The Representation of Audiovisual Regularities in the Human Brain

Julien Besle\textsuperscript{1,2,3}, Zahra Hussain\textsuperscript{3}, Marie-Hélène Giard\textsuperscript{1,2}, and Olivier Bertrand\textsuperscript{1,2}

Abstract

Neural representation of auditory regularities can be probed using the MMN, a component of ERPs generated in the auditory cortex by any violation of that regularity. Although several studies have shown that visual information can influence or even trigger an MMN by altering an acoustic regularity, it is not known whether audiovisual regularities are encoded in the auditory representation supporting MMN generation. We compared the MMNs elicited by the auditory violation of (a) an auditory regularity (a succession of identical standard sounds), (b) an audiovisual regularity (a succession of identical audiovisual stimuli), and (c) an auditory regularity accompanied by variable visual stimuli. In all three conditions, the physical difference between the standard and the deviant sound was identical. We found that the MMN triggered by the same auditory deviance was larger for audiovisual regularities than for auditory-only regularities or for auditory regularities paired with variable visual stimuli, suggesting that the visual regularity influenced the representation of the auditory regularity. This result provides evidence for the encoding of audiovisual regularities in the human brain.

INTRODUCTION

There is evidence for widespread multisensory processing within the human brain: Cortical regions devoted to a particular sensory modality can be activated or have their activity modulated by information from another modality (reviews in Bulkin & Groh, 2006; Ghazanfar & Schroeder, 2006). In the auditory domain, electrophysiological studies have shown the influence of visual stimuli manifesting early in auditory processing (Besle et al., 2008; Besle, Fort, Delpuech, & Giard, 2004; Giard & Peronnet, 1999).

What is the impact of early audiovisual interactions on the brain’s representation of the acoustic environment? The neural traces of auditory regularities in the auditory system can be probed using the MMN, an evoked potential component generated in the auditory cortex when an incoming sound violates a perceived auditory regularity (review in Näätänen, Paavilainen, Rinne, & Alho, 2007; Sussman, 2007). Furthermore, the MMN can be elicited by the violation of regularities beyond simple tone repetition, for instance, by violations of abstract rules such as in pitch relations (e.g., Korzyukov, Winkler, Gumenyk, & Alho, 2003; Saarinen, Paavilainen, Schoger, Tervaniemi, & Näätänen, 1992) and feature conjunctions (e.g., Takegata et al., 2005), even when the regularity is too abstract and complex for participants to report (Paavilainen, Arajarvi, & Takegata, 2007). Because visual information can have an early influence on auditory processing, it is likely that the neural representation of audiovisual events underlying MMN generation will include visual information.

Several previous studies have shown that visual deviance can influence or trigger an auditory MMN. Behaviorally, incongruent visual information can modify the perception of an auditory input (e.g., in the ventriloquist illusion: Thomas, 1941, or the McGurk illusion: McGurk & McDonald, 1976). When these types of incongruent stimuli are embedded in a stream of congruent audiovisual standard stimuli, they can trigger an auditory MMN even though there is no acoustic difference between the standard and deviant auditory stimuli (McGurk effect: Colin, Radeau, Soquet, & Deltenre, 2004; Colin, Radeau, Soquet, Dachy, & Deltenre, 2002; Colin, Radeau, Soquet, Demolin, et al., 2002; Möttönen, Krause, Tiippana, & Sams, 2002; Sams et al., 1991; ventriloquist illusion: Hertrich, Mathiak, Lutzenberger, Menning, & Ackermann, 2007; Saint-Amour, De Sanctis, Molholm, Ritter, & Foxe, 2007; Stekelenburg, Vroomen, & de Gelder, 2004). MMN-like components have also been found when visual cues provide conflicting information about an upcoming standard auditory stimulus (Aoyama, Endo, Honda, & Takeda, 2006; Ullsperger, Erdmann, Freude, & Dehoff, 2006; Yumoto et al., 2005; Widmann, Kujala, Tervaniemi, Kujala, & Schröger, 2004). As well, using synchronous auditory and visual stimuli, we have shown that an audiovisual
event deviating on both auditory and visual dimensions elicits an MMN that is different from the sum of MMNs elicited by unimodal auditory and visual deviances (Besle, Fort, & Giard, 2005).

