Published in final edited form as:
Brain Res. 2007 May 4; 1144: 127–135.

Visual cues can modulate integration and segregation of objects in auditory scene analysis

Torsten Rahne[a,*], **Martin Böckmann**[a], **Hellmut von Specht**[a], and **Elyse S. Sussman**[b]

[a] *Department of Experimental Audiology and Medical Physics, Otto-von-Guericke-University Magdeburg, Germany*

[b] *Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA*

Abstract

The task of assigning concurrent sounds to different auditory objects is known to depend on temporal and spectral cues. When tones of high and low frequencies are presented in alternation, they can be perceived as a single, integrated melody, or as two parallel, segregated melodic lines, according to the presentation rate and frequency distance between the sounds. At an intermediate distance, in the 'ambiguous' range, both percepts are possible. We conducted an electrophysiological experiment to determine whether an ambiguous sound organization could be modulated toward an integrated or segregated percept by the synchronous presentation of visual cues. Two sets of sounds (one high frequency and one low frequency) were interleaved. To promote integration or segregation, visual stimuli were synchronized to either the within-set frequency pattern or to the across-set intensity pattern. Elicitation of the mismatch negativity (MMN) component of event-related brain potentials was used to index the segregated organization, when no task was performed with the sounds. MMN was elicited only when the visual pattern promoted the segregation of the sounds. The results demonstrate cross-modal effects on auditory object perception in that sound ambiguity was resolved by synchronous presentation of visual stimuli, which promoted either an integrated or segregated perception of the sounds.

Keywords

Auditory perception; Auditory scene analysis; Cross-modal perception; Audiovisual interaction; Mismatch negativity (MMN)

1. Introduction

Perception of coherent objects in the environment often relies on the integration of information from multiple sensory modalities. Cross-modal interactions can enhance perception of objects in the environment by speeding detection (Schröger and Widmann, 1998) and by resolving ambiguities (Watanabe and Shimojo, 2001). Many studies have shown that information from one modality can influence object perception in another modality (Besle et al., 2005;Guttman et al., 2005;King and Calvert, 2001;Remijn et al., 2004;Shams et al., 2000;Vroomen et al., 2001). Visual perception can be influenced by auditory information, such as when the number of light flashes is altered by the number of simultaneous auditory beeps (Shams et al., 2000). Similarly, auditory perception can be influenced by visual information, such as in the misinterpretation of place of articulation (the 'McGurk effect', McGurk and McDonald,

[*] Corresponding author. Universitäts-HNO-Klinik, Abt. für Experimentelle Audiologie, Leipziger Str. 44, D-39120 Magdeburg, Germany. Fax: +49 391 6713888. E-mail address: torsten.rahne@medizin.uni-magdeburg.de (T. Rahne).

1976) or in the misperception of auditory spatial location that is shifted by visual input (the ‘ventriloquist effect’, Bertelson et al., 2000).

The current study focuses on two important issues of cross-modal interaction: (1) whether visual input influences complex auditory processes, namely auditory stream segregation, when the sound input is irrelevant to the direction of focused attention; and (2) the stage of processing reflected in cross-modal interactions of auditory and visual input. To test these questions, we presented sound sequences containing tones of varying frequencies that could be represented as either one or two auditory streams. We investigated the influence of visual information on the neural representation of ambiguous auditory input to determine whether the ambiguity could be resolved by passive viewing of simultaneously presented visual inputs within this complex auditory scene.

The mismatch negativity (MMN) component of event-related potentials (ERPs) is well suited to address questions related to how sounds are organized in memory, the stage of processing influenced by cross-modal interactions, as well as the influence of attention on the representation of sounds in memory. The MMN, which is generated bilaterally within auditory cortices (Alho, 1995;Giard et al., 1990), reflects the output of a deviance detection process. The deviance detection process is based upon the extraction of regularities (‘standard’) within the auditory input, from which detected violations (‘deviants’) elicit MMN as a result of a comparison process (Näätänen et al., 2001;Sussman, 2005). Thus, MMN can be used to determine which regularities (individual features or pattern of sounds) are represented in sensory memory at the time the ‘deviant’ occurs. In this way, MMN can be used to evaluate the representation of the organization of a set of sounds as integrated or segregated (e.g., Müller et al., 2005;Ritter et al., 2006;Sussman et al., 1999,2005;Sussman and Steinschneider, 2006;Yabe et al., 2001).

Any influence of visual input on auditory neural representations would be reflected in the MMN output only if the interactions between sensory modalities occurred prior to or parallel with the cortical processes that elicit MMN (Besle et al., 2005). The auditory regularities that form the basis for the deviance detection process eliciting MMN would already have been influenced by cross-modal interactions for MMN to index changes in the neural representation of the sound input. Evidence of early level interactions among the senses is consistent with findings that multisensory interactions do not depend upon attention (Bertelson et al., 2000;Fort and Giard, 2004;Watkins et al., 2006;Vroomen et al., 2001). Because MMN elicitation does not require participants to actively detect the deviant sounds, it can be used to determine whether the cross-modal influence occurs without direct attentional influence.

To assess cross-modal effects, a set of ambiguously organized high and low tones were presented in two conditions of visual influence: one intended to segregate the sounds to two streams (*Two-streams-segregated* condition) and one intended to integrate the sounds to one stream (*Two-streams-integrated* condition). The logic is that only when the high and low sounds are segregated in the underlying memory representation can the repeating within-stream patterns emerge and deviant patterns be detected, eliciting MMN (Sussman et al., 1999). Thus, if the visual information can passively influence segregation of the auditory input then MMN should be elicited when the visual information induces segregation (*Two-streams-segregated*) but not when it induces integration (*Two-streams-integrated*). On the other hand, if cross-modal influence requires a specific focus of attention, then there should be no difference in MMN elicitation because the set of sounds and the focus of attention both do not differ across the experimental conditions.

