

Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity

Stekelenburg, J.J.; Vroomen, J.; de Gelder, B.

Published in:
Neuroscience Letters

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

Publication date:
2004

[Link to publication](#)

Citation for published version (APA):
Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters*, 357(3), 163-166.

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.



ELSEVIER

Neuroscience Letters 357 (2004) 163–166

Neuroscience
Letters

www.elsevier.com/locate/neulet

Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity

Jeroen J. Stekelenburg*, Jean Vroomen, Beatrice de Gelder

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

Received 28 October 2003; received in revised form 3 December 2003; accepted 4 December 2003

Abstract

The ventriloquist illusion arises when sounds are mislocated towards a synchronous but spatially discrepant visual event. Here, we investigated the ventriloquist illusion at a neurophysiological level. The question was whether an illusory shift in sound location was reflected in the auditory mismatch negativity (MMN). An ‘oddball’ paradigm was used whereby simultaneously presented sounds and flashes coming from the same location served as standard. The deviant consisted of a sound originating from the same source as the standard together with a flash at 20° spatial separation, which evoked an illusory sound shift. This illusory sound shift evoked an MMN closely resembling the MMN evoked by an actual sound shift. A visual-only control condition ruled out that the illusory-evoked MMN was confounded by the visual part of the audiovisual deviant. These results indicate that the crossmodal interaction on which the ventriloquist illusion is based takes place automatically at an early processing stage, within 200 ms after stimulus onset.

© 2003 Elsevier Ireland Ltd. All rights reserved.

Keywords: Crossmodal interaction; Ventriloquist illusion; Mismatch negativity

In daily life, information coming from several modalities impinges simultaneously on the sensory receptors. Events taking place in one modality may influence perceptual processing of other modalities. One of the most striking manifestations of such crossmodal interactions is the so-called ventriloquist illusion, referring to the observation that discrepancies in the spatial location of synchronized auditory and visual events can lead to a bias of the perceived auditory location towards the visual one [1]. For example, when subjects are required to indicate the location of a sound in a situation of audiovisual spatial conflict, their response is typically displaced in the direction of the visual stimulus [1,5]. The explanation is that the perceived location of the sound is shifted in the direction of the visual stimulus, thereby reducing the spatial conflict [1]. By doing so, the perceptual system integrates the intermodal discordant signals into a unitary multisensory percept. The ventriloquist illusion is basically considered to be a perceptual phenomenon [1,2,5,15], but as yet little is known about its time course. If the ventriloquist illusion has a perceptual origin one expects that crossmodal

integration takes place at the early processing stages. In the current study we used one specific event-related brain potential (ERP) called the mismatch negativity (MMN) as a tool to trace the time course of the ventriloquist illusion.

The MMN is typically evoked by an occasional auditory change (deviant) in a homogenous sequence of auditory stimuli (standards). It is considered to reflect the outcome of a preattentive comparison process between the neural representation of the incoming deviant and the neural trace of the standard [10]. If the neural trace of the standard is violated, an MMN is evoked. It has been suggested that the mechanism underlying the MMN may trigger an involuntary attention switch to novel auditory stimuli [14]. The auditory MMN appears as a negative wave, acquired by subtracting the standard waveform from the deviant waveform. It peaks at 100–250 ms after the onset of the sound change and has a frontocentral scalp distribution. A wide range of deviations in sound features may elicit the MMN of which changes in sound location [5,11,12] is the critical one for this study.

Here, our goal was to investigate whether an *illusory* shift in sound location induced by a ventriloquist illusion is reflected in the ‘auditory’ location MMN. Recently, Colin et al. [5] demonstrated that the MMN may indeed be

* Corresponding author. Tel.: +31-13-466-8171; fax: +31-13-466-2067.
E-mail address: j.j.stekelenburg@uvt.nl (J.J. Stekelenburg).

sensitive to the ventriloquist illusion. In their study, occasional speech sounds (/pi/) coming from 20° spatial separation from the auditory standards (/pi/) at central location elicited an MMN. The same deviants, however, elicited no MMN when the deviant stimulus was accompanied by a simultaneously presented face articulating /pi/, coming from the location of the standard. It was argued that the visual stimulus attracted the apparent location of the deviant sound. As a result, no spatial separation between standard and deviant was perceived (because of the ventriloquist effect), thereby preventing an MMN to occur.

