

Visual Anticipatory Information Modulates Multisensory Interactions of Artificial Audiovisual Stimuli

Jean Vroomen and Jeroen J. Stekelenburg

Abstract

■ The neural activity of speech sound processing (the N1 component of the auditory ERP) can be suppressed if a speech sound is accompanied by concordant lip movements. Here we demonstrate that this audiovisual interaction is neither speech specific nor linked to humanlike actions but can be observed with artificial stimuli if their timing is made predictable. In Experiment 1, a pure tone synchronized with a deformation of a rectangle induced a smaller auditory N1 than auditory-only presentations if the temporal occurrence of this audiovisual event was made predictable by two moving disks that touched the rectangle. Local autoregressive average source estimation indicated that this audiovisual interaction may be related to integrative processing in auditory areas. When the moving disks

did not precede the audiovisual stimulus—making the onset unpredictable—there was no N1 reduction. In Experiment 2, the predictability of the leading visual signal was manipulated by introducing a temporal asynchrony between the audiovisual event and the collision of moving disks. Audiovisual events occurred either at the moment, before (too “early”), or after (too “late”) the disks collided on the rectangle. When asynchronies varied from trial to trial—rendering the moving disks unreliable temporal predictors of the audiovisual event—the N1 reduction was abolished. These results demonstrate that the N1 suppression is induced by visual information that both precedes and reliably predicts audiovisual onset, without a necessary link to human action-related neural mechanisms. ■

INTRODUCTION

Speech is the example “par excellence” that human information processing is multisensory. One of the most striking demonstrations of multisensory processing is the discovery by McGurk and MacDonald that when listeners hear /aba/ and lip-read /aga/, they often report to “hear” /ada/, the so-called “McGurk effect” (McGurk & MacDonald, 1976). This audiovisual (AV) illusion shows that at some processing stage, the auditory and the visual streams are combined into a multisensory representation. A key issue for any behavioral, neuroscientific, and computational account of multisensory integration is to know when and where in the brain the sensory-specific information streams merge.

Hemodynamic studies have shown that multisensory cortices (STS/gyrus; Callan et al., 2004) and “sensory-specific” cortices (Callan et al., 2003; Calvert et al., 1999) are involved in multisensory integration of speech. Moreover, EEG and magneto-encephalography (MEG) studies have shown that AV speech interactions occur in the auditory cortex between 150 and 250 msec using the MMN paradigm (Colin et al., 2002; Möttönen, Krause, Tiippana, & Sams, 2002; Sams et al., 1991), whereas others have reported that as early as 100 msec, the auditory-evoked N1 component is attenuated (Besle, Fort, Delpuech, & Giard, 2004) and

speeded up (van Wassenhove, Grant, & Poeppel, 2005) when auditory speech is accompanied by concordant lip-read information.

N1 suppression by lip-read speech was originally thought to be based on a speech-specific mechanism (Besle, Fort, Delpuech, et al., 2004) because simplified AV nonspeech stimuli like pure tones and geometrical shapes (Fort, Delpuech, Pernier, & Giard, 2002b; Molholm et al., 2002; Giard & Peronnet, 1999) and spoken and written forms (Rajj, Uutela, & Hari, 2000) were associated with superadditive auditory N1 interactions. Although it would seem that the subadditive N1 interactions to AV speech are inconsistent with one of the basic principles of multisensory integration, namely superadditivity (Stein & Meredith, 1993), it should be noted that, in general, multisensory interactions in the ERP may consist of multiple interactions, expressed as new neural activities and as modulations (increase and decrease) in sensory-specific and polysensory regions (Fort et al., 2002b; Giard & Peronnet, 1999). Here, the relevant question is what is the key factor causing the difference between subadditive and superadditive auditory N1 interactions? Furthermore, what is the functional significance of the observed suppression of auditory N1 during AV stimulation? One factor that may be crucial is that the nonspeech stimuli in the above mentioned studies are presented synchronously, whereas in AV speech visual speech naturally precedes the acoustical signal by tens to a few

hundreds of milliseconds (van Wassenhove et al., 2005). Possibly, N1 suppression occurs because leading lip-read information allows to predict the auditory signal thereby reducing signal uncertainty and lowering computational demands for auditory brain areas (Besle, Fort, Delpuech, et al., 2004).