2. Results

2.1. Behavioral results

On average, participants responded correctly on 83% of the targets (SD=10%). Average false alarm rate (FAR) for all participants was below 1% (M: 0.14%, SD=0.37%). Mean reaction time to visual red target stimulus was 483 ms (SD=109 ms).

2.2. ERP results

Fig. 1 displays the grand mean ERP waveforms elicited by the deviants and the standards in all conditions separately. The overall amplitude of the ERPs was small due to the rapid stimulus presentation rate. The auditory evoked P1 waveforms (*a*P1) elicited by the standard and deviant tones can be discerned at a latency range of about 50 ms in all conditions except for the *Visual-alone* condition where no auditory P1 is seen. A clear visual evoked N1–P2 waveform (*v*N1–*v*P2) is seen with a peak latency of 50 ms for *v*N1 and 150 ms for *v*P2 in all conditions that presented squares and with a peak latency of 80 ms for *v*N1 and 160 ms for *v*P2 in the condition that presented circles (*Two-streams-integrated* condition).

The amplitude of the grand mean ERPs elicited by the standards and deviants, measured in the latency range of the MMN component, as well as the corresponding differences is reported in Table 1. A three-way repeated measures ANOVA with factors of electrode (Fz, F3, F4, FC1, FC2, Cz, and Pz), condition (all conditions) and stimulus type (standard vs. deviant) revealed a main effect of stimulus type ($F(1,6)=30.15$, $p<0.01$) and an interaction between stimulus type and condition ($F(4,28)=4.55$, $p<0.05$). No other main effects or interactions were found. Post hoc analysis showed that the amplitudes of the ERP's for the deviant stimuli were more negative than for the standard stimuli only in the *One-stream*, *Two-streams-segregated* and *Visual-alone* conditions. Thus, MMN was elicited in the conditions when the visual stimuli corresponded with segregation of the sound input and in the condition with visually presented sequential pattern reversal deviants.

There were no significant differences between deviant and standard stimuli in the *Auditory-alone* and *Two-streams-integrated* conditions ($p>0.05$). No MMN was observed in these conditions.

Grand-averaged difference potentials (see Data processing for description) for all conditions of the experiment are depicted in Fig. 2. A prominent negativity in the *One-stream* and *Two-streams-segregated* conditions is observed at a latency of 140 ms and 145 ms, respectively, at the Fz electrode site, which is identified as the MMN.

In the *One-stream* and *Two-streams-segregated* conditions, in which auditory and visual stimuli were presented and MMN was elicited, a prominent negativity was observed at both the frontal and occipital electrode sites (Fz and Oz, respectively). In comparison, a prominent negativity peaking at 142 ms can be seen in the *Visual-alone* condition at Oz but not at Fz, which may be a *v*MMN.

Additionally, to confirm the presence and absence of the MMN in the various conditions, the mean voltage of the difference potentials was compared with a two-way ANOVA using factors of electrode (Fz, F3, F4, FC1, FC2, Cz, and Pz) and condition (all five conditions). This analysis revealed a main effect of condition ($F(1,4)=28.63$, $p<0.001$), a main effect of electrode ($F(1,6)=6.10$, $p<0.001$), and no interaction ($F(1,24)= 1.0$). Post hoc analyses (Tukey HSD) revealed that the amplitude of the difference potentials was significantly larger in the conditions where MMN was elicited compared to the conditions in which no MMN was observed, and a significantly larger amplitude in the *Two-streams-segregated* condition (where MMN was elicited) than in the *Two-streams-integrated* and *Auditory-alone* conditions (where no MMNs were elicited by the three-tone pattern reversals). There was no significant difference in mean amplitude between the two conditions in which MMN was not observed (*Two-streams-integrated* and *Auditory-alone* conditions). Also the amplitude in the *One-stream* condition (MMN was elicited) was larger than in the *Visual-alone* condition. Furthermore, there was no significant difference in the mean amplitude among the conditions that MMN was elicited or among those in which it wasn't.

Post hoc analyses (Tukey HSD) for the main effect of electrode revealed that the amplitude of the difference potential at Pz was larger than the other electrodes. No significant differences in the mean amplitude among the electrodes other than Pz were found.

Fig. 3 displays the scalp voltage and scalp current density (SCD) maps. Only in the *Two-streams-segregated* and *One-stream* conditions, in which MMN was elicited, bilateral foci at frontal electrode sites F3 and F4 are observed, consistent with bilateral generators of MMN. Additionally, bilateral posterior foci at occipital electrodes O1 and O2 can be seen. In the *Visual-alone* condition only the bilateral occipital foci at electrode sites O1 and O2 are seen, consistent with the visual evoked potentials. In the *Auditory-alone* and *Two-streams-integrated* conditions, in which no MMNs were observed there is a broad distribution of the voltage across the scalp without any main foci. This difference clearly supports the presence of auditory MMN in the segregated conditions and its absence in the integrated and auditory alone conditions.

3. Discussion

The results of this study demonstrate cross-modal effects of visual information on auditory object formation: ambiguous auditory input was modified by synchronized presentation of visual input toward one or two auditory streams. The MMN component, which depends upon the memory representation of sound regularities, was elicited by within-stream pattern deviants only when the visual input was coordinated with the segregated (two-stream) organization of the sound mixture. Thus, we show an interaction between auditory and visual input, occurring prior to the instantiation of the memory representation used in the auditory MMN deviance detection process. These results provide evidence of an early level interaction between visual and auditory modalities, consistent with previous studies (Besle et al., 2005;Giard and Peronnet, 1999;Fort et al., 2002), and which is not dependent upon attention (Bertelson et al., 2000).