In contrast with Colin et al. [5], we did not test whether the ventriloquist illusion could *eliminate* the MMN (which is basically a null-effect), but rather whether an MMN could be *evoked* by an illusory sound change. This set-up also allowed us to compare an MMN to illusory sound shifts with one to actual sound shifts. Furthermore, Colin et al. [5] used realistic real-life stimuli, thereby introducing a role for cognitive factors, referring to the knowledge that faces and voices are likely to originate from the same source. Such experience-based variables may have contributed to the elimination of the MMN. Here, we tested whether the effect could be observed with non-speech stimuli in situations in which the subject's familiarity with the bimodal stimuli is less obvious. This was realized by using meaningless, simplified visual (flash) and auditory (beep) stimuli [1], with the purpose to reveal whether the contribution of sensory factors as similarity in temporal onset are sufficient to cause the crossmodal effect.

We designed a ventriloquist-MMN paradigm in which the standard consisted of a synchronized beep and flash originating from the same spatial location. The deviant stimulus consisted of a beep coming from the same location as the standard, and a simultaneously presented flash with 20° separation. As confirmed in a control experiment (see below), we expected 'capture' of the sound by the spatially discrepant visual deviant, leading to the illusion that the beep was displaced in the direction of the flash. Our prediction was that such an illusory sound shift would elicit an 'auditory' MMN, despite the fact that the auditory component remained physically unchanged. Besides the audiovisual (AV) condition, there was an auditory-only condition (A) to confirm that spatially discrepant auditory signals evoked an MMN, and a visual-only condition (V) to control for purely visual contributions to the MMN in the audiovisual condition.

Twenty-one healthy volunteers (19–36 years, mean 21.6 years) with normal hearing and normal or corrected-to-normal vision participated in the experiment. Conform the procedure used in the study of Colin [5], only those participants ($n = 16$) with a clearly discernible auditory MMN ($>1 \mu\text{V}$) in the auditory-only condition were selected for further analysis. Stimuli were presented by two display units at a viewing distance of 140 cm whose centers were spatially separated by 20° (10° to the left and

10° to the right of straight ahead). A display unit consisted of a loudspeaker and a red LED (300 cd/m² luminance) attached to the center of the speaker cone. The set-up was hidden from the participant's view by a black, acoustically transparent curtain, occluding the speakers but allowing the LED to pass. Auditory stimuli were pure tones of 600 Hz at an intensity of 50 dB(A) SPL (80-ms duration, including 10-ms rise and fall times). Simultaneously presented visual stimuli also had a duration of 80 ms. Inter-stimulus interval (from stimulus onset) was 700 ms. In each condition (A, V, and AV), standard stimuli ($P = 0.90$) were always presented at one location (left or right of medium), while deviant stimuli ($P = 0.10$) were presented at the other location. The order of stimuli was randomized with the restriction that at least two standards preceded each deviant. Participants were administered four blocks per condition (A, V, and AV), each containing 500 trials. In half of the blocks, standard stimuli came from the left location, while in the other half they came from the right location. Block order was quasi-randomized across participants. As is standard practice in an MMN paradigm, a task was included to direct attention away from the test stimuli. The task was to fixate a green LED (8 cd/m² luminance) that was centrally positioned between the display units. Infrequently (ten times per block), the LED was turned off for 200 ms. This event was timed between two standards and occurred never within two trials before or after the presentation of the deviant stimulus. Participants had to react as fast as possible to the offset of the LED by pressing a key attached to the right arm rest of the chair.