Although the preceding studies demonstrate that visual information can enter early bottom-up auditory processing, they leave open the question of whether visual information is truly integrated into the representation of auditory regularities. Previous experiments were designed such that visual information altered the perception of only the deviant stimuli and not the auditory regularity to which it is compared. Hence, it is possible that the altered auditory percept of the deviant, compared with the representation of the auditory regularity, generated an MMN but that the representation of the auditory regularity per se was left intact.

To show unequivocally that visual information enters the representation of auditory regularities, one must show that the MMN differs for audiovisual versus auditory-only regularities. In this study, we compare MMNs elicited by the same auditory deviance when regularities were either auditory or audiovisual: In the auditory condition, standard (A) and deviant (A') auditory stimuli were presented. In the constant audiovisual condition, auditory standards and deviants were presented synchronously with the same visual stimulus so that there were standard audiovisual stimuli (AVc) and infrequent audiovisual stimuli deviating only on the auditory dimension (AVc). We predicted that the MMNs would be different, reflecting differences in the representation of the auditory and audiovisual regularities. To control for the possibility that any difference in MMNs may be because of the mere presence of visual information, we added a variable audiovisual condition in which all stimuli were audiovisual but there was no visual regularity. This was achieved by systematically varying the visual features independently of the standard or deviant status of the auditory stimulus (AV1, AV2, AV3, AV4, AV1, AV2, AV3, and AV4). In this control condition, visual information might influence the auditory processing of the deviant, but the regularity on which MMN depends is only auditory and should generate an MMN identical to the purely auditory condition.

Therefore, our hypothesis was that, if the representation of an audiovisual regularity in the auditory system includes its visual dimension, then the MMN elicited by the auditory deviance from a constant audiovisual stimulus (AVc) should differ from the MMN elicited by the same auditory deviance both from an auditory-only stimulus (Aud) and from a variable audiovisual stimulus (AVv).

**METHODS**

**Participants**

Sixteen right-handed adults (nine women, mean age = 24 years, SD = 2.5 years) were paid to participate in the study; for which, they gave a written informed consent in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). The study was approved by the local ethics committee of INSERM (Comité de Protection des Personnes). All participants were free of neurological disease and had normal hearing and normal or corrected-to-normal vision.

**Stimuli**

The stimuli were adapted from those previously used by our group in various experimental paradigms, which have revealed a variety of cross-modal interactions in the first 200 msec of processing (Fort, Delpuech, Pernier, & Giard, 2002a, 2002b; Giard & Peronnet, 1999). They were identical to stimuli used in our other studies (Besle et al., 2005, 2007).

Visual stimuli consisted of the deformation of a yellow circle on a black background into an ellipse either in a horizontal (V1), vertical (V2), or 45° oblique direction (V3 and V4; see Figure 1). The basic circle had a diameter of 4.55 cm and was displayed on a CRT monitor (refresh rate of 100 Hz) placed 130 cm in front of the participants’ eyes, subtending a visual angle of 2°. The amount of deformation in either direction relative to the diameter of the circle was 33% and lasted 140 msec. Between each deformation, the circle remained present on the screen; a white cross at its center served as the fixation point. The fixation cross never overlapped with the deforming circle (see Figure 1).

Auditory stimuli were rich tones (the fundamental and the second and fourth harmonics), the fundamental frequency of which shifted from 500 to 540 Hz (A1) or from 500 to 600 Hz (A2). Visual stimuli were circles deforming in either of four directions: horizontal (V1), vertical (V2), first diagonal (V3), or second diagonal (V4). In the actual stimuli, the circle was yellow and the fixation cross was white on a black background.

