Time Course of Early Audiovisual Interactions during Speech and Nonspeech Central Auditory Processing: A Magnetoencephalography Study

Ingo Hertrich[1], Klaus Mathiak[2], Werner Lutzenberger[1], and Hermann Ackermann[1]

Abstract

■ Cross-modal fusion phenomena suggest specific interactions of auditory and visual sensory information both within the speech and nonspeech domains. Using whole-head magnetoencephalography, this study recorded M50 and M100 fields evoked by ambiguous acoustic stimuli that were visually disambiguated to perceived /ta/ or /pa/ syllables. As in natural speech, visual motion onset preceded the acoustic signal by 150 msec. Control conditions included visual and acoustic nonspeech signals as well as visual-only and acoustic-only stimuli. (a) Both speech and nonspeech motion yielded a consistent attenuation of the auditory M50 field, suggesting a visually induced ''preparatory baseline shift'' at the level of the auditory cortex. (b) Within the temporal domain of the auditory M100 field, visual speech and nonspeech motion gave rise to different response patterns (nonspeech: M100 attenuation; visual /pa/: left-hemisphere M100 enhancement; /ta/: no effect). (c) These interactions could be further decomposed using a six-dipole model. One of these three pairs of dipoles (V270) was fitted to motion-induced activity at a latency of 270 msec after motion onset, that is, the time domain of the auditory M100 field, and could be attributed to the posterior insula. This dipole source responded to nonspeech motion and visual /pa/, but was found suppressed in the case of visual /ta/. Such a nonlinear interaction might reflect the operation of a binary distinction between the marked phonological feature ''labial'' versus its underspecified competitor ''coronal.'' Thus, visual processing seems to be shaped by linguistic data structures even prior to its fusion with auditory information channel. ■

INTRODUCTION

Visual information has a significant impact upon speech perception, for instance, enhancing intelligibility in noisy environments (Sumby & Pollack, 1954). Furthermore, perceived facial movements may elicit auditory illusions such as the McGurk effect (MacDonald & McGurk, 1978). The perceptual fusion of visual and auditory speech features depends on the temporal relation between the two channels and is most pronounced if, as in natural speech, visual cues precede the acoustic signal by 100–150 msec (Van Wassenhove, Grant, & Poeppel, 2007). Visually induced activation of even primary auditory regions has been demonstrated in various studies (Schroeder & Foxe, 2005). Furthermore, monkey experiments using audiovisual (AV) face/voice stimuli (Ghazanfar, Maier, Hoffman, & Logothetis, 2005), as well as a study in human sign language users (Petitto et al., 2000), suggest that the cortical information flow from the visual toward the auditory system is characterized by adaptations for species-specific communicative demands.

Electrophysiological analysis of evoked neural cerebral activity is one possibility to assess early visual influences on central auditory processing at a high temporal resolution. Recordings of event-related brain responses to AV speech signals revealed cross-modal interactions to arise as early as the time domain of the electroencephalographic (EEG) P50/N1 potentials or their magnetoencephalographic (MEG) counterparts, the M50/M100 fields. For example, Lebib, Papo, De Bode, and Baudonniere (2003) demonstrated that videos of a speaker uttering a vowel induce an attenuation of the P50 component evoked by the respective auditory event. Whereas the P50/M50 components, emerging between 30 and 80 msec after the onset of an auditory stimulus, are considered an index of sensory gating processes (Boutros & Belger, 1999), the N1/M100 complex has been found sensitive to distinct signal features such as periodicity and spectral shape and, therefore, appears to represent less stereotypic neural activity (Tiitinen, Mäkelä, Mäkinen, May, & Alku, 2005; Näätänen & Winkler, 1999). The available studies of AV interactions within the N1/M100 time domain either report visually induced enhancement of the auditory N1 component (Giard & Peronnet, 1999) or no effects (Miki, Watanabe, & Kakigi, 2004) if visual events were

[1]University of Tübingen, Germany, [2]RWTH University of Aachen, Germany

exactly synchronized with the acoustic signal. Regarding natural AV syllables characterized by a specific time delay between the onset of motion and the acoustic speech signal, visually induced dampening of auditory N1/P2 potentials concomitant with a shortened N1 peak latency could be observed (Van Wassenhove, Grant, & Poeppel, 2005). Similarly, Besle, Fort, Delpuech, and Giard (2004) noted cross-modal hypoadditive [AV < (A + V)] event-related potentials in response to AV speech stimuli within a time domain of 120–190 msec after acoustic signal onset (visually induced decrease of auditory N100). Intracortical recordings in animals were able to document additive, hypoadditive, and hyperadditive responses to AV stimuli in the central auditory system, depending, among others, on the relative timing of the acoustic and visual stimuli (Bizley, Nodal, Bajo, Nelken, & King, 2007) as well as the on the information domain, for instance, the integration of species-specific voice/face communication elements (Ghazanfar et al., 2005).

A previous MEG study (Hertrich, Mathiak, Lutzenberger, Menning, & Ackermann, 2007) found visual displays of the spoken syllables /pa/ and /ta/ to differentially influence auditory-evoked M100 fields: Enhanced responses to a tone signal were found in association with visual /pa/ as compared to visual /ta/ events. Kinematic parameters such as the extent and speed of lip movements were assumed to account for these differences. As an alternative explanation, however, nonlinear phonological distinctions might contribute to the observed effects. Phonetic features (e.g., ''voicing,'' ''place of articulation'' . . .) represent the basic information-bearing elements of speech sounds. Recent phonological models postulate an explicit specification of place of articulation in case of labial phonemes such as /b/ or /m/, whereas their alveolar (coronal) cognates (/d/ and /n/, respectively) seem to have an ''underspecified'' and less ''marked'' structure (De Lacy, 2006; Wheeldon & Waksler, 2004; Harris & Lindsey, 1995; Avery & Rice, 1989). Conceivably, the enhanced impact of visual /pa/ as compared to /ta/ upon the auditory M100 field reflects differences in the phonetic–phonological structure of these syllables, even if a meaningful fusion with the acoustic (nonspeech) tone signal was not possible. In other words, visible articulatory gestures might be phonologically encoded even in the absence of a congruent acoustic signal.

The present MEG study was designed to further elucidate early AV interactions within the time domain of M50/M100 fields.

(a) In order to distinguish between an unspecific impact of visual motion upon central auditory processing and speech-related operations such as the fusion of auditory and visual information into a common phonetic–phonological representation, the experiment encompasses all four combinations (see Table 1) of stimulus type (speech/nonspeech) and sensory modality (visual/acoustic).

Table 1. Stimulus Design: Four Audiovisual Conditions (ATYP × VTYP) Were Assessed in Different Runs

Visual Type (VTYP)	*Acoustic Type (ATYP)*	
	Speech	*Nonspeech*
Speech	Static face	Static face
	Video /ta/	Video /ta/
	Video /pa/	Video /pa/
	Static face + Syl	Static face + Tone
	Video /ta/ + Syl	video /ta/ + Tone
	Video /pa/ + Syl	video /pa/ + Tone
Nonspeech	Static circles	Static circles
	Small motion	Small motion
	Large motion	Large motion
	Static circles + Syl	Static circles + Tone
	Small motion + Syl	Small motion + Tone
	Large motion + Syl	Large motion + Tone

Each run comprised multiple repetitions of six stimulus types presented in randomized order, three levels of visual motion (static, small or /ta/, large or /pa/) paired with silence (visual-only conditions) or with an acoustic signal (Syl or Tone). The silent static condition can be considered as an ''empty'' stimulus because it is identical with the display during the interstimulus and baseline intervals.

Syl = synthetic acoustic syllable, ambiguous between /ta/ and /pa/; Tone = acoustic tone signal.

(b) The visual stimuli varied across three levels of movement range: no movement, small movement (or /ta/ in case of speech), and larger movement (or /pa/), respectively (see Table 1). If the impact of visual information within the time domain of the auditory M100 field is shaped by phonetic–categorical distinctions such as, for example, the specification or underspecification of phonetic features, nonlinear categorical effects of visual motion could be expected: Visual /pa/, signalling the fully specified phonological feature ''labial,'' can be expected to elicit a significant impact upon event-related brain activity, whereas responses to visual /ta/, representing an underspecified feature, might be suppressed.

(c) To separate AV interactions from superimposed additive effects, speech and nonspeech stimulus configurations restricted to a single sensory modality each were included as well. In contrast to previous electrophysiological studies (Colin, Radeau, Soquet, & Deltenre, 2004; Möttönen, Krause, Tiippana, & Sams, 2002), single-modality stimuli were interspersed among the AV events, rather than being presented in separate runs, to establish the same attentional setting across all trials.

(d) Because previous studies had reported AV interactions within the auditory system to depend upon

attention directed to an auditory event such as an imagined speech signal (Pekkola et al., 2006; Jäncke & Shah, 2004; MacSweeney et al., 2000; Calvert et al., 1997) or complex sound (Bunzeck, Wuestenberg, Lutz, Heinze, & Jäncke, 2005), the present study included an auditory recognition task by asking the subjects to detect an upward or downward going pitch shift at the end of each acoustic stimulus. Similarly, a monkey experiment had shown that the impact of nonauditory events onto auditory processing depends upon attention toward the auditory modality (Brosch, Selezneva, & Scheich, 2005).

It was expected that this experimental design allows, first, for a differentiation of speech-related and speech-independent AV interactions within the time domain of the M50/M100 complex and, second, for the analysis of nonlinear effects of visual motion cues bound to phonological processing (/pa/ vs. /ta/).