EEG was recorded from 16 locations using active Ag-AgCl electrodes (BioSemi Active 2) mounted in an elastic cap: Fz, F3, F4, FC3, FC4, Cz, C3, C4, Pz, P3, P4, Oz, O1, O2, left and right mastoids (M1, M2). The reference was the tip of the nose. Horizontal EOG was recorded from two electrodes placed at the outer canthi of both eyes. Vertical EOG was recorded from electrodes on the infraorbital and supraorbital regions of the right eye in line with the pupil. EEG signals were band-pass filtered (0.1–30 Hz, 24 dB/octave) at a sample rate of 256 Hz. EOG artifacts were corrected according to the procedure described by Gratton, Coles, and Donchin [7]. The raw data were segmented into epochs of 600 ms, including a 100-ms prestimulus baseline. Epochs with an amplitude change exceeding $\pm 70 \mu\text{V}$ at any channel were automatically rejected. For all three conditions (A, V, and AV), ERPs were averaged separately for the standards and deviants.

Difference waveforms were computed by subtracting the averaged ERP elicited by the standard from that of the deviant. The difference wave in the AV condition may be composed of overlapping components pertaining to the illusory sound shift and to the change in location of the flash. To suppress ERP activity evoked by the visual shift, the difference waveform of V was subtracted from the difference waveform of AV. This AV-V difference wave (henceforth, the ventriloquist MMN) thus represents the

EEG activity evoked by the illusory sound shift without the contribution of the visual component. Peak MMN amplitude was scored as the maximal negative value in a window of 150–250 ms post-stimulus. Peak MMN latency was scored at Fz using the same time window.

Before testing, a behavioral control experiment with 14 participants (18–24 years, mean 19.6 years) was conducted in order to investigate the effectiveness of the experimental set-up in eliciting a ventriloquist illusion. Sounds and flashes either emanated from the same location (left or right display unit) or from a different location (sound left/flash right, sound right/flash left). Fifty trials were run for each stimulus combination, amounting to a total of 200 self-paced, randomized trials. Participants indicated whether the sound came from the same location as the flash (right button) or from a different location (left button). In 75% (SD = 15.7%) of the different location trials, an (erroneous) ‘same’ response was given, which was far above chance level, $t(13) = 5.95$, $P < 0.01$. This indicates that the experimental set-up was successful in eliciting the ventriloquist illusion.

The performance on the detection task in the EEG experiment was high (98.4% correct detection of turning-off the central LED, SD = 1.6%) indicating that attention was indeed paid to the participant’s main task. Fig. 1 shows the ERPs at Fz evoked by the standards and deviants and their corresponding difference waves for each condition. The deviant in the AV condition elicited a negativity between 130 and 200 ms. The task-irrelevant visual deviant in the V condition did not elicit a negativity at Fz, but a P3a peaking at about 280 ms, a finding similar to a study of Berti and Schröger [3]. Furthermore, it can be seen that in the auditory-only control condition an MMN was obtained for auditory spatial deviance, confirming previous reports [5,11,12]. The auditory MMN and the AV-V difference wave (ventriloquist MMN) are presented in Fig. 2. The side of the location did not have a significant effect on peak amplitude or peak latency of the difference waves. The data were therefore pooled across stimulation side. Visual inspection of Fig. 2 reveals that the ventriloquist MMN and the auditory MMN are very similar in terms of waveform morphology, timing, and scalp distribution. This was confirmed by the statistical analysis of peak latency and peak amplitude of both difference waves. Peak latency of the auditory MMN (202 ms) did not significantly differ from

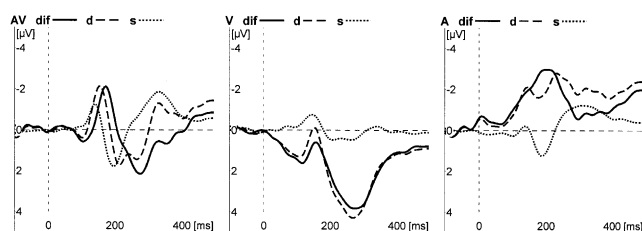


Fig. 1. Grand average ERPs at Fz elicited by the deviants (d) and standards (s), and their corresponding difference waves (dif) in the audiovisual (AV), visual-only (V), and audio-only (A) condition.

