RESEARCH ARTICLE

Neural correlates of audiovisual motion capture

Jeroen J. Stekelenburg · Jean Vroomen

Abstract Visual motion can affect the perceived direction of auditory motion (i.e., audiovisual motion capture). It is debated, though, whether this effect occurs at perceptual or decisional stages. Here, we examined the neural consequences of audiovisual motion capture using the mismatch negativity (MMN), an event-related brain potential reflecting pre-attentive auditory deviance detection. In an auditory-only condition occasional changes in the direction of a moving sound (deviant) elicited an MMN starting around 150 ms. In an audiovisual condition, auditory standards and deviants were synchronized with a visual stimulus that moved in the same direction as the auditory standards. These audiovisual deviants did not evoke an MMN, indicating that visual motion reduced the perceptual difference between sound motion of standards and deviants. The inhibition of the MMN by visual motion provides evidence that auditory and visual motion signals are integrated at early sensory processing stages.

Keywords Audiovisual motion capture · Multisensory integration · ERP · MMN

Introduction

It is generally acknowledged that human perception is inherently multisensory. Signals from different modalities are effortlessly integrated into coherent multisensory representations. This is evident from cross-modal illusions in which sensory cues in one modality influence the perception of other modalities. One of the best-known examples is the 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 toward the visual one (Bertelson 1999). Visual capture of auditory space has also been found for objects in motion as demonstrated in a illusion called 'dynamic visual capture' in which visual motion can attract the perceived direction of auditory motion (Mateeff et al. 1985; Kitajima and Yamashita 1999; Soto-Faraco et al. 2002, 2004b, 2005; Sanabria et al. 2007). The opposite effect (auditory capture of visual motion) has also been demonstrated (Meyer and Wuerger 2001; Wuerger et al. 2003; Alais and Burr 2004; Meyer et al. 2005). Furthermore, cross-modal dynamic capture has been found for auditory-tactile (Soto-Faraco et al. 2004a) and visual-tactile stimuli (Bensmaïa et al. 2006; Craig 2006).

Despite the fact that the ventriloquist effect is considered to be a perceptual effect (Bertelson 1999; Vroomen et al. 2001; Colin et al. 2002a; Stekelenburg et al. 2004), it still remains to be established at what processing stage audiovisual motion integration occurs. In most studies the effect of visual motion on auditory motion perception has been measured online (i.e., observed in presence of the conflict), but this raises the question whether these immediate effects are a consequence of perceptual integration per se or are due to post-perceptual corrections. Interpretation of immediate cross-modal effects can be problematic because—due to the transparency of the cross-modal conflict situation—participants may adopt specific response strategies to satisfy the demands of the particular laboratory task (de Gelder and Bertelson 2003). Participants may for example occasionally report, despite instructions not to do so, the direction of the to-be ignored visual stimulus rather than the direction of the target sound. If so, then at least part of the visual-capture

J. J. Stekelenburg (✉) · J. Vroomen
Department of Psychology, Tilburg University,
P.O. Box 90153, 5000 LE Tilburg, The Netherlands
e-mail: J.J.Stekelenburg@uvt.nl

phenomenon could be attributed to confusion between target and distractor modality (Vroomen and de Gelder 2003). A number of studies may suggest that motion signals are initially processed independently in the auditory and the visual pathways and are subsequently integrated at a later processing (decisional) stage because—although auditory and visual motion integration induce response biases—there is no increase in sensitivity for motion detection (Meyer and Wuerger 2001; Wuerger et al. 2003; Alais and Burr 2004). In contrast with these models of late multisensory integration, though, studies specifically designed to disentangle perceptual from post-perceptual processes using psychophysical staircases (Soto-Faraco et al. 2005) and adaptation after-effects (Kitagawa and Ichihara 2002; Vroomen and de Gelder 2003) have indicated that auditory and visual motion might be integrated at early sensory levels.

To further explore the processing stage of audiovisual motion integration, we tracked the time-course of dynamic visual motion capture using the mismatch negativity (MMN) component of event-related potentials (ERPs). The MMN signals an infrequent discernible change in an acoustic feature in a sound sequence and reflects pre-attentive auditory deviance detection, most likely generated in the primary and secondary auditory cortex (Näätänen 1992). The generation of the MMN is not volitional; it does not require attentive selection of the sound and is elicited irrespective of the task-relevance of the sounds (Näätänen et al. 1978). The MMN is measured by subtracting the ERP of the standard sound from the deviant one and appears as a negative deflection with a fronto-central maximum peaking around at 150–250 ms from change onset. The MMN has been successfully used to probe the neural mechanisms underlying audiovisual integration. Typically, in these studies audiovisual conflict situations are created such as the ventriloquist effect (Colin et al. 2002a; Stekelenburg et al. 2004) or the McGurk effect (referring to the illusion that observers report to 'hear' /ada/ when presented with auditory /aba/ and visual /aga/) (Sams et al. 1991; Colin et al. 2002b; Möttönen et al. 2002; Saint-Amour et al. 2007; Kislyuk et al. 2008) in which lipread information affects the heard speech sound thereby either evoking or inhibiting the MMN. The modulation of the MMN by audiovisual illusions is taken as evidence that activity in the auditory cortex can be modulated by visual stimuli before 200 ms.

