

Tilburg University

Lexical effects on auditory speech perception

van Linden, S.; Stekelenburg, J.J.; Tuomainen, J.; Vroomen, J.

Published in:
Neuroscience Letters

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
van Linden, S., Stekelenburg, J. J., Tuomainen, J., & Vroomen, J. (2007). Lexical effects on auditory speech perception: An electrophysiological study. *Neuroscience Letters*, 420(1), 49-52.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Lexical effects on auditory speech perception: An electrophysiological study

Sabine van Linden^a, Jeroen J. Stekelenburg^a,
Jyrki Tuomainen^b, Jean Vroomen^{a,*}

^a *Psychonomics Laboratory, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands*

^b *University College London, Human Communication Science, UK*

Received 22 January 2007; received in revised form 28 March 2007; accepted 3 April 2007

Abstract

Lexical information can bias categorization of an ambiguous phoneme and subsequently evoke a shift in the phonetic boundary. Here, we explored the extent to which this phenomenon is perceptual in nature. Listeners were asked to ignore auditory stimuli presented in a typical oddball sequence in which the standard was an ambiguous sound halfway between /t/ and /p/ embedded in a Dutch word normally ending in /t/ ('vloot', meaning 'fleet') or /p/ ('hoop', meaning 'hope'). As deviant served the non-ambiguous sound /t/ embedded in the same context. The amplitude of the MMN-response, indexing the perceptual difference between the ambiguous sound and unambiguous /t/ was bigger for the p-word 'hoop' than the t-word 'vloot'. This result is taken as an indication that lexical information actually reached down to early perceptual processing stages.

© 2007 Elsevier Ireland Ltd. All rights reserved.

Keywords: Mismatch negativity (MMN); Lexical processing; Speech perception; Ambiguous words

Identification of a speech sound is influenced by word context, especially if the sound is ambiguous or degraded. For example, an ambiguous sound that might be a /g/ or a /k/ is more likely to be identified as a /g/ if followed by 'ift' and as a /k/ if followed by 'iss' [3]. Presumably, this bias effect occurs because 'gift' and 'kiss' are words in English, but not 'kift' and 'giss'. What is less known, is that next time listeners hear the same ambiguous sound, they have learned from the past and now perceive the initially ambiguous 'g/k' as /g/ or /k/ right away [6,15,18]. The occurrence of such a lexically induced aftereffect is taken as an indication that listeners have adjusted, or *recalibrated*, the phonetic categories of their language so as to adapt to the new situation. Here, we explored the extent to which these phenomena are truly perceptual in nature rather than reflecting a post-lexical decision stage.

Interactive approaches argue that bias effects are due to a direct lexical influence on pre-lexical representations [8]. They predict that lexical information actually reaches down and changes the momentary activation of the sound that is heard. Other proposals, in which speech recognition is seen as a more

autonomous, bottom-up process, propose that lexical contextual information does not change the activation of the pre-lexical representation per se [15] because that will harm speech recognition proper. Lexical bias effects occur, on this view, on a post-lexical phonemic decision stage. Nevertheless, autonomous accounts leave open the possibility that pre-lexical levels are affected, but in an indirect way, via recalibration [18]. The notion is that lexical information induces a shift in the boundary between two phonetic categories, and to the extent that this shift occurs at a perceptual level, one may observe that lexical information affects early processing stages. Both accounts can therefore predict that lexical information penetrates mechanisms of perception at early pre-lexical levels, and thus affect the way a sound is heard. Here, we tested this prediction, for the first time, using recordings of human brain event-related potentials (ERPs) focusing on the mismatch negativity (MMN).

The MMN is an ERP component that signals an infrequent discriminable change in an acoustic or phonological feature of a repetitive sound [12]. The behavioural discriminability of the stimuli is usually correlated with the amplitude and latency of the MMN-response [7]. The MMN-generating process is not volitional, it does not require attentive selection of the sound (although it can be diminished under high attentional load [17]), and it is elicited whether or not the sounds are relevant

* Corresponding author. Tel.: +31 13 466 2394; fax: +31 13 466 2370.
E-mail address: J.Vroomen@uvt.nl (J. Vroomen).