The hypothesis of speech specificity of the suppression of auditory N1 was refuted by recent evidence showing that natural human actions—containing anticipatory visual movement—such as the sight of clapping hands also reduce auditory N1 (Stekelenburg & Vroomen, 2007). N1 suppression is not affected by whether the auditory information and the visual information are congruent or incongruent (e.g., hearing /ba/ while lip-reading /fu/), but, importantly, no N1 suppression is observed when there is no visual anticipatory information about sound onset as was the case in a video recording of a moving saw (Stekelenburg & Vroomen, 2007). Here, although sight and sound were synchronous, vision did not predict sound onset. This suggests that auditory processing is affected by visual information provided that vision precedes and predicts sound onset. This inference, if correct, would have important consequences for multisensory research in general because most ERP studies have been conducted with artificial stimuli without visual anticipatory information (Talsma & Woldorff, 2005; Molholm, Ritter, Javitt, & Foxe, 2004; Fort, Delpuech, Pernier, & Giard, 2002a; Fort et al., 2002b; Molholm et al., 2002; Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002; Giard & Peronnet, 1999). The assumption is that AV synchrony is the natural default situation, but this notion is strongly biased because AV speech and, indeed, many other natural events do contain visual anticipatory information about sound onset with potentially important consequences for neural processing.

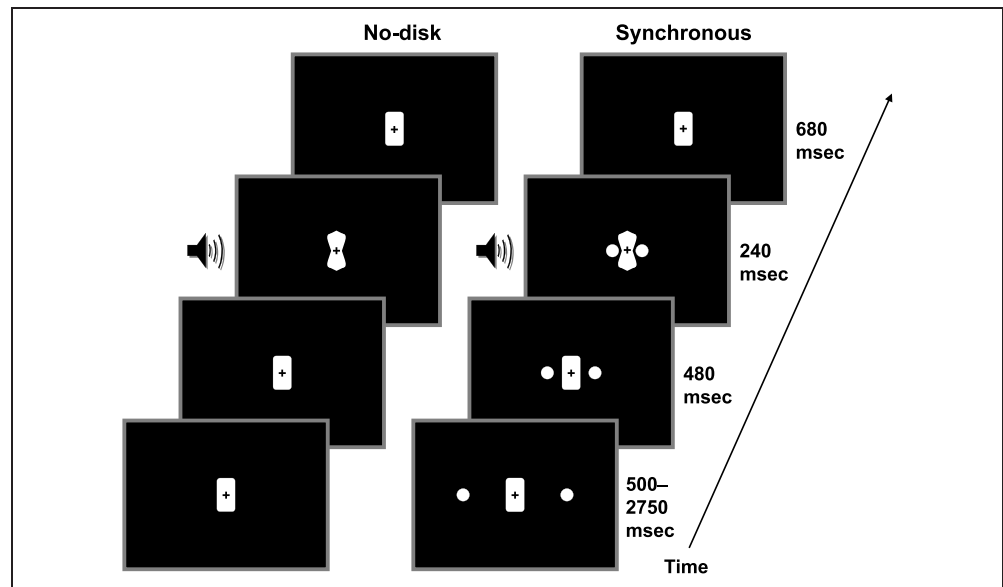
The current study investigates whether the subadditive AV ERP effects reflect a general or a more specific human action-related multisensory integrative mechanism. Previous studies reporting N1 suppression used natural actions like speaking faces or clapping hands (Besle, Fort, Delpuech, et al., 2004), but others using non-human-related natural events containing predictive visual motion like a drop of water hitting the water surface failed to obtain suppression of the auditory N1 (Senkowski, Saint-Amour, Kelly, & Foxe, 2007). A potential explanation for this difference is that speech and meaningful goal-directed human actions have been associated with activation of the mirror neuron system in motor regions of planning and execution (Broca's area, premotor cortex, and anterior insula; Ojanen et al., 2005; Skipper, Nusbaum, & Small, 2005; Callan et al., 2003), whereas the mirror neuron system is not critically involved in nonhuman events. The presumed function of the mirror neuron system is to mediate imitation and aid action and understanding. Broca's area is not only involved in speech production, silent lip-reading, and passive listening of auditory speech but is also responsive to perception and imitation of meaningful goal-directed hand

