frequent particular th-substitutions occur in the listeners' own L2 speech. In this study, Dutch and German listeners heard L2 English words in which word-initial [θ] was substituted (e.g., *theft* pronounced as *teft* and *seft*). Dutch and German L2 speakers vary with respect to the frequency with which they produce these substitutions. While hearing the auditory probe, participants looked at a computer screen displaying four printed words: the English th-word (e.g., *theft*), a phonological competitor (e.g., *left*), and two unrelated distracters (e.g., *kiss*, *mask*). The amount of looks to printed words in this task is assumed to reflect the strength of lexical activation during spoken-word recognition (e.g. Allopenna, Magnuson & Tanenhaus 1998). It was found that L2 listeners looked most often at the th-words when the auditory prime represented the accent-specific predominant substitute. That is, Dutch listeners looked more at *theft* when hearing *teft* than when hearing *seft*, but German listeners for whom the predominant substitute is [s] (Hancin-Bhatt 1994; Hanulíková & Weber 2010) looked more often at *theft* when hearing *seft* than when hearing *teft*. This is evidence that language-specific experience with mispronunciations influences lexical processing. Listeners recognise L2 words more easily when they are pronounced in the way that they are familiar with from their own accent, which is in line with frequency-based accounts of phonological variation (Connine 2004; Mitterer & Ernestus 2006).

The present study seeks to find neurophysiological evidence for the question of whether experience with mispronunciations, in the form of segmental substitutions in L2 speech, already influences the prelexical level of processing. A well-established method to investigate experience-based auditory memory traces is the

Mismatch Negativity (MMN) paradigm. With respect to language-specific features, the MMN, an Event-Related Potential (ERP) component discovered by Näätänen et al. (1978), so far has been used to study differences in neural representations for within- and across-category stimuli in L1 speech. The MMN is an early, automatic brain response that becomes visible as a negative component of an ERP upon the detection of a deviant feature in a stimulus. Relating sounds to meaning involves mapping sounds to phonological categories, and this process is strongly influenced by experience-based phonetic representations. Language-specific experience affects the processing of speech sounds, which has been shown in EEG research before (e.g. Näätänen 1997; Jacobsen 2003; Kirmse et al. 2008; Dehaene-Lambertz 1997; Brunellière et al. 2009). The specific aim of this study was to test whether experience with accent-specific th-mispronunciations in English influences the activation of memory traces in an MMN design. More specifically, we wanted to examine whether the presentation of the predominant th-substitution [t] results in smaller MMNs or longer MMN latencies for Dutch listeners than the presentation of the less frequent substitution [s]. Dutch participants were presented with an English pseudoword with initial [θ] ([thond]) as the standard condition, and with two pseudowords starting with either [t] or [s] ([tond] and [sond]) as the deviant conditions. If familiarity with th-substitutions influences prelexical processing, the less frequent mispronunciation [sond] should result in a larger MMN amplitude and/or shorter MMN latencies, while the more frequent variant [tond] should result in a smaller MMN amplitude and/or longer MMN latencies. If familiarity does not influence prelexical processing, then we predict that the deviant with [s], which is perceptually more similar to [θ], should result in a smaller MMN than the deviant with [t], which is perceptually more distinct from [θ] (Cutler et al. 2004).

2. RESEARCH DESIGN AND MATERIALS

Sixteen native listeners of Dutch participated in this study (mean age of 23, SD 3.3, nine females), all right-handers, with good hearing, with no speech or language problems, and with a good command of English.

The standard stimulus was the English monosyllabic pseudoword 'thond' [θond], and the two deviants were 'tond' [tond] and 'sond' [sond]. The stimuli were recorded by a native speaker of American English. Using Praat (Boersma & Weenink 2002), the stimuli were cross- and identity-spliced to avoid the elicitation of MMNs owing to small differences in features other than the beginning phoneme in the recorded materials. Variation in the stimuli was added by the creation of pitch changes of +6, +12, -6 and -12 Hz. These pitch variations made it possible to abstract away from the specific acoustic properties of one token (see Bien et al. 2009). Application of the pitch variations resulted in five tokens per pseudoword, thereby creating fifteen different tokens. The length of the stimuli was 593 ms for [θond], 499 ms for [tond] and 609 ms for [sond].