Table 1
Experimental design

	Standard	Deviant
t-Word ‘vloot’	/vlo?/	/vlot/
p-Word ‘hoop’	/ho?/	/hot/

for the participant’s task [9,14]. Furthermore, the MMN is not only sensitive to acoustic changes, but also to learned language-specific auditory deviancy [11]. For example, in a cross-linguistic study of Hungarian and Finnish, Winkler et al. [19] used within- and across-category phoneme contrasts that were reversed for the two languages. By means of this crossed design, they demonstrated that the MMN-generating process simultaneously operates both on the basis of auditory sensory memory and categorical phonetic stimulus representations (see also [2,13]). These results suggest that linguistic information triggers additional processes, which may prepare the auditory system for detecting language-specific auditory deviations. The pre-attentional and automatic nature of the MMN [9,10] together with its sensitivity to phonetic contrasts and stimulus discriminability therefore makes it suitable to investigate whether lexical information can affect early pre-lexical processing stages. If it can be demonstrated that lexical information indeed changes the MMN while acoustic factors are strictly controlled for, it would naturally strengthen the idea that lexical information affects pre-lexical processes, be it direct via top-down lexical activation, or indirect via recalibration.

Here, we presented Dutch listeners a word normally ending in /t/ (‘vloot’, meaning ‘fleet’ in English) or /p/ (‘hoop’, meaning ‘hope’), whereby the final consonant (/t/ or /p/) was replaced by an ambiguous sound halfway between /t/ and /p/ (henceforth /?/). This thus resulted in the t-word /vlo?/ and the p-word /ho?/ (note that ‘vloop’ and ‘hoot’ do not exist in Dutch). In a previous study [18], we confirmed that these words evoked a lexical bias in phoneme categorization (i.e., listeners judged /?/ in /vlo?/ as more t-like than in /ho?/) and a recalibration effect (i.e., listeners were more likely to categorize /?/ as /t/ after hearing /vlo?/ than after hearing /ho?/). The t- and p-words were presented in a typical oddball paradigm (Table 1). The standard stimulus was either the t- or the p-word containing the ambiguous sound /?/, while on infrequent deviant trials /?/ was replaced by non-ambiguous /t/. Listeners thus heard /vlot/ as deviant in the t-word condition (i.e., the word that is in congruence with the lexical information ‘vloot’) and /hot/ in the p-word condition (i.e., a pseudoword that is incongruent with the lexical information ‘hoop’). Crucially, the acoustic change from the standard /?/ to the deviant /t/ was in the p- and t-word conditions exactly the same, as these two words only differed in their initial consonants. Interactive accounts, though, predict the MMN to be smaller in the t-word than p-word because the t-word increased activation of /t/, while the p-word increased activation of /p/. Similar predictions can be made for accounts that instantiate recalibration at an early perceptual level [18]. If the shift in the phoneme boundary as evoked by the lexical information is perceptual in nature, one expects the perceptual difference between the /?/ and /t/ to be smaller in the t-word than the p-word, because /?/ is recal-

ibrated towards /t/, which in turn should yield a smaller MMN amplitude.

Given that the deviant in the t-word condition is a word (‘vloot’), while in the p-word condition it is a pseudoword (‘hoot’), one might ask whether a smaller MMN for t-words might reflect a change in the lexical status of the deviant rather than a change in the way the ambiguous sound is heard. At present, there is mixed evidence about the role of the lexical status of the deviant. Pulvermüller et al. [16] argued that words engage a lexical representation in addition to the acoustic and phonetic representations activated by pseudowords, and word deviants will therefore always evoke a larger MMN than pseudoword deviants irrespective of the lexical status of the standard. This hypothesis is thus in the opposite direction of our prediction (i.e., the t-word condition with a word as deviant will have a smaller MMN). Jacobsen et al. [5], though, argued that the lexical status of the deviant is irrelevant for the MMN because they found no difference between word or pseudoword deviants when acoustic and language factors were controlled. Whichever of these two accounts is correct, here it seems safe to conclude that a smaller MMN in the t-word condition is unlikely to be caused by the fact that the deviant in this condition is a word rather than a pseudoword, because previous studies suggest that the MMN should either be bigger [16], or not be affected [5].

