Published in final edited form as:
Brain Topogr. 2009 May ; 21(3-4): 207–215. doi:10.1007/s10548-009-0094-5.

Mismatch Negativity with Visual-only and Audiovisual Speech

Curtis W. Ponton,
Compumedics/Neuroscan, Inc., 6605W W.T. Harris Blvd Suite F, Charlotte, NC 28269, USA

Lynne E. Bernstein, and
Neuroscience Graduate Program, University of Southern California, Los Angeles, CA 90089, USA

Edward T. Auer Jr
Department of Speech-Language-Hearing, Dole Human Development Center, University of Kansas, 1000 Sunnyside Avenue, Room 3001, Lawrence, KS 66045-7555, USA

Abstract

The functional organization of cortical speech processing is thought to be hierarchical, increasing in complexity and proceeding from primary sensory areas centrifugally. The current study used the mismatch negativity (MMN) obtained with electrophysiology (EEG) to investigate the early latency period of visual speech processing under both visual-only (VO) and audiovisual (AV) conditions. Current density reconstruction (CDR) methods were used to model the cortical MMN generator locations. MMNs were obtained with VO and AV speech stimuli at early latencies (approximately 82-87 ms peak in time waveforms relative to the acoustic onset) and in regions of the right lateral temporal and parietal cortices. Latencies were consistent with bottom-up processing of the visible stimuli. We suggest that a visual pathway extracts phonetic cues from visible speech, and that previously reported effects of AV speech in classical early auditory areas, given later reported latencies, could be attributable to modulatory feedback from visual phonetic processing.

Keywords

EEG; MMN; Audiovisual speech processing; Visual speech processing

Introduction

The functional organization of cortical speech processing is thought to be hierarchical, increasing in complexity (e.g., from phonetic cues or features to consonant and vowel segments) and proceeding from primary sensory areas centrifugally (Scott and Johnsrude 2003). Therefore, evidence for visual speech feed forward effects in primary auditory cortex with visual-only (VO) or audiovisual (AV) speech would imply a role for the auditory system in visual phonetic stimulus analysis (Kislyuk et al. 2008; Möttönen et al. 2002; Sams et al. 1991). In contrast, evidence for visual feedback effects would imply that input from visual areas modulates ongoing auditory feature processing (Bernstein et al. 2008a; Calvert et al. 1999). Evidence that the phonetic information in visible speech is processed outside of classical auditory areas (Bernstein et al. 2008a; Calvert and Campbell 2003; Capek et al. 2008; Santi et al. 2003) at temporally early latencies would lend credence to the view that auditory cortex activation is due to modulatory processes. The current study used the mismatch negativity

e-mail: E-mail: lbernstein@hei.org.

A previous publication (Bernstein et al. 2008a) reported on responses to only the stimulus standards in the data set presented here.

(MMN) (Näätänen et al. 1978) obtained with electrophysiology (EEG) to investigate the early latency period of visual speech processing under VO and AV conditions.

The MMN is an attractive tool for temporally and spatially localizing the site(s) of perceptual stimulus processing. The classical auditory MMN is generated by the brain's automatic response to a change in repeated stimulation that exceeds a threshold corresponding approximately to the behavioral discrimination threshold, whether the stimuli are speech or non-speech, and whether they are attended or not (Näätänen et al. 1978, 2005, 2007). The auditory MMN waveforms are attributed to two processes, a bilateral supratemporal process and a predominantly frontal right-hemispheric process (Giard et al. 1990; Molholm et al. 2005; Näätänen et al. 2007). The supratemporal process is considered to be a pre-perceptual memory-based discriminative response, and the frontal right-hemispheric process is attributed to an obligatory attention-switching response. Importantly, the auditory MMN generating system is considered to maintain a representation of stimulus-specific acoustic regularities (Molholm et al. 2005).

The type of stimulus change that can result in an auditory MMN is not fixed, ranging from the level of quasi-steady-state acoustic features to that of the conjunction of features into unitary sounds, to higher-order spatiotemporal patterns, and to speech features and segments (Näätänen et al. 2005, 2007). MMN responses have also been obtained with non-speech visual stimuli (Astikainen et al. 2008; Czigler et al. 2007; Pazo-Alvarez et al. 2003). The visual MMN (vMMN) is frequently recorded over occipital areas of the cortex, when stimuli violate an established regularity, and even when the regularity is not related to ongoing behavior (Czigler et al. 2007). Because the MMN is attributed to maintenance of stimulus-specific representations and not to feedback, an early latency vMMN outside of classical auditory areas in response to visible speech cues would imply that classical auditory areas are not sufficient for processing the information in visual phonetic stimuli. This result would be consistent with the possibility that AV speech activations in classical auditory areas are due to modulatory feedback (Bernstein et al. 2008a; Calvert et al. 1999).

The MMN approach in AV speech integration research has generally been combined with the so-called *McGurk effect* (McGurk and MacDonald 1976). An example of the effect is said to have occurred when a visual “ga” and an auditory “ba” stimulus are presented together, and perceivers report hearing “da.” Because, theoretically, localizing an MMN generator is identical to localizing representations of stimulus feature regularities (Molholm et al. 2005; Näätänen et al. 2007), evidence of a discriminative response between matched and mismatched stimuli can suggest the latency and location of AV integration. If a change in the visual part of an AV stimulus results in an MMN that appears to have been generated in a classical auditory temporal site—particularly, a hierarchically and temporally early site—an implication is that the auditory and visual stimuli were integrated at or before that site (Kislyuk et al. 2008; Möttönen et al. 2002; Sams et al. 1991). Examples of latencies reported for AV integration implied by the presence of an MMN in the auditory cortex are approximately 200 ms (Sams et al. 1991), 130-300 ms (Möttönen et al. 2002), 300 ms (Lebib et al. 2004), and 266-316 ms (Saint-Amour et al. 2007).