The MMN deviance detection process is highly context dependent. The regularities stored in memory provide the auditory context from which deviance detection takes place. Thus, the representation of sounds in memory (e.g., as one or two streams) forms the basis for evaluating incoming sound information (Sussman, 2005;Sussman and Steinschneider, 2006). Deviance detection is determined on the basis of the stored regularities. Modification of the MMN output in this study therefore indicates an effect of visual input on the neural representation of sounds in memory. That is, the memory representation of the sounds had to have already been subject to cross-modal effects in order for differences in MMN elicitation to be observed. Thus, the visual inputs had to bias or influence the organization of the sounds prior to the extraction of the regularities to memory. Given that the composition and presentation rate of the sounds was identical in the conditions with two auditory and one visual stream, the elicitation of MMN in the *Two-streams-segregated* but not in the *Two-streams-integrated* condition indicates that the visual stimulation biased the representation of the sounds toward one or two sound streams, thus changing the regularities detected in the input.

The key finding of this study is that MMN was elicited in the *Two-streams-segregated* but not in the *Two-streams-integrated* condition. This difference is remarkable because the auditory stimuli of these main conditions of the experiment were ambiguously organized. In the *Two-streams-segregated* condition, the visual stream illustrated the within-stream three-tone rising pattern of the low-frequency sounds, whereas in the *Two-streams-integrated* condition, the visual stream illustrated an across-stream, integrated pattern. In the *Two-streams-segregated* condition, the within-stream three-tone repeating regularity could be detected and therefore, the deviant, reversal of the three-tone rising pattern, elicited MMN. In contrast, in the *Two-streams-integrated* condition, the alternation of the high and low tones interfered with the emergence of the low tone pattern. Thus, there was no regularity to serve the basis for deviance detection. MMN was elicited only when the visual input biased the sounds toward the two-stream organization. The results demonstrate a dramatic effect of visual influence on the representation of sound streams in memory.

The control conditions provide further support for the main finding. When the ambiguous sound organization was presented without simultaneous visual input, in the *Auditory-alone* condition, a significant MMN was not observed. One explanation is that the ambiguity of the sound organization led to a flipping back and forth between a one- and two-stream organization (when participants had no task with the sounds). MMNs that would have been elicited 50% of the time when the sounds were segregated would attenuate below significance by the grand mean. An alternative explanation would be that no MMN was elicited in this condition because when the sounds are ambiguously organized, and ignored, the streams do not emerge. For either reason, the results of this control condition demonstrate that the segregation of the sounds and elicitation of a significant MMN in the *Two-streams-segregated* condition was fully induced by the synchronously presented visual stimuli.

Further evidence that the visual input induced segregation is obtained by comparison with the other two control conditions. The MMN elicited in the *One-stream* condition, in which the sounds were unambiguously presented as a single coherent stream of the rising frequency pattern, was comparable in latency and scalp distribution to the MMN obtained in the *Two-streams-segregated* condition. Additionally, the *Visual-alone* condition revealed similar scalp topography in the occipital region to that in the *Two-streams-segregated* condition but not to that in the *Two-streams-integrated* condition. It is also interesting to note the significant difference between ERPs elicited by the standard and deviant patterns in this *Visual-alone* condition because it suggests the presence of a visual MMN (vMMN, Czigler et al., 2006;Maekawa et al., 2005;Tales et al., 1999). A vMMN may have been evoked by the reversal of the rising size of the three-stimulus visual pattern.

The current data are consistent with the results of Besle et al. (2005), who found that the scalp distribution of the MMNs elicited by audio-visual deviants was contributed by both supratemporal and occipital areas. Our results demonstrate that the auditory memory representation used in the MMN deviance detection process was altered by the visual input. The bimodal interaction occurred prior to the MMN process. Similar visual evoked activity was observed when MMN was present, either when the visual stimuli were presented alone or when synchronized with auditory stimulation. This suggests that there may be both integrated and independent contributions from both modalities that give rise to the unified perceptions of synchronized cross-modal input.

4. Experimental procedure

4.1. Subjects

Ten right-handed healthy adults (24–30 years, 4 males) with normal hearing were paid for their participation in the experiment. Participants gave written informed consent in accordance with the guidelines of the Internal Review Board of the Albert Einstein College of Medicine after the procedures were explained to them. The study conforms to The Code of Ethics of the World Medical Association (Declaration of Helsinki). The data from one of the participants were excluded from analysis due to an inability to perform the task and the data from another participant were excluded due to excessive electrical artifact.

4.2. Stimuli

Auditory stimuli—Six pure tones (50 ms duration, 7.5 ms rise/fall time), occupying two distinct frequency ranges, were created with Neuroscan STIM software and presented binaurally through insert earphones (EAR-tone 3A™). The low-frequency range included three tones (L1: 830, L2: 880, and L3: 932 Hz) and the high-frequency range included three tones (H1:1174, H2:1244, and H3:1318 Hz). Within each range, the tones were separated by one semitone (a frequency ratio of about 1.06). Frequency separation between the highest, low-tone and the lowest, high-tone was four semitones. At a four semitone frequency separation between tone ranges, listeners perceive one integrated stream of sounds 50% of the time and two segregated streams 50% of the time (Bregman, 1999;Sussman et al., 2007). In other words, the sound stream organization was ambiguous.

The high and low tones were alternated at a constant stimulus onset asynchrony (SOA) of 110 ms. One set of tones (e.g., the low tones) consisted of a frequently occurring three-tone rising pattern (L1–L2–L3, called the ‘standard’) and an infrequently (15%), randomly occurring three-tone falling pattern (L3–L2–L1, called the ‘deviant’). The other set of tones (e.g., the high tones) were presented equiprobably (H1 (33.3%), H2 (33.3%), and H3 (33.3%)) and randomly within the set of high tones in the sequence. The intensity value of the tones occurred in a three-tone intensity pattern (70–70–85 dB SPL) on successive tones in the sequence, across frequency range (see Fig. 4).