Sixteen native speakers of Dutch (4 males, 12 females) with normal hearing and normal or corrected-to-normal vision participated in the experiment after giving written informed consent. Their age ranged from 18 to 25 years (mean age 19.5 years). The experiment was conducted in accordance with the Declaration of Helsinki. The experiment took place in a dimly lit, sound attenuated and electrically shielded room. Stimulus creation started with digital recording of /vlo/ and /hop/ by a male Dutch speaker. The final vowel and consonant of the two words were replaced by /o?/. The ambiguous sound /?/ was created with Praat [1] from another recording of /ot/ and /op/ in which the second (F2) and third (F3) formant were changed. The steady-state value of the F2 in the vowel was 950 Hz (72 ms in duration), and the offset frequency of the transition (45 ms duration) was 928 Hz. The steady-state value of F3 in the vowel was 2400 Hz, and the offset frequency of the transition was 2265 Hz. There was 40 ms of silence before the final release of the stop consonant. The aspiration part of the final release of /p/ and /t/ (134 ms) were mixed from natural /p/ and /t/ bursts in relative proportions to each other. The total duration of the words were /hoo?/ = 531 ms, /hoot/ = 495 ms, /vlo?/ = 664 ms and /vloot/ = 628 ms. Stimuli were presented from a loudspeaker located in front (90 cm) of the participant with a peak intensity of 70 dB(A).

Pre-tests confirmed that the ambiguous phoneme /?/ was perceived by each participant as halfway between /p/ and /t/ when presented in the neutral pseudoword /so?/. We also obtained a behavioural measure of lexical bias on phoneme categorization by asking participants at the end of the MMN experiment to rate on a 7-point Likert scale the quality of the final phoneme /?/ when embedded in the t-word /vlo?/ and p-word /ho?/. As expected, /?/ in the t-word /vlo?/ was rated as more t-like than when embedded in the p-word /ho?/ (3.83 versus 5.80 units on the 7-point Likert scale, $t(1, 12) = 6.31$, $p < 0.001$).

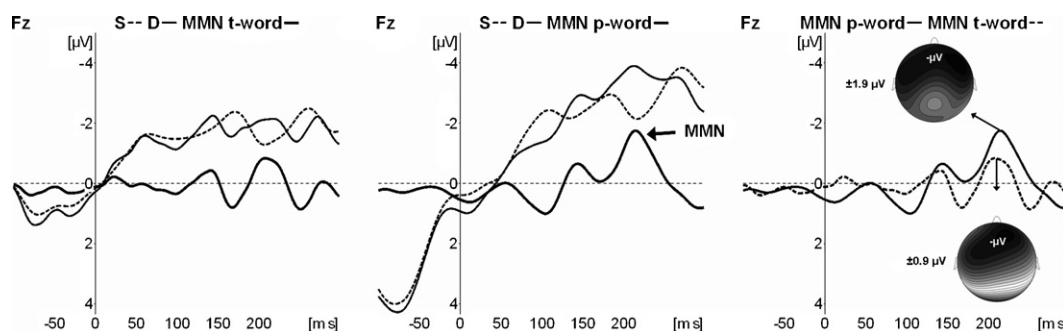


Fig. 1. Grand-averaged waveforms of the standard (S), deviant (D) and MMN at electrode Fz for the t-word condition (left panel) and p-word condition (middle panel). The right panel shows the MMNs and their scalp topographies for both conditions. The range of the voltage maps in μV are displayed left to each map. The y-axis marks the onset of the acoustic deviation between /ʋ/ and /t/.

In the MMN experiment, stimuli were presented in a typical unattended oddball paradigm (standard 82%, deviant 18%). The order of stimuli was randomized with the restriction that at least two standards preceded each deviant. The stimulus onset asynchrony was 1250 ms. During stimulus presentations, participants fixated on a small white cross on a monitor – placed directly above the loudspeaker – and detected an occasional catch trial (11% of the standard trials). Their task was to indicate by a button press when the colour of the fixation point changed. Participants were administered two blocks per word type condition, each consisting of 440 trials, which amounted to a total 720 standards (including 80 catch trials) and 160 deviants. Presentation order of the four blocks was counterbalanced across participants.