Although the vMMN has been established for non-speech stimuli (Astikainen et al. 2008; Czigler et al. 2007; Pazo-Alvarez et al. 2003), a vMMN for speech has been elusive (Colin et al. 2004; Colin et al. 2002; Kislyuk et al. 2008; Möttönen et al. 2002; Saint-Amour et al. 2007). A possible explanation for difficulty recording a speech-related vMMN has been that previously obtained results incorporated the obligatory stimulus-specific exogenous responses, because the MMN was calculated using different stimuli for the standard than for the deviant (Colin et al. 2002, 2004; Möttönen et al. 2002). That is, if different neural populations respond to different stimuli (Molholm et al. 2005), a MMN calculated on responses with different stimuli could include stimulus-specific activity as well as change detection activity. Another possibility is that the subtraction of standard from deviant responses results in relatively low-amplitude and noisy, less reliable signals (Ponton et al. 1997). To reduce effects of the obligatory response to physically different stimuli, the MMN can be formed by subtracting the event-related potential (ERP) to a given stimulus in the role of standard (i.e., frequently presented) from the same stimulus when in the role of deviant (i.e., rarely presented) (Horvath et al. 2008; Näätänen et al. 2007). In order to overcome the intrinsic noise in the derived MMN responses, an integrated MMN (MMN_i) can be used (Ponton et al. 1997, and see section "Methods"). Another potentially important methodological factor to consider in a study of the vMMN is the duration and timing of the visible speech. Frequently, visible speech face movements precede auditory stimulus onset by tens or hundreds of milliseconds. However, typically, the EEG recording epochs are timed such that those preceding movements are in what is considered to be the pre-stimulus period, and the onset of the evoked response is considered to coincide with the auditory stimulus onset. Here, the visible mouth opening was close in time to the auditory stimulus onset. However, we consider in more detail the visible speech in the preceding time period.

In the current study, we investigated the visual speech MMN under both VO and AV conditions: Only the visual stimulus changed in the AV condition. A previous study (Bernstein et al. 2008a) that examined the spatial and temporal dynamics of the responses to standard stimuli in the current study found evidence for extensive occipital, parietal, and posterior temporal activation in response to visual-only "ba" and "ga" stimuli. Those areas could potentially be generators of vMMN responses. To remove the contribution of stimulus-specific activity, the vMMN in the current study was calculated using responses to the same stimulus presented under standard and deviant conditions. Then current density reconstruction (CDR) (Fuchs et al. 1999) models were computed on MMN time waveforms and MMN_i waveforms. CDR represents spatiotemporal cortical response patterns using a large number of distributed dipole sources for which no prior assumptions are made regarding number or dynamical property of the cortical dipoles. The approach, without a priori assumptions, seemed well-suited to this study, given that the literature (Colin et al. 2002, 2004; Möttönen et al. 2002; Saint-Amour et al. 2007; Sams et al. 1991) has reported sparse evidence for speech vMMNs, although EEG evidence has been reported for non-speech face motion (Puce et al. 1998; Puce and Perrett 2003).

Materials and Methods

Participants

Twelve right-handed adults (mean age 30, range 20-37 years) were pre-screened for susceptibility to McGurk effects (McGurk and MacDonald 1976). In the screening test, 48 stimuli were presented that combined a visual token of "tha," "ga," "ba," "da" and auditory token of each of the same tokens. For the classic McGurk stimulus with auditory "ba" and visual "ga," all the participants responded with a non-"ba" response on 50% or more of the trials (mean non-"ba" response of 90%). All had normal pure-tone auditory thresholds. Prior to testing, the purpose of the study was explained to each participant, and informed consent was obtained from all participants in accordance with the St. Vincent's Institutional Review Board. All participants were paid.

Stimuli

The stimuli were based on natural productions of the AV syllables "ba" and "ga."[1] Video was recorded at a rate of 29.97 frames/s. The "ba" and "ga" stimulus tokens were selected so that

[1] Note: Auditory conditions were also tested, including/d$\hat{=}$/and/b$\hat{=}$/, but they are not reported here.

the video end frames could be seamlessly dubbed from “ba” to “ba” or “ga” to “ba” (i.e., start and end frames of each token were highly similar), thus preventing a visually evoked response to the trial onset. In order to reduce the video stimulus durations, alternate frames were removed from the quasi-steady state portion of the vowel, resulting in a total of 20 video frames (667 ms) per video trial. The setting for the 0-ms point of the EEG sweeps coincided with the 200-ms point in the trial, which was the onset point for the acoustic “ba” signal.

The major segment of lip opening approximately coincided with the 0-ms point of the EEG sweep for the two stimuli. However, the video tokens, because they were different syllables, had different temporal dynamics. Compression of the lips began in the first frame of the “ba” video, but the lips did not part, and the jaw did not drop until the 6th video frame. The “ga” stimulus was static during its first two video frames. Then the jaw dropped with visible movements between frames 2 and 3, frames 5 and 6, and frames 7 and 8. For the congruent AV (AVc) “ba” stimulus, the natural relationship between the visible speech movement and the auditory speech was maintained. For the incongruent AV (AVi) auditory “ba” and visual “ga,” the acoustic “ba” signal was dubbed so that its onset was at the onset of the original acoustic “ga” for that token. For the VO “ba” and “ga” conditions, the auditory portion of the stimuli was muted. In order to guarantee the audio-visual synchrony of the stimuli, they were dubbed to video tape using an industrial betacam SP video tape deck, thus locking their temporal relationships. The audio was amplified through a Crown amplifier for presentation via earbuds. In order to guarantee synchrony for data averaging of the EEG, a custom trigger circuit was used to insert triggers from the video tape directly into the Scan™ acquisition system.

Procedure

Participants were tested in an electrically shielded and sound-attenuated booth. All of the EEG recordings were obtained on a single day. The data were collected during a mismatch negativity paradigm in which standards were presented on 87% of trials pseudo-randomly ordered with 13% of deviant trials. Each stimulus was tested as both a standard and a deviant. The different conditions and deviants were presented in separate runs. For example, VO “ba” occurred as a standard versus VA “ga” as a deviant, and vice versa in another run. Thus, there were two runs for each stimulus. Two thousand two hundred trials were presented per participant per condition. Visual stimuli were viewed at a distance of 1.9 m from the screen. Participants were not required to respond behaviorally to the stimuli. Testing took approximately 4.5 h per participant and rests were given between runs.