METHODS

Subjects

Twenty-five right-handed subjects (age = 26 years, *SD* = 7 years; 14 women), all of them native speakers of German, participated in this MEG experiment. Self-reported right-handedness was confirmed by means of a short questionnaire (German version of the Edinburgh Handedness Inventory; Oldfield, 1971), predicting hemispheric left-lateralization for language functions in well above 90% of right-handers (Pujol, Deus, Losilla, & Capdevila, 1999). None of the subjects reported a history of any relevant neurological or audiological disorders. The study had been approved by the Ethics Committee of the University of Tübingen. Anatomical magnetic resonance imaging (MRI) datasets could be obtained from 17 out of the total of 25 participants.

Stimuli

The design of the investigation was based upon four different AV configurations (Acoustic/Visual modality × Speech/Nonspeech events; see Table 1) applied during different runs of the experiment. Figure 1 displays the temporal structure of the AV speech stimuli.

Acoustic Speech Stimuli

A synthetic syllable comprising a voiceless stop consonant followed by the vowel /a/ (formant frequencies F1–F5 = 800, 1240, 2300, 3800, and 4500 Hz) was generated by means of a formant synthesizer (Hertrich & Ackermann, 1999, 2007). The spectral characteristics of the initial consonant during the burst and aspiration phase (voice onset time = 60 msec) represent an acoustically intermediate and, thus, ambiguous event between /t/ and /p/. A subsequent listening experiment verified that, indeed, the synthesized signal is perceived either as /pa/ or /ta/, depending upon whether a synchronized video of a speaker displays a /pa/ or /ta/ utterance. In other words, orofacial speech movements disambiguate the acoustic signal of the AV stimulus configurations. Therefore, the speech videos can be expected to have a significant influence on auditory phonetic processing. Previous experiments have demonstrated that the impact of visual information on speech perception is particularly high in case of ambiguous acoustic signals, for example, when noise is added to the speech recordings (Sekiyama, Kanno, Miura, & Sugita, 2003; Sekiyama & Tohkura, 1991). Fundamental frequency (F0) of the speech signal amounted to 120 Hz during the initial part of the vowel, extending across a time interval of up to 200 msec after stimulus onset. Following this stationary phase, F0 either began to rise or to fall (randomized) by six semitones to either 170 or 85 Hz at stimulus offset (syllable duration = 300 msec). These stimulus-final pitch movements

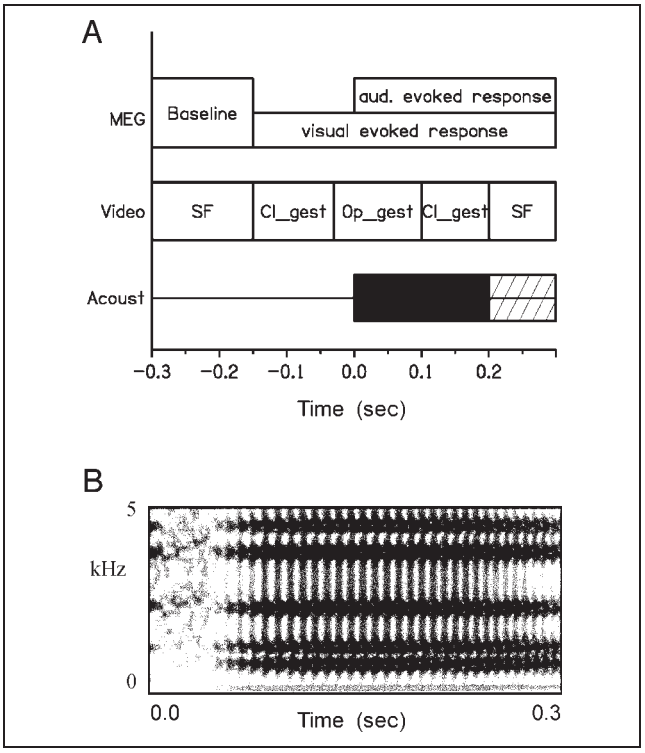

Figure 1. (A) Time course of a singe AV speech trial. Bottom (Acoust): duration of the acoustic signal, the hatched part corresponds to the final pitch movement. Middle panel (Video): SF = static face, continuously displayed between the stimuli; Cl_gest, Op_gest = duration of the visible mouth closing and opening gestures. Top (MEG): Baseline = prestimulus interval serving as the baseline of the MEG sensor data. Note that the onset of visual motion precedes the acoustic signal by 150 msec. (B) Spectrogram of the synthetic acoustic speech signal representing an ambiguous event which could be perceived either as a /ta/ or a /pa/ syllable.

approximately correspond to the range of natural intonation of a male voice during speech production.

Acoustic Nonspeech Stimuli

Periodic signals consisting of repetitions of single-formant sweeps served as the acoustic nonspeech stimuli of the MEG experiment. Within each pitch period, a formant was down-tuned from 2000 to 500 Hz and dampened to zero at its offset. Because of their periodic structure, these signals give rise to a strong pitch percept, lacking, however, any resemblance to speech sounds. Similar sounds have proved valuable in previous pitch processing studies (Hertrich et al., 2007; Hertrich, Mathiak, Lutzenberger, & Ackermann, 2004). Again, F0 amounted to 120 Hz (pitch frame duration = 9 msec) across the initial time interval of 200 msec of these events, followed by pitch movements comparable to the ones of the speech stimuli.

Visual Speech Stimuli

The visual speech condition comprised two different video sequences showing a male speaker uttering the syllable /pa/ or /ta/, respectively. The size of the display was approximately adapted to the original size of the speaker's head, and the distance from the subjects amounted to 1.2 m. These video sequences of a duration of 300 msec each were embedded into a larger frame extending across a time interval of 1.4 sec (=onset-to-onset interstimulus interval during the experiment). In other words, a static (immobile) display of the same speaker's face preceded and followed the /pa/ and /ta/ sequences. As a consequence, the visual speech stimuli could be concatenated into larger runs without any visible discontinuities of the video displays.

Visual Nonspeech Stimuli

During the visual nonspeech condition, contraction/expansion of concentric circles (light blue on a black background) served as an analogue to orofacial motion during the /pa/ and /ta/ utterances. As in the speech condition, the same static picture preceded and succeeded the movement sequences. The diameter of the movement structure was adapted to the size of the speaker's mouth on the display. Two motion sequences of a duration of 300 msec each were created (contraction and expansion time = 150 msec each), differing in size and velocity in analogy to the /pa/ and /ta/ video sequences. Although the range of vertical lower lip movement during /pa/ and /ta/ utterances approximately differed by the factor of two, the large nonspeech excursion had to be scaled by factor of four as compared to the small movements in order to create the subjective contrast of a small and a double-sized motion.

Experimental Design

The entire experiment encompassed eight runs (two repetitions of the four AV stimulus configurations each; see Table 1). Within each run, nine different combinations (3 movement levels of the visual signal × 3 acoustic conditions) of the basic AV constellation were presented at an equal probability in randomized order (27 repetitions each):

(1) large movement (or /pa/)–acoustic signal with rising pitch
(2) large movement (or /pa/)–acoustic signal with falling pitch
(3) large movement (or /pa/)–no acoustic signal
(4) small movement (or /ta/)–acoustic signal with rising pitch
(5) small movement (or /ta/)–acoustic signal with falling pitch
(6) small movement (or /ta/)–no acoustic signal
(7) static picture–acoustic signal with rising pitch
(8) static picture–acoustic signal with falling pitch
(9) static picture–no acoustic signal

In Table 1, the two acoustic pitch conditions are not explicitly listed because the F0 movements were outside the analysis window of the MEG experiment and were just included to direct attention to the acoustic channel. For the analysis of evoked MEG responses, the data from each subject were pooled across both pitch conditions. Altogether, thus, each run comprised 9 × 27 = 243 stimuli. During half of the eight runs, listeners had to detect the rising pitch and during the other half they had to respond to the falling pitch. In both cases, responses had to be performed by simultaneously pressing two buttons of an optoelectronic device with both index fingers. The measurements of evoked magnetic fields were restricted to a time interval preceding the utterance-final pitch movements, and the F0 manipulation (rise or fall) was independently randomized from the remaining stimulus characteristics. Thus, apart from unspecific expectation effects, no direct impact of the pitch changes upon the evoked M50 and M100 fields must be expected.

The behavioral data were evaluated with respect to the percentage of correct responses and reaction time. The former parameter showed a ceiling effect (ca. 90% correct responses) without any significant impact of the various experimental conditions. Similarly, mean reaction time (663 msec after acoustic onset, i.e., 463 msec after the onset of the pitch change) did not show any significant main effects of the various AV conditions (acoustic speech versus nonspeech, visual speech versus nonspeech, visual motion, high versus low pitch target).

These findings indicate that task difficulty was comparable across conditions.

MEG Measurements and Data Processing

Using a whole-head device (CTF, Vancouver, Canada; 151 sensors, sampling rate = 312.5 Hz, anti-aliasing filter cutoff = 120 Hz), evoked magnetic fields were recorded across a time interval of 550 msec, starting 150 msec prior to the onset of orofacial speech movements or nonspeech motion of the video display. The initial interval of 150 msec served as a prestimulus baseline. MEG offset was removed from each sensor signal by subtracting its mean baseline value. An automatic software procedure allowed for the detection of eyeblink artifacts (subspace projection onto a prototypical eyeblink dipole structure, threshold = 50 nA), and the respective trials (ca. 5–10%) were discarded from analysis.