Visual stimuli—Five visual stimuli were created with Neuroscan STIM software, two circles (C) and three squares (S). The two circle stimuli were white unfilled circles presented on a black background; one was large (8 cm in diameter) and the other was smaller (3 cm in diameter). The three squares were three different sizes from smaller to larger (edge length: 2.5 cm, 5.5 cm, and 10 cm). The square stimuli were presented as white unfilled squares against a black background (see Fig. 4, ‘visual’ panel). Visual stimuli were presented on a monitor with a subtended visual angle of 1–4°.

4.3. Conditions

The experiment consisted of five conditions (*Two-streams-segregated*, *Two-streams-integrated*, *One-stream*, *Auditory-alone*, and *Visual-alone*). In three conditions, auditory and visual stimuli were presented simultaneous and in the other two conditions, stimuli were presented unimodally (see Table 2). The unimodal conditions and the *One-stream* condition served as controls for comparison. Participants were instructed to ignore the auditory stimuli in all of the conditions. In each condition containing visual stimuli, participants were instructed to attend to the visual stimuli and press the response key every time they detected a red shape, which occurred randomly and rarely (1% of the visual stimuli were red). The visual task was unrelated to the organization of the sounds and nothing was told to participants about the organization of the stimuli.

In the *Two-streams-segregated* condition, the sounds were presented as described above, with a square stimulus presented synchronously with each low-frequency (or high-frequency) tone. The rising and falling frequency patterns of the three-tone sequences were represented by increasing or decreasing the size of the square corresponding with the rising or the falling frequency of the tones. This was intended to induce segregation between the low and high tones by defining the rising within-stream tone pattern (see Fig. 4A).

In the *Two-streams-integrated* condition the auditory stimuli were the same as in the *Two-streams-segregated* condition except that the visual stimuli were intended to induce *integration* of the high and low tones to a single stream of alternating sounds. To do this, the visual circle stimuli were presented synchronously with the higher-intensity (85 dB) tones, which crossed frequency ranges. The size of the circles corresponded to the frequency of these higher-intensity tones. The smaller circles were synchronously presented with the low-frequency tones and the larger circles were synchronized with the high-frequency tones (see Fig. 4B). Thus, the across-stream intensity pattern was demarcated by the visual stimuli to induce integration of the acoustic information into one stream.

In the *One-stream* condition only the low-frequency tones were presented synchronously with the squares. This condition served as a control condition to obtain ERP responses to the low tones as a single stream presented at the same rate as occurred within the *Two-streams-segregated* condition.

In the *Auditory-alone* condition, the auditory stimuli (high-and low-frequency tones) were presented without the visual stimuli (Fig. 4, ‘auditory’ panel). The purpose for this condition was to provide a control to determine whether or not one of the two organizations would be represented while subjects ignored the sounds, with no influence from the visual stimuli. Participants were instructed to watch a silent, captioned video during presentation of the sounds.

In the *Visual-alone* condition, the visual stream shaped by the squares (see Fig. 4, top ‘visual’ panel) was presented alone. The purpose for this condition was to obtain ERPs elicited by the visual stimuli when they were presented without the simultaneous auditory stimuli, at the same stimulus presentation rate as within the *Two-streams-segregated* condition.

Prior to the ERP experiment, we asked four individuals who did not participate in the ERP experiment to tell whether they heard one or two sound streams when we presented them samples of the *Two-streams-segregated* and *Two-streams-integrated* conditions. Pilot subjects reported hearing two streams when the *Two-streams-segregated* condition was presented and reported hearing one stream when the *Two-streams-integrated* condition was presented.

4.4. Procedures

The *Auditory-alone* condition was presented first to all participants, who were comfortably seated in an acoustically dampened room, to obtain the ERP responses before the possibility of the visual input influencing the sound processing. The remaining conditions were presented in a randomized order, counterbalanced across participants. For half of the participants, the low tones comprised the three-tone rising frequency pattern and for the other half of the participants the high tones comprised the three-tone rising frequency pattern. Each condition was presented in four consecutive blocks of 4 min each, yielding 55 deviants in each, with a total of 220 deviants per condition. Including electrode placement and rest breaks between conditions, the session took about 3 h.

4.5. Electroencephalogram (EEG) recording

EEG was continuously recorded with a Neuroscan Synamps AC coupled amplifiers (0.05–200 Hz bandwidth; sampling rate: 1 kHz) using a 32-channel electrode cap (Electrocap Co.) placed according to the International 10–20 System (Fpz, Fz, Cz, Pz, Oz, Fp1, Fp2, F7, F8, F3, F4, Fc5, Fc6, Fc1, Fc2, T7, T8, C3, C4, Cp5, Cp6, Cp1, Cp2, P7, P8, P3, P4, O1, O2) plus the left and right mastoids, and referred to the nose. Impedances were kept below 5 kΩ. Vertical electrooculogram (VEOG) was recorded with a bipolar electrode configuration using Fp2 and an external electrode placed below the right eye. Horizontal electrooculogram (HEOG) was recorded with the F7 and F8 electrodes.