Here, we examined whether the MMN as induced by changes in sound motion can be modulated by visual motion that captures auditory motion. We used a paradigm in which the cross-modal effect renders auditory deviant stimuli to be perceptually identical to the standard stimuli thereby inhibiting MMN. This paradigm has proven to be a valid procedure when applied to the ventriloquist effect (Colin et al. 2002a) and the McGurk effect (Kislyuk et al. 2008). For example, in the case of the ventriloquist effect

(Colin et al. 2002a), a deviant sound with a 20° spatial separation from the centrally presented standard evoked a clear MMN. Crucially, when a centrally presented visual stimulus was synchronized with the sounds no MMN was elicited presumably because the visual stimulus attracted the apparent location of the distant deviant, eliminating the perceived spatial discrepancy between standard and deviant. This paradigm may be preferred to the one in which an illusionary auditory change elicits the MMN (Sams et al. 1991; Colin et al. 2002b; Möttönen et al. 2002; Saint-Amour et al. 2007) in the audiovisual condition because the audiovisual MMN has to be corrected for pure visual effects to isolate the cross-modal effect. As processing of visual changes can be modulated by auditory signals the response recorded in the visual-only odd-ball condition could be a poor estimate of the contribution of visual processing to the MMN recorded in the audiovisual condition (Kislyuk et al. 2008).

Here, we used a 200-ms white noise sound that was cross-faded between two loudspeakers thereby inducing auditory apparent motion from left to right or vice versa. Relevant for the purpose of the current study is that an MMN can be evoked by change in sound motion (Altman et al. 2005). We therefore expected an MMN to an occasional change in auditory motion direction. In the audiovisual condition, this auditory oddball sequence was accompanied by a moving bar that always moved in the same direction as the auditory standard (Fig. 1). We hypothesized that—if audiovisual motion is integrated early (i.e., before the MMN generation process)—the dynamic visual capture of auditory motion of the deviant will result in similar neural codes of the standard and the deviant. As a consequence, we expected no MMN in the AV condition. On the other hand, if dynamic visual capture
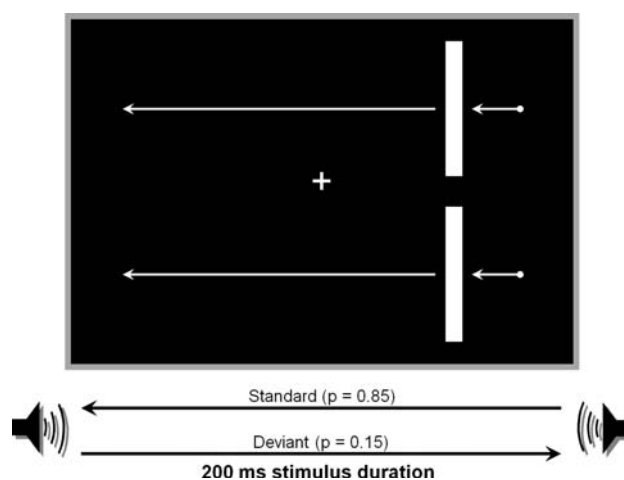

Fig. 1 Experimental set-up of the MMN experiment. In the auditory-only condition standards consisted of *leftward* auditory motion and deviants consisted of *rightward* motion. In the audiovisual condition the auditory standards and deviants were synchronized with a visual stimulus that moved in the same direction as the auditory standards

reflects integration at the decision level, no early integration effects were to be expected and the MMN of the auditory and audiovisual conditions should thus be equal.

Methods

Participants

Fifteen healthy participants (4 males, 11 females) with normal hearing and normal or corrected-to-normal vision participated after giving written informed consent (in accordance with the Declaration of Helsinki). Their age ranged from 18 to 38 years with mean age of 21 years.