movements (Koski et al., 2002; Grezes, Costes, & Decety, 1999; Iacoboni et al., 1999). Activation of the mirror neuron system may, therefore, constitute the link between the auditory and the visual input that suppresses the sound of speaking faces and clapping hands (Stekelenburg & Vroomen, 2007) but not of falling water drops (Senkowski et al., 2007). Alternatively, though, on closer inspection of the AV stimuli containing falling drops, it appeared that there was only very little anticipatory information (visual to auditory onset of ~ 100 msec) that might not have been enough to induce auditory N1 suppression. If so, one might obtain N1 suppression with nonhuman artificial stimuli if only there is enough anticipatory information, thus not necessitating the involvement of the mirror neuron system. The use of artificial stimuli rather than natural and humanlike ones allows for examining the generality of the underlying neural mechanism of auditory N1 suppression. To this end, we measured ERPs of perceivers viewing an animated rectangle being “squeezed” while hearing a sound synchronized with the deformation of the rectangle. Typically, combining a sound with the deformation of a geometrical shape does not result in a suppression of auditory N1 (Fort et al., 2002a, 2002b; Giard & Peronnet, 1999). However, in the present study, we made the onset of the AV event predictable. The AV event occurred when two disks moving toward the rectangle touched the rectangle (see Figure 1). In a second experiment, we further investigated the role of visual temporal predictability on N1 suppression by manipulating the predictive power conveyed by the moving disks. AV events occurred at the moment (“synchronous”), before (too “early”), or after (too “late”) the disks collided on the rectangle. It was expected that when temporal asynchrony between the AV event and the collision of moving disks varied across trials, visual anticipatory information would become unreliable in predicting the AV onset thus resulting in no or less N1 suppression.

EXPERIMENT 1

Participants watched a video animation of a rectangle being squeezed while a sound was played whose onset was synchronized with the squeeze. In the “synchronous” condition, the AV event in turn was synchronized with the moment at which two disks that moved toward the rectangle hit the rectangle. The onset of the AV event could thus be predicted from the time that the moving disks touched the rectangle (Figure 1). In a control (“no-disk”) condition, there were no disks and the timing of the AV event was therefore unpredictable. ERPs were collected and AV interactions were determined by comparing auditory-only (A) responses to AV minus visual-only responses (AV – V) in both the “synchronous” condition (containing anticipatory visual information from the disks) and the “no-disk” condition. The difference between A and (AV – V) at any given time in the ERP is interpreted as an

Figure 1. Experimental design. In the synchronous condition, two disks moved after a variable interval of 500–2750 msec from trial start toward a central rectangle in 480 msec. At the moment of impact, the rectangle was deformed (“squeezed”), and a sound was played concurrently. In the no-disk condition, the AV event occurred without the moving disks at the same variable interval from trial start as in the synchronous condition.



integration effect between the two modalities (Besle, Fort, Delpuech, et al., 2004; Klucharev, Möttönen, & Sams, 2003; Fort et al., 2002b; Molholm et al., 2002; Teder-Sälejärvi et al., 2002; Giard & Peronnet, 1999). We expected auditory N1 suppression for AV stimuli in the synchronous condition in which disks preceded and reliably predicted the AV event, whereas no N1 suppression was expected in the no-disk condition where the AV event was not preceded by visual anticipatory information.

Methods

Participants

Fourteen healthy participants (4 men, 10 women) with normal hearing and normal or corrected-to-normal vision participated after giving written informed consent. Their age ranged from 18 to 38 years with mean age of 21 years. The study was conducted with approval of the local ethics committee of Tilburg University.

Stimuli and Procedure

The experiment took place in a dimly lit, sound-attenuated, and electrically shielded room. Visual stimuli were presented on a 19-in. monitor positioned at eye level, 70 cm from the participant’s head. During a trial, participants fixated on a central cross (+ of 4 × 4 mm). Sounds were presented from a central loudspeaker directly below the monitor.

In the synchronous condition, a white rectangle (30 cd/m², 3.5 × 2 cm) on a dark background (3 cd/m²) was presented in the middle of the screen. Two white disks (10-mm diameter) were shown on each side of the rectangle at median level and at 6.8 cm from the center. After

a variable interval (500–2750 msec), the disks would start to move at a constant speed (8.5 degrees/s) toward the rectangle and touch it after 480 msec. At that moment, a sound (a 65-dB(A) pure tone of 1000 Hz with a duration of 240 msec, including 10-msec rise/fall times) was played and the rectangle was deformed by indenting the vertical sides and by expanding the horizontal sides (by 5 pixels; see Figure 1). The deformation lasted for 240 msec after which the rectangle returned to normal for 680 msec. The no-disk condition was the same as the synchronous condition, except that there were no disks. For the ERP analyses, it was necessary to subtract a visual-only ERP from the AV one and to compare the resulting difference wave with the auditory-only ERP. Both conditions were therefore also presented in a visual-only (silent) mode where no sound was present and in auditory-only mode. In the auditory-only mode, a static rectangle was shown throughout the whole trial while the sound was played with the same timing as in the AV case. The intertrial interval varied randomly between 800 and 1800 msec during which the screen was black.