Analysis of evoked magnetic fields was performed by dipole analysis using the iterative ''DipoleFit'' procedure of the CTF software (spatio-temporal analysis minimizing residual variance within the entire analysis window considered, fixed orientation). Dipole analysis was based on two approaches.

First, a two-dipole model was fitted, focusing on central auditory processing at the level of the supratemporal plane. One major advantage of the two-dipole model, in comparison to multidipole models, is that comparable dipole pairs can be fitted to each individual dataset separately. Furthermore, this method allows for comparisons with previous studies regarding visually induced effects on auditory M50/M100 fields.

As a second step, a six-dipole model was implemented in order to separate event-related fields arising within the central auditory system from activity bound to other sources. This analysis was performed on group data because it was not possible to fit homologous dipoles in response to visual motion consistently across all subjects. In order to keep spatial errors as small as possible, all subjects were positioned within the MEG device comfortably and consistently in the same way. A further aspect regarding group averaging is that, in the CTF system, sensor locations are registered in head-based coordinates. In order to avoid an arbitrary assignment of sensor locations to the group average, for example, by using a ''standard'' head, the individual sensor locations were averaged across subjects as well, using an external MATLAB routine. In order to estimate the spatial error of source location after group averaging, an auditory two-dipole model was also derived from the group average, allowing for a comparison with the averaged individual coordinates.

The group analysis was restricted to a number of six dipoles (three pairs) as the amount of variance associated with further magnetic sources was too low to provide a sufficient signal-to-noise ratio for statistical analysis. Based on the two- and six-dipole models, two- and six-dimensional response curves (dipole strength across time) were obtained by means of subspace analysis, projecting the MEG sensor data on the variance components associated with the respective dipole sources. In order to avoid an overestimation of common variance among the dipole sources, all dipoles of the respective model were entered synchronously.

Anatomical MRI Data

Anatomical MRI datasets could be obtained from 17 out of the total of 25 participants and were transformed into the head-related coordinates of the MEG device (orthogonal axes based on two preauricular points and the nasion, resolution = 1 mm, 256 × 256 × 256 voxels). As a head model for the six-dipole analysis, all 17 MRI datasets were pooled (voxelwise averaging and gray-scale normalization to the dynamic range of the display program; ''MRIViewer,'' CTF, Vancouver). Despite obvious individual variability of head size and shape, the MRI group average still displays the relevant anatomic structures.

RESULTS

Auditory Dipole Source Analysis

On the basis of the acoustic-only trials, a two-dipole model was created, representing left- and right-hemisphere

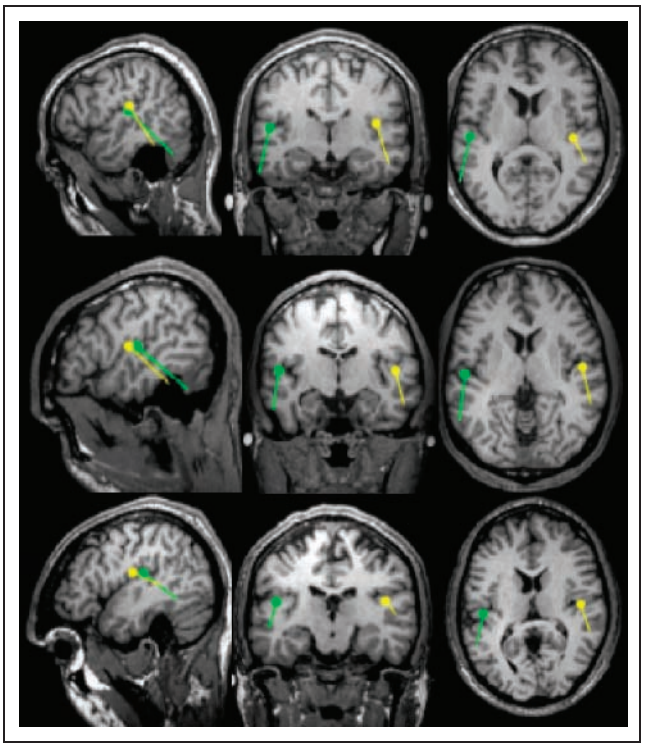

Figure 2. Auditory dipole source location (small circles) and orientation (tails) in three subjects, projected onto their anatomical MRI scans. The selected MRI planes correspond to the level of the left-hemisphere dipoles (green), the right dipoles might be localized within different slices, but are displayed as if the brain was transparent.

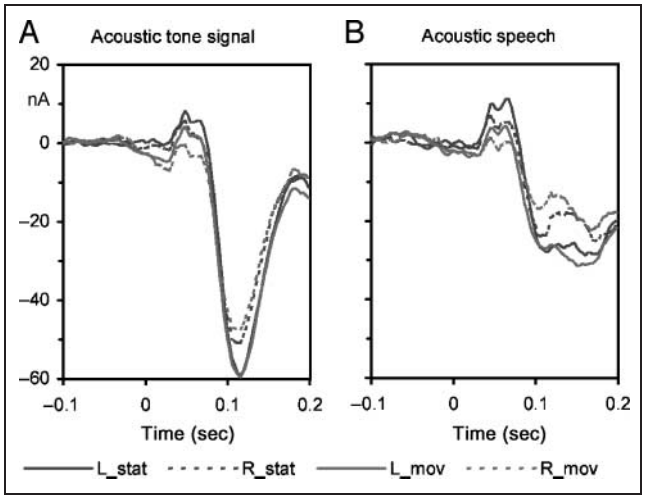

Figure 3. Time course of subspace projections onto the left (solid line) and right (broken line) dipoles of the auditory two-dipole model: Effect of large movement excursions (red) in comparison to the static conditions (blue), pooled across the visual speech and nonspeech conditions (A = acoustic nonspeech condition; B = acoustic speech condition). Zero on the time scale indicates the onset of the acoustic signal. L_, R_ = left, right hemisphere; stat = static display; mov = moving stimulus.

auditory source locations. A single pair of dipoles was derived from each individual's average across all acoustic-only trials, minimizing residual variance within a time interval extending from 30 to 130 msec after acoustic stimulus onset, thus encompassing both the M50 as well as the M100 fields. Consistently, dipole sources could be assigned to the supratemporal plane as exemplified in Figure 2. Subspace projection of the 151 MEG channels onto these sources (Figure 3) yielded a typical response pattern, comprising a double-peaked (ca. 50 and 75 msec) M50 deflection and a subsequent M100 component of an inverse polarity (see e.g., Ackermann, Hertrich, Mathiak, & Lutzenberger, 2001). In the following sections, the impact of the various experimental conditions upon these response curves will be analyzed. For the sake of statistical comparison of the M50 and M100 dipole moments, two time windows were selected, extending from 45 to 85 msec (M50) and from 100 to 140 msec (M100), respectively. For each time window, the average dipole moment was computed and then used as the dependent variable.

Impact of Visual Motion on Auditory-evoked M50 and M100 Fields

In order to assess the influence of visual motion on auditory M50 and M100 fields, an ANOVA was performed including all trials with an acoustic stimulus component. *ATYP* (= acoustic signal type: speech vs. nonspeech), *VTYP* (= visual signal type: speech vs. nonspeech), *MOT* (3 stages of visual motion: [1] no movement, i.e., static face or circle pattern, [2] small movements of the mouth (/ta/) or the circles, and [3] larger excursions of the mouth (/pa/) or the circles), and *SIDE* (hemisphere) were entered as the independent factors. Assuming the impact of visual information upon M50/M100 to represent a linear function of movement range, the contrast of a static display with visual /pa/ or the large circle movements can be expected to elicit a stronger effect by about the factor two as compared to the respective small movement excursions. Based upon these suggestions, the factor *MOT* was decomposed using a polynomial model in order to assess, first, the impact of large movements versus the static display on M100 strength (linear component, *MOT.1*) and, second, the deviation of the M100 deflection evoked by small speech and nonspeech motion from a linearly scaled intermediate response between the static and the large movement conditions (nonlinear component; *MOT.2*).

M50 Field

The acoustic speech signals elicited a stronger M50 field as compared to the acoustic tone signals (main effect of ATYP; compare Figure 3A and B). Visual motion gave rise to a consistent attenuation of the M50 field (main effect of MOT.1). Furthermore, a three-way MOT.1 × ATYP × SIDE interaction could be observed in that acoustic speech stimuli elicited a more pronounced left-hemisphere motion-induced M50 suppression (solid lines in Figure 3B) as compared to the nonspeech condition (Figure 3A). This effect might be due to the particularly enlarged left-hemisphere M50 amplitude under the acoustic speech condition in the absence of motion. Separate post hoc analyses revealed a significant MOT.1 × ATYP interaction over the left hemisphere [$F(1, 24) = 5.90$, $p = .023$], but not the right hemisphere [$F(1, 24) = 0.04$, $p > .1$]. MOT.2 did not yield any significant main effects or interactions within the time domain of the M50 fields, indicating that small movements (or /ta/ in case of the face videos) did not give rise to a significant deviation from an intermediate response between the static and the large (or /pa/) condition.