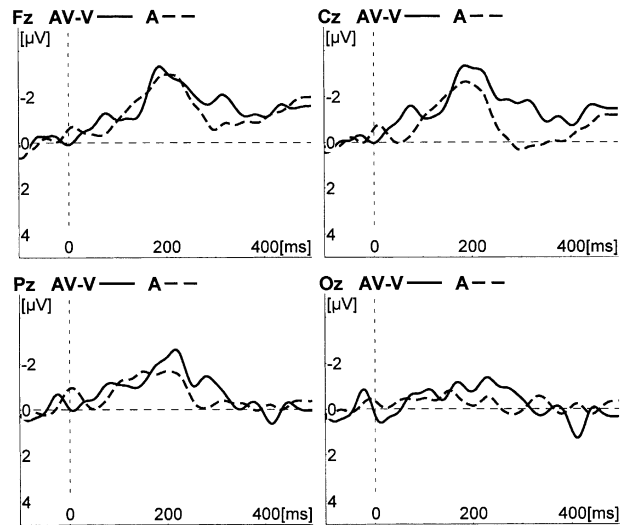


Fig. 2. Grand average difference waves at midline sites for the ventriloquist MMN (AV-V) and the auditory MMN (A).

peak latency of the ventriloquist MMN (193 ms). Peak amplitude of the difference waves at Fz was tested against the zero voltage baseline to identify the existence of an MMN. Both auditory and the ventriloquist MMN were significantly different from zero, $t(15) = 5.44$, $P < 0.001$ and $t(15) = 4.34$, $P < 0.001$, respectively. Peak amplitude of the ventriloquist MMN did not differ significantly from peak amplitude of the auditory MMN at Fz. The scalp distribution of the ventriloquist MMN was compared with that of the auditory MMN by testing the interaction between Condition (A vs. AV-V) and Lead (16 levels), using a multivariate analysis of variance procedure for repeated measures. No interaction was found, $F(1, 15) = 3.73$, $P = 0.39$, indicating that the scalp distribution of the ventriloquist MMN did not differ significantly from that of the auditory MMN.

To summarize, we observed that the MMN to an illusory sound shift was very similar to the MMN to an actual sound shift. This result suggests that the ventriloquist illusion arises at an early perceptual level, within 200 ms after stimulus onset. Moreover, the fact that the ventriloquist illusion was manifested in the MMN, which indexes the detection of acoustic changes at a preattentive level [10,14], suggests that the ventriloquist illusion is based on a mechanism operating at a preattentive, automatic processing stage. The existence of the ventriloquist MMN is in itself remarkable, since it occurred in the absence of any acoustical change. This suggests that the illusory sound shift induced by the ventriloquist illusion in the deviant trials resulted in a change in the acoustic sensory trace, causing the activation of the preattentive MMN generators.

Our findings are in accordance with those of Colin et al. [5] who also found that the MMN is sensitive to the ventriloquist illusion. The new finding here is that we now demonstrate that the ventriloquist MMN is very similar to the auditory MMN. In addition, we demonstrated that

familiar associations between articulatory movements of a face and the corresponding utterances, as used by Colin et al. [5], are not necessary to elicit the ventriloquist MMN. The fact that the ventriloquist MMN was elicited with simplified stimuli implies that sensory factors are sufficient to bring it about [1,2].

Audiovisual conflicts of non-spatial origin have also been shown to have an effect on the auditory MMN. One of them is the McGurk effect. It is the illusion that articulatory movements that are incongruent to the utterances modify the auditory percept [8]. For example, when people see someone articulating /ka/, while an auditory /pa/ is presented, they often report hearing /ta/ [13]. In a study of Sams et al. [13], infrequently presented incongruent audiovisual utterances (visual /ka/ paired with auditory /pa/) among congruent audiovisual standards (auditory and visual /pa/) elicited an (neuromagnetic) MMN. A similar effect was demonstrated in a study of crossmodal perception of emotions [6]. Incongruous face-voice pairs (angry voice and sad face) presented among congruous face-voice pairs (both angry voice and face) elicited an MMN. Note that these studies show that the MMN can be evoked in spite of the fact that there was no acoustical difference between the deviants and standards. This implies that the MMN is sensitive to a visual change that induces an apparent change of the acoustic stimulus with which it interacts.