Stimuli and procedure

The experiment took place in a dimly-lit, sound-attenuated and electrically shielded room. Visual stimuli were presented on a 19-inch monitor positioned at eye-level, 70 cm from the participant's head. Sounds were delivered from two loudspeakers positioned at the two sides of the monitor with a 59-cm inter-speaker distance. The auditory stimuli were 200-ms white noises. Apparent sound motion was induced by cross-fading a 63 dB(A) white noise of 200 ms (including 5 ms rise-fall times) between the loudspeakers. For leftward-moving sounds, the initial intensity of the sound on the left speaker started at 80% of the original intensity, and then decreased linearly to 20% in 200 ms, while at the same time the intensity of the right speaker started at 20% and then increased linearly to 80%. The opposite arrangement was used for rightward-moving sounds. Visual motion stimuli consisted of two light gray vertically oriented bars (RGB values of 100,100,100; 8 cd/m^2 luminance, against a black background) of 2.3 cm × 12.3 cm (subtending 1.5° × 7.8° visual angle) with a 1-cm separation between them. The bars moved in horizontal direction from one end of the screen (37 cm) to the other end in 200 ms (112°/s). There were two conditions comprising auditory-only and audiovisual stimulus presentations. In both conditions the standard stimulus was a rightward-moving sound (85% probability) and the deviant a leftward-moving sound (15% probability). In the audiovisual condition both the standard and deviant sounds were accompanied by a rightward-moving bar. The inter-stimulus interval was 1,000 ms during which the screen was black. For both auditory-only and audiovisual conditions 1,020 standards and 180 deviants were administered across three identical blocks per condition. Trial order was randomized with the restriction that at least two standards preceded a deviant. The order of the six blocks (three auditory-only, three audiovisual) was varied quasi-randomly across participants. The task for the participants

was to fixate on a light grey central cross (+). To ensure that participants were indeed looking at the monitor during stimulus presentation, they had to detect, by key press, the occasional occurrence of catch trials (3.75% of total number of trials). During a catch trial, the fixation cross-changed from '+' to '×' for 120 ms. Catch trials occurred only for the standards.

ERP recording and analysis

The electroencephalogram (EEG) was recorded at a sampling rate of 512 Hz from 49 locations using 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 active electrode) and ground (Driven Right Leg passive electrode). Two electrodes (FP2 and Oz) were discarded from analysis because of hardware failure. EEG was referenced offline to an average of left and right mastoids and band-pass filtered (1–30 Hz, 24 dB/octave). The raw data were segmented into epochs of 600 ms, including a 100-ms prestimulus baseline. ERPs were time-locked to auditory onset. After EOG correction (Gratton et al. 1983), epochs with an amplitude change exceeding ±100 μV at any EEG channel were rejected. ERPs of the non-catch trials were averaged for standard and deviant, separately for the A-only and AV blocks. MMN for the A-only and AV condition was computed by subtracting the averaged standard ERP from the averaged deviant ERP. MMN was subsequently low-pass filtered (8 Hz, 24 dB/octave). To test for the onset of the MMN, point-by-point two-tailed t tests were performed on the MMN at each electrode in a 1–400 ms window after stimulus onset. Using a procedure to minimize type I errors (Guthrie and Buchwald 1991), the difference wave was considered significant when at least 12 consecutive points (i.e., 32 ms when the signal was resampled at 375 Hz) were significantly different from 0.

Behavioral experiment

To examine whether our stimuli indeed induced visual capture of auditory motion, we also ran a behavioral control experiment in which the same participants judged the direction of an auditory motion stimulus. The same 200-ms white noise and visual moving bars were used as in the ERP experiment. The degree of auditory motion was varied by varying the amount of cross-fading from 90/10% (leftward-motion) to 90/10% (rightward motion) in steps of 10% (leftward 90/10, 80/20, 70/30, 60/40; stationary 50/50; rightward 60/40, 70/30, 80/20, 90/10). All nine auditory stimuli were combined with bars moving either leftward or

rightward. As a base-line, auditory stimuli were presented without visual stimuli (auditory-only condition). Each of the nine auditory motion stimuli was presented 16 times for auditory-only, leftward and rightward visual motion amounting to 432 trials, randomly administered across two identical blocks. The participant's task was to fixate on a central fixation cross and to identify the direction of auditory motion by pressing a left key for leftward motion and a right key for rightward motion. The next trial started after 1 s after the response. A block of 32 trials served as practice.