M100 Field

Auditory nonspeech signals elicited stronger M100 fields than the acoustic speech stimuli (main effect of ATYP; compare Figure 3A and B) and were found to be stronger over the left as compared to the right hemisphere (main effect of SIDE; Figures 3 and 4, dashed vs. solid lines). In contrast to M50, visual motion did not show a consistent main effect on the M100 field, but interacted with SIDE and with the visual condition VTYP (face vs. circle pattern), in addition to a main effect of VTYP (Table 2). Figure 4 portrays the direction of these effects and inter-

Table 2. Repeated Measures ANOVA: Effects of AV Conditions and Visual Motion on M50 and M100 Amplitude

Field Component	*Effect*	*F(1, 24)*	*p*
M50	ATYP	7.13	.013
	MOT.1	11.00	.003
	ATYP × MOT.1 × SIDE	4.26	.049
M100	ATYP	108.28	<.001
	VTYP	9.93	.004
	SIDE	9.77	.005
	VTYP × MOT.1	6.00	.022
	MOT.1 × SIDE	11.69	.002
	ATYP × MOT.1 × SIDE	5.33	.029
	VTYP × MOT.2 × SIDE	4.21	.051

Independent factors: ATYP (acoustic speech versus nonspeech); VTYP (visual circle pattern versus face video); MOT.1: effect of large movement (or /pa/ video) versus static video display; MOT.2: deviation of responses to small motion (or /ta/ video) from an intermediate response between the static and the large condition; SIDE: left versus right hemisphere.

actions, displaying the responses pooled across the two acoustic (speech and nonspeech) conditions. The right panel (Figure 4B, visual face condition) demonstrates a stronger overall M100 field as compared to the nonface condition (Figure 4A). The effect visual /pa/ articulation versus static face can be described as a left-hemisphere M100 enhancement (solid red line in Figure 4B), whereas the moving circle pattern gave rise to a predominant attenuation of M100 over the right hemisphere (Figure 4A). Post hoc analyses revealed significant MOT.1 × SIDE interactions for both the face [$F(1, 24) = 9.53$, $p = .005$] and the circle conditions [$F(1, 24) = 7.81$, $p = .010$]. Furthermore, a three-way ATYP × MOT.1 × SIDE interaction could be noted (Table 2), suggesting differential lateralization effects of visual motion, depending on whether the videos were paired with an acoustic speech or a nonspeech signal. Post hoc analyses revealed significant MOT.1 × SIDE interactions for both the acoustic speech [$F(1, 24) = 15.66$, $p < .001$] and the nonspeech tone conditions [$F(1, 24) = 5.74$, $p = .025$]. The directions of these interactions are shown in Figure 3. M100 attenuation was most pronounced over the right hemisphere during the acoustic speech condition (dashed red line in the right panel of Figure 3).

Regarding the expected nonlinear impact of visual motion upon the auditory M100 field, Table 2 shows a tendency of a three-way MOT.2 × VTYP × SIDE interaction. In order to further elucidate this interaction, a post hoc analysis was performed, considering separately speech and nonspeech visual motion. The small nonspeech excursions (green lines in Figure 5) gave rise to a bilateral M100 amplitude intermediate between the static and the large condition (upper panels of Figure 5). Accordingly, neither the MOT.2 effect nor the MOT.2 × SIDE interaction reached significance ($p > .4$). By contrast, the articulating face showed an asymmetric nonlinear response pattern as indicated by a significant MOT.2 × SIDE interaction [$F(1, 24) = 4.54$, $p = .044$]: Visual /pa/ or /ta/ articulation did not change the M100 amplitude over the right hemisphere as compared to the static face condition, whereas the left hemisphere showed visual motion-induced M100 enhancement only for /pa/, but not for /ta/ syllables (lower left panel of Figure 5).

Taken together, visual motion had a differential impact on auditory M50 and M100 fields. Whereas visual speech and nonspeech motion were found to consistently dampen the M50 amplitude, these two conditions had a differential impact upon the M100 field. It should also be mentioned that no significant VTYP × ATYP interactions were observed, indicating that the effects of speech versus nonspeech motion are not specifically associated

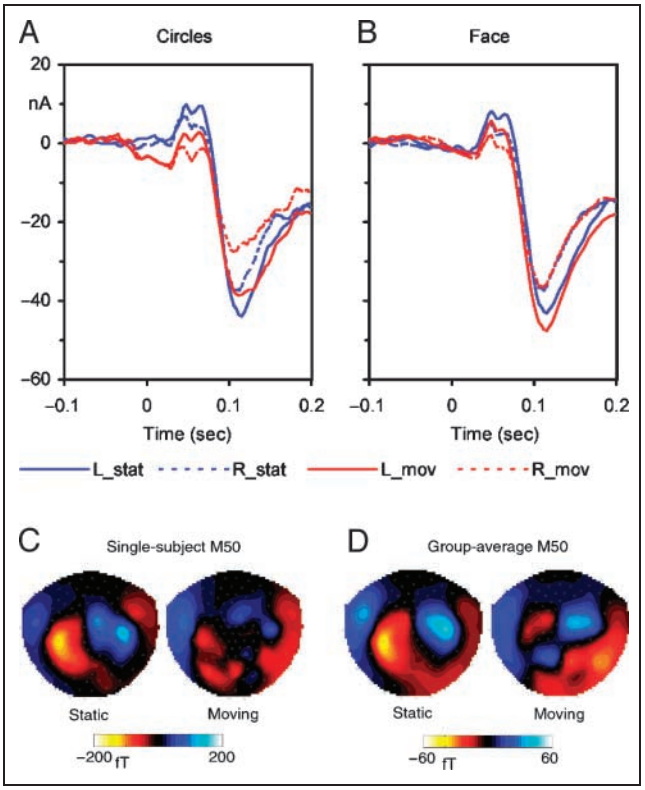

Figure 4. (A and B) Time course of subspace projections onto the left (solid lines) and right (broken lines) dipoles of the auditory two-dipole model: Effect of large movement excursions (red) in comparison to the static conditions (blue), pooled across the acoustic speech and nonspeech conditions (A = visual nonspeech condition; B = visual speech condition). Zero on the time scale indicates the onset of the acoustic signal. L_, R_ = left, right hemisphere; stat = static display; mov = moving stimulus. (C and D) M50 surface maps (45–85 msec after acoustic signal onset) demonstrating motion-induced M50 suppression: (C) Single subject, moving circle pattern versus static display, acoustic conditions pooled; (D) Group average, acoustic speech signal, visual speech and nonspeech conditions pooled.

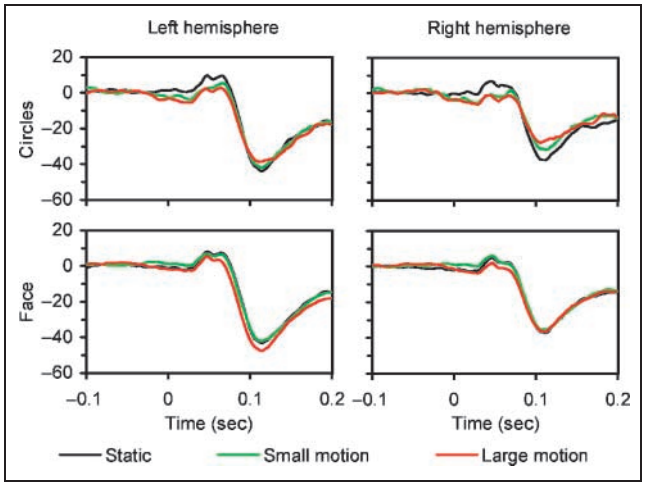

Figure 5. Time course of subspace projections onto the left (left panels) and right (right panels) dipoles of the auditory two-dipole model: Effect of large (red) and small (green) movement excursions in comparison to the static conditions (black), pooled across the acoustic speech and nonspeech conditions. Upper panels: visual nonspeech condition; lower panels: visual speech condition. Zero on the time scale indicates the onset of the acoustic signal.

with acoustic signal type (syllable vs. tone signal). Finally, the sole interaction of visual motion with ATYP (i.e., the MOT.1 × ATYP × SIDE interaction), was due to right- rather than left-hemisphere effects.

Effects of Visual Motion on Auditory Dipoles in the Absence of an Acoustic Signal

The auditory dipole model was also applied to the silent visual-only trials with large motion cues (or /pa/, respectively) in order to address the question in how far the impact of visual signals upon M50 and M100 fields depends on the presence of an acoustic signal. Within the M50 time window, a consistent effect of visual motion with a polarity opposite to the auditory M50 could be observed (one-sample t test: $p = .013$; signed-rank test: $p < .001$), and repeated measures ANOVA did not yield any significant effects of ATYP, VTYP, and SIDE. Thus, a significant influence of visual signals upon the M50 field (in terms of the applied dipole model) emerged even in the absence of an acoustic signal.

Within the time domain of the M100 field, the moving circle pattern and the speaking face had different effects on the auditory dipole moment, as indicated by a significant main effect of VTYP [$F(1, 24) = 6.74$, $p = .016$]: Motion effects of visual /pa/ had the same polarity as the acoustically evoked M100 field, whereas nonspeech motion gave rise to a response with an opposite polarity. Considering hemispheric lateralization effects, a more complex pattern was observed in that SIDE [main effect: $F(1, 24) = 5.31$, $p = .030$] interacted with VTYP [$F(1, 24) = 5.31$, $p = .030$] and with ATYP [$F(1, 24) = 10.38$, $p = .004$]. Furthermore, the VTYP × ATYP interaction achieved significance [$F(1, 24) = 6.29$, $p = .019$]. Because the stimuli underlying this analysis did not have an acoustic component, this interaction reflects differences with respect to an omitted acoustic speech versus nonspeech signal (implemented in different runs of the MEG session). In line with the assumption of a left-hemisphere mechanism of phonetic encoding, the strongest effect with a polarity in the direction of the auditory M100 was observed over the left hemisphere under the visual /pa/ condition in association with the expectation of an acoustic speech signal. In order to exclude the possibility that expectation alone, in absence of any visual motion, might have caused a deviation from baseline just as a result of the regular interstimulus interval, the ''empty'' trials (nonmoving silent stimuli) were analyzed as well. In this case, no relevant deviation from baseline could be observed, and statistical analysis did not yield any significant effects of the experimental factors (static face vs. circle pattern; omitted speech vs. nonspeech signal).