Electrophysiological Recordings and Analyses

Thirty silver/silver-chloride electrodes were placed on the scalp at locations based on the International 10/20 recording system (Jasper 1958). A reference electrode was placed on the forehead at Fpz, with a ground electrode located 2 cm to the right and 2 cm up from Fpz. Vertical and horizontal eye movements were monitored on two differential recording channels. Electrodes located above and below the right eye were used to monitor vertical eye movements. Horizontal eye movements were recorded by a pair of electrodes located on the outer canthus of each eye. For each stimulus condition, the EEG was recorded as single epochs, filtered between DC and 200 Hz and sampled at a rate of 1.0 kHz. Recording was initiated 100 ms prior to the acoustic onset and for 500 ms following the onset. Recordings obtained for the VO stimuli used the same recording onset and offset as for the AV stimuli, that is, relative to the temporal point of the acoustic onset. Off-line, the individual EEG single-sweeps were baseline corrected over the pre-stimulus interval and subjected to an automatic artifact rejection algorithm. A regression-based eye blink correction algorithm was applied to the accepted single sweeps (at least 1500 per participant per condition), which were then averaged. The averages were filtered from 1 to 70 Hz and average-referenced. For each stimulus, data from all 12 subjects were used to generate grand average waveforms.

Grand mean waveforms were computed separately for the standard and deviant, and then the standard grand mean was subtracted from the deviant grand mean, on a per-stimulus basis. These MMNs are referred to as the *MMN time waveforms*. Then the integrated MMN (MMNi) was computed, which represents an almost noise-free estimate of MMN magnitude (Ponton et al. 1997). MMNi waveforms were computed by simple discrete mathematical integration (i.e., running summation) of the individual difference waveforms.

An integrated surface-recorded evoked potential represents the shape of the compound membrane potential of the group of synchronously active pyramidal cells that generate the MMN (Ponton et al. 1997). The individual short-duration peaks in the time waveform that are produced by random physiological noise are essentially cancelled out in the integrated waveform, resulting in smooth and relatively noise-free data. Integration effectively acts as a low pass filter, enhancing ERP components in the MMN frequency range (4-12 Hz). MMN difference waveforms that are not integrated can produce unstable solutions (see Figs. 2, 3). This instability can be attributed directly to low SNR, given that the difference waveforms have relatively small deflections relative to the noise. When the integrated waveform comprises an MMN, the continuously increasing negative deflection reaches a maximum when the MMN terminates, that is, when the time difference waveform returns to baseline. The peak of the MMNi is later than the peak of the MMN difference waveform, due to integration.

The MMN time waveforms and the MMN_i waveforms were submitted to the Curry™ 6.0 software (Neuroscan, NC) for generation of current density reconstruction (CDR) models based on the standardized low resolution electric tomography analysis (sLORETA) (Pascual-Marqui 2002). This technique is an extension of the minimum norm least squares models for distributed dipoles in which the current density solution is standardized against the background noise in the model.

With forward solution dipole models, optimization techniques allow the user to define constraints that reduce the space of possible solutions, which is an advantage if prior knowledge is available to localize activity (Scherg 1990). In contrast, CDR models solve the inverse problem, which is the relationship between the cortical sources and the resulting potentials or fields (Darvas et al. 2001). CDR uses regularization to constrain the forward solution to be the one with minimum activity. An advantage with CDR methods is that they require no a priori knowledge of the activation sites. Here, we sought to model activation without a priori knowledge. Although estimates of spatial resolution vary, the analyses here are commensurate with ones in the literature that suggest spatial resolution of 1-2 cm (Darvas et al. 2001; Fuchs et al. 1999; Yvert et al. 1997). As implemented within Curry, CDR solutions with SNRs < 1.0 are simply not accepted. In our results, solutions with SNRs < 8.0 are not presented.

CDR computation utilized a three-shell, spherical head, volume conductor model with an outer radius of 9 cm. Analyses were constrained to the cortical surface of a segmented brain (Wagner et al. 1995). The CDRs were computed on every millisecond of ERP data, thus, resolving events at the same resolution as the underlying ERPs due to the linearity of the CDR computations. However, in the case of the MMNi waveforms, the integration reduces arbitrary changes due to noise from millisecond to millisecond (Darvas et al. 2001).

The CDR models were examined in the temporal window of the time waveform MMNs, as well as the time window of the MMN_is. The mean global field power (MGFP) (Lehmann and Skrandies 1980) signal-to-noise ratios (SNRs), and the fit strengths are reported. The individual CDR dipole in each visualized model (see Figs. 2, 3), represents the center of gravity of the instantaneous orientation, strength, and location of the distributed dipole field. The interpretation of the MMN activity is focused primarily on the field and not on the CDR dipole, because the dipole simply indexes the center of gravity of the field.

Results and Discussion

Figure 1 shows the MMN time waveforms in a butterfly plot, as well as the associated MGFP SNR function and the peak MGFP SNR for the responses to each stimulus. The figure clearly shows that the visual “ba” and “ga” resulted in differently structured MMN time waveforms, across VO and AV conditions. The MMN waveforms for conditions with visual “ba” show a prominent region of increased activity centered at 82 ms—VO “ba” and at 87 ms—AVc “ba.” These peaks were selected as the peaks of the MMN. In contrast, the VO “ga” and AVi stimuli did not result in a unique prominent region of increased activity, and the highest amplitude activity was at much longer latencies (i.e., 161 ms, VO “ga;” 185 and 244 ms AVi). Also, the MGFP SNRs for the “ga” VO and AVi stimuli were considerably lower than for the VO and AVc “ba.”

Although CDR models were computed for the MMNs associated with all four stimuli, acceptable fit statistics were not obtained with the responses to VO and AVi “ga” stimuli. This failure was attributable to the relatively little structure in the MMN time waveforms and the relatively low SNRs. Further explanations for the poorer “ga” results were sought in a careful examination of the stimuli, and we discuss those explanations in the General Discussion. Here, we focus on the VO and AVc “ba” results (see Figs. 2, 3, respectively).