The EEG recordings were made with the actiCAP system (Brain Products GmbH, Munich, Germany). Impedances were kept below 20 kΩ throughout the recordings, and a total of 38 electrodes were placed on standard electrode sites. The electrodes were referenced to the right mastoid. In addition, one electrode was placed underneath the left eye to monitor the vertical oculogram. The horizontal oculograms were monitored by electrodes F9 and F10. EEG signals were amplified with BrainAmp DC amplifiers.

2.1. Procedure

Participants were equipped with a 38-electrode cap on a common connector, and were placed in front of a computer screen in a soundproof, electrically shielded EEG-booth. The auditory stimuli were presented over speakers at approximately 60 dB. Participants were instructed to watch a self-selected silent film while the stimuli were being presented, and were instructed to ignore the auditory stimuli. Stimuli were presented in two different blocks. The first block was a multideviant block, wherein the standard [θond] was presented for 80% and the deviants [tond] and [sond] each for 10% of the time. The stimuli were presented in a random order, with the criteria that a block started with at least eight standards, and that deviants were always presented with at least two standards in between. Stimuli were presented at 1000 ms inter-stimulus intervals (ISIs). The multideviant block consisted of 1000 stimuli and lasted for 26 minutes.

After a short break, the second block with an equiprobable design started. This block, in which [θond], [tond], and [sond] were each presented for 33.3 % of the time, functioned as a baseline for comparison of the standard and deviants from the first block. It furthermore served as a control for the alignment of [t] with [θ] and [s]. When producing the plosive [t], the airflow is first stopped before it is released with a burst. Since it is impossible to measure the closure phase for isolated words, the MMN latencies for [t] in the equiprobable block would allow a control for MMN latencies in the multideviant block. The equiprobable block consisted of 600 stimuli, with each stimulus randomly presented 200 times, and lasted for 19 minutes.

After the EEG recordings, participants completed a discrimination ABX-task, that was set up to confirm that participants could perceptually discriminate the three stimuli [θond], [tond] and [sond] in an offline task. On twelve trials, the three stimuli were presented in random order at the A and B positions, with the X matching either the A or B stimulus on each trial. The participants' task was to match the X position to one of the two preceding stimuli. On average, participants responded on 84% of the trials correctly.

3. RESULTS

3.1. EEG-Recordings

The analyses were carried out with Advanced Source Analyses (ASA) software (Advanced Neuro Technology, Enschede, The Netherlands). The raw data were first re-referenced to the right mastoid (M2). To monitor the vertical and horizontal oculograms (VEOG and HEOG), bipolar channels were then added by subtracting Fp1 from V1 for VEOG monitoring and by subtracting F10 from F9 for HEOG monitoring. The data were then 35 Hz low pass filtered with a slope of 12 dB/oct. Artefacts were excluded for the -75 to 75 µV range. For all epochs, a baseline correction of -250 ms was employed. Segmentation took place for all epochs of interest, for a time window of -250 to 550 ms. The averages of the two blocks and three conditions were first computed for each participant, and the signal of each participant was visually inspected to detect possible bad electrodes or other abnormalities. The grand averages (GAs) of all conditions were computed and, for the analysis of the first block, the GA elicited by the standard [θond] was subtracted from those of the deviants [tond] and [sond]. For the second block, the procedure was the same, but since [θond], [tond] and [sond] were presented equally often, they were all considered as standard here.