Six-dipole Model of Visual and Auditory Activations

Using group data, a six-dipole model of AV activations was created, providing the basis for the following analyses. In order to obtain an estimate of the spatial error associated with the computation of group dipoles, Table 3 compares the averages across individual dipole fits from the above two-dipole analysis with dipole locations and orientations based on group-averaged MEG data. As shown in Table 3, the spatial difference amounted to 1–2 mm within each coordinate. Across all six parameters, the group dipole deviates from the mean of individual dipoles by approximately one standard error only ($n = 25$), suggesting that under the recording conditions of the present study, the spatial error of source locations derived from group averages can be tolerated.

Source modeling of the six-dipole model was performed in three steps in an interactive way by inspection of brain maps and overlaid response curves from all sensors. First, a pair of auditory dipoles (A110) was fitted to the M100 peak of averaged responses to acoustic-only stimuli (ca. 110 msec after acoustic signal onset, pooled across speech and nonspeech). Residual variance amounted to less than 3%, and the location of these dipoles was found to be nearly identical to the sources of the auditory two-dipole model.

Second, MEG responses to visual-only stimuli with large movement excursions, pooled across the speech (/pa/) and nonspeech (large circle movements) conditions, were averaged across all subjects. A strong visual motion-induced MEG activity characterized by a dipolar field distribution could be observed, peaking ca. 170 msec after the onset of visual motion (V170; Figure 6A), and thus,

Table 3. Auditory Dipole Location and Orientation, Group Means, and Standard Error of Dipole Parameters Fitted for Each Subject Separately (n = 25) versus Group Dipole Fitted on the Basis of MEG Data Averaged across Subjects

Parameter	*Group Mean*	*Standard Error*	*Group Dipole*
Left x	−0.03	0.26	−0.22
Left y	4.45	0.33	4.56
Left z	6.09	0.26	5.90
Left d(x)	0.57	0.11	0.63
Left d(y)	−0.19	0.08	−0.10
Left d(z)	0.73	0.07	0.77
Right x	0.30	0.24	0.19
Right y	−4.47	0.30	−4.61
Right z	6.26	0.35	6.09
Right d(x)	0.51	0.13	0.57
Right d(y)	0.26	0.06	0.18
Right d(z)	0.74	0.07	0.80

x, y, and z are the anterior–posterior, lateral, and vertical dimensions of the head-related coordinate system based on the nasion and the two preauricular reference points (in cm); d(x), d(y), and d(z) refer to the respective orientations. Note that the deviance of the group dipoles from the group mean of individual dipoles is in the order of one standard error.

preceding the time domain of the M50 response evoked by acoustic stimuli by about 40 msec. This field pattern could be modeled by a pair of dipoles within the anterior medial region of the occipital cortex, accounting for ca. 85% of the variance of the data.

Thirdly, the second-strongest response to visual motion approximately coincided with the time window of the auditory M100 field, peaking ca. 270 msec after the onset of visual motion (V270; Figure 6B). This activity could be modeled (ca. 90% variance) by a pair of dipoles located within posterior parts of the insula. When separate dipole fits were performed with responses to visual /pa/ and fast nonspeech motion, quite similar dipole locations were found (differences less than ca. 0.5 mm in each dimension).

All three dipole pairs (A110, V170, and V270) were combined to a six-dipole model, individual response curves were derived by subspace projection (Figure 7), and statistical analysis, as in case of the two-dipole analysis, considered mean dipole moments within the M50 and M100 time windows as dependent variables.

AV Interactions on M50 and M100 Strength Based upon the Six-dipole Model

Because no significant ATYP × VTYP interactions had been observed in the preceding two-dipole analysis, only the runs with congruent AV constellations (i.e., talking face and moving cycles) were considered for analysis.

M50 Field

In line with the two-dipole analysis, the auditory dipoles still showed a highly significant motion effect on the A110 source [MOT.1: $F(1, 24) = 19.94$, $p < .001$]. In addition, the V170 [MOT.1: $F(1, 24) = 4.49$, $p = .045$] and V270 [MOT.1: $F(1, 24) = 4.49$, $p = .045$] dipole moments also exhibited visual motion effects above the significance threshold ($p < .05$) within this time window. Figure 7 shows the time course of motion effects (difference between black and red line) in the subspace projection onto the A110, V170, and V270 dipoles,

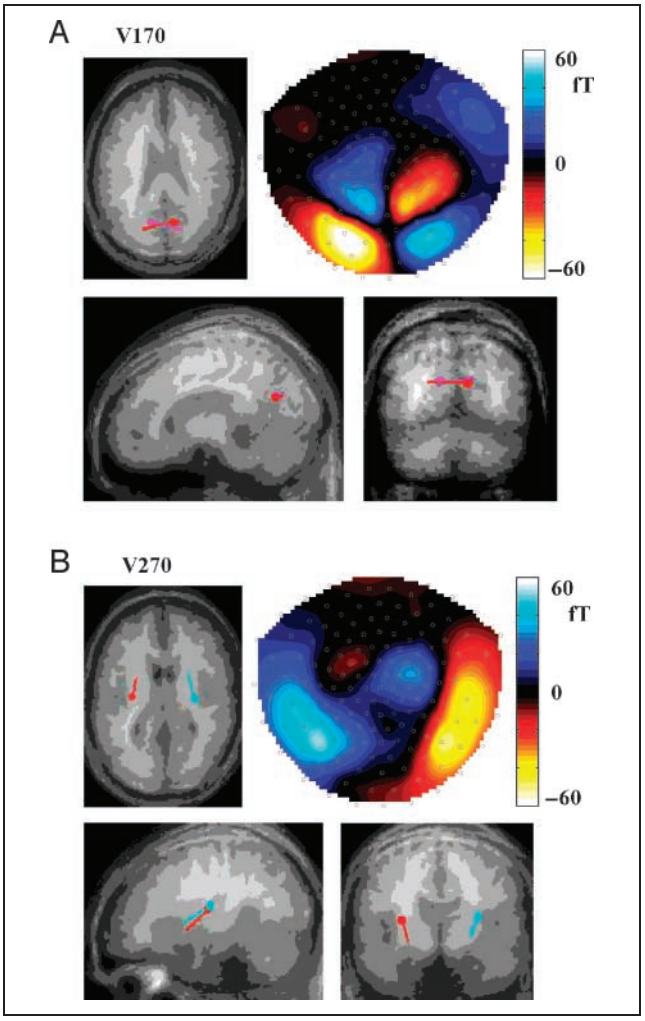

Figure 6. MEG brain maps (upper right panels) of (A) the V170 (170 msec after motion onset) and (B) the V270 fields (270 msec after motion onset), concomitant with the anatomical location of source dipoles (small circles = position; tails = orientation). Data are based on group averages (n = 25) pooled across all visual-only conditions. The anatomical MRI pictures were also averaged across subjects (n = 17). The displayed slices correspond to the left-hemisphere dipoles.

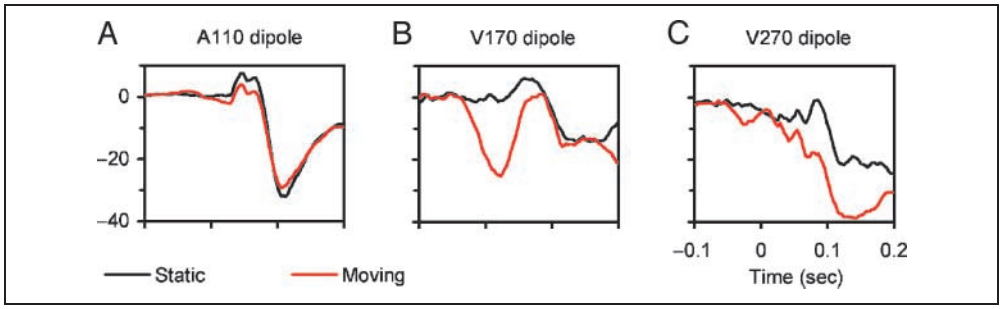

Figure 7. Time course of subspace projections onto the A110 (A), the V170 (B), and the V270 dipoles (C) of the six-dipole model (pooled across both hemispheres): Effect of large visual motion (red) in comparison to the static conditions (black), pooled across the speech and nonspeech AV conditions. Zero on the time scale indicates the onset of the acoustic signal.

averaged across both hemispheres and the speech and nonspeech conditions.