4.6. Data processing

The EEG at Fz, Cz, F3, F4, and the EOG channel were examined automatically for artifacts and rejected if the voltage step per sample point was larger than 20 μV or the absolute voltage difference in an interval of 100 ms was larger than 100 μV and then digitally filtered (1–15 Hz bandpass, 24 dB/oct. rolloff). A second artifact criterion was set at ±50 μV using the same set of electrodes, after the epochs were baseline corrected on a prestimulus time range of 100 ms. The EEG was epoched relative to reference marker positions according to the stimulus type (standard and deviant), separately for each condition. Epochs were 600 ms, starting from 100 ms before and ending 500 ms after the onset of the stimulus. The 'standard' ERP was created as an average of the first tone of the frequent (rising) three-tone pattern (L1), and the 'deviant' ERP was created as an average of the first tone of the infrequent (falling) pattern reversals (L3) (Sussman et al., 1998). Difference waveforms were obtained by subtracting the ERPs elicited by the standard stimuli from the ERPs elicited by the deviant stimuli. Two 'standard' ERPs elicited by the three-tone rising patterns were excluded when they immediately followed a 'deviant' three-tone falling pattern.

To statistically measure the MMN, the peak latency was determined from the grand-mean difference waveform at the Fz electrode site as the maximum voltage difference potential between 100 and 200 ms after stimulus onset for the *One-stream* and *Two-streams-segregated* conditions in which MMNs were observed. A 40-ms interval centered on the peak was used to obtain mean voltages of the standard and deviant ERPs for every subject, separately in each condition. In the conditions that no MMN was observed (*Two-streams-integrated*, *Auditory-alone*, and the *Visual-alone* conditions), the 40-ms interval was chosen from the peak latency of the MMN elicited in the *Two-streams-segregated* condition.

Presence of the MMN was determined using a three-way analysis of variance for repeated measures (ANOVA) with the factors of electrode (Fz, F3, F4, FC1, FC2, Cz, and Pz)×stimulus type (standard and deviant) × condition (all five conditions). Comparison of the mean voltage of the difference potentials was made using a two-way ANOVA with factors of electrode site (Fz, Cz, Pz, Oz) × condition. Greenhouse–Geisser corrections for sphericity were applied as appropriate and the *p* values were reported.

A reference-free measure of the scalp current density (SCD) was also used to depict the MMN. Maps showing scalp voltage topography and SCD were computed on the mean amplitude difference waveforms for each condition, corresponding to the peak latency of the MMN (142 ms). An estimate of the second spatial derivative of the voltage potential (the Laplacian) was performed using BESA2000 software. This SCD analysis of topography sharpens the differences in the scalp fields, providing better information about the cortical generators and about the occurrence of components in each hemisphere.

Behavioral responses were measured by calculating hit rate and FAR to the visual target stimuli. Responses were considered correct if they occurred within an interval of 20–1000 ms from the onset of the target stimulus.

Acknowledgements

We thank the Magdeburger Forschungsverbund "Neurowissenschaften" for providing funding to collect data during a research stay of the author in the Cognitive Neuroscience Laboratory of the Albert Einstein College of Medicine in New York, USA. This research was supported by the National Institutes of Health (DC 004263).

References

Alho K. Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. Ear Hear 1995;16:38–51. [PubMed: 7774768]

Bertelson P, Vroomen J, de Gelder B, Driver J. The ventriloquist effect does not depend on the direction of deliberate visual attention. Percept Psychophys 2000;62:321–332. [PubMed: 10723211]

Besle J, Fort A, Giard MH. Is the auditory sensory memory sensitive to visual information? Exp Brain Res 2005;166:337–344. [PubMed: 16041497]

Bregman, AS. Auditory Scene Analysis: The Perceptual Organisation of Sounds. The MIT Press: Cambridge, Massachusetts; 1999.

Czigler I, Weisz J, Winkler I. ERPs and deviance detection: visual mismatch negativity to repeated visual stimuli. Neurosci Lett 2006;401:178–182. [PubMed: 16600495]

Fort, A.; Giard, MH. Multiple electrophysiological mechanisms of audio-visual integration in human perception. In: Calvert, G.; Spence, C.; Stein, B., editors. The Handbook of Multi-sensory Processing. MIT press; Cambridge: 2004.

Fort A, Delpuech C, Pernier J, Giard MH. Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. Cereb Cortex 2002;12:1031–1038. [PubMed: 12217966]

Giard MH, Peronnet F. Auditory-visual integration during multimodal object recognition in humans: a behavioural and electrophysiological study. J Cogn Neurosci 1999;11:473–490. [PubMed: 10511637]

Giard MH, Perrin F, Pernier J, Bouchet P. Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. Psychophys 1990;27:627–640.

Guttman SE, Gilroy LA, Blake R. Hearing what the eyes see: auditory encoding of visual temporal sequences. Psychol Sci 2005;16:228–235. [PubMed: 15733204]

King AJ, Calvert GA. Multisensory integration: perceptual grouping by eye and ear. Curr Biol 2001:11.

Maekawa T, Goto Y, Kinukawa N, Tanikawa T, Kanba S, Tobimatsu S. Functional characterization of mismatch negativity to a visual stimulus. Clin Neurophysiol 2005;116:2392–2402. [PubMed: 16122977]

McGurk H, McDonald J. Hearing lips and seeing voices. Nature 1976;264:746–748. [PubMed: 1012311]

Müller D, Widmann A, Schröger E. Auditory streaming affects the processing of successive deviant and standard sounds. Psychophys 2005;42:668–676.

Näätänen R, Tervaniemi M, Sussman ES, Paavilainen P, Winkler I. "Primitive intelligence" in the auditory cortex. Trends Neurosci 2001;24:283–288. [PubMed: 11311381]

Remijn GB, Ito H, Nakajima Y. Audiovisual integration: an investigation of the 'streaming-bouncing' phenomenom J. Physiol Anthropol Appl Hum Sci 2004;23:243–247.

Ritter W, De Sanctis P, Molholm S, Javitt DC, Foxe JJ. Preattentively grouped tones do not elicit MMN with respect to each other. Psychophys 2006;43:423–430.

Schröger E, Widmann A. Speeded responses to audiovisual signal changes result from bimodal integration. Psychophys 1998;35:755–759.