Although the currently adopted ERP method shows that the ventriloquist illusion occurs at an early stage within the perceptual system, it does not allow to exactly pinpoint the brain regions involved. However, the fact that we did not find a topographical difference between the auditory and ventriloquist MMN suggests similar generators for both conditions. This would imply that the auditory cortex participates in the evocation of the ventriloquist MMN, since MMN generators that are responsive to physical acoustic differences are located in the supratemporal plane of the auditory cortex [10]. Furthermore, the neural source of the McGurk-like MMN has been localized in the supratemporal auditory cortex as well [4,9,13] suggesting that, in situations of audiovisual conflict, visual information has access to the auditory cortex. The hypothesis that the ventriloquist MMN originates in the auditory cortex is consistent with the perceptual experience of actually

‘hearing’ sounds coming from spatially discrepant visual inputs.

References

- [1] P. Bertelson, B. de Gelder, The psychology of multimodal perception, in: C. Spence, J. Driver (Eds.), *Crossmodal Space and Crossmodal Attention*, Oxford University Press, Oxford, 2003.
- [2] P. Bertelson, J. Vroomen, B. de Gelder, J. Driver, The ventriloquist effect does not depend on the direction of deliberate visual attention, *Percept. Psychophys.* 62 (2000) 321–332.
- [3] S. Berti, E. Schröger, A comparison of auditory and visual distraction effects: behavioural and event-related indices, *Cogn. Brain Res.* 10 (2001) 265–273.
- [4] G.A. Calvert, R. Campbell, M.J. Brammer, Evidence from functional magnetic resonance imaging of crossmodal binding in the heteromodal cortex, *Curr. Biol.* 10 (2000) 649–657.
- [5] C. Colin, M. Radeau, A. Soquet, B. Dachy, P. Deltenre, Electrophysiology of spatial scene analysis: the mismatch negativity (MMN) is sensitive to the ventriloquism illusion, *Clin. Neurophysiol.* 113 (2002) 507–518.
- [6] B. de Gelder, K.B.E. Böcker, J. Tuomainen, M. Hensen, J. Vroomen, The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses, *Neurosci. Lett.* 260 (1999) 133–136.
- [7] G. Gratton, M.G. Coles, E. Donchin, A new method for off-line removal of ocular artifact, *Electroenceph. clin. Neurophys.* 55 (1983) 468–484.
- [8] H. McGurk, J. MacDonald, Hearing lips and seeing voices, *Nature* 383 (1976) 746–748.
- [9] R. Möttönen, C.M. Krause, K. Tiippana, M. Sams, Processing of changes in visual speech in the human auditory cortex, *Cogn. Brain Res.* 13 (2002) 417–425.
- [10] R. Näätänen, *Attention and Brain Function*, Erlbaum, Hillsdale, NJ, 1992.
- [11] P. Paavilainen, M.L. Karlsson, K. Reinikainen, R. Näätänen, Mismatch negativity to change in spatial location of an auditory stimulus, *Electroenceph. clin. Neurophys.* 73 (1989) 129–141.
- [12] T. Ruusuvirta, From spatial acoustic changes to attentive behavioral responses within 200 ms in humans, *Neurosci. Lett.* 275 (1999) 49–52.
- [13] M. Sams, R. Aulanko, M. Hämäläinen, R. Hari, O.V. Lounasmaa, S.-T. Lu, J. Simola, Seeing speech: visual information from lip movements modifies activity in the human auditory cortex, *Neurosci. Lett.* 127 (1991) 141–145.
- [14] E. Schröger, On the detection of auditory deviants: a pre-attentive activation model, *Psychophysiology* 34 (1997) 245–257.
- [15] J. Vroomen, P. Bertelson, B. de Gelder, The ventriloquist effect does not depend on the direction of automatic visual attention, *Percept. Psychophys.* 63 (2001) 651–659.