M100 Field

Repeated measures ANOVAs were performed separately for the speech and nonspeech conditions. Intrasubject factors included ACU (= presence or absence of an acoustic signal), MOT.1, MOT.2, and SIDE, and the strength of the A110, V170, and V270 dipoles within the M100 time window (100–140 msec) served as dependent variables. Significant motion effects and interactions are summarized in Table 4. Although the V170 dipole shows a strong overall motion effect at the time of its peak amplitude [150–190 msec after motion onset, i.e., 0–40 msec after acoustic onset; MOT.1: $F(1, 24) = 20.24$, $p < .001$; mid panel of Figure 7], subspace projection onto this dipole component within the M100 time window did not yield any relevant motion effects. A110 dipole strength showed an ACU × MOT.1 interaction both under the speech and the nonspeech conditions. The pattern of this interaction, however, varied across conditions. Visual motion caused an attenuation of the A110 deflection (Figure 8, upper left panel) under the nonspeech condition, depending upon the presence of an acoustic signal. By contrast, visual speech did not influence the A110 deflection in the presence of an acoustic signal (Figure 8, lower left panel), and silent speech motion stimuli yielded an effect on A110 dipole strength with the same polarity as the auditory M100. V270 dipole strength showed a different pattern of motion effects within the auditory M100 time window. In case of both nonspeech and speech, the V270 showed a strong effect of the large movements or visual /pa/, respectively (effect of MOT.1 in Table 4, red lines in the right panels of Figure 8). Furthermore, intermediate motion levels (smaller circle movement or smaller lip movement in case of visual /ta/) yielded significant deviations from an intermediate response under both conditions (MOT.2 effects in Table 4), but in different directions. In case of nonspeech, the intermediate movement gave rise to a response similar to the one

Table 4. Repeated Measures ANOVAs: Linear (MOT.1) and Nonlinear (MOT.2) Effects of Visual Motion and Interactions with Hemisphere (SIDE) and the Presence or Absence of an Acoustic Signal (ACU) on A110, V170, and V270 Dipole Strength within the Time Window of the Auditory M100 (100–140 msec)

Condition	*Dipole*	*Effect*	*F(1, 24)*	*p*
(a) Nonspeech	A110	ACU × MOT.1	5.10	.033
	V170	MOT.2	4.35	.048
	V270	MOT.1	8.04	.009
		MOT.2	4.40	.047
(b) Speech	A110	ACU × MOT.1	5.67	.026
	V270	MOT.1	16.62	<.001
		MOT.2	6.87	.015
		MOT.2 × SIDE	8.42	.008
		ACU × MOT.2 × SIDE	4.82	.038

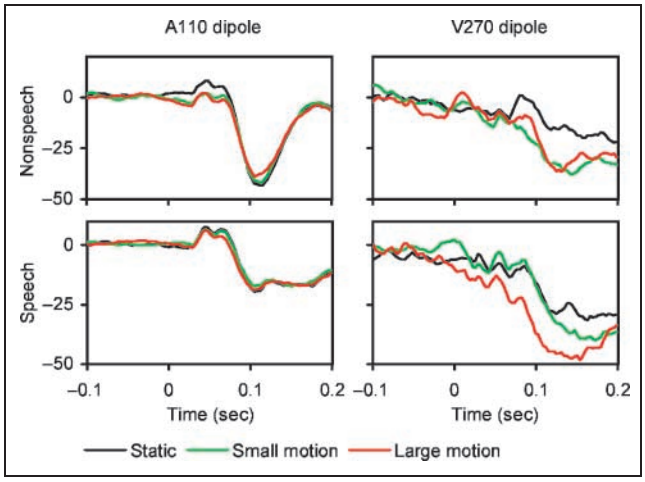

Figure 8. Time course of subspace projections onto the A110 (left panels) and the V270 dipoles (right panels) of the six-dipole model: Effect of large (red) and small (green) movement excursions in comparison to the static conditions (black), under the AV nonspeech (upper panels) and the AV speech conditions (lower panels). Zero on the time scale indicates the onset of the acoustic signal.

following the large movement, whereas in case of visual /ta/ no motion effect was present (compared to the acoustic-only condition). Thus, in both cases, motion had a nonlinear effect. Considering the speech condition, this latter effect interacted, in addition, with hemisphere and the presence or absence of an acoustic signal as indicated by the ACU × MOT.2 × SIDE interaction listed in Table 4. Post hoc analyses were performed separately for the silent and the AV trials: In the presence of an acoustic signal, a bilateral-symmetric response pattern was observed similar to the one displayed by the lower right panel of Figure 8, grouping visual /ta/ with the static condition (suppression of visual /ta/ motion effects). In the absence of an acoustic signal. a significant MOT.2 × SIDE interaction emerged [$F(1, 24) = 8.25$, $p = .008$]. Left-hemisphere V270 in response to /ta/ was suppressed, whereas the right hemisphere showed an intermediate response to /ta/ between /pa/ and the empty condition. The differential motion effects on MEG surface maps are shown in Figure 9 for the silent motion conditions, demonstrating that responses to visual /ta/ are weaker than the responses to /pa/ and to the two nonspeech movements.

In summary, the six-dipole model allowed for a further decomposition of the impact of visual motion upon the M50/M100 complex. Regarding the M50 time window, visual effects were distributed across various brain regions, but still had a significant impact on the auditory dipole source, with an inverse polarity as compared to the auditory-evoked M50. A significant impact of visual motion upon the auditory dipoles also emerged within the M100 analysis window. These latter effects differed between the speech and nonspeech conditions and interacted in both cases with the presence or absence of an acoustic signal (nonspeech: M100 suppression only in the presence of an acoustic signal; speech: effect with the same polarity as the auditory M100 only in case of silent motion stimuli). Furthermore, nonlinear effects of visual motion could be assigned to the V270 dipoles, indicating selective suppression of responses to visual /ta/.

DISCUSSION

Motion-induced M50 Suppression

In line with previous data (Lebib et al., 2003), temporally correlated visual information consistently elicited an attenuation of the auditory M50 field (M50 suppression). In the present study, the onset of visual motion preceded the associated auditory event by an interval of 150 msec. By contrast, Lebib et al. (2003) had introduced an acoustic delay more than twice as large (320 msec). This broad temporal ''tolerance'' of AV asynchrony indicates that the human central auditory system is capable to adapt to regular time delays of up to more than 300 msec (Fujisaki, Shimojo, Kashino, & Nishida, 2003, 2004). Because larger orofacial and nonspeech movements had a stronger effect on the M50 fields than smaller ones, this response of the central auditory system seems to be scaled by movement parameters of the visually displayed motion stimuli.

As compared to the motion-induced M50 suppression of the two-dipole model, the subsequent six-dipole analysis revealed similar modulations of auditory dipole strength (A110 dipole). Therefore, the impact of visual information upon cortical activity within the temporal domain of the M50 can be assumed to arise, at least partially, at the cortical level of the central auditory system. Basically, the same motion effects could be observed in response to visual-only stimuli, that is, in the absence of any associated acoustic stimulus. Thus, early AV interactions are not restricted to ''sensory gating'' of afferent auditory input at a subcortical level of the ascending auditory pathways or to a modulation of cortical responses to an external acoustic signal. Rather, assuming that the visual signal acts as an anticipatory cue for acoustic signal onset, the visual effect on the M50 field might reflect a ''preparatory baseline shift,'' that is, preactivation of cortical sensory areas prior to stimulus onset, in case ''the observer prepares to attend to an anticipated stimulus'' (Driver & Frith, 2000). Such an effect has been demonstrated, for example, for the human visual system by means of functional magnetic resonance imaging [fMRI] (Kastner, Pinsk, De Weerd, Ungerleider, & Desimone, 1999). A variety of animal experiments provided evidence for an influence of visual signals on both the primary and secondary auditory cortex. For example, electrophysiological recordings in several macaque species showed multisensory convergence within the acoustic cortex posterior to A1 (Schroeder & Foxe, 2005). Furthermore, a study on AV processing in ferrets, based upon both intracortical recordings as well as neural tracer techniques, suggested direct inputs from the visual into the auditory cortex as a potential source of origin for visually

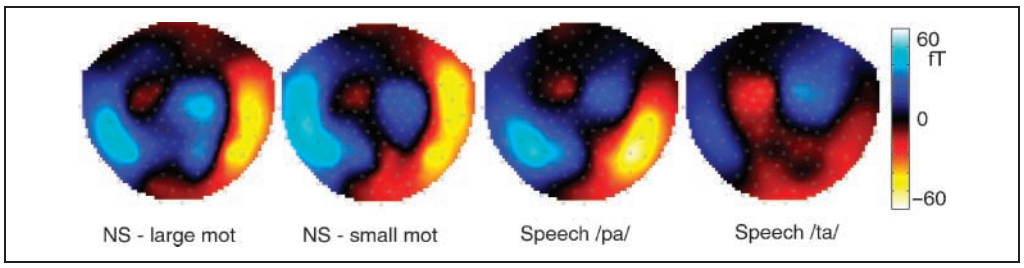

Figure 9. Surface maps of MEG responses to silent visual motion (mot) stimuli within the time window of the auditory M100 response (i.e., 250–290 msec after visual motion onset). From left to right: large and small nonspeech (NS) motion, visual presentation of /pa/ and /ta/.

induced responses within the auditory system (Bizley et al., 2007).

Impact of Visual Motion upon the M100 Field: Two-dipole Model

Van Wassenhove et al. (2005) found the visual component of AV speech signals to dampen the EEG N1 response to the paired acoustic syllables. The present study documented a similar influence of nonspeech(!) motion stimuli upon the M100 field, the MEG equivalent of the N1 deflection, whereas orofacial speech gestures failed to elicit a comparable effect. Conceivably, first, spatial variation of the electromagnetic sources of EEG N1 deflections and magnetic M100 fields and, second, differences in experimental design and task demands might contribute to these discrepancies. Nonetheless, motion-induced attenuation of the N1 and the M100 components, obviously, does not appear to be specifically linked to speech processing.

The differential effects of speech and nonspeech motion on M100 deflections as observed in the present study might reflect a highly automatized impact of perceived orofacial articulatory gestures upon the central auditory system as these movements are tightly bound, in our daily life, to a distinct acoustic signal. By contrast, the presentation of shrinking and expanding circles is not inherently associated with a familiar sound source. It remains to be settled, thus, whether the observed motion-induced M100 enhancement is specifically related to speech perception or whether any authentic, that is, ecologically valid visual signal, representing a natural sound source, may give rise to this effect.