Figure 1a shows for the multideviant block the time course of the topography of the difference in brain responses from 31 until 415 ms after onset, with each frame corresponding to a time window of 32 ms. The top row shows the topography of the deviant [sond] minus the standard [θond], the bottom row depicts the deviant [tond] minus the standard [θond]. The accordant topography for the equiprobable block can be found in Figure 1b. As can be seen in Figures 1a and b, both deviants [tond] and [sond] showed an early negativity, followed by a strong positivity, which again changes into a stronger negativity. This resembles the transition from N1 to P2 and to MMN, as also becomes visible in Figure 2. Furthermore, a difference in latency can be seen in Figures 1a and b: the responses for [tond] always occurred about one or two frames earlier than those for [sond], indicating a difference of at least 32 ms. The maps in Figure 1c show the difference waves for identical stimuli, of [sond] in the multideviant block minus the same stimulus [sond] in the equiporable block in the top row, and the difference of [tond] in the multideviant block minus [tond] in the equiprobable block in the bottom row. Time windows are chosen from 102 to 486 ms, with an interval of 32 ms. Again, in Figure 1c a latency difference can be seen as main effect. The latency effect, that is always earlier for [tond] than for [sond], suggests that it is not experience that is being measured here. Rather, it seems that acoustic similarity between phonemes determines processing.

For all conditions, peak scores were extracted according to the relevant MMN time windows. These were determined by visual inspection, and by calculation of the point of maximum amplitude for the separate conditions. For [tond], the relevant time window was set at 200 to 300 ms after stimulus onset, and for [sond] at 250 to 350 after stimulus onset. Electrodes that were included in the repeated measures analysis of variance (ANOVA) were visually selected on the basis of their relevance to MMN analysis. Of the 38 electrodes used for recording, nineteen were selected: Fp1, Fp2, F3, F4, C3, C4, F7, F8, T7, Fz, Cz, FC1, FC2, CP1, CP2, FC5, FC6, CP5 and CP6.

The repeated measures ANOVA for the multideviant block showed significant effects for the MMNs of both [tond] and [sond]. Also, the MMN of [tond] minus [θond] vs. the MMN of [sond] minus [θond] was significant for maximum amplitude ($F(1, 15) = 7.9$, $p<0.02$), maximum latency ($F(1,15) = 63.4$, $p<0.01$), and area measured in µV per ms ($F(1,15) = 6.5$, $p<0.05$). Figure 2 shows the difference waves of [tond] and

[sond] minus [θond] for the single electrode Fz, with N1, P2 and MMN marked. For the equiprobable block the difference waves were also significant different regarding maximum amplitude ($F(1,15) = 11.2$, $p<.005$), maximum latency ($F(1,15) = 53.4$, $p<.001$), and area ($F(1,15) = 12.5$, $p<.004$).

Figure 1a: Deviant minus standard block 1 (2 μV). Top row: /s/ - /th/. Bottom row: /t/ - /th/.

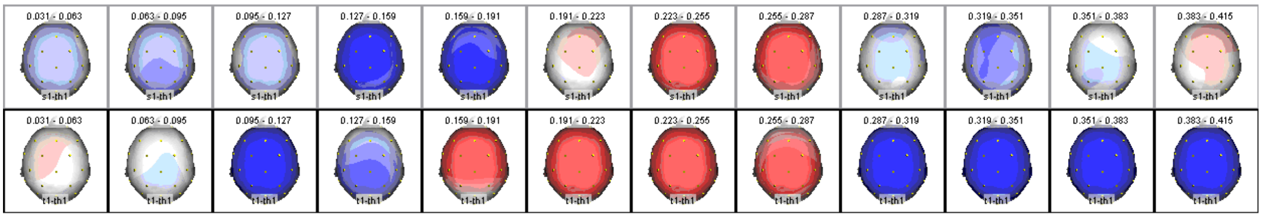


Figure 1b: Deviant minus standard block 2 (2 μV). Top row: /s/ - /th/. Bottom row: /t/ - /th/.