**Figure 1.** Auditory and visual stimuli. Auditory stimuli were rich tones, the fundamental frequency of which shifted from 500 to 540 Hz (A1) or from 500 to 600 Hz (A2). Visual stimuli were circles deforming in either of four directions: horizontal (V1), vertical (V2), first diagonal (V3), or second diagonal (V4). In the actual stimuli, the circle was yellow and the fixation cross was white on a black background.
frequency of which shifted linearly either from 500 to 540 Hz (A1) or from 500 to 600 Hz (A2). Tone duration was also 140 msec, including 14-msec rise/fall times. Tones were presented through loudspeakers located beside the CRT monitor at a comfortable hearing level.

For audiovisual presentations, auditory and visual stimuli were presented in synchrony. Stimuli were presented using custom-made software running on MS-DOS. Actual audiovisual synchrony was insured before the experiments by placing a photodiode on the monitor and feeding its signal as well as the audio output to the EEG recording system. Over thousands of trials, the first frame of the visual stimulus was always within 2 msec of the start of the sound.

Procedure

Participants were seated in a dark, sound-attenuating room and were given instructions describing the task along with an audiovisual practice block. There were three conditions presented in separate blocks (Aud, AVc, and AVv). In each condition, 1,600 stimuli, including 320 (20%) deviants, were presented over eight blocks of 200 stimuli. Stimuli were presented at a fixed ISI of 560 msec. The 24 blocks were presented in a random order that changed from participant to participant. In all three conditions, participants had to perform the same visual distractor task on the fixation cross and ignore all the other visual and auditory stimuli: The participants’ task was to monitor the fixation cross and click the mouse button each time the cross disappeared. Disappearance occurred unpredictably for 120 msec in 13% of the trials, only within standard trials not preceding a deviant trial, and at a random time after the trial onset (i.e., desynchronized from the standard onset).

In half of the blocks, A1 was the standard sound and A2 was the deviant, and in the other half, this was reversed. In the auditory-only condition (Aud), only sounds were presented (apart from the fixation cross); standards and deviants are respectively named A and A’ in this condition. In the other two conditions, sounds were presented synchronously with deformations of the circle. In the constant audiovisual condition (AVc), the same visual stimulus was presented with the standard and deviant sounds (V1 in half of the blocks and V2 in the other half, with A1V1 as the standard and A2V1 as the deviant in half of the blocks and A2V2 as the standard and A1V2 as the deviant in the other half). Standards and deviants are respectively named Avc and A’vc in this condition. In the variable audiovisual condition (AVv), different stimulus types (A, AVc, AVv, A’, A’vc, and A’Vv), over a period of 800 msec including 300-msec prestimulus. Trials including disappearance of the fixation cross and standard trials following a deviant were not included. The mean numbers of averaged trials (across participants) were 741, 755, 826, 279, 284, and 317 for A, AVc, and AVv standards and A’, A’vc, and A’Vv deviants, respectively.

Finally, ERPs were digitally filtered (bandwidth: 1.5–30 Hz, slope: 24 dB/octave). The mean amplitude over the (−100 to 0 msec) prestimulus period was taken as the baseline for all amplitude measurements. In each of the conditions, the MMN was estimated in the sample-wise difference between the responses to the deviant and to the standard stimulus.

Scalp potential maps were generated using two-dimensional spherical spline interpolation and radial projection from T3, T4, or Oz (left, right, and back views, respectively), which respects the length of the meridian arcs. Scalp current densities (SCDs) were obtained by computing the second spatial derivative of the spline functions used in interpolation (Perrin, Pernier, Bertrand, & Echallier, 1989; Perrin, Pernier, Bertrand, & Giard, 1987).