CDR models using the MMN time waveforms for VO “ba” were computed at the peak, 82 ms, and at 40 and 120 ms. CDR models using the MMN_i waveforms were computed at 82 ms, the peak MMN_i at 158 ms, and at 190 ms. Both sets of computations resulted in right-lateralized activity. The distributions of activity were less stable with MMN time waveforms than with the MMN_i waveforms. This is expected, because the time waveforms had lower MGFP SNRs. The temporal extent of the MMN_i was longer and with later latencies, as expected given the integration. CDR models based on time and integrated MMN waveforms resulted in activity in the region of right STG, STS, MTG, and parietal cortex. The CDR dipoles, representing the center of gravity of the dipole fields, were obtained in the region of the posterior or middle, lateral STS and MTG. MMN_i produced more posterior CDR dipole locations than did MMN time waveforms. Notably, neither MMN nor MMN_i waveforms resulted in left-hemisphere activity. This is in contrast with results previously reported based on a symmetrical forward dipole model for the responses evoked by the *standard* stimulus, which demonstrated stronger left than right hemisphere activation, particularly for AV stimulus conditions (Ponton et al. 2002).

CDR models using the MMN time waveforms for the AVc “ba” were computed at the peak, 87 ms, and at 40 and 120 ms (see Fig. 3). CDR models using the MMN_i were computed at 87 ms, the peak MMN_i (252 ms), and 320 ms. Both sets of computations resulted in primarily right-lateralized activity and some bilateral superior parietal activity, which was also seen in analyses of the standard responses only (Bernstein et al. 2008a). The distributions of activity were less stable with MMN time waveforms than with MMN_i waveforms. The temporal extent of the MMN_i was longer. CDR models based on MMN_i waveforms were consistent with activation of the right STG, STS, MTG, inferior temporal gyrus, and the parietal cortex. CDR models based on time waveforms resulted in temporal lobe activity centered in the inferior temporal gyrus. Comparison of CDR models for MMN_i waveforms across Figs. 2b and 3b demonstrates remarkably similar temporal and spatial activation patterns, taking into account the shift towards the inferior temporal gyrus and the addition of superior parietal activations.

General Discussion

We computed MMNs on responses to VO and AV speech stimuli. However, CDR modeling was reliable only for VO and AVc "ba." VO MMNs were mostly right-lateralized to the regions of the posterior and lateral STG, STS, and MTG. Notably, the MMNs were at short latencies, with the waveform VO MMN peak at 82 ms and the AVc MMN peak at 87 ms (see Fig. 1). The center of AVc cortical activation was localized somewhat inferiorly to that with VO "ba," which is not discussed further here due to the lack of precise spatial resolution.

The differences in the MMN waveforms for VO "ba" and "ga" stimuli could be explained by the different stimulus temporal dynamics (see Stimuli). A likely explanation for the less distinct MMN with VO "ga" is that the stimulus contained three early rapid discontinuities in visible movement of the jaw, each of which might have generated its own C1, P1, and N1 visual responses, resulting in the oscillatory appearance of the MMN waveforms (see Fig. 1b, d). The differences in responses to the "ba" and "ga" suggest that the ability to demonstrate a vMMN to speech depends crucially on the internal structure of the visual stimuli.

The early latencies of the AVc and VO "ba" and MMNs were quite similar. These latencies can be attributed to gestures in the "ba" stimulus. Mouth opening began at approximately 0.0 ms in the EEG; however, the talker produced visible lip compression beginning with the second video frame of the trial. These face motions prior to the temporal point of acoustic speech onset are natural in speech production. Subtle but visible and linguistically relevant face motion around the lips beginning at approximately -133 ms, when added to the 82 ms peak in the MMN, is well within the latency range that has been reported for EEG responses to non-speech face movement (Puce et al. 1998; Puce and Perrett 2003).

Bilateral N170s to mouth opening have been reported, with earlier N170s on the right (Puce et al. 2000). Right-lateralized biological motion activation has been reported in an fMRI study (Grossman et al. 2000). But abundant evidence has been presented showing both hemispheres capable of processing human movement stimuli (Puce and Perrett 2003).[2] One possibility is that the vMMNs reported here is not specific to speech, and this might explain the strong right-lateralization of the vMMN. Although different vMMN waveforms were obtained across "ba" and "ga," in order to ascertain exactly what visual features are processed by the vMMN generators, further research will be needed to compare speech with non-speech, and also different visible speech tokens of the same phonemes.

The MMN latencies reported here are much earlier than MMN latencies typically reported in the left auditory cortex for AV speech (e.g., Möttönen et al. 2002; Saint-Amour et al. 2007; Sams et al. 1991). In a previous analysis (Ponton et al. 2002) of the responses to the *standard* stimuli in this study, equivalent current dipole models (Scherg 1990) were applied. Two dipole pairs were symmetrically fixed in occipital and temporal cortices of each hemisphere. That analysis demonstrated enhanced left-hemisphere activity to the standard AV stimuli by 100 ms; however, there was not a differential effect for *standard* AVi versus AVc stimuli, suggesting that the early responses were not due to feature integration but to modulation (see also, Bernstein et al. 2008a). Thus, activation of the left auditory temporal cortex has been

[2] In an fMRI study (Bernstein et al., Visual phonetic processing localized using speech and non-speech face gestures in video and point-light displays, in revision for publication), to isolate cortical sites with responsibility for processing visible speech features, speech and non-speech face gestures were presented in natural video and point-light displays during fMRI scanning at 3.0T. Participants with normal hearing and varied lipreading ability viewed the stimuli. Independent of stimulus media (i.e., point-light versus video), bilateral regions of the superior temporal sulcus, the superior temporal gyrii, and the middle temporal gyrii were activated by speech gestures. These regions were more activated in good versus poor lipreaders, consistent with an interpretation that they are important areas in the processing of visible speech.

shown with the stimuli in the current study, although that hemisphere appears not to be a generator of the MMN reported here.