Results

Behavioral experiment

Figure 2 shows that the proportion of 'rightward'-responses for the auditory-only, leftward, and rightward visual motion conditions. In the auditory-only condition, a typical psychometric curve was found with more 'rightward'-responses with increasing rightward motion. As apparent from Fig. 2, visual motion strongly influenced auditory motion detection. Participants reported more frequently rightward auditory motion when the visual stimulus moved from left-to-right and vice versa less rightward responses with leftward visual motion. To statistically test audiovisual motion capture the mean proportion of right responses were calculated for auditory-only, leftward and rightward visual motion conditions and subjected to a MANOVA for repeated measures with Condition (auditory-only, leftward and rightward visual motion) as the within subject variable. A significant effect of Condition was found, $F(2,13) = 20.31$, $P < 0.001$. Pair-wise post hoc test revealed that the proportion of right responses for rightward visual motion was higher than for A-only and leftward visual motion

while the proportion of right responses for leftward visual motion was lower than for A-only presentations (all P values <0.001).

ERP experiment

For the ERP experiment, the 20/80 (leftward) and 80/20 (rightward) sounds were used. In the behavioral experiment these two sounds were clearly distinguishable in the auditory-only condition (15% vs. 87% rightward responses, a 72% difference). However, this perceptual difference was much smaller (14% vs. 42%, a 28% difference) when the same sounds where combined with visual leftward moving bars, $t(14) = 9.21$, $P < 0.001$. Figure 3 shows that an MMN was generated by deviant sounds in the auditory-only condition but, crucially, there was *no* MMN in the AV condition. The A-only MMN was maximal at the (pre)frontal electrodes and slightly lateralized to the left. The scalp topography of the MMN in the current study is similar to the MMN in other MMN studies and is consistent with neural generators in the supratemporal plane. As apparent from Fig. 3 no MMN was evoked when identical standard and deviant sounds were accompanied by visual bars moving in the direction of the standard. Running t test analysis performed to explore the time-course of the MMN at each electrode demonstrates that the MMN for A-only stimuli differed significantly from baseline in a 150–320 ms post-stimulus interval at the fronto-central electrodes. For the AV presentations no significant MMN was found in this temporal window. To directly compare A-only MMN with AV MMN mean activity in a 170–300 ms interval at electrode Fz (where a robust A-only MMN was found) was tested between conditions. A-only MMN was significantly more negative ($-0.84 \mu V$) than AV MMN ($0.20 \mu V$), $t(14) = 3.76$, $P < 0.01$. Testing MMN against 0 revealed that the difference wave in the A-only condition was significantly

Fig. 2 a Mean proportion of auditory rightward responses as a function of the auditory motion stimulus for the auditory-only (*A-only*) and audiovisual conditions with leftward (*V to left*) and rightward (*V to right*) visual motion. **b** Mean proportion of auditory rightward responses averaged across all levels of the of auditory motion stimulus for the auditory-only and audiovisual conditions
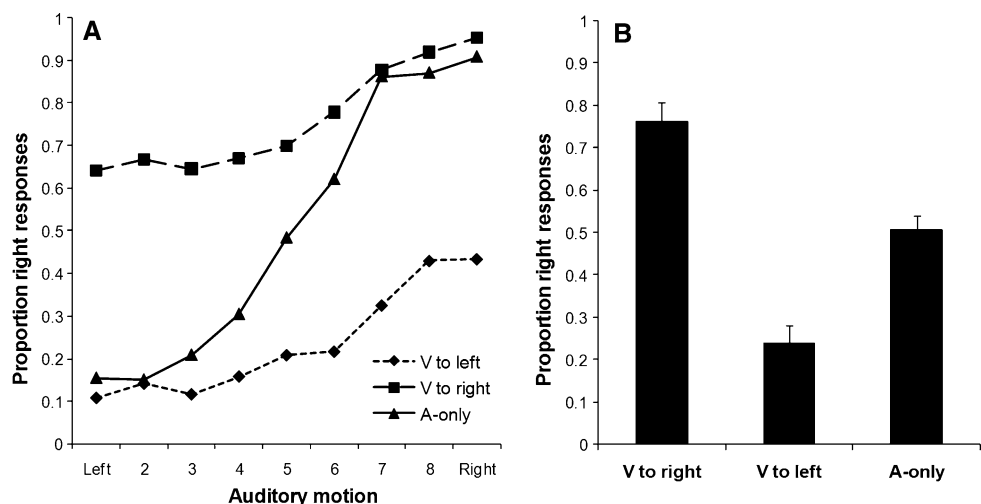
Fig. 3 a Grand-average ERPs recorded at Fz of the standard, the deviant and the difference wave (deviant—standard) of the auditory-only (*A-only*) and audiovisual (*AV*) conditions. **b** Point-wise *t*–tests on the difference wave of the auditory-only and audiovisual conditions at every electrode in a 1–400 ms post-stimulus window. *Shaded areas* indicate significant deviance from 0. **c** The scalp topographies are displayed for the mean activity in a 170–300 ms interval of the difference waves of the auditory-only and audiovisual conditions. The range of the voltage maps in microvolts are displayed below each map
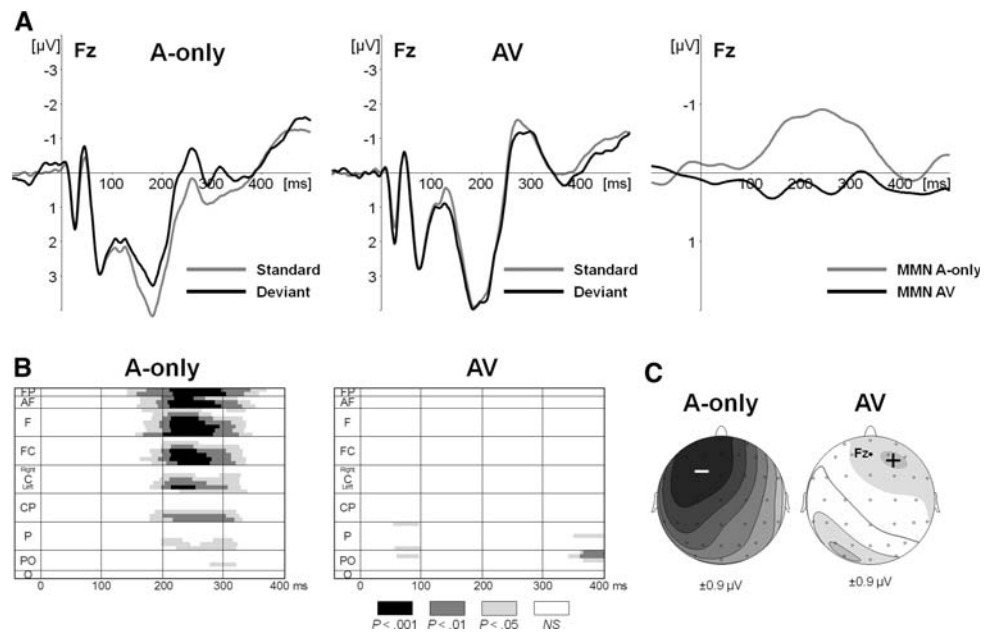
different from 0, $t(14) = 4.85$, $P < 0.001$, whereas in the AV condition the difference wave did not differ from 0 ($t < 1$).