Both synchronous and no-disk conditions were subdivided into three identical blocks. In each block 40 AV, visual-only and auditory-only trials were presented in random order, which amounted to a total of 120 trials per modality per condition. Blocks of the no-disk condition were alternated with blocks of the synchronous condition. For half of the participants, the experiment started with a block of the synchronous condition, the other half started with a block of the no-disk condition. The testing lasted about 2 hours (including short breaks between the blocks). To ensure that participants were looking at the monitor during stimulus presentation, they had to detect, by key-press, the occasional occurrence of catch trials (16.7% of total number of trials). During a catch trial, the fixation

cross changed from “+” to “x” for 120 msec. This change occurred quasi-randomly between 400 msec after trial onset and 500 msec before trial offset. Catch trials occurred equally likely in all conditions.

ERP Recording and Analysis

The EEG was recorded at a sampling rate of 512 Hz from 49 locations using active Ag-AgCl electrodes (BioSemi, Amsterdam, The Netherlands) mounted in an elastic cap and two mastoid electrodes. Electrodes were placed according to the extended International 10-20 system. Two additional electrodes served as reference (common mode sense active electrode) and ground (driven right leg passive electrode). EEG was referenced off-line to average reference and band-pass filtered (0.5–30 Hz, 24 dB/octave). The raw data were segmented into epochs of 600 msec, including a 100-msec prestimulus baseline. ERPs were time locked to the onset of the AV event. After EOG correction (Gratton, Coles, & Donchin, 1983), epochs with an amplitude change exceeding $\pm 100 \mu\text{V}$ at any EEG channel were rejected. ERPs of the noncatch trials were averaged per modality (AV, A, and V), separately for the no-disk and synchronous condition. AV interactions in both conditions were investigated by subtracting the visual ERP from the AV ERP and then compare the (AV – V) difference wave with the auditory-only (A) ERP. The additive model (AV – (A + V)) assumes that the neural activity evoked by AV stimuli is equal to the sum of activities of A and V if the unimodal signals are processed independently. This assumption is valid for extracellular media and is based on the law of superposition of electric fields (Barth, Goldberg, Brett, & Di, 1995). If the bimodal response differs (supra-additive or subadditive) from the sum of the two unimodal responses, this is attributed to the interaction between the two modalities. However, the additive model approach can lead to spurious interaction effects if common activity like anticipatory slow wave potentials (which continue for some time after stimulus onset) or late common potentials like N2 and P3 are found in A, V, and AV because this common activity will be present in A but is removed in the AV – V subtraction (Besle, Fort, & Giard, 2004; Teder-Sälejärvi et al., 2002). To reduce common anticipatory processes that might lead to spurious early interactions, we used a variable intertrial interval and a variable interval from trial start to visual motion onset (Besle, Fort, & Giard, 2004) and high-pass filtered (2 Hz, 24 dB/octave) the AV – (A + V) difference waves (Teder-Sälejärvi et al., 2002). To circumvent potential problems of late common activity, we restricted our analysis to the early stimulus processing components (<300 msec).

The auditory N1 and P2 had a central maximum, and analyses were therefore conducted at the central electrode Cz. The N1 was scored in a window of 70–150 msec, P2 was scored in a window of 120–250 msec. Topographic analysis of N1 and P2 comprised vector-normalized amplitudes of all (49) electrodes. Because amplitude dif-

ferences between conditions can bias the topographic comparisons, the vector normalization transform was applied to correct the data for the overall amplitude difference between conditions (McCarthy & Wood, 1985). For the vector normalization, the amplitude at each location was divided by the condition specific vector length. Vector length is defined by the square root of the sum of squared voltages over all electrode locations. Violations of sphericity in topographic analysis were adjusted by the Huynh-Feldt correction.