Shams L, Kamitani Y, Shimojo S. Illusions: what you see is what you hear. Nature 2000;408:788. [PubMed: 11130706]

Sussman ES. Integration and segregation in auditory scene analysis. J Acoust Soc Am 2005;117:1285–1298. [PubMed: 15807017]

Sussman E, Steinschneider M. Neurophysiological evidence for context-dependent encoding of sensory input in human auditory cortex. Brain Res 2006;1075:165–174. [PubMed: 16460703]

Sussman ES, Ritter W, Vaughan HG Jr. Attention affects the organisation of auditory input associated with the mismatch negativity system. Brain Res 1998;789:130–138. [PubMed: 9602095]

Sussman ES, Ritter W, Vaughan HG Jr. An investigation of the auditory streaming effect using event-related brain potentials. Psychophysiology 1999;36:22–34. [PubMed: 10098377]

Sussman ES, Bregman AS, Wang WJ, Khan FJ. Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. Cogn Affect Behav Neurosci 2005;5:93–110. [PubMed: 15913011]

Sussman E, Wong R, Horváth J, Winkler I, Wang W. The development of the perceptual organization of sound by frequency separation in 5–11 year-old children. Hear Res. in pressAvailable online 20 January 2007

Tales A, Newton P, Troscianko T, Butler S. Mismatch negativity in the visual modality. NeuroReport 1999;10:3363–3367. [PubMed: 10599846]

Vroomen J, Bertelson P, de Gelder B. The ventriloquist effect does not depend on the direction of automatic visual attention. Percept Psychophys 2001;63:651–659. [PubMed: 11436735]

Watanabe K, Shimojo S. When sound affects vision: effects of auditory grouping on visual motion perception. Psychol Sci 2001;12:109–116. [PubMed: 11340918]

Watkins S, Shams L, Tanaka S, Haynes JD, Rees G. Sound alters activity in human V1 in association with illusory visual perception. NeuroImage 2006;31:1247–1256. [PubMed: 16556505]

Yabe H, Winkler I, Czigler I, Koyama S, Kakigi R, Sutoh T, Hiruma T, Kaneko S. Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. Brain Res 2001;897:222–227. [PubMed: 11282382]

Abbreviations

ANOVA	analysis of variance
dB	decibel
EEG	electroencephalogram
ERP	event-related potential
Hz	Hertz
MMN	mismatch negativity
SCD	scalp current density
SD	standard deviation
SOA	stimulus onset asynchrony

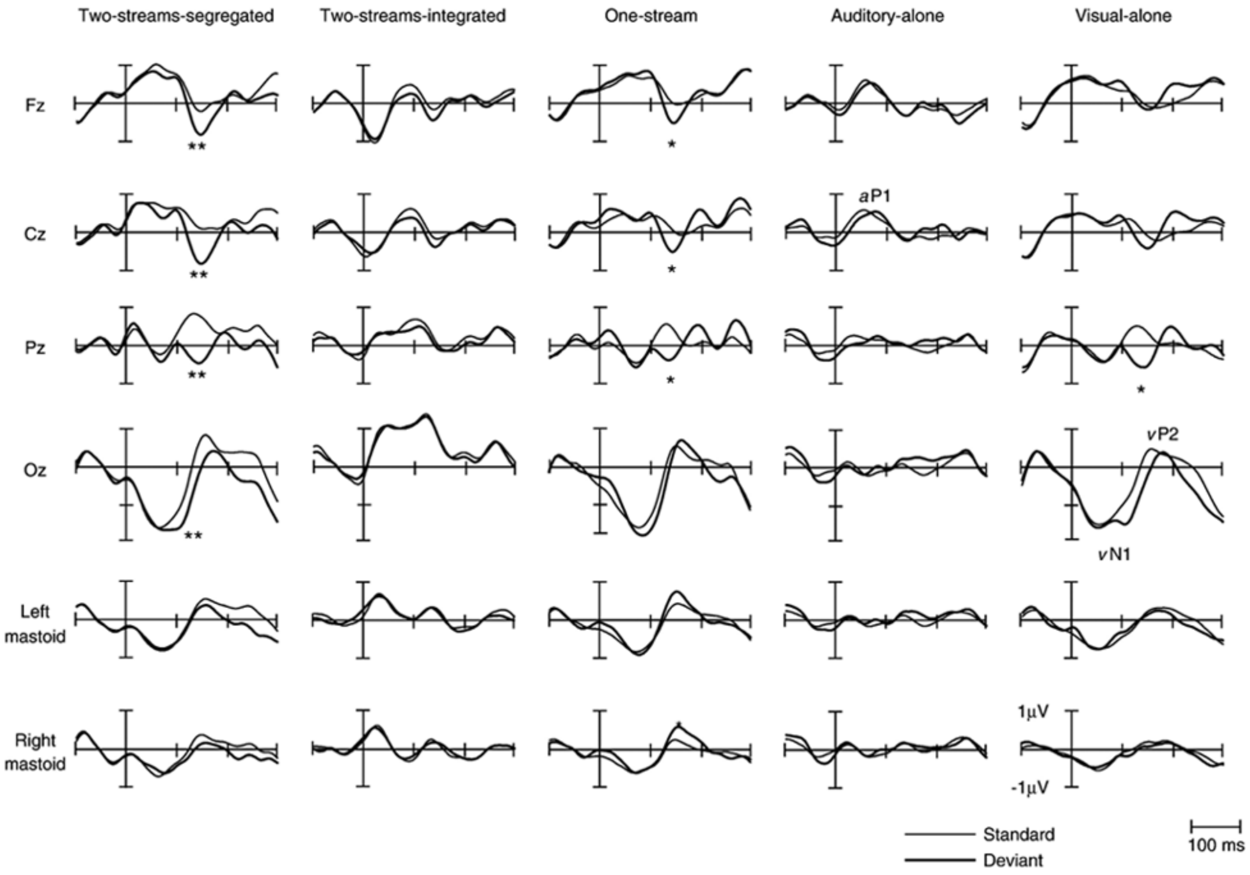

Fig 1.