MMN amplitudes in the three, AVc, Aud, and AVv, conditions were compared using planned orthogonal contrasts in a repeated-measures ANOVA. To avoid multiple comparisons in time and across electrodes, we computed

**EEG Recording**

EEG was recorded continuously via a Neuroscan Compu-medics system through Synamps AC coupled amplifiers (analogue bandwidth = 0.1–200 Hz, sampling rate = 1 kHz) from 36 Ag–AgCl scalp electrodes referenced to the nose and placed according to the International 10–20 System: Fz, Cz, Pz, POz, and Iz; Fp1, F3, F7, FT3, FC1, T3, C3, TP3, CP1, T7, P3, P7, PO3, and O1; and their counterparts on the right hemiscalep, Ma1 and Ma2 (left and right mastoids, respectively) and IMa and IMb (midway between Iz–Ma1 and Iz–Ma2, respectively). Electrode impedances were kept below 5 kΩ. Horizontal eye movements were recorded from the outer canthus of the right eye; eye blinks and vertical eye movements were measured in channels Fp1 and Fp2.

**Data Analysis**

EEG analysis was undertaken with the ELAN Pack software developed by the CRNL-DYCOG Team (Lyon, France). Trials with signal amplitudes exceeding 100 μV at any electrode from 300 msec before time zero to 500 msec after time zero were automatically rejected to discard responses contaminated by eye movements or muscular activities.

ERPs were averaged off-line separately for the six different stimulus types (A, AVc, AVv, A’, A’vc, and A’Vv), over a period of 800 msec including 300-msec prestimulus. Trials including disappearance of the fixation cross and standard trials following a deviant were not included. The mean numbers of averaged trials (across participants) were 741, 755, 826, 279, 284, and 317 for A, AVc, and AVv standards and A’, A’vc, and A’Vv deviants, respectively.

Finally, ERPs were digitally filtered (bandwidth: 1.5–30 Hz, slope: 24 dB/octave). The mean amplitude over the (−100 to 0 msec) prestimulus period was taken as the baseline for all amplitude measurements. In each of the conditions, the MMN was estimated in the sample-wise difference between the responses to the deviant and to the standard stimulus.
the mean amplitude in a 30-msec window around the peak latency of the MMN (as measured in the Aud condition) and across a cluster of fronto-central electrodes (F3, F4, Fz, FC1, FC2, Cz, C3, C4, CP1, and CP2). Given our hypotheses, we tested two orthogonal planned comparisons: The first contrast ([0 1 −1]) tested whether the MMN amplitude differed between the Aud and AVv conditions. The second contrast ([1 −0.5 −0.5]) tested whether the MMN amplitude differed between the AVc condition and the average of the Aud and AVv conditions. Standard effect sizes were computed as the absolute contrast value (i.e., the mean difference between the conditions of interest) divided by the pooled participant and Participant × Condition standard deviation (Maxwell & Delaney, 2004, p. 549). To examine the topography of the difference between MMNs, we also tested these contrasts at each electrode. We tested SCD amplitudes using the same methods, but as there were no significant differences, these results will not be reported.

RESULTS

Behavioral Results
The average RTs in the visual distractor task were 334, 345, and 348 msec in the Aud, Avc, and AVv conditions, respectively (SDs = 52, 53, and 50 msec). The effect of condition (tested in a repeated-measures ANOVA) was significant \[F(2,45) = 11.92, p < 10^{-4}\] and was because of the fact that participants were faster in the auditory condition compared with both audiovisual conditions (Aud vs. AVc condition: \[T(15) = 3.21, p < .003\]; Aud vs. AVv condition: \[T(15) = 5.24, p < 10^{-5}\]). The miss rate was very low and not significantly different between conditions (1.17%, 1.00%, and 1.25%, respectively).

Figure 2. ERPs to standard and deviant stimuli and difference curves in the three conditions. (A) ERPs to standard and deviant tones in the auditory-only experiment. (B) ERPs to the standard and deviant audiovisual stimuli in the constant audiovisual condition. (C) ERPs to standard and deviant audiovisual stimuli in the variable audiovisual condition. (D) Difference curves (MMNs) between deviant and standard ERPs in each of the three conditions. The scale is identical for all panels. Negativity is upward.
Electrophysiological Results