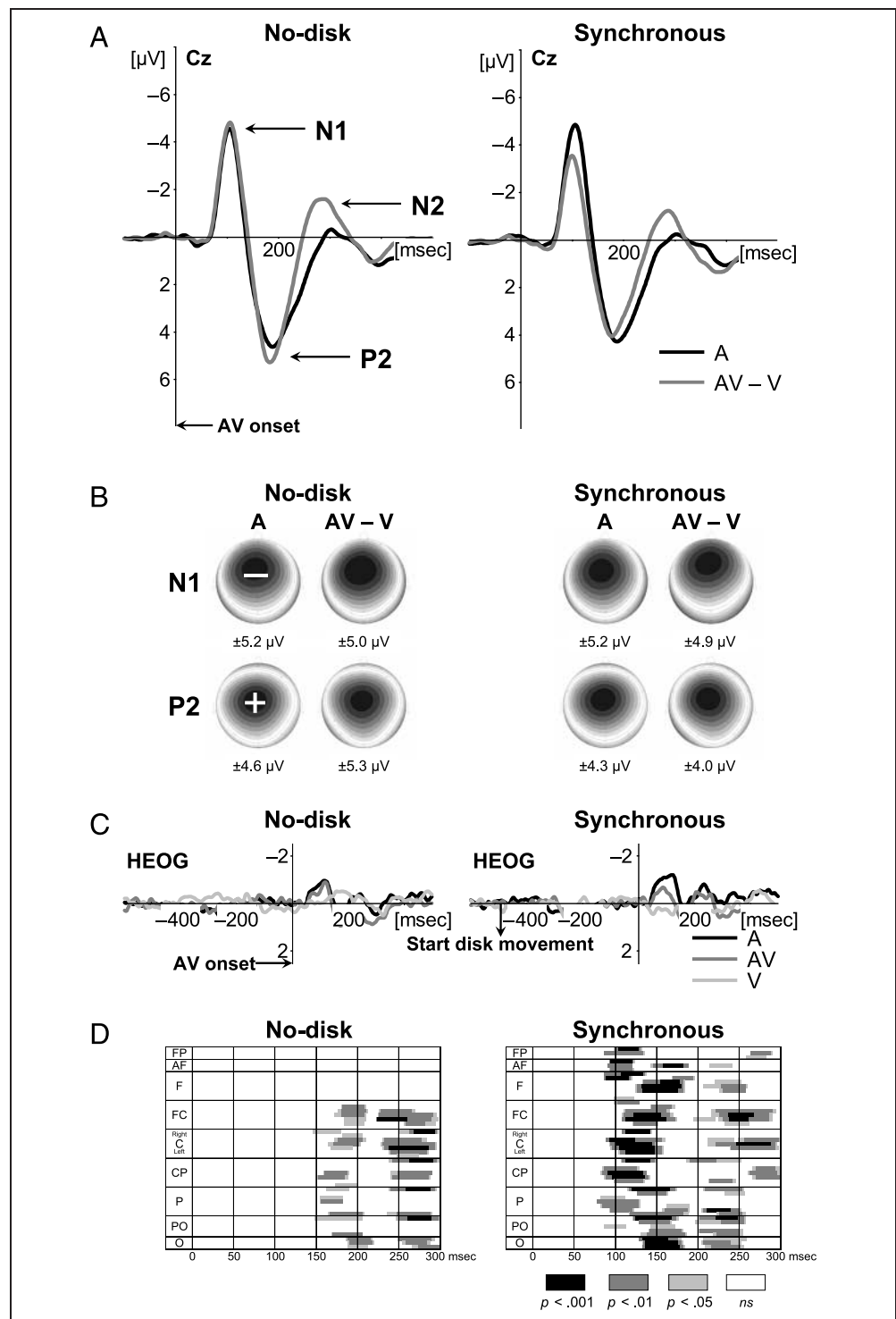
In the second analysis, the neural sources underlying AV N1 interactions were estimated by using a linear distributed inverse solution based on a local autoregressive average (LAURA) model of the unknown current density in the brain (Grave de Peralta Menendez, Gonzalez Andino, Lantz, Michel, & Landis, 2001). LAURA estimates three-dimensional current density distributions calculated on a realistic head model with 4024 solution nodes equally distributed in the gray matter of the average Montreal Neurological Institute brain. LAURA makes no a priori assumptions regarding the number of sources or their locations and can deal with multiple simultaneously active sources. This analysis was implemented using Cartool software by Denis Brunet (<http://brainmapping.unige.ch/Cartool.htm>). Using LAURA, plausible source estimations have been reported earlier for AV interactions (Meylan & Murray, 2007; Mishra, Martinez, Sejnowski, & Hillyard, 2007; Senkowski et al., 2007). In the current article, the resulting source estimations provide visualization of the likely underlying sources and do not represent a statistical analysis. Another consideration is whether with the current number of electrodes (49) reliable source imaging can be obtained. Adequate localization precision is to be expected based on a simulation study (Michel et al., 2004) where it was shown that with 49 electrodes, the percentage of sources with a dipole localization error of less than two grid points (within an equally spaced grid of 1152 solution points) is about 80%.