Grand-averaged ERP's elicited by the first stimulus of the three-tone standard (thin line) and deviant (thick line) sequences are displayed for all five conditions at midline electrodes and mastoids. Auditory P1 waveform (*a*P1) is labeled in the *Auditory-alone* condition at Cz. The visual N1–P2 waveforms (*v*N1–*v*P2) are labeled at Oz in the *Visual-alone* condition. Significant differences between standard and deviant stimuli, denoting the MMN, are also shown (*p<0.05, **p<0.01).

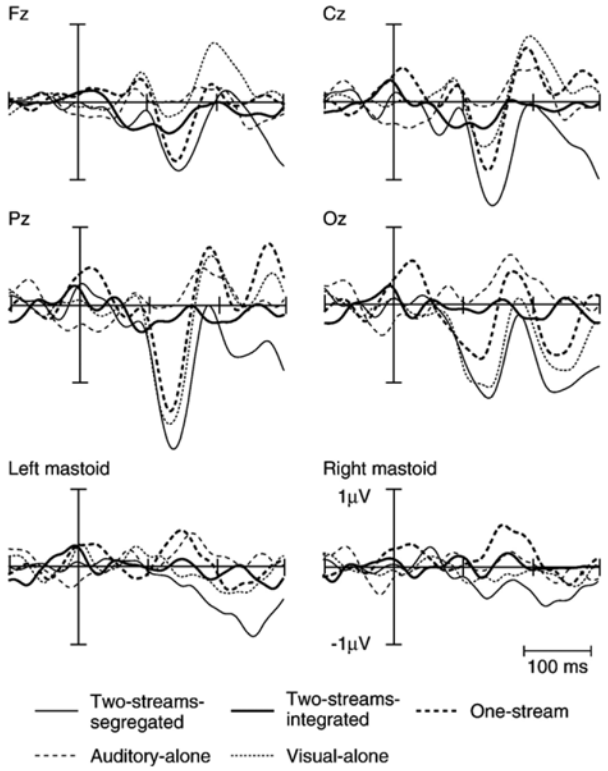

Fig 2.

Grand-averaged difference potentials are displayed for all conditions *Two-streams-segregated* (thin line), *Two-streams-integrated* (thick line), *One-stream* (thick dotted), *Visual-alone* (thin dotted, small space) and *Auditory-alone* (thin dotted, large space) at the midline electrodes and the mastoids. The prominent negative waveform at the Fz electrode site, peaking at about 140 ms, is identified as the MMN. An inversion of polarity at the mastoids in the *One-stream* condition is discernable. A prominent negative waveform at the Oz electrode in the *Visual-alone* condition is likely a *v*MMN.

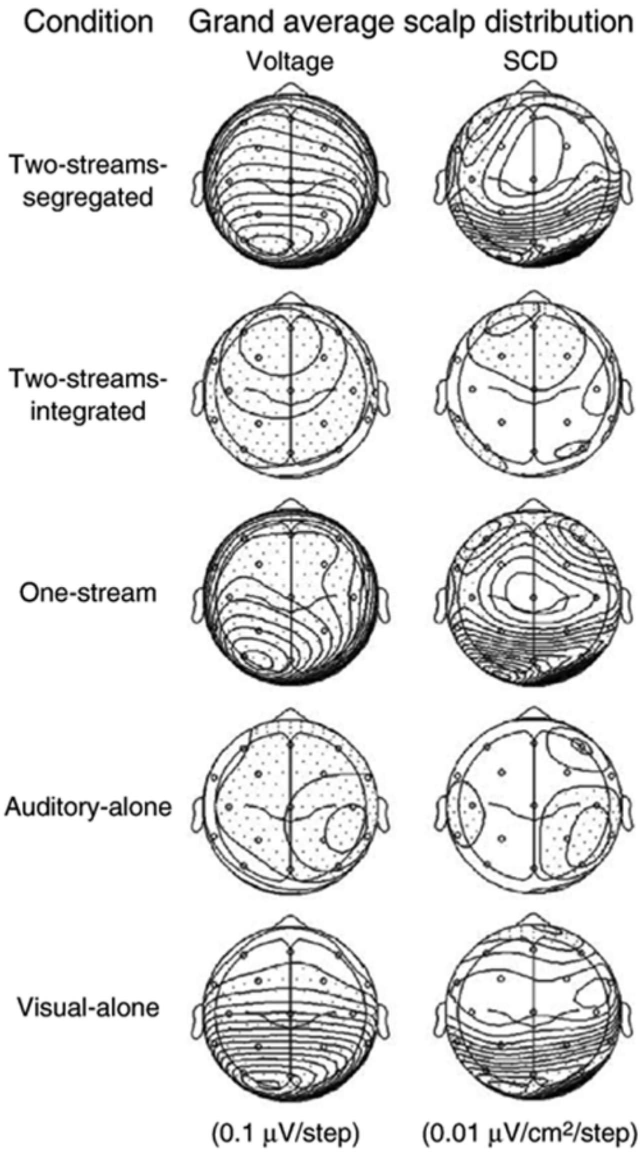

Fig 3.

Scalp voltage distribution (left) and source current density (SCD) maps (right) are displayed for all conditions of the experiment (top view) calculated on the grand-mean difference potential at a latency of 142 ms. Dotted areas indicate negative values. *Two-streams-segregated* and *One-stream* conditions show bilateral foci at F3 and F4 conditions (consistent with bilateral generators of the MMN) as well as a unilateral focus at P3. In the *Visual-alone* condition, only posterior sinks are discernable.