Besides N1 amplitude, Van Wassenhove et al. (2005) found the visual component of AV stimuli to influence N1 latency as well. However, visual inspection of the data of the present study (see Figure 3) does not indicate any impact of visual motion upon the temporal characteristics of the M100 field. Therefore, latency was not considered for further analysis. These discrepancies might again reflect differences in the sources of EEG N1 and MEG M100 deflections.

The absence of a significant interaction between visual and acoustic signal type (VTYP × ATYP) in the present study indicates that at the stage of the auditory M50/M100 field visual motion did not have differential effects on MEG responses to congruent versus incongruent AV signal types. Thus, the influence of visual motion up to the time domain of the M100 does not seem to reflect speech-specific AV fusion effects. Similarly, Hertrich et al. (2007) documented basically the same visual speech motion effects on auditory M100, irrespective of whether these stimuli were paired with acoustic speech or nonspeech signals. Thus, despite obvious early AV interactions at the level of the cortical auditory system, the AV fusion of phonetic features, giving rise to auditory perceptual illusions such as the McGurk effect, seems to occur at a later stage of processing. Also the fact that the sole interaction of visual motion with acoustic signal type (ATYP × MOT.1 × SIDE in Table 2) was due to right- rather than left-hemisphere effects argues against the assumption that visual effects on M100 reflect speech-specific mechanisms of phonetic fusion. Nevertheless, because visual speech and nonspeech motion effects were actually different and, furthermore, interacted with the expectation of an acoustic speech versus nonspeech signal in case of silent motion stimuli, visual effects on auditory M100 cannot be considered just as an unspecific impact of visual motion.

Impact of Visual Motion on the M100 Field: Six-dipole Model

As a second step of analysis, a six-dipole model was created in order to separate central auditory responses (A110 dipoles) from (the bulk of) visual motion-related MEG activity (V170 and V270 dipoles). Even this more fine-grained analysis revealed a significant modulation of the auditory dipole moments by visual motion. These findings support the assumption of an influence of visual signals on the central auditory system. Besle et al. (2004) had reported cross-modal hypoadditive effects on event-related potentials in response to AV speech stimuli. Similarly, in the present study, the impact of visual /pa/ on auditory dipole strength was larger in the absence than in the presence of an acoustic signal. Conceivably, auditory events give rise to saturation effects under these conditions. As an alternative, acoustic stimulation could ''protect'' the central auditory system within the time window of the M100 field against direct visually induced activation. In the absence of acoustic signals, visual stimuli may have access to those brain areas, resulting, for example, in auditory imagery phenomena.

A variety of studies reported visual motion to elicit a characteristic MEG deflection at a latency of about 170 msec (Miki, Watanabe, Kakigi, & Puce, 2004; Tsuda & Ueno, 2000; Ahlfors et al., 1999), bound to an occipito-temporal source (area MT) of a more lateral location as compared to the V170 dipoles of the present study. In contrast to the V270 source (see below), the V170 component of evoked magnetic fields did not exhibit any significant AV interactions within the time domain of auditory-evoked M100 fields.

The strength of the V270 dipoles showed a significant main effect of motion and significant interactions between movement extent (large vs. small), on the one hand, and the speech/nonspeech distinction, on the other: Large and small excursions of the nonspeech stimuli yielded similar bilateral effects, significantly different from the static condition. In case of visual speech signals, /pa/ utterances also yielded a significant activation, whereas the response to visual /ta/ was found to be suppressed. Presumably, this suppression effect reflects the phonological status of visual /ta/, differing from /pa/

in terms of underspecification or markedness (see below). Because the experimental design of the present study required pitch detection rather than phoneme recognition, it might be expected that the phonetic structure of the speech stimuli had no impact upon the evoked brain activity and that both speech and nonspeech signals just operate as predictors of acoustic signal onset. However, speech recognition is a highly automatized process, and various studies based on the mismatch paradigm (Phillips et al., 2000; Näätänen et al., 1997) have shown that explicit attention to single features is not required for early stages of phonological encoding.

Hertrich et al. (2007) suggested AV fusion of categorical speech information, such as the integration of visible labial movements into the auditory percept of the syllable /pa/, to occur at a quite late level of processing (ca. 275 msec after acoustic onset). In the light of these data, the present nonlinear effects of visual /pa/ versus /ta/ upon V270 dipole strength appear to pertain to a computational stage preceding the fusion of auditory and visual information into a common phonetic representation. Because the strength of the A110 source of the six-dipole model did not show comparable interactions, the source of this visual motion effect can be assumed to be localized outside the supratemporal plane. To be more specific, based upon the averaged MRI images from 17 subjects, the V270 dipoles could be attributed to the posterior insular region. Because this dipole fit was characterized by a larger residual variance than the auditory-evoked M100 fields, any attempt to localize this source must be considered with some precautions and awaits further confirmation, for instance, by means of fMRI studies. Nevertheless, several studies reported a contribution of intrasylvian cortex to AV interactions (Fort & Giard, 2004; Bushara, Grafman, & Hallett, 2001; Calvert, 2001; Calvert, Hansen, Iversen, & Brammer, 2001). For example, Bushara et al. (2001) suggest the insula to mediate ''temporally defined auditory–visual interaction at an early stage of cortical processing, permitting phenomena such as the ventriloquist and the McGurk illusions.'' Furthermore, an fMRI study by Noesselt, Shah, and Jäncke (2003) provided evidence for a participation of the posterior insula in top–down modulated segregation of phonetic information. As a further support for the engagement of the insula in higher-order visual operations, intracortical recordings in monkeys revealed neurons within the posterior insula to be sensitive to reward-predicting visual cues and to differentially respond to go/no-go trials (Asahi et al., 2006). Thus, this area appears to support some kind of binary stimulus evaluation.

An essential processing stage of speech perception is the transformation of ''analogue'' data structures, representing stimulus parameters, for example, within tonotopic maps or as phase-locked periodic activity, to a categorical code in terms of phonological units, that is, abstract information-bearing elements. A variety of phonological processes such as assimilation, epenthesis, and reduction indicate that phonetic features, that is, the elementary information units of speech sounds are organized in an asymmetric and hierarchical manner, some units being ''marked'' and others being characterized by an ''underspecified'' structure (De Lacy, 2006; Wheeldon & Waksler, 2004). As a rule, a higher rank within the markedness hierarchy has been assigned to the labial feature for place of articulation as compared to its coronal competitor (De Lacy, 2006; Harris & Lindsey, 1995; Avery & Rice, 1989). Although the integration of auditory and visual input into common sound categories and, subsequently, into a unique auditory percept, appears to be bound to a time window succeeding the M100 field (see above), the findings of the present study suggest that visual information might be shaped by phonological–linguistic structures even prior to its fusion with the auditory channel. Presumably, the nonlinear impact of visible speech upon the V270 dipole moment reflects the working characteristics of a threshold detector sensitive to phonological features, that is, a filtering process separating ''unmarked'' from ''marked'' information, and thus, mapping movement parameters, such as the range of lip excursions, on a binary phonetic–linguistic distinction. Such mechanisms may contribute to the human ability of fast categorization of continuous visual speech input during lipreading.

Methodological Considerations

The assignment of MEG responses to cortical sources is a critical issue because, in some cases, it is not possible to find unambiguous dipole solutions accounting for an observed surface pattern. Furthermore, dipole analysis is based on the assumption that sources have a point-like structure which may actually not be the case. Therefore, any attempt to localize brain activity on the basis of MEG data has to be considered with some precautions. The present two-dipole model was fitted carefully to each subject's MEG data, and in the 17 participants with anatomical MRI data these two sources consistently could be assigned to the supratemporal plane. However, it cannot be ruled out that additional sources outside the cortical auditory system also project part of their variance onto these structures, giving rise to the erroneous suggestion of a direct influence of visual motion on central auditory processing. Thus, an attempt was made to separate the contribution of auditory and nonauditory sources to visually induced modulation of M50 and M100 fields. Because preliminary analysis of individual datasets had indicated that consistent multi-dipole analyses cannot be applied at the level of single subjects, this analyses was based upon pooled group data. As a validation, the spatial coordinates and orientations of the individual auditory dipole sources were averaged across subjects and were compared to dipole

specifications derived from group-averaged MEG data. As shown in Table 3, the locations and orientations were quite similar. Thus, a tolerable spatial error, for instance, due to individual differences in head positions, can be assumed. Because, admittedly, inter-subject variability in source locations results in a spatially smoothed averaged field pattern, the group-based equivalent dipole might be located systematically deeper in the brain than the average of the true sources. Such a tendency, however, cannot be seen in Table 3 where the group dipole had even a slightly more lateral position as compared to the mean of individual dipoles. Regarding the V270 source location, the performed dipole fit minimized residual variance. When the dipole fit procedure was repeatedly performed with different starting positions across the entire cortex, consistently, the same dipole locations were found. In fact, by manually posing dipoles in lateral temporo-occipital regions, somewhat similar lead fields could be obtained, but in these cases, the residual variance was clearly larger, and if these lateral sources were additionally entered into the model, their dipole moment was considerably lower as compared to the V270 dipoles. Of course, this does not definitely exclude the potential contribution of other visual sources to the observed field patterns. In this respect, further experiments using different methodology may be required to refine the present assumptions regarding functional anatomy.