Finally, the spatio-temporal dynamics of the AV interaction were explored by conducting point-by-point two-tailed *t* tests on the AV – (A + V) difference wave at each electrode in a 1- to 300-msec window after AV onset. Using a procedure to minimize type I errors (Guthrie & Buchwald, 1991), AV interactions were considered significant when at least 12 consecutive points (i.e., 24 msec when the signal was resampled at 500 Hz) were significantly different from zero. This analysis allowed for detection of the earliest time where AV interactions occurred.

Results

Participants detected 98% of the catch trials, indicating that they were indeed watching the screen during a trial. Figure 2 shows that the amplitude of the auditory-evoked N1 was reduced to bimodal stimuli in the synchronous condition, whereas there was no difference in the no-disk

Figure 2. (A) ERPs time locked to AV onset at electrode Cz for both synchronous condition (AV stimulus occurred when two moving disks hit the visual stimulus) and no-disk condition (AV stimulus occurred without anticipatory visual motion). (B) The scalp topographies of auditory-only and AV – V N1 and P2. The range of the voltage maps in microvolts is displayed below each map. (C) Horizontal EOG (HEOG) for auditory, visual, and AV presentations starting at the onset of the movement of the two disks in the synchronous condition or at the corresponding time stamp in the no-disk condition. HEOG shows that participants did not make eye movements to the moving disks in the synchronous condition. (D) Pointwise *t* tests on the difference wave AV – (A + V) evaluating AV interactions at every electrode in a 1- to 300-msec poststimulus window.



condition. The amplitude of the AV – V P2 was larger than the A-only one in the no-disk condition but did not substantially differ in the synchronous condition. These observations were tested using a MANOVA for repeated measures. First, latency and amplitude difference scores for N1 and P2 at electrode Cz were computed by subtracting the AV – V scores from the A-only scores (i.e., $A - (AV + V)$). Second, the difference scores were submitted to

the MANOVA with Condition (no-disk vs. synchronous) and Modality (testing the overall difference from zero) as within-subject factors.

For N1 amplitude, there was an effect of Condition, $F(1, 13) = 24.72, p < .001$. Post hoc analysis to explore whether there was an effect within each condition between the A-only and the AV – V ERPs revealed no significant difference between the bimodal and the A-only N1 amplitude in

the no-disk condition ($p = .22$), whereas in the synchronous condition the bimodal N1 was $1.4 \mu\text{V}$ smaller than the A-only N1 ($p < .001$). For N1 latency, there was an effect of Condition, $F(1, 13) = 11.48, p < .01$. Post hoc analysis revealed that in the synchronous condition, the AV - V N1 was 5.6 msec shorter than the A-only N1, ($t(13) = 3.65, p < .01$). In the no-disk condition, N1 latency did not differ between conditions ($p = .23$). P2 amplitude was not significantly affected by Condition, $F(1, 13) = 3.86, p = .07$, nor Modality ($F < 1$). For P2 latency, an effect of Condition was found, $F(1, 13) = 6.10, p < .05$. Further testing showed that P2 latency to bimodal stimuli was shortened for 16 msec in the synchronous condition, $t(13) = 2.86, p < .05$, but not in the no-disk condition ($t < 1$). Figure 2 also shows that in both conditions, there were late frontal-central auditory-visual interactions starting at peak P2 and ending after peak N2. We tested these interactions at peak N2 at electrode FCz where AV interactions were most pronounced. N2 amplitude was $1.3 \mu\text{V}$ larger for the bimodal presentations, $F(1, 13) = 10.91, p < .01$, than for the A-only ones but did not differ between conditions ($F < 1$). N2 latency was unaffected by Condition, $F(1, 13) = 2.01, p = .18$, or Modality ($F < 1$).

Topographies of N1 and P2 were tested using within-subjects variables Condition (no-disk vs. synchronous), Modality (A vs. AV - V), and Electrode (49 electrodes). A significant Condition \times Modality \times Electrode interaction was found, $F(48, 624) = 8.85, p < .001$. Simple effect test revealed that topography of N1 in the no-disk condition did not significantly differ between A-only and bimodal stimuli ($F < 1$). In the synchronous condition, there was a Modality \times Electrode interaction for N1, $F(48, 624) =$

$13.22, p < .001$). As depicted in Figure 2, the topography of N1 was slightly more anterior for bimodal presentations. For both no-disk and synchronous conditions, P2 topography did not differ between bimodal and A-only presentations ($F < 1$).