The electroencephalogram (EEG) was recorded at a sample rate of 512 Hz from 43 active Ag–AgCl electrodes (BioSemi, Amsterdam, The Netherlands) mounted in an elastic cap and two mastoid electrodes. Electrodes were placed according to the extended International 10–20 system. Two additional electrodes served as reference (Common Mode Sense [CMS] active electrode) and ground (Driven Right Leg [DRL] passive electrode). EEG was re-referenced offline to an average of left and right mastoids and band-pass filtered (0.1–30 Hz, 24 dB/octave). The electrooculogram (EOG) measuring horizontal and vertical eye-movements were recorded using electrodes at the outer canthus of each eye as well as above and below the right eye. The raw data were segmented into epochs of 500 ms including a 100 ms pre-stimulus baseline. After eye movement correction [4], epochs with an amplitude change exceeding $\pm 100 \mu\text{V}$ at any channel (except EOG) were rejected (4% of the deviant trials). ERPs of the standard and deviant non-catch trials were averaged separately for t- and p-words. The standard ERP was subtracted from the deviant ERP to obtain the MMN. To match the timing of the ERP components in the t-word condition with those in the p-word condition, ERPs were time-locked to the onset of the final phoneme of the standard and deviant word (i.e., the point where /ʋ/ and /t/ started to deviate). Based on visual inspection of the grand average waveforms at electrode Fz, the MMN was identified as a negative deflection in a 150–250 ms window after the onset of the final phoneme. MMN amplitude was calculated as a 50 ms mean amplitude centred on the individual peak latency of MMN. The MMN was tested at electrodes F3, Fz, F4, FC3, FCz, FC4 with a MANOVA for repeated measures with as within-subject factors Condition (p-word versus t-word),

Hemisphere (left, middle, right) and Anterior–Posterior (frontal versus fronto-central). A one-tailed test for Condition was used because there was a clear prediction about the direction of the lexical effect on the MMN. Two participants were excluded from the analysis because strong alpha waves prevented reliable scoring of the MMN. Data of one participant were discarded because of hardware failure.

Fig. 1 depicts the ERPs elicited by standards and deviants and the difference waves at Fz for both p- and t-conditions. MMN peaked at 215 ms at Fz with no difference between p- and t-words in their timing ($t < 1$). As predicted, the amplitude of the MMN was larger for p-words than for t-words, $F(1, 12) = 3.62$, $p < 0.05$. Post hoc analysis showed that the MMN for p-words ($-1.6 \mu\text{V}$) significantly deviated from zero, $F(1, 12) = 22.74$, $p < 0.001$, whereas MMN amplitude for t-words ($-0.5 \mu\text{V}$) did not differ significantly from zero ($p = 0.18$). Testing the scalp distribution of MMN revealed no Condition \times Hemisphere, Condition \times Anterior–Posterior, or Condition \times Hemisphere \times Anterior–Posterior interactions (all $F_s < 1$), indicating that the scalp distribution of the MMN did not differ between conditions.

The results thus show that, with acoustic factors being controlled for, the perceptual change from /ʋ/ to /t/ was smaller in the t-word than in the p-word. This finding is in line with accounts that attribute the lexical context effects in speech perception to a pre-lexical level rather than a post-lexical phonemic decision stage. Interactive accounts might argue that the lexical representation of the t-word ‘vloot’ increased the activation of /t/ via feedback connections, while the p-word ‘hoop’ increased activation of /p/. A recalibration account might suggest that upon hearing the t-word /vloʋ/, the phoneme boundary is shifted such that next time /ʋ/ is presented, it is heard as /t/. The perceptual change from /ʋ/ to /t/ is therefore smaller in ‘vloot’ than in ‘hoop’, eliciting on its turn an MMN of smaller amplitude.

In the present study, a visual task (i.e., detection of an occasional change in fixation) was used to draw attention away from auditory stimulation. It is unknown, though, to which extent participants ignored the auditory stimuli. Given that the MMN can be modified by attention [17], future studies might manipulate the attentional load to determine whether the lexically induced MMN reflects auditory processing at a pre-attentive stage.

Future work is also needed to further elucidate the link between behavioural performance and the underlying cogni-

tive processes and neural system that support it. For example, behaviourally, we observed that the ambiguous phoneme /ʔ/ was rated more t-like when embedded in the t-word /vloʔ/ than the p-word /hoʔ/. However, the size of this lexical bias effect on phoneme categorization did not correlate with the amplitude of the MMN ($r_s = -0.19$, $p = 0.54$). There was thus no simple relation such that participants who had a large lexical bias effect also had a strong MMN. This aspect of the results will be further investigated.