The MMNs reported here at early latencies (<100 ms) can be attributed to feed forward visual processing and support the possibility that later AV effects in classical areas of auditory cortex (e.g., Möttönen et al. 2002; Saint-Amour et al. 2007; Sams et al. 1991) are attributable to modulatory feedback (Bernstein et al. 2008a; Calvert et al. 1999). One distinct likelihood is that several processes are ongoing in parallel. Early AV modulatory effects could be due to the mere presence of the visual stimulus (Lebib et al. 2004) and not due to integration with visual phonetic stimulus features or visual feature processing (Reale et al. 2007). Previously, we argued that to distinguish between modulatory effects that up- or down-regulate ongoing sensory/modality-specific responses and AV stimulus feature integration effects, experimental methods must afford the possibility of obtaining responses specifically sensitive to stimulus features (Bernstein et al. 2008a, b). The MMN paradigm theoretically provides that opportunity. As computed here holding stimulus constant, the MMNs indicate only the discriminative response to stimulus features. Here, we find evidence for a discriminative response only at the early latencies, on the right, and outside of classical auditory areas.

Previous studies of the MMN with AV speech used MMNs computed from responses to different stimuli. Sams et al. (1991) reported the first MMNm (MMN with magnetoencephalography, MEG) results with AV speech stimuli. Auditory “pa” was presented on all trials, and visual “pa” or “ka” were presented in the frequent or deviant roles. The MMNm was recorded over the left supratemporal region only. Responses to the frequent stimuli were subtracted from responses to the infrequent ones. No MMNm response was obtained in a VO condition with two participants. If the vMMN is right-lateralized, as shown here, the failure to record a VO MMN in the Sams et al. study is understandable given the sensor placement only over the left auditory temporal cortex. But a mismatch response was obtained in the AV context beginning around 180 ms, with 0.0 ms at the auditory stimulus onset.

Colin et al. (2002) presented VO and AV stimuli. Stimuli were tokens of auditory and visual “bi” and “gi.” Activity was recorded from six electrodes only, including ones on the mastoids as well as F_z and O_z, and a VO MMN was not obtained at any location. The AV condition did result in MMN waveforms for visual changes, and the responses differed depending on the visual stimulus. Notably, the polarity between F_z and M_1 or M_2 electrodes was not reversed, which would have been expected had there been activation on the supratemporal plane. A follow-up experiment produced substantially similar results (Colin et al. 2004). The sparsely placed electrodes and different standard versus deviant stimuli might have precluded observing vMMNs.

With whole-head MEG, Möttönen et al. (2002) presented congruent AV ‘ipi,’ ‘iti,’ ‘ivi,’ and incongruent A-’ipi’-V-’iti’ in an MMN oddball design. A VO experiment was also carried out. Equivalent current dipole (ECD) models were fitted using a fixed subset of 28 magnetometers over each temporal lobe. At least 65% goodness-of-fit was required, along with orientations consistent with the auditory EEG MMN response. Thus, an a priori hypothesis for the location of the AV and VO MMNm was implicit in the analytic approach. Bilateral MMNm were obtained with AV stimuli. Bilaterally, lower amplitude VO MMNs at longer latencies (approximately 245-410 ms), deeper, more posterior, and with different *z*-coordinates (see Table 1, 2002) than for AV stimuli were obtained for the ECDs, which also had lower goodness-of-fit statistics. The latter dipoles probably do not correspond to those reported here, given the Möttönen et al. analytic focus on the supratemporal plane, their physical stimulus change between deviant and standard, and their relatively lenient goodness-of-fit criterion.

Saint-Amour et al. (2007) presented “ba” and “va” stimuli in an AV experiment with the goal of eliminating from the MMN the obligatory exogenous activity due to changes in the visual speech stimuli. Essentially, the VO MMN was subtracted from the AV MMN to obtain the AV mismatch response. The video stimuli began more than 300 ms ahead of the acoustic stimuli and differed in temporal dynamics, as did the early EEG responses to the two stimuli. No VO MMN was obtained following the 0-ms EEG recording point. However, Fig. 1 in their report suggests that there was a greater negativity to the deviant VO stimulus in the pre-stimulus EEG traces.

In summary, the current study combined the following methodological approaches that varied from previous studies: It used short duration visual speech stimuli, held the stimulus constant in calculating the MMN, transformed the MMN to MMN_i (a relatively noise-free representation), computed CDR models, which involved no a priori determination of activity location/s, and accepted only large goodness-of-fit values for models that were used for interpretation. The CDR models showed that right lateral middle to posterior temporal cortex was activated at short duration latencies in response to VO and AVc stimuli, suggesting a role for this temporal region in the representation of visible speech. The latencies of the MMNs obtained here are earlier than the latencies reported elsewhere for integrative AV effects in classical temporal auditory areas (e.g., Möttönen et al. 2002; Saint-Amour et al. 2007; Sams et al. 1991).

We suggest that evidence for bottom-up visual phonetic feature analysis and evidence for concurrent (Bernstein et al. 2008a) or later AV modulatory effects is not contradictory. The AV stimulus context could condition modulatory effects that are not specific to speech feature integration, and such effects could arise in parallel with bottom-up sensory/modality-specific stimulus feature processing. Future studies will be needed for the replication and elaboration of our results. In particular, additional studies are needed to compare responses to visual speech versus non-speech face gestures, and to different phonemes and different tokens of the same phoneme. Study designs are needed that can differentiate between feature integration and modulation, and this requires carefully controlled visual and auditory stimulus generation, with attendant consideration of the temporal dynamics of each type of stimulus.

Acknowledgments

Ewen Chao and Betty Kwong are thanked for their assistance. This research was supported by a grant from NIH/NIDCD DC006035 and DC008308, and by NSF IIS9996088.