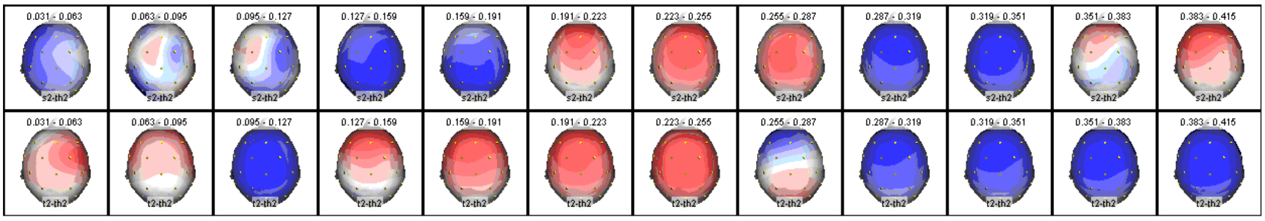


Figure 1c: Difference map block 1 / block 2 (2 μV). Top row: /s1/ - /s2/. Bottom row: /t1/ - /t2/.

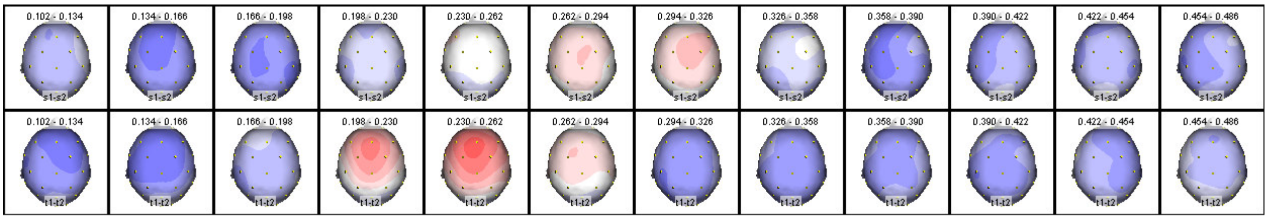
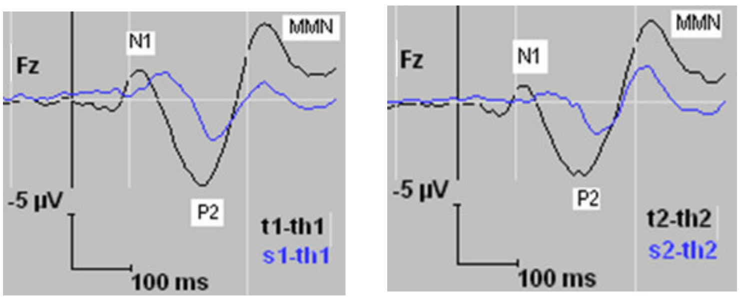


Figure 2: Difference waves for Fz in the multideviant block (left) and in the equiprobable block (right).



Analysis of the difference waves of [tond] and [sond] as deviants in block 1 versus their difference as standards in block 2 revealed no significant interactions for maximum amplitude ($F(1, 15) = .4$, $p>.5$) or area ($F(1,15) = .120$, $p>.7$). There was, however, a significant interaction for maximum latency ($F(1,15) = 44.4$, $p<.001$). As [tond] evokes an earlier MMN than [sond], this indicates that it is not experience but acoustical similarity that is observed here. Since the observed P2 seemed rather large, and P2 effects have been found in studies on language-specific effects (Dehaene-Lambertz 1997; Brunellière et al. 2009) it was analysed here as well. The time window of the P2 for [tond] in block 1 minus [tond] in block 2 was set at 100 to 200 ms, and it was compared with the P2 of [sond] in block 1 minus [sond] in block 2, set at a time window of 110 to 210 ms. No significant interactions were found for maximum amplitude, maximum latency, and area.