References

- [1] P. Boersma, D. Weenink, Praat: Doing Phonetics by Computer, vol. 2002, Amsterdam, 2002.
- [2] G. Dehaene-Lambertz, Electrophysiological correlates of categorical phoneme perception in adults, *Neuroreport* 8 (1997) 919–924.
- [3] W.F. Ganong, Phonetic categorization in auditory word perception, *J. Exp. Psychol. Hum. Percept. Perform.* 6 (1980) 110–125.
- [4] G. Gratton, M.G. Coles, E. Donchin, A new method for off-line removal of ocular artifact, *Electroencephalogr. Clin. Neurophysiol.* 55 (1983) 468–484.
- [5] T. Jacobsen, J. Horváth, E. Schröger, S. Lattner, A. Widmann, I. Winkler, Pre-attentive auditory processing of lexicality, *Brain Language* 88 (2004) 54–67.
- [6] T. Kraljic, A.G. Samuel, Perceptual learning for speech: is there a return to normal? *Cogn. Psychol.* 51 (2005) 141–178.
- [7] H. Lang, T. Nyrke, M. Ek, O. Aaltonen, I. Raimo, R. Näätänen, Pitch discrimination performance and auditory event-related potentials, in: C.H.M. Brunia, A.W.K. Gaillard, A. Kok, G. Mulder, M.N. Verbaten (Eds.), *Psychophysiological Brain Research*, vol. 1, Tilburg University Press, Tilburg, The Netherlands, 1990, pp. 294–298.
- [8] J.L. McClelland, D. Mirman, L.L. Holt, Are there interactive processes in speech perception? *Trends Cogn. Sci.* 10 (2006) 363–369.
- [9] R. Näätänen, *Attention and Brain Function*, Hillsdale, 1992.
- [10] R. Näätänen, The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm), *Psychophysiology* 38 (1999) 1–21.
- [11] R. Näätänen, The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent, *Psychophysiology* 38 (2001) 1–21.
- [12] R. Näätänen, A.W.K. Gaillard, S. Mäntysalo, Early selective-attention effect in evoked potential reinterpreted, *Acta Psychol.* 42 (1978) 313–329.
- [13] R. Näätänen, A. Lehtokoski, M. Lenneberg, M. Cheour, M. Huottilainen, A. Livonen, M. Vainio, P. Alku, R.J. Ilmoniemi, A. Luuk, J. Allik, J. Sinkkonen, K. Alho, Language-specific phoneme representations revealed by electric and magnetic brain responses, *Nature* 385 (1997) 432–434.
- [14] R. Näätänen, P. Paavilainen, H. Tiitinen, D. Jiang, K. Alho, Attention and mismatch negativity, *Psychophysiology* 30 (1993) 436–450.
- [15] D. Norris, J.M. McQueen, A. Cutler, Perceptual learning in speech, *Cogn. Psychol.* 47 (2003) 204–238.
- [16] F. Pulvermüller, T. Kujala, Y. Shtyrov, J. Simola, H. Tiitinen, P. Alku, K. Alho, S. Martinkauppi, R.J. Ilmoniemi, R. Näätänen, Memory traces for words as revealed by the mismatch negativity, *Neuroimage* 14 (2001) 607–616.
- [17] E. Sussman, I. Winkler, M. Huottilainen, W. Ritter, R. Näätänen, Top-down effects can modify the initially stimulus-driven auditory organization, *Brain Res. Cogn. Brain Res.* 13 (2002) 393–405.
- [18] S. van Linden, J. Vroomen, Recalibration of phonetic categories by lipread speech versus lexical information, *J. Exp. Psychol. Hum. Percept. Perform.*, in press.
- [19] I. Winkler, T. Kujala, Y. Shtyrov, J. Simola, H. Tiitinen, P. Alku, A. Lehtokoski, I. Czigler, R.J. Ilmoniemi, R. Näätänen, Brain responses reveal the learning of foreign language phonemes, *Psychophysiology* 36 (1999) 638–642.