Discussion

The present study shows that both at the behavioral and at the neuronal level visual motion affects auditory motion perception. In line with previous studies on cross-modal motion capture (Mateeff et al. 1985; Kitajima and Yamashita 1999; Soto-Faraco et al. 2002; Sanabria et al. 2007) visual motion biased the subjective reports of auditory motion. Here, we demonstrate that this effect has also neural consequences. The central finding of the current study is that dynamic visual capture was represented at the neural level as a modulation of the auditory MMN. In the A-only condition deviant sound motion evoked a clear MMN starting at approximately 150 ms, which is in line with an earlier MMN study (Altman et al. 2005) showing an MMN to changes of auditory motion direction based on a variable interaural time delay. The MMN to an infrequent change of the direction of auditory motion was inhibited when the sounds were accompanied by visual motion congruent with the auditory motion of the standard. Our behavioral findings support the idea that this cross-modal effect on the neural level was caused by visual capture of auditory motion of the deviant which induced an illusionary auditory motion shift in the same direction as the standard. As a result, the auditory system considered the direction of auditory motion of the deviant not to be different from the standard and no stimulus deviance was therefore detected and no MMN was evoked.

It might be argued, unlikely, that the inhibition of the MMN was not a consequence of the audiovisual fusion per se, but rather that it was instead induced by the mere presentation of the visual stimulus. On this account, moving bars or indeed any other stimulus that attracts attention, would lead to a suppression of auditory deviance detection. There are, though, at least four arguments against this notion. First, although the MMN can be somewhat attenuated when attention is strongly focused on a concurrent *auditory* stimulus stream (e.g., Woldorff et al. 1991), there is no consistent attenuation of MMN (Näätänen et al. 2007) and sometimes even augmentation when *visual* attentional load increases (Zhang et al. 2006). One exception is a MMN study (Yucel et al. 2005) that found a negative perceptual load effect on MMN amplitude, however, only under a high demanding visual task whereas the visual stimuli in our experiment were not task relevant. Moreover, although MMN amplitude in the Yucel et al. (2005) study was diminished a clear MMN was preserved, whereas in our study the MMN was completely abolished. Second, if the suppression of MMN amplitude resulted from synchronized visual events capturing attention one would also expect the performance on the catch trials to be worse in the AV condition than in the A-only condition because of visual capture. However, the percentage of detected catch trials did not differ between conditions (both 97%, $t(14) = 0.5$, $P = 0.63$). Third, MMN studies on the ventriloquist illusion (Colin et al. 2002a) and the McGurk effect (Kislyuk et al. 2008) which used the same experimental paradigm as in the current study also support the notion that the attenuation of MMN does not result from visual distraction per se but is indeed induced by audiovisual illusions. In