References

Astikainen P, Lillstrang E, Ruusuvirta T. Visual mismatch negativity for changes in orientation—a sensory memory-dependent response. Eur J NeuroSci 2008;28:2319–2324. [PubMed: 19019200]

Bernstein LE, Auer ET Jr, Wagner M, Ponton CW. Spatiotemporal dynamics of audiovisual speech processing. NeuroImage 2008a;39:423–435. [PubMed: 17920933]

Bernstein LE, Lu ZL, Jiang J. Quantified acoustic-optical speech signal incongruity identifies cortical sites of audiovisual speech processing. Brain Res 2008b;1242:172–184. [PubMed: 18495091]

Calvert GA, Campbell R. Reading speech from still and moving faces: the neural substrates of visible speech. J Cogn Neurosci 2003;15:57–70. [PubMed: 12590843]

Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS. Response amplification in sensory-specific cortices during crossmodal binding. NeuroReport 1999;10:2619–2623. [PubMed: 10574380]

Capek CM, MacSweeney M, Woll B, Waters D, McGuire PK, David AS, Brammer MJ, Campbell R. Cortical circuits for silent speechreading in deaf and hearing people. Neuropsychologia 2008;46:1233–1241. [PubMed: 18249420]

Colin C, Radeau M, Soquet A, Demolin D, Colin F, Deltenre P. Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. Clin Neurophysiol 2002;113:495–506. [PubMed: 11955994]

Colin C, Radeau M, Soquet A, Deltenre P. Generalization of the generation of an MMN by illusory McGurk percepts: voiceless consonants. Clin Neurophysiol 2004;115:1989–2000. [PubMed: 15294201]

Czigler I, Weisz J, Winkler I. Backward masking and visual mismatch negativity: electrophysiological evidence for memory-based detection of deviant stimuli. Psychophysiology 2007;44:610–619. [PubMed: 17521378]

Darvas F, Schmitt U, Louis AK, Fuchs M, Knoll G, Buchner H. Spatio-temporal current density reconstruction (stCDR) from EEG/MEG-data. Brain Topogr 2001;13:195–207. [PubMed: 11302398]

Fuchs M, Wagner M, Kohler T, Wischmann H-A. Linear and nonlinear current density reconstruction. J Clin Neurophysiol 1999;16:267–295. [PubMed: 10426408]

Giard MH, Perrin F, Pernier J, Bouchet P. Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. Psychophysiology 1990;27:627–640. [PubMed: 2100348]

Grossman E, Donnelly M, Price R, Pickens D, Morgan V, Neighbor G, Blake R. Brain areas involved in perception of biological motion. J Cogn Neurosci 2000;12:711–720. [PubMed: 11054914]

Horvath J, Czigler I, Jacobsen T, Maess B, Schroger E, Winkler I. MMN or no MMN: no magnitude of deviance effect on the MMN amplitude. Psychophysiology 2008;45:60–69. [PubMed: 17868262]

Jasper HH. The ten-twenty electrode system of the international federation. Electroencephalogr Clin Neurophysiol 1958;10:371–375.

Kislyuk DS, Möttönen R, Sams M. Visual processing affects the neural basis of auditory discrimination. J Cogn Neurosci 2008;20:2175–2184. [PubMed: 18457500]

Lebib R, Papo D, Douiri A, de Bode S, Dowens MG, Baudonniere P-M. Modulations of 'late' event-related brain potentials in humans by dynamic audiovisual speech stimuli. Neurosci Lett 2004;372:74–79. [PubMed: 15531091]

Lehmann D, Skrandies W. Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalogr Clin Neurophysiol 1980;48:609–621. [PubMed: 6155251]

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

Molholm S, Martinez A, Ritter W, Javitt DC, Foxe JJ. The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. Cereb Cortex 2005;15:545–551. [PubMed: 15342438]

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

Näätänen R, Gaillard AWK, Mäntysalo S. Early selective-attention effect on evoked potential reinterpreted. Acta Psychol 1978;42:313–329.

Näätänen R, Jacobsen T, Winkler I. Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. Psychophysiology 2005;42:25–32. [PubMed: 15720578]

Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin Neurophysiol 2007;118:2544–2590. [PubMed: 17931964]

Pascual-Marqui R. Standardized low resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find Exp Clin Pharmacol 2002;24D:5–12. [PubMed: 12575463]

Pazo-Alvarez O, Cadaveira F, Amenedo E. MMN in the visual modality: a review. Biol Psychol 2003;63:199–236. [PubMed: 12853168]

Ponton CW, Don M, Eggermont JJ, Kwong B. Integrated mismatch negativity (MMNi): a noise-free representation of evoked responses allowing single-point distribution-free statistical tests. Electroencephalogr Clin Neurophysiol 1997;104:143–150. [PubMed: 9146480]

Ponton, CW.; Auer, ET., Jr; Bernstein, LE. Neurocognitive basis for audiovisual speech perception: evidence from event-related potentials; 7th international congress of spoken language processing; Denver, CO. 2002; p. 1697-1700.

Puce A, Perrett D. Electrophysiology and brain imaging of biological motion. Philos Trans R Soc B-Biol Sci 2003;358:435–445.

Puce A, Allison T, Bentin S, Gore JC, McCarthy G. Temporal cortex activation in humans viewing eye and mouth movements. J Neurosci 1998;18:2188–2199. [PubMed: 9482803]

Puce A, Smith A, Allison T. ERPs evoked by viewing facial movements. Cogn Neuropsychol 2000;17:221–239.

Reale RA, Calvert GA, Thesen T, Jenison RL, Kawasaki H, Oya H, Howard MA, Brugge JF. Auditory-visual processing represented in the human superior temporal gyrus. Neuroscience 2007;145:162–184. [PubMed: 17241747]

Saint-Amour D, De Sanctis P, Molholm S, Ritter W, Foxe JJ. Seeing voices: high-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. Neuropsychologia 2007;45:587–597. [PubMed: 16757004]

Sams M, Aulanko R, Hamalainen M, Hari R, Lounasmaa OV, Lu ST, Simola J. Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. Neurosci Lett 1991;127:141–145. [PubMed: 1881611]

Santi A, Servos P, Vatikiotis-Bateson E, Kuratate T, Munhall K. Perceiving biological motion: dissociating visible speech from walking. J Cogn Neurosci 2003;15:800–809. [PubMed: 14511533]

Scherg, M. Fundamentals of dipole source analysis. In: Grandori, F.; Hoke, M.; Romani, GL., editors. Auditory evoked magnetic fields and electric potentials. Karger; Basel: 1990. p. 40-69.