Conclusion

The present study allows for a further characterization of early AV interactions within the time domain of the M50/M100 components (Figure 10 illustrates the various stages of AV information processing):

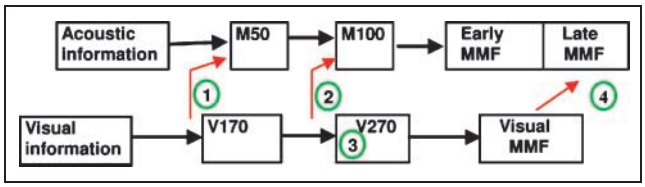

Figure 10. Sequence of interactions between the auditory and visual information streams, derived from the MEG data of the present and a preceding study (the numbers within the green circles refer to distinct effects of visual motion on evoked magnetic fields during AV speech perception): (1) Bilateral speech-unspecific attenuation of auditorily evoked M50 fields (preparatory baseline shift). (2) AV interactions at the level of the auditorily evoked M100 component, indicating a pre-representational differential impact of visual speech and nonspeech information (speech signals = hypoadditive M100 enhancement; nonspeech signals = M100 attenuation). (3) Phonetic–linguistic weighting of visual input outside the auditory system: left-hemisphere suppression of /ta/. (4) Cross-modal sensory memory operations developing into a fused phonetic percept as indicated by a speech-specific visually induced left-lateralized late (275 msec) mismatch field (MMF) component (based on Hertrich et al., 2007).

(a) Both speech and nonspeech motion stimuli elicited an attenuation of auditory-evoked M50 activity. Because similar visual influences emerged within this time domain also in the absence of acoustic stimulation, visual events might act as a precue eliciting preparatory baseline shifts at the level of the central auditory system.

(b) Within the temporal domain of the M100 field, visual speech and nonspeech motion stimuli yielded different response patterns (speech = M100 enhancement, nonspeech = M100 attenuation).

(c) Nonlinear visual effects, indicating sensitivity of evoked magnetic responses to phonetic–categorical information such as the presence or absence of a labial feature, were localized outside the auditory system. Thus, categorical–phonetic information appears already to shape the visual processing stream prior to its fusion with the auditory channel. Presumably, this ability of a fast categorization of continuous visual speech input is engaged during lipreading.

Acknowledgments

This study was supported by the German Research Foundation (DFG; SFB 550/B1). We thank Maike Borutta for excellent technical assistance.

Reprint requests should be sent to Ingo Hertrich, Department of General Neurology, University of Tübingen, Hoppe-Seyler-Str. 3, D-72076 Tübingen, Germany, or via e-mail: ingo.hertrich@uni-tuebingen.de.

REFERENCES

Ackermann, H., Hertrich, I., Mathiak, K., & Lutzenberger, W. (2001). Contralaterality of cortical auditory processing at the level of the M50/M100 complex and the mismatch field: A whole-head magnetoencephalography study. *NeuroReport, 12,* 1683–1687.

Ahlfors, S. P., Simpson, G. V., Dale, A. M., Belliveau, J. W., Liu, A. K., Korvenoja, A., et al. (1999). Spatiotemporal activity of a cortical network for processing visual motion revealed by MEG and fMRI. *Journal of Neurophysiology, 82,* 2545–2555.

Asahi, T., Uwano, T., Eifuku, S., Tamura, R., Endo, S., Ono, T., et al. (2006). Neuronal responses to a delayed-response delayed-reward go/nogo task in the monkey posterior insular cortex. *Neuroscience, 143,* 627–639.

Avery, P., & Rice, K. (1989). Segment structure and coronal underspecification. *Phonology, 6,* 179–200.

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience, 20,* 2225–2234.

Bizley, J. K., Nodal, F. R., Bajo, V. M., Nelken, I., & King, A. J. (2007). Physiological and anatomical evidence for multisensory interactions in auditory cortex. *Cerebral Cortex, 17,* 2172–2189.

Boutros, N. N., & Belger, A. (1999). Midlatency evoked potentials attenuation and augmentation reflect different aspects of sensory gating. *Biological Psychiatry, 45,* 917–922.

Brosch, M., Selezneva, E., & Scheich, H. (2005). Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. *Journal of Neuroscience, 25,* 6797–6806.
Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H. J., & Jäncke, L. (2005). Scanning silence: Mental imagery of complex sounds. *Neuroimage, 26,* 1119–1127.
Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory–visual stimulus onset asynchrony detection. *Journal of Neuroscience, 21,* 300–304.
Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex, 11,* 1110–1123.
Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science, 276,* 593–596.
Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage, 14,* 427–438.
Colin, C., Radeau, M., Soquet, A., & Deltenre, P. (2004). Generalization of the generation of an MMN by illusory McGurk percepts: Voiceless consonants. *Clinical Neurophysiology, 115,* 1989–2000.
De Lacy, P. (2006). *Markedness: Reduction and preservation in phonology (Cambridge Studies in Linguistics, Vol. 112)*. Cambridge: Cambridge University Press.
Driver, J., & Frith, C. (2000). Shifting baseline in attention research. *Nature Reviews Neuroscience, 1,* 147–148.
Fort, A., & Giard, M. H. (2004). Multiple electrophysiological mechanisms of audio-visual integration in human perception. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 503–513). Cambridge: MIT.
Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2003). Recalibration of audiovisual simultaneity by adaptation to a constant time lag. *Journal of Vision, 3,* 34a.
Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience, 7,* 773–778.
Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Journal of Neuroscience, 25,* 5004–5012.
Giard, M. H., & Peronnet, F. (1999). Auditory–visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11,* 473–490.
Harris, J., & Lindsey, G. (1995). The elements of phonological representation. In J. Durand & F. Katamba (Eds.), *Frontiers of phonology: Atoms, structures, derivations* (pp. 34–79). Harlow, Essex: Longman.
Hertrich, I., & Ackermann, H. (1999). A vowel synthesizer based on formant sinusoids modulated by fundamental frequency. *Journal of the Acoustical Society of America, 106,* 2988–2990.
Hertrich, I., & Ackermann, H. (2007). Modelling voiceless speech segments by means of an addditive procedure based on the computation of formant sinusoids. In P. Wagner, J. Abresch, S. Breuer, & W. Hess (Eds.), *Proceedings of the 6th ISCA Workshop on Speech Synthesis* (pp. 178–181). Bonn: University of Bonn.
Hertrich, I., Mathiak, K., Lutzenberger, W., & Ackermann, H. (2004). Transient and phase-locked evoked magnetic fields in response to periodic acoustic signals. *NeuroReport, 15,* 1687–1690.
Hertrich, I., Mathiak, K., Lutzenberger, W., Menning, H., & Ackermann, H. (2007). Sequential audiovisual interactions during speech perception: A whole-head MEG study. *Neuropsychologia, 45,* 1342–1354.
Jäncke, L., & Shah, N. J. (2004). Hearing syllables by seeing visual stimuli. *European Journal of Neuroscience, 19,* 2603–2608.
Kastner, S., Pinsk, M. A., De Weerd, P., Ungerleider, L. G., & Desimone, R. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22,* 751–761.
Lebib, R., Papo, D., De Bode, S., & Baudonniere, P. M. (2003). Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation. *Neuroscience Letters, 341,* 185–188.
MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24,* 253–257.
MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., et al. (2000). Silent speechreading in the absence of scanner noise: An event-related fMRI study. *NeuroReport, 11,* 1729–1733.
Miki, K., Watanabe, S., & Kakigi, R. (2004). Interaction between auditory and visual stimulus relating to the vowel sounds in the auditory cortex in humans: A magnetoencephalographic study. *Neuroscience Letters, 357,* 199–202.
Miki, K., Watanabe, S., Kakigi, R., & Puce, A. (2004). Magnetoencephalographic study of occipitotemporal activity elicited by viewing mouth movements. *Clinical Neurophysiology, 115,* 1559–1574.
Möttönen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research, 13,* 417–425.
Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385,* 432–434.
Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin, 125,* 826–859.
Noesselt, T., Shah, N. J., & Jäncke, L. (2003). Top–down and bottom–up modulation of language related areas—An fMRI study. *BMC Neuroscience, 4,* 13.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113.
Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I. P., Mottonen, R., & Sams, M. (2006). Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. *Human Brain Mapping, 27,* 471–477.
Petitto, L. A., Zatorre, R. J., Gauna, K., Nikelski, E. J., Dostie, D., & Evans, A. C. (2000). Speech-like cerebral activity in profoundly deaf people processing signed languages: Implications for the neural basis of human language. *Proceedings of the National Academy of Sciences, U.S.A., 97,* 13961–13966.
Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience, 12,* 1038–1055.
Pujol, J., Deus, J., Losilla, J. M., & Capdevila, A. (1999). Cerebral lateralization of language in normal left-handed people studied by functional MRI. *Neurology, 52,* 1038–1043.
Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, ''unisensory'' processing. *Current Opinion in Neurobiology, 15,* 454–458.
Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory–visual speech perception examined by fMRI and PET. *Neuroscience Research, 47,* 277–287.

Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America, 90,* 1797–1805.
Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26,* 212–215.
Tiitinen, H., Mäkelä, A. M., Mäkinen, V., May, P. J., & Alku, P. (2005). Disentangling the effects of phonation and articulation: Hemispheric asymmetries in the auditory *N1*m response of the human brain. *BMC Neuroscience [electronic resource], 6,* 62.
Tsuda, R., & Ueno, S. (2000). Source localization of visually evoked magnetic fields to stimuli in apparent motion. *IEEE Transactions on Magnetics, 36,* 3727–3729.
Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, U.S.A., 102,* 1181–1186.
Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory–visual speech perception. *Neuropsychologia, 45,* 598–607.
Wheeldon, L., & Waksler, R. (2004). Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain and Language, 90,* 401–412.