these studies, the MMN was attenuated only when auditory and visual signals were expected to be integrated on the deviant trials, whereas deviant audiovisual stimuli that failed to elicit audiovisual illusions (e.g., because of a too large spatial discrepancy between auditory signals (60°) in the case of the ventriloquist illusion (Colin et al. 2002a) or when in the case of the McGurk effect the visual stimulus comprised an ellipse pulsating at the same rhythm as the auditory speech stimuli instead of the talking face (Kislyuk et al. 2008)) did not attenuate the MMN despite the presence of synchronized visual stimulation. The fourth argument is that visual stimuli not only attenuate, but also induce audiovisual illusions thereby *evoking* an MMN. This has been shown in an odd-ball paradigm in which the audio part of the audiovisual standard and deviant are identical whereas the visual part of the deviant is incongruent with the auditory stimulus. The deviant—intended to induce the audiovisual illusion—then evokes an illusory sound change which in turn gave rise to an MMN (Sams et al. 1991; Colin et al. 2002b; Möttönen et al. 2002; Stekelenburg et al. 2004; Saint-Amour et al. 2007). If visual attention would have interfered with auditory deviance detection, no elicitation of the MMN was to be expected in these cases. Taken together, there is thus quite strong evidence favoring the idea that attenuation of the MMN is the consequence of audiovisual integration rather than visual distraction.

What is the neural network underlying visual dynamic capture? Although the present study indicates that visual input modifies activity in the auditory cortex it is still largely unknown how the link (by direct or via higher multisensory areas) between visual and auditory cortex is realized. Visual motion may have affected auditory motion processing via feedforward or lateral connections. Support for direct auditory-visual links comes from electrophysiological studies showing very early (<50 ms) interaction effects (Giard and Peronnet 1999; Molholm et al. 2002). These integration effects occur so early in the time course of sensory processing that purely feedback mediation becomes extremely unlikely (Foxe and Schroeder 2005). Anatomical evidence for early multisensory interactions comes from animal studies showing direct cross-connections between the visual and auditory cortex (Falchier et al. 2002; Smiley et al. 2007). Although processing of co-localized audiovisual stimuli is linked to very early interactions integration of spatially disparate audiovisual stimuli is relatively late. The earliest location-specific audiovisual interactions were found at 140–190 ms (Gondan et al. 2005; Teder-Salejarvi et al. 2005) whereas visual capture of auditory space is associated with even later AV interactions (230–270 ms) (Bonath et al. 2007). These data suggest that visual modulation of the perception of auditory space depends on long latency neural interactions. The same may hold for visual modulation of auditory motion given the relatively long latency of the audiovisual interactions in auditory cortex as reflected in the inhibition of MMN. It should be noted though that because the MMN puts an indirect upper bound on the timing of multisensory integration (Besle et al. 2004), it cannot be determined with the current paradigm when exactly in the pre-MMN window audiovisual integration occurs.

The time-course of audiovisual motion integration suggests that the currently observed inhibition of audiovisual MMN may result from feedback inputs from higher multisensory convergence zones where unisensory signals of multisensory moving objects are initially integrated. These areas may include intraparietal sulcus, the anterior middle fissure, the anterior insula regions (Lewis et al. 2000), and the superior temporal cortex, supra marginal gyrus and the superior parietal lobule (Baumann and Greenlee 2007). In a recent fMRI study on visual dynamic capture neural activity on incongruent audiovisual trials in which the cross-modal capture was experienced was compared to trials in which the illusion was not reported (Alink et al. 2008). In the illusionary trials activation was relatively reduced in auditory motion areas (AMC) and increased in the visual motion area (hMT/V5+), ventral intra parietal sulcus and dorsal intraparietal sulcus. The activation shift between auditory and visual areas was interpreted as representing competition between senses for the final motion percept at an early level of motion processing. In trials in which visual motion capture was experienced vision wins the competition between the senses. The fact that activity of early visual and auditory motion areas are affected by visual dynamic capture suggests that perceptual stage stimulus processing is involved in the integration of audiovisual motion.