Figure 2 presents the ERPs elicited by all types of stimuli at one central electrode (Cz), two mastoid electrodes, and two occipital electrodes (O1 and O2). In the Aud condition (Figure 2A), standard sounds elicited the characteristic succession of auditory-evoked components: P50 peaking around 60 msec, N1 peaking around 100 msec, and P2 peaking around 155 msec. (The very small N1 peak amplitude at Cz can be explained by the constant and rapid rate of sound presentation and perhaps by the fact that attention was away from the auditory modality). The ERPs evoked by deviant sounds were similar for the first 120-msec poststimulus; then, deviants elicited a large response of negative polarity (MMN) peaking around 195 msec at Fz. Figure 2D illustrates the deviant-minus-standard difference curves isolating the MMN component (termed MMN\(_{AVc}\) in the Aud condition). As can be seen, the MMN\(_{AVc}\) reversed polarity at both mastoid sites.

In the two audiovisual conditions (Figure 2B and C), ERPs to the standard and deviant stimuli were a combination of typical auditory and visual sensory responses: Superimposed on the auditory responses described above, a negative component peaking bilaterally around 170 msec was recorded at occipital electrodes O1 and O2, followed by a positive component peaking bilaterally around 220 msec. Note that the auditory P2 was not clearly identifiable on the audiovisual conditions because these occipital responses spread to central electrodes. As in the Aud condition, the differences in ERPs between standards and deviants began at about 120 msec, and the MMNs isolated in the deviant-minus-standard difference curves (MMN\(_{AVc}\) and MMN\(_{AVv}\)) had the same morphologies and peak latencies as MMN\(_{AVc}\) (Figure 2D).

Indeed, the same visual features were balanced between the standard and deviant stimuli, and the only physical differences between standards and deviants were in the auditory modality and were similar in the three conditions.

Figure 2D suggests that the MMN was larger in amplitude in the AVc condition (MMN\(_{AVc}\) at Fz: minimum amplitude = −3.02 μV, SD = 1.20 μV) compared with the Aud and AVv conditions (at Fz: minimum amplitude = −2.76, SD = 0.84 μV and minimum amplitude = −2.71, SD = 0.99 μV, respectively). Planned contrasts confirmed that (1) the amplitude of the MMN averaged over fronto-central electrodes did not significantly differ between the Aud and AVv conditions (−1.86 vs. −1.87 μV; F(1, 30) = 0.055, p = .94; absolute standardized effect size = 0.01) and (2) the MMN amplitude in the AVc condition significantly differed from the average of the A and AVv conditions (−2.19 vs. −1.87 μV; F(1, 30) = 5.2276, p = .029; absolute standardized effect size = 0.41). To ensure that our effect was not driven by a subset of participants, we examined the distribution of the unstandardized effect size across all participants.

Only one participant was a significant outlier (more than 2 SDs removed from the mean) in the direction opposite to the mean effect. We recomputed the contrasts after excluding this participant. With the outlier excluded, the MMN amplitude in the AVc condition again significantly differed from the average of the A and AVv conditions (−2.26 vs. −1.84 μV; F(1, 28) = 9.245, p = .005), and the absolute standardized effect size was larger (0.53); the amplitude of the MMN in Aud and AVv conditions did not differ (−1.87 vs. −1.84 μV; F(1, 28) = 0.052, p = .82; absolute standardized effect size = 0.09).

Figure 3A-C shows that the topography of the MMN at its peak latency was very similar in the three conditions, with polarity reversals over the temporal scalp areas, both in the potential and SCD maps. Figure 3D and E shows the topography and statistical significance of the above contrasts. As can be seen in Figure 3D, the amplitude of the difference between the MMNs in the Aud and AVv conditions was close to zero and nonsignificant. In contrast, the topography of the difference between the MMN in the AVc...
condition and the average of the MMNs in the Aud and AVv conditions (Figure 3E) showed a central topography that was significant at electrodes FC2, Cz, C3, C4, CP1, and CP2 (Figure 3D).