Source analysis was applied to estimate the neural source of N1 suppression to bimodal presentations in the synchronous condition. First, the difference wave AV - (A + V) was calculated per participant. LAURA source estimations of the difference wave were estimated per participant over a period of 90–120 msec and subsequently averaged across participants. LAURA revealed neural generators of N1 suppression mainly in the left posterior auditory cortex (Brodmann's area 22). To examine whether the location of N1 suppression corresponds to the neural source of auditory N1, the source of A-only N1 was estimated in the same temporal window. As is evident from Figure 3, the location of the AV interaction corresponds well with the neural source of auditory-only N1.

Running t test analysis was used to explore the time course of AV interactions (Figure 2). In the no-disk condition, there were no AV interactions at N1 latency. Reliable AV interactions were found in a 150- to 220-msec window corresponding to the superadditive interactions at P2 latency. Broadly distributed late interactions were found in a 220- to 300-msec window at fronto-central, central, parietal, and occipital electrodes. In the synchronous condition, sub-additive N1 effects were found at central and central-parietal electrodes in a 80- to 130-msec window. Late interactions were found between 190 and 300 msec at frontal-central and parietal-occipital electrodes. The timing and the location of the AV interactions corresponded to the modulation

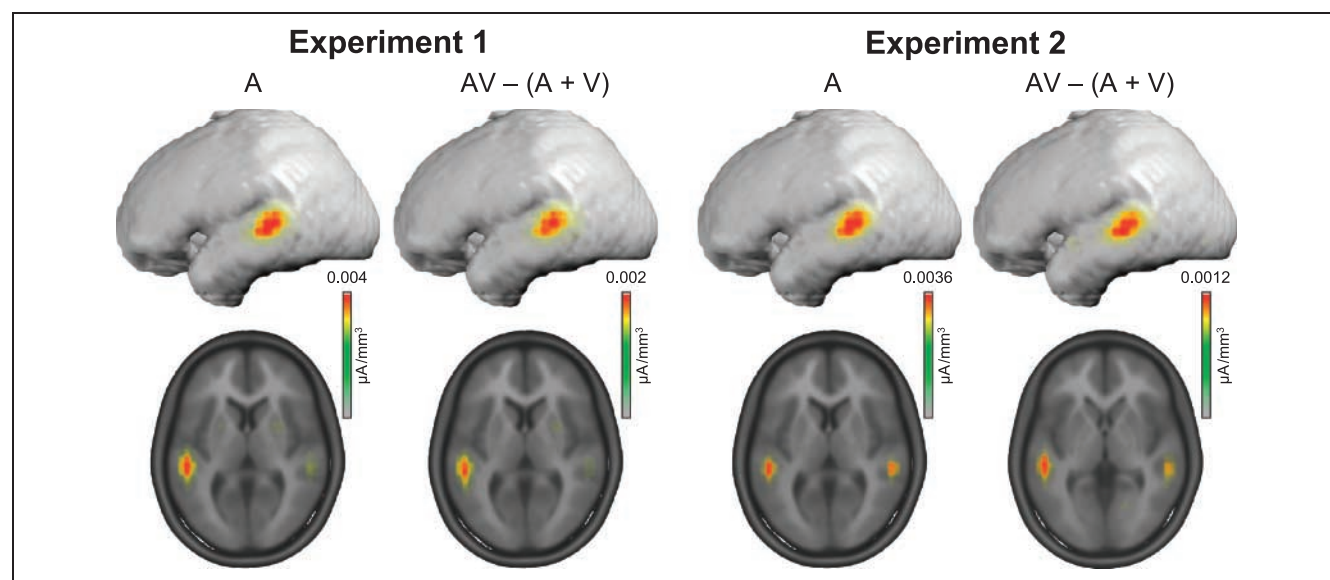


Figure 3. LAURA source estimations for the synchronous condition (AV stimulus occurred when two moving disks hit the visual stimulus) for Experiment 1 and Experiment 2 (fixed blocks) on the A-only ERP and the AV - (A + V) difference wave in a 90- to 120-msec interval wherein significant suppression of auditory-evoked N1 was found to bimodal presentations. Note that the because of $AV < (A + V)$, the neural activation of the difference wave represents subadditive interaction effects.

of both the auditory N1 and P2/N2 in the synchronous condition and P2/N2 in the no-disk condition.