Scott SK, Johnsrude IS. The neuroanatomical and functional organization of speech perception. Trends Neurosci 2003;26:100–107. [PubMed: 12536133]

Wagner, M.; Fuchs, M.; Wischmann, H-A.; Ottenberg, K.; Dössel, O. Cortex segmentation from 3D MR images for MEG reconstructions. In: Baumgartner, C.; Deecke, L.; Stroink, G.; Williamson, SJ., editors. Biomagnetism: fundamental research and clinical applications, vol 7 studies in applied electromagnetics and mechanics. Elsevier Science IOS Press; Amsterdam, The Netherlands: 1995. p. 433-438.

Yvert B, Bertrand O, Thevenet M, Echallier JF, Pernier J. A systematic evaluation of the spherical model accuracy in EEG dipole localization. Electroencephalogr Clin Neurophysiol 1997;102:452–459. [PubMed: 9191589]

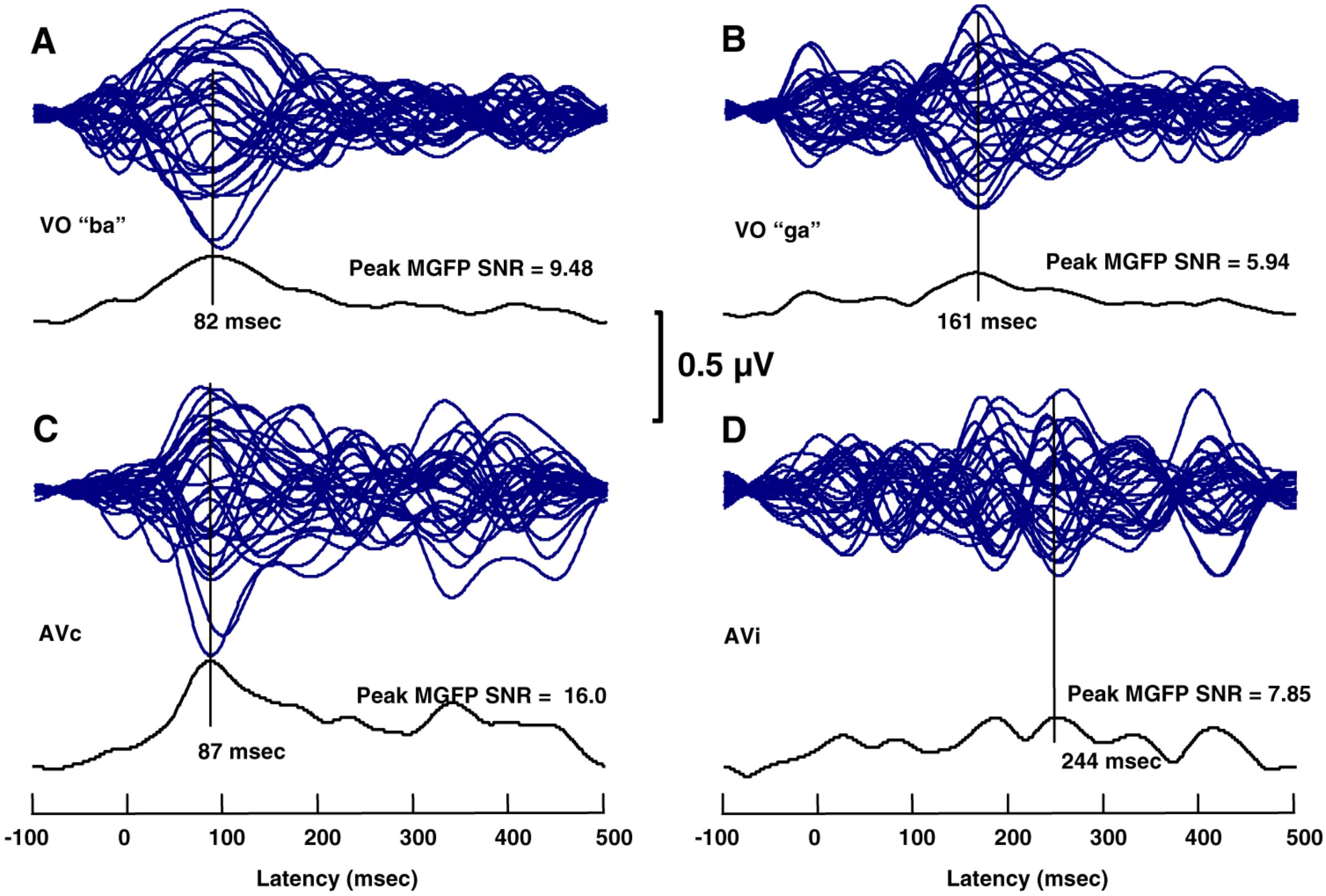

Fig. 1.

MMN time waveforms and mean global field power signal-to-noise ratio (MGFP SNR). **a-d** Butterfly plots of the MMN time waveforms for all electrodes for each of the VO and AV stimuli. *Vertical lines* are drawn at the peaks of the MGFP SNR curves