DISCUSSION

We showed that the amplitude of the MMN is larger when standard sounds have been consistently associated with a standard visual stimulus (AVc condition) than when it was either not associated with any visual stimulus (Aud condition) or inconsistently associated with variable and equiprobable visual stimuli (AVv condition). The deviations were only auditory and identical in all three conditions, hence this difference cannot be accounted for by purely auditory mechanisms. Furthermore, the visual stimuli accompanying the auditory standards and deviants were identical and balanced in each condition, hence the differences in MMNs cannot be explained by purely visual mechanisms. Therefore, the differences in auditory MMNs must be because of some cross-modal effect, more specifically to the effect of differences in visual stimulation on the auditory MMN. In general, differences in the MMN may arise either from differences in the deviance or from differences in the representation of the regularity. If our effect was because of visual information altering the processing of the auditory deviant, then we should have found a difference between the Aud and AVv conditions. We therefore conclude that the observed effect is because of the influence of consistent audiovisual pairing on the representation of the auditory regularity. In other words, the neural representation of the audiovisual regularity underlying the MMN generation differed from that of the auditory regularity. Hence, these results demonstrate that the human brain can encode audiovisual regularities. Our experimental design did not constrain the stage of processing at which the visual stimulus influenced the representation of the auditory regularity. Audiovisual interactions responsible for the effect may have occurred at an early stage of auditory processing (Besle et al., 2004, 2008; Giard & Peronnet, 1999) or in a higher, abstract representation where regularities in the structure of sensory input are encoded (e.g., Sussman, Winkler, Huotilainen, Ritter, & Näätänen, 2002).

Adding visual motion to sounds in the AVc and AVv conditions resulted in a significant increase of participants’ RTs on the distractor task. This suggests that the irrelevant visual circle, or perhaps its transient deformations, may have taken participants’ attention away from the fixation cross in these conditions, resulting in unbalanced visual attention between the auditory-only and both audiovisual conditions. Visual attention has been shown to have an effect on the amplitude of the auditory MMN (Muller-Gass, Stelmack, & Campbell, 2006; Valtonen, May, Makinen, & Tiitinen, 2003; Otten, Alain, & Picton, 2000; Dittmann-Balcari, Thienel, & Schall, 1999) and may have affected the MMN amplitudes in the audiovisual conditions relative to the auditory-only condition, but it cannot explain why the MMN was selectively increased in the AVc condition and not in the AVv condition.

Two previous MMN studies reported an effect of standard visual stimuli (written syllables) on the MMN elicited by an auditory deviance of speech sounds. Froyen, Van Atteveldt, Bonte, and Blomert (2008) showed that the auditory MMN to phonological deviants was increased compared with an auditory-only condition when standard written syllables were presented. Mittag, Takegata, and Kujala (2011) showed that the auditory MMN to some (but not all) types of deviance from standard speech sounds differed depending on the type of standard visual stimuli presented concurrently (written syllables vs. scrambled written syllables). Although both studies found a detrimental effect of audiovisual asynchrony on their effect, compatible with integration of visual information to the representation of auditory regularities, neither studies included a critical control equivalent to our variable audiovisual condition. In fact, when RTs in a visual distractor task were measured (Mittag et al., 2011), significant differences in RT were found between different audiovisual conditions, suggesting that their results might be explained by differences in the allocation of visual attention.

In a previous study, we used a different strategy to test whether visual information was included in the representation of auditory regularities (Besle et al., 2007): two audiovisual regularities, A1V1 and A2V2, were presented, whereas deviants consisted of the infrequent conjunctions A1V2 and A2V1. According to the logic of conjunction deviance (Gomes, Bernstein, Ritter, Vaughan, & Miller, 1997), an MMN to the conjunction deviants can only be elicited if the specific audiovisual regularities A1V1 and A2V2 are represented in memory. In that study (Besle et al., 2007) we found an MMN-like component, but not in all participants. The drawback of deviance conjunction designs is that they require two regularities to be maintained in memory, which has been shown to weaken their strength (Winkler, Paavilainen, & Näätänen, 1992) and could explain the absence of consistent MMN to audiovisual conjunction deviance. In this study, on the other hand, only one regularity was presented and elicited a clear MMN that was modulated by consistent audiovisual pairing.