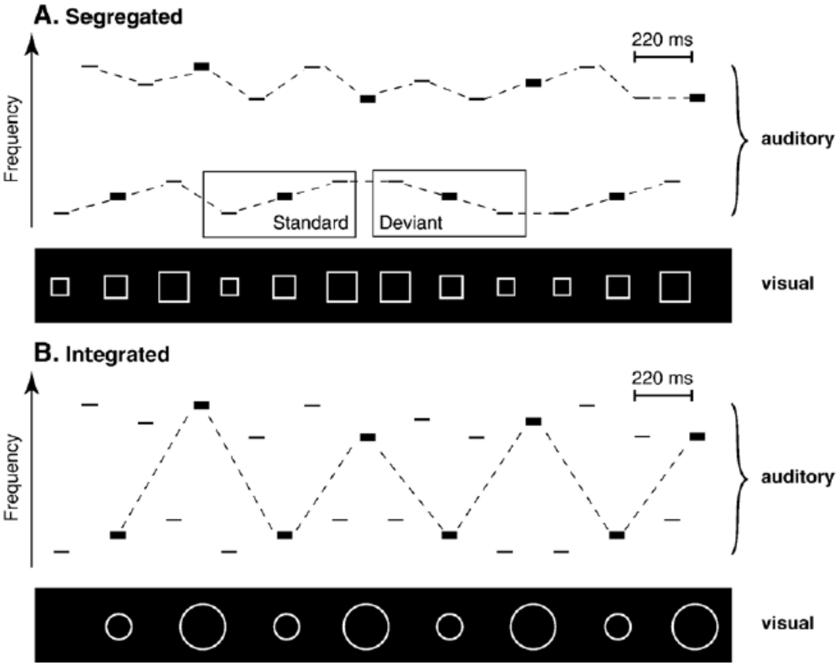

Fig 4.

Stimulus paradigm. The rectangles represent tones. The width of the rectangles denotes stimulus intensity (bold: 85 dB SPL, thin: 70 dB SPL). The axes represent time (abscissa) and frequency (ordinate). The auditory panels of the schematic show a rising three-tone frequency pattern within the low set of tones, alternating with a random presentation of high tones (see text, Section 4.3 for further description). The visual panel shows the sequential presentation of square stimuli, synchronized with the rising frequency pattern (A). The standard (rising sequence) and deviant (falling sequence) of the low sequence are pointed out in the boxes. The presentation of the circle stimuli is synchronized with the 85 dB tones, which cross frequency ranges (B). The dashed lines indicate the expected organization induced by the visual stimuli (i.e., frequency pattern when segregated and intensity pattern when integrated).

Table 1

Mean (M) amplitude and standard deviation (SD) in μV of the standard and deviant ERP waveforms and the difference potential in the latency region used to measure MMN for the five conditions

Two-streams-segregated							**Two-streams-integrated**						
	Standard		**Deviant**		**Difference**			**Standard**		**Deviant**		**Difference**	
	M	**SD**	**M**	**SD**	**M**	**SD**		**M**	**SD**	**M**	**SD**	**M**	**SD**
Fz	−0.04	0.75	−1.17	0.74	−1.13	0.67**	Fz	−0.19	0.96	−0.53	0.98	−0.34	0.55
Cz	0.29	0.71	−1.18	0.71	−1.47	0.95**	Cz	−0.10	0.73	−0.40	1.08	−0.29	0.56
Pz	0.99	0.88	−0.72	0.78	−1.71	1.05**	Pz	0.28	1.01	−0.03	0.71	−0.32	0.62
Oz	0.36	1.24	−0.16	1.18	−0.52	1.49	Oz	1.39	2.25	1.19	1.72	−0.20	0.65
LM	0.38	0.68	0.37	0.91	−0.01	0.49	LM	0.39	0.89	0.39	0.66	0.00	0.71
RM	0.32	0.63	0.17	0.66	−0.15	0.71	RM	0.29	0.88	0.24	0.47	−0.06	0.54
One Stream							Auditory-alone						
Fz	0.42	0.75	−0.65	1.21	−1.07	0.99*	Fz	−0.22	0.68	−0.11	0.78	0.11	0.61
Cz	0.24	0.71	−0.72	1.06	−0.96	0.78*	Cz	−0.13	0.67	−0.01	0.94	0.12	1.05
Pz	0.40	0.88	−0.65	0.92	−1.06	0.91*	Pz	−0.05	0.46	0.15	0.80	0.20	1.21
Oz	−0.08	1.65	0.05	1.19	0.13	1.26	Oz	−0.12	0.57	0.35	0.73	0.47	1.13
LM	0.11	0.83	0.57	0.84	0.46	0.60	LM	0.02	0.38	0.21	0.57	0.19	0.58
RM	−0.05	0.75	0.43	0.82	0.49	0.55*	RM	0.07	0.45	0.19	0.74	0.12	0.77
Visual-alone													
Fz	0.08	0.82	−0.24	0.75	−0.31	0.80							
Cz	0.03	0.78	−0.50	1.09	−0.53	0.97							
Pz	0.45	0.96	−0.65	1.36	−1.10	0.97*							
Oz	−0.08	1.33	−0.32	1.31	−0.24	1.74							
LM	0.02	0.41	0.23	0.56	0.20	0.55							
RM	0.17	0.54	0.17	0.74	0.00	0.58							

One-sample, two-tailed Student's *t*-test.

[*] *p*<0.05.

[**] *p*<0.01.

Table 2

Experimental conditions

Condition	Auditory streams		Visual streams		Video film
	Low (three-tone pattern)	High (random pitch)	Squares	Circles	
Two-streams-segregated	x	x	x		
Two-streams-integrated	x	x		x	
One-stream	x		x		
Auditory-alone	x	x			x
Visual-alone			x		