How do the current electrophysiological results relate to dynamic visual capture at the behavioral level? Behavioral studies found evidence for perceptual as well as post-perceptual contributions to audiovisual motion integration. We found that a purely sensory ERP component was modified by audiovisual motion integration. This implies that perceptual components are involved in audiovisual motion integration. However, this does not mean that post-perceptual influences may not also play a role in the interactions between auditory and visual motion. Indeed, a behavioral study demonstrated both shifts in response criterion and changes in perceptual sensitivity for detection and classification of audiovisual motion stimuli (Sanabria et al. 2007). The fMRI study of Alink et al. (2008) also provides evidence for the co-existence of both perceptual and decisional components involved in audiovisual motion processing because next to visual and auditory motion areas frontal areas were involved in dynamic visual capture of auditory motion(Alink et al. 2008).

A relevant question is whether such late decisional processes (associated with frontal activity) are involved in the

amplitude modulation of the MMN in the AV condition. We consider this possibility, though, to be unlikely. First, in the study of Alink et al. (2008) frontal activity associated with decisional processes was already present *before* stimulus onset and was linked to the specific task, namely subjects were required to detect auditory motion among coherent or conflicting visual motion. In our experiment, though, the task (detection of a visual transient) was completely irrelevant for the critical aspect of the situation. Second, whenever MMN amplitude is affected by non-sensory factors, typically accessory tasks are involved in which task demands, attention and/or workload are manipulated (Woldorff and Hillyard 1991; Muller-Gass et al. 2005; Yucel et al. 2005; Zhang et al. 2006). In our case, though, the task at hand was very easy and only intended to ensure that subjects watched the screen. Most importantly, the task was identical for the auditory and audiovisual conditions, and task performance was in both conditions identical and virtually flawless (both conditions 97% correct), so no differential task effects modulating the MMN of the AV condition are likely to be involved.

To conclude, the current study investigated at what processing stage visual capture of auditory motion occurs by tracking its time-course using the MMN. We showed that MMN to auditory motion deviance is inhibited by concurrently presented visual motion because of visual capture of auditory motion. Because MMN reflects automatic, pre-attentive signal processing we interpret the inhibition of MMN as providing evidence that auditory and visual motion signals are integrated during the sensory phase of stimulus processing before approximately 200 ms.

References

Alais D, Burr D (2004) No direction-specific bimodal facilitation for audiovisual motion detection. Brain Res Cogn Brain Res 19:185–194

Alink A, Singer W, Muckli L (2008) Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. J Neurosci 28:2690–2697

Altman JA, Vaitulevich SP, Shestopalova LB, Varfolomeev AL (2005) Mismatch negativity evoked by stationary and moving auditory images of different azimuthal positions. Neurosci Lett 384:330–335

Baumann O, Greenlee MW (2007) Neural correlates of coherent audiovisual motion perception. Cereb Cortex 17:1433–1443

Bensmaïa SJ, Killebrew JH, Craig JC (2006) Influence of visual motion on tactile motion perception. J Neurophysiol 96:1625–1637

Bertelson P (1999) Ventriloquism: a case of crossmodal perceptual grouping. In: Aschersleben G, Bachmann T, Müsseler J (eds) Cognitive contributions to the perception of spatial and temporal events. Elsevier, Amsterdam, pp 347–362

Besle J, Fort A, Delpuech C, Giard MH (2004) Bimodal speech: early suppressive visual effects in human auditory cortex. Eur J Neurosci 20:2225–2234

Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, Hillyard SA (2007) Neural basis of the ventriloquist illusion. Curr Biol 17:1697–1703

Colin C, Radeau M, Soquet A, Dachy B, Deltenre P (2002a) Electrophysiology of spatial scene analysis: the mismatch negativity (MMN) is sensitive to the ventriloquism illusion. Clin Neurophysiol 113:507–518

Colin C, Radeau M, Soquet A, Demolin D, Colin F, Deltenre P (2002b) Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. Clin Neurophysiol 113:495–506

Craig JC (2006) Visual motion interferes with tactile motion perception. Perception 35:351–367

de Gelder B, Bertelson P (2003) Multisensory integration, perception and ecological validity. Trends Cogn Sci 7:460–467

Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22:5749–5759

Foxe JJ, Schroeder CE (2005) The case for feedforward multisensory convergence during early cortical processing. Neuroreport 16:419–423

Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11:473–490

Gondan M, Niederhaus B, Rösler F, Röder B (2005) Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. Percept Psychophys 67:713–726

Gratton G, Coles MG, Donchin E (1983) A new method for off-line removal of ocular artifact. Electroencephalogr Clin Neurophysiol 55:468–484

Guthrie D, Buchwald JS (1991) Significance testing of difference potentials. Psychophysiology 28:240–244