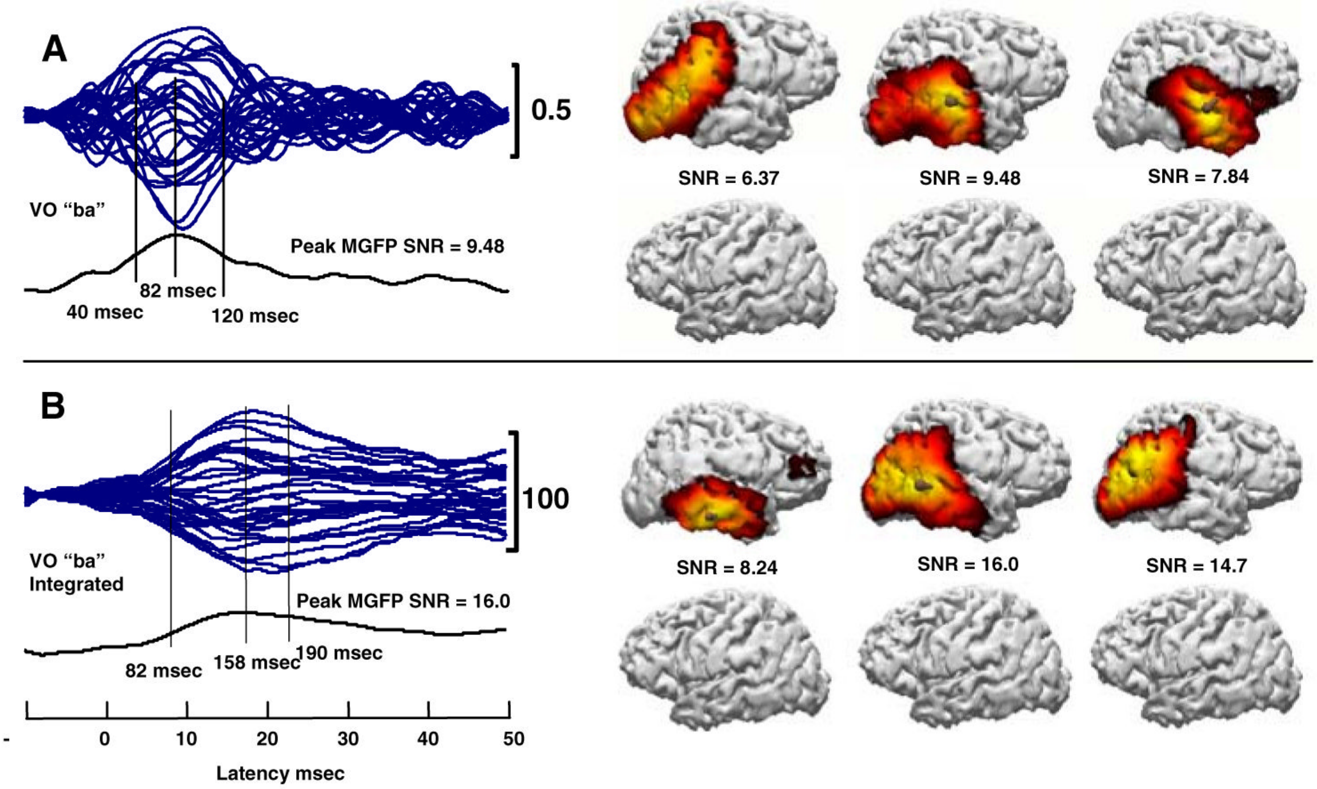

Fig. 2.

VO “ba” **a** MMN time waveform butterfly plot and CDR models, and **b** MMN$_i$ waveform butterfly plot and associated CDR models. CDR models on the right of each panel show the left and right hemispheres at times corresponding sequentially to the *vertical lines* on the *left* of the panel. SNRs are the MGFP SNR for the particular point in time

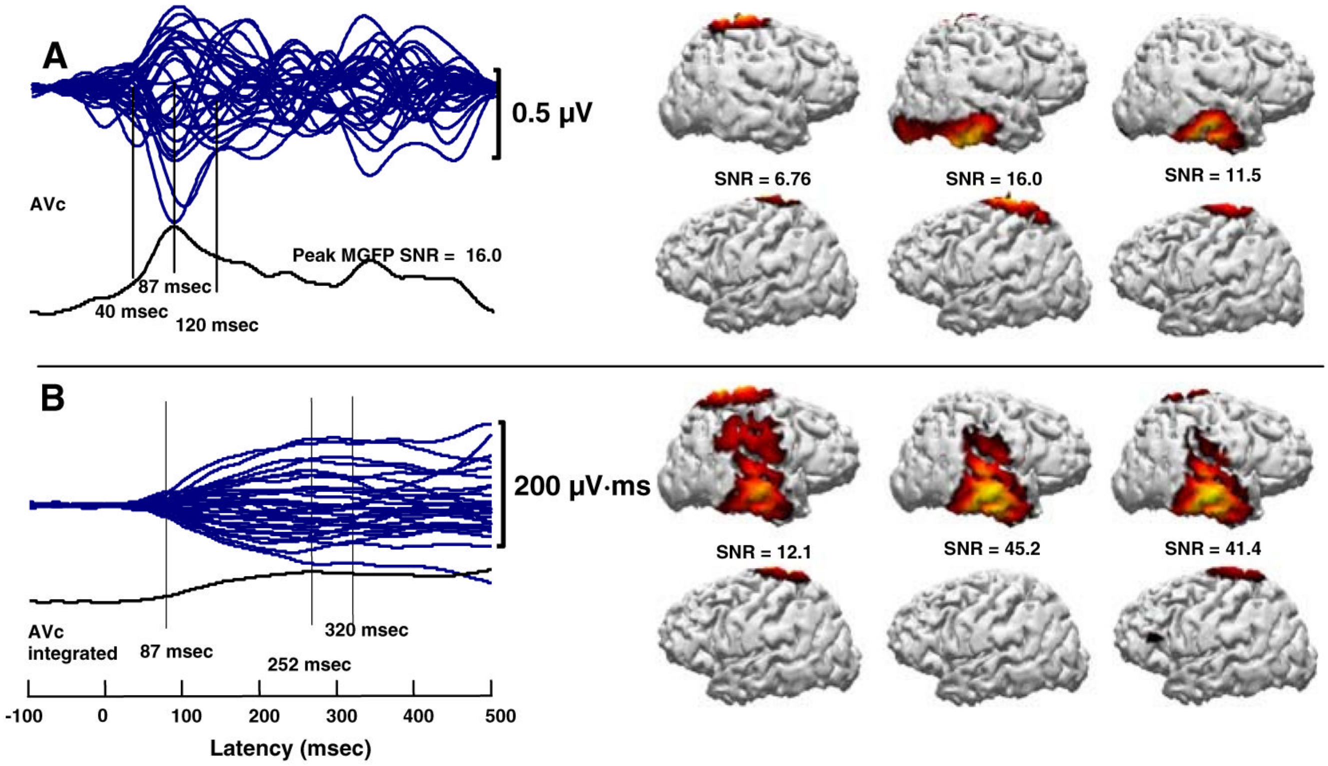

Fig. 3.

AVc “ba” **a** MMN time waveform butterfly plot and CDR models, and **b** MMN_i waveform butterfly plot and associated CDR models. CDR models on the right of each panel show the left and right hemispheres at times corresponding sequentially to the *vertical lines* on the *left* of the panel. SNRs are the MGFP SNR for the particular point in time