4. DISCUSSION AND CONCLUSION

The prelexical processing of the highly preferred th-substitution [t] and the less preferred th-substitutions [s] in English was investigated in an MMN study with Dutch listeners. The elicitation of MMNs was clearly shown when [θond] was the standard and [tond] and [sond] functioned as deviants, with [tond] eliciting a larger MMN and a shorter latency than [sond]. This result pattern contradicts predictions based on language experience, and is rather in line with an explanation based on perceptual similarities. In terms of perceptual similarity, [s] is closer to [θ] than [t] for Dutch listeners (Cutler et al. 2004). Since [tond] elicited a larger MMN than [sond], the results suggest that perceptual similarity was the driving force for the present findings. This is consistent with other findings showing that the MMN is correlated with acoustic and perceptual distance (e.g. Shtyrov et al. 2007) and that MMN latencies decrease as a function of increased auditory discrimination performance (e.g. Näätänen et al. 1993). It is not consistent, however, with some further MMN studies that did find effects of experience (e.g. Näätänen 1997; Jacobsen 2003; Kirmse et al 2008; Dehaene-Lambertz 1997; Brunellière et al. 2009). However, in contrast to the present study, these experience-based effects were always observed for native listeners. In the earlier studies, phoneme contrasts were also usually chosen based on a within-category versus an across-category difference. Conversely, in the present study the contrast was a th-substitution that is often produced by Dutch learners of English, namely [t], versus the less frequent substitution, [s]. Such an effect of preferences in foreign-accented speech has not been investigated before. It is possible that the experience-based memory representations for the preferred phoneme substitutions are not as well established as native language-specific memory representations.

Nevertheless, prior experience with th-substitutions has been shown to interact with the lexical level of processing (Hanulíková & Weber, in prep.). Before concluding now that, at least for L2 listeners, effects of experience are not influencing prelexical processing, we want to point out some factors that may have affected the observed pattern. One difference to the Hanulíková & Weber study is that the stimuli in the eye-tracking study were spoken by Dutch and German learners of English, but in the present study stimuli were spoken by a native speaker of American English. A native speaker rather than a Dutch speaker was chosen in order to avoid accent-specific advantages for a future comparison with German participants listening to the same stimuli. However, our choice of speaker may have created an English-language context for the listeners, and it remains possible that speaker identity interfered with the application of experience gathered from listening to Dutch learners of English. Moreover, Dutch and English /t/ also differ.

The durational differences of the stimuli should also be considered for their influence on the MMN-amplitudes. As noted, the durations of [θond], [tond], and [sond] were 593, 499 and 609 ms, respectively. Research on durational differences in an MMN paradigm, however, usually focuses on *vowel* length (e.g. Kirmse et al. 2008), whereas here, the durational differences originated from *consonants*. In addition, in studies on vowel length, stimuli are spectrally kept constant, whereas we employed different consonants. Therefore, it is not likely that durational differences alone could have accounted for the amplitude and latency differences found here. Effects of stimulus length are furthermore unexpected because of the relatively small durational differences; effects are generally found when stimuli are at least 40% different in duration (e.g. Jacobsen et al. 2003; Kirmse et al. 2008). In comparison, [tond] was 15% shorter than [θond] in the present study, making a duration-based account of MMN elicitation unlikely.

Surprisingly, the equiprobable block also showed significant interactions between the difference waves of [tond]-[θond] versus [sond]-[θond]. This was unexpected, because this design is well-attested as a control condition. Presenting the three stimuli equally often was expected to result in three 'standard' waveforms - hence, no MMNs were expected. One explanation for the elicitation of MMNs in the equiprobable block could be that the order of the blocks (with the equiprobable block being second) was never reversed. It could be that the exposure to 800 [th] standard stimuli in the first block created a very solid neural representation of the standard. The second block, in which a further 200 occurrences of the standard were presented, may have further strengthened this standard representation. The deviants, which were each presented only 100 times in the first block, may not have had enough instances to function as a standard as well. Another possible explanation for the results of the second block in this study may be that the stimuli did not have the exact same frequency as the deviant in the first block. This procedure has been adapted by some studies (see e.g. Maess et al. 2007), although others do report the procedure adopted here (e.g. Horváth et al. 2008).