Kislyuk DS, Mottonen R, Sams M (2008) Visual processing affects the neural basis of auditory discrimination. J Cogn Neurosci

Kitagawa N, Ichihara S (2002) Hearing visual motion in depth. Nature 416:172–174

Kitajima N, Yamashita Y (1999) Dynamic capture of sound motion by light stimuli moving in three-dimensional space. Percept Mot Skills 89:1139–1158

Lewis JW, Beauchamp MS, DeYoe EA (2000) A comparison of visual and auditory motion processing in human cerebral cortex. Cereb Cortex 10:873–888

Mateeff S, Hohnsbein J, Noack T (1985) Dynamic visual capture: apparent auditory motion induced by a moving visual target. Perception 14:721–727

Meyer GF, Wuerger SM (2001) Cross-modal integration of auditory and visual motion signals. Neuroreport 12:2557–2560

Meyer GF, Wuerger SM, Rohrbein F, Zetzsche C (2005) Low-level integration of auditory and visual motion signals requires spatial co-localisation. Exp Brain Res 166:538–547

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res Cogn Brain Res 14:115–128

Möttönen R, Krause CM, Tiippana K, Sams M (2002) Processing of changes in visual speech in the human auditory cortex. Brain Res Cogn Brain Res 13:417–425

Muller-Gass A, Stelmack RM, Campbell KB (2005) "…and were instructed to read a self-selected book while ignoring the auditory stimuli": the effects of task demands on the mismatch negativity. Clin Neurophysiol 116:2142–2152

Näätänen R (1992) Attention and brain function. Erlbaum, Hillsdale

Näätänen R, Gaillard AW, Mantysalo S (1978) Early selective-attention effect on evoked potential reinterpreted. Acta Psychol 42:313–329

Näätänen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin Neurophysiol 118:2544–2590

Saint-Amour D, De Sanctis P, Molholm S, Ritter W, Foxe JJ (2007) Seeing voices: high-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. Neuropsychologia 45:587–597

Sams M, Aulanko R, Hämäläinen M, Hari R, Lounasmaa OV, Lu ST, Simola J (1991) Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. Neurosci Lett 127:141–145

Sanabria D, Spence C, Soto-Faraco S (2007) Perceptual and decisional contributions to audiovisual interactions in the perception of apparent motion: a signal detection study. Cognition 102:299–310

Smiley JF, Hackett TA, Ulbert I, Karmas G, Lakatos P, Javitt DC, Schroeder CE (2007) Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. J Comp Neurol 502:894–923

Soto-Faraco S, Lyons J, Gazzaniga M, Spence C, Kingstone A (2002) The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. Brain Res Cogn Brain Res 14:139–146

Soto-Faraco S, Spence C, Kingstone A (2004a) Congruency effects between auditory and tactile motion: extending the phenomenon of cross-modal dynamic capture. Cogn Affect Behav Neurosci 4:208–217

Soto-Faraco S, Spence C, Kingstone A (2004b) Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities. J Exp Psychol Hum Percept Perform 30:330–345

Soto-Faraco S, Spence C, Kingstone A (2005) Assessing automaticity in the audiovisual integration of motion. Acta Psychol 118:71–92

Stekelenburg JJ, Vroomen J, de Gelder B (2004) Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. Neurosci Lett 357:163–166

Teder-Salejarvi WA, Di Russo F, McDonald JJ, Hillyard SA (2005) Effects of spatial congruity on audio-visual multimodal integration. J Cogn Neurosci 17:1396–1409

Vroomen J, de Gelder B (2003) Visual motion influences the contingent auditory motion aftereffect. Psychol Sci 14:357–361

Vroomen J, Bertelson P, de Gelder B (2001) Directing spatial attention towards the illusory location of a ventriloquized sound. Acta Psychol 108:21–33

Woldorff MG, Hillyard SA (1991) Modulation of early auditory processing during selective listening to rapidly presented tones. Electroencephalogr Clin Neurophysiol 79:170–191

Woldorff MG, Hackley SA, Hillyard SA (1991) The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. Psychophysiology 28:30–42

Wuerger SM, Hofbauer M, Meyer GF (2003) The integration of auditory and visual motion signals at threshold. Percept Psychophys 65:1188–1196

Yucel G, Petty C, McCarthy G, Belger A (2005) Graded visual attention modulates brain responses evoked by task-irrelevant auditory pitch changes. J Cogn Neurosci 17:1819–1828

Zhang P, Chen X, Yuan P, Zhang D, He S (2006) The effect of visuospatial attentional load on the processing of irrelevant acoustic distractors. Neuroimage 33:715–724