An alternate interpretation of our result is that a larger MMN in the constant audiovisual condition reflects a superimposed visual MMN (e.g., Czigler, 2007) induced by the auditory deviance, the same way previous studies showed that an auditory MMN can be elicited by an illusory auditory change arising from visual input (Hertrich et al., 2007; Saint-Amour et al., 2007; Colin et al., 2004; Colin, Radeau, Soquet, Dachy, et al., 2002; Colin, Radeau, Soquet, Demolin, et al., 2002; Möttönen et al., 2002; McGurk effect: Stekelenburg et al., 2004; Sams et al., 1991). This possibility is unlikely for two reasons. First, we have shown that the visual MMN elicited by dynamic stimuli similar to those...
used in this study has a bilateral occipital topography (Besle et al., 2005) that is inconsistent with the topography of the MNH difference in this study. Second, cross-modal MMN elicitations have been reported only in cases of strong audiovisual illusions such as the McGurk or ventriloquist effect, and there was no evidence that our stimuli elicited such illusion.

The MNH is thought to be mainly generated in the auditory cortex (e.g., Kropotov et al., 1995), but it is likely that multiple cortical areas outside the auditory cortex contribute to the MNH, including frontal generators (Giard, Perrin, Pernier, & Bouchet, 1990). It may be that the difference we observed between the MNHs elicited by the violation of auditory and audiovisual regularities corresponds to nonauditory generators. In favor of this possibility, the topography of the observed difference is more posterior than those of both MNHs, which suggests the contribution of parietal, perhaps multisensory, generators (see also Winkler, Horvath, Weisz, & Trejo, 2009). This topography is also consistent with our previous MEG data showing a posterior MMH-like component in response to an audiovisual conjunction deviance (Besle et al., 2007).

For many years, the neural representation underlying MMN generation has been interpreted in terms of sensory memory trace (Näätänen, 1992). More recently, it has been proposed that the MMN represents an increase in prediction error between the predicted sensory model based on a detected regularity and the deviant (Wacongne et al., 2007; Baldeweg, 2007; Winkler, 2007). In this predictive coding framework, our effect would reflect a top–down cross-modal effect of the visual information that has been included into the predictive model of auditory regularities. Similar prediction effects based on visual cues have been reported in the context of audiovisual speech perception: Modulatory effects of visual speech cues on speech processing have been interpreted as an error prediction between a predictive model based on visual speech temporally leading to auditory speech (Arnal, Morillon, Kell, & Giraud, 2009).

An alternative account of the auditory MMN is that it is not because of a higher-order memory process but of the fact that neurons adapt to the presentation of the standard hence respond more to the presentation of a new stimulus (the deviant) than to the standard (the refractoriness or adaptation hypothesis: May & Tiitinen, 2009). Although this explanation has been refuted for several types of auditory deviancies (Jacobsen & Schröger, 2001, 2003; Schröger & Wolff, 1996), the refractoriness hypothesis might apply to our study: A multisensory cell population could adapt specifically to the association between the auditory and visual dimensions of the standard audiovisual event (see also van Atteveldt, Blau, Blomert, & Goebel, 2010), and because the association between the auditory deviant and the visual standard is infrequent, the response to this rare association would elicit a larger response compared with the standard association. Refractoriness and memory/predictive accounts are not mutually exclusive, and the refractoriness account can be seen as a memory account at a lower physiological level than classically assumed for MMN (May & Tiitinen, 2009; but see Wacongne et al., 2012).

To conclude, this study is consistent with evidence that the brain encodes the regularities of the acoustic environment in its complexity (Paavilainen et al., 2007; Takegata et al., 2005; Korzyukov et al., 2003; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001; Saarinen et al., 1992), specifically audiovisual regularities. This finding fits with the current view that virtually all levels of cortical processing are capable of integrating information from different modalities, although to different degrees (Schroeder & Foxe, 2005).

Reprint requests should be sent to Julien Besle, Visual Neuroscience Group, School of Psychology, University of Nottingham, Nottingham NG17DQ, United Kingdom, or via e-mail: julien.besle@nottingham.ac.uk.

REFERENCES