In sum, we investigated the prelexical processing of English th-substitutions in an MMN-paradigm with Dutch L2 listeners. Our main goal was to test whether the deviant stimulus [tond], which is a highly typical mispronunciation for Dutch learners of English, is processed differently from the less typical deviant [sond]. The results clearly show that word initial [t] and [s] lead to different brain responses in terms of latency when they are compared with word-initial [θ], with [t] producing later MMNs effects than [s]. Thus, although Dutch listeners are highly familiar with [t]-substitutes from their L2 speech, this experience was not reflected in the MMN effects (in this case, the opposite pattern of results should have been observed). Rather perceptual similarity between the substitutes and the English interdental fricative [θ] can explain the findings. It remains possible, however, that experience still modulates the MMN effects even though it is not the main force for the effects. A comparison of Dutch listeners in the present study with German listeners, for whom the opposite experience pattern holds (i.e., German learners substitute [θ] mostly with [s] and less often with [t]), will help to clarify this possibility in the future.

5. REFERENCES

Advanced Source Analysis Software, ANT Software, Enschede, The Netherlands. [Computer Programme]

Allopenna, P.D., Magnuson, J.S., Tanenhaus, M.K. 1998. Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language* 38. 419-439.

Bien, H., Lagemann, L., Dobel, C., Zwitserlood, P. 2009. Early automatic categorization of speech sounds is not misled by changes in the surface form - a dissociation of behavioural and neurophysiological data.

Boersma, P. and Weenink, D. 2002. Praat 5.0: a system for doing phonetics with the computer (version 5.0) [Computer Programme]

Brunellière, A., Dufour, S., Nguyen, N. and Frauenfelder, U. H. 2009. Behavioral and electrophysiological evidence for the impact of regional variation on phoneme perception. *Cognition* 111. 390-396.

Coenen, E., Zwitserlood, P., Bölte, J. 2001. Variation and assimilation in German: consequences for lexical access and representation. *Language and Cognitive Processes* 16. 535-564.

Connine, C.M. 2004. It's not what you hear but how often you hear it: on the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review* 11. 1084–1089.

Cutler, A., Weber, A., Smits, R., Cooper, N. (2004). Patterns of English phoneme confusions by native and non native listeners. *Journal of the Acoustical Society of America* 116. 3668-3678.

Dehaene-Lambertz, G. 1997. Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport, 8,* 919-924.

Hancin-Bhatt, B.J. 1994. Phonological transfer in second language perception and production. Doctoral Dissertation, University of Illinois.

Hanulíková, A., Weber, A. In prep. Sink positive: Linguistic experience with th-substitutions influences non-native word recognition.

Hanulíková, A., Weber, A. Production of English interdental fricatives by Dutch, German, and English speakers. Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech, New Sounds 2010, Poznan.

Hautus, M.J., Meng, X. 2002. Decision strategies in the ABX psychophysical task. *Perception & Psychophysics* 64. 89-106.

Horváth, J., Czigler, I., Jacobsen, T., Maess, B., Schröger, E., Winkler, I. 2008. MMN or no MMN: no magnitude of deviance effect on the MMN amplitude. *Psychophysiology* 45. 60-69.

Jacobsen, T., Horváth, 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. 2003. Measuring duration mismatch negativity. *Clinical Neurophysiolog* 114. 1133-1143.

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. *International Journal of Psychophysiolog* 67. 131-143.

Maess, B., Jacobsen, T., Schröger, E., Friederici, A.D. 2007. Localizing pre-attentive auditory memory-based comparison: magnetic mismatch negativity to pitch change. *NeuroImage* 37. 561-571.

Mitterer, H., Ernestus, M. 2006. Listeners Recover /t/s that speakers reduce: evidence from /t/-lenition in Dutch. *Journal of Phonetics* 34. 37-103

Näätänen, R., Gaillard, A.W., Mantysalo, 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., 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. *Letters to 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

Shtyrov, Y., Pulvermüller, F. 2007. Language in the mismatch negativity design. motivations, benefits, and prospects. *Journal of Psychophysiology* 21. 176-187.

Wester, F, Gilbers, D., Lowie, W. 2007. Substitution of dental fricatives in English by Dutch L2 speakers. *Language Sciences* 29. 477-491.