Discussion of Experiment 1

The main result of Experiment 1 is that in the no-disk condition, the amplitude of the auditory N1 to a pure tone is not different from the N1 to the sound when synchronized with a deformation of a rectangle, whereas in the synchronous condition, the same AV event induced decrement of N1 amplitude. We argue that the suppression of N1 is the consequence of increased temporal predictability of the onset of the artificial AV event by the two critically timed moving disks. Without anticipatory visual motion information, there was no N1 reduction, which shows that visual information has to precede and predict the AV event. Experiment 1 demonstrates furthermore that the N1 suppression is not crucially dependent on humanlike biological motion stimuli that might be linked to the mirror neuron system.

A critical issue is whether the N1 suppression reflects a genuine cross-modal interaction effect rather than a spurious effect generated by differences in eye movement behavior. The moving disks may have induced horizontal eye movements with the result that participants were not focusing on the rectangle that might lead to suboptimal conditions for AV integration and as a consequence subadditive interactions. There are, however, two arguments that dismiss this possibility. First, the results of the behavioral task show that the participants were almost flawless in detecting a transient change in the fixation cross (98% correct) that was located in the middle of the rectangle, indicating that they were focusing on the rectangle. Second, the EOG shows that participants made no horizontal eye movements during the course of a trial (see Figure 2).

Another alternative account of the suppression of N1 is that mere presentation of the moving disks (and thus independent of whether they provided temporal information about the AV event) or the collision of the disks with the rectangle might have summoned involuntary transient attention to the visual modality, thereby depleting attentional resources of the auditory modality. This might then be reflected in suppression of the auditory-related N1. At this stage, we cannot rule out this alternative explanation because only the synchronous condition contained visual moving stimuli and a collision. We therefore conducted a second experiment in which the temporal prediction account was put to further test.

EXPERIMENT 2

In Experiment 2, the onset of the AV event now occurred before, at, or after the disks would touch the rectangle, and this asynchrony was either fixed within a block of trials or varied from trial to trial. If temporal prediction

is crucial, one would expect to find N1 suppression only if the disks were synchronized with the AV event (i.e., the AV event occurred at the time the moving circles hit the rectangle) and the asynchrony was fixed because only then one can precisely predict when the AV event occurs. However, if suppression of N1 is induced by involuntary attention to the moving or colliding visual stimuli, one would expect N1 suppression no matter whether the asynchrony were fixed or varied from trial to trial. Finally, it should be noted that the comparison between the fixed and the mixed conditions also allowed us to rule out any other potential stimulus artifact because the stimuli were identical in these conditions.

Methods

Participants

Nineteen new participants (4 men, 15 women) with mean age of 19 years (range = 18–25 years) participated in Experiment 2.

Stimuli and Procedure

Stimuli and procedures were as in Experiment 1, except that the onset of the AV event was either at 240 msec before the disks would hit the rectangle (“early”), at 0 msec (“synchronous,” as the synchronous condition of Experiment 1), or at 240 msec after the disks would hit the rectangle (“late”). The same no-disk condition (without moving disks) as in Experiment 1 was included. In the early condition, the AV event (i.e., the squeezing of the rectangle and the presentation of the sound) occurred in middle of the trajectory of the moving disks, whereas in the late condition it occurred after the disks had already hit the rectangle. The asynchronies of -240 and $+240$ msec were considered to be sufficiently different from the synchronous one so that adaptation to AV asynchrony was prevented (Vroomen, Keetels, de Gelder, & Bertelson, 2004). The asynchronies were either kept constant within a block of trials or they varied randomly from trial to trial. There were 100 trials per modality per condition, grouped into 13 blocks. In 5 (mixed) blocks, the asynchronies varied randomly from trial to trial so that the timing of the AV event was unpredictable. For the other 8 (fixed) blocks, the conditions (no-disk, synchronous, early, and late) were run in separate blocks (two per condition). As in Experiment 1, the order of the blocks was varied quasi-randomly across participants.

Results

Participants detected 93% of the catch trials. Latency and amplitude difference scores $AV - (A + V)$ of N1 and P2 at electrode Cz were submitted to a MANOVA with Condition (no-disk, synchronous, early, and late) and Modality (testing the overall difference from zero) as within-subject

factors. Data were analyzed separately for fixed and mixed blocks.

Fixed Blocks

When the asynchronies were fixed in blocks of trials, it is apparent from the upper part of Figures 4 and 5 that the main results of Experiment 1 were replicated. Auditory N1 amplitude was reduced only if the onset of the AV stimulus could be accurately predicted by visual anticipatory motion (disks synchronized with the AV event and asynchrony fixed). There was no N1 suppression in the no-disk, early, and late conditions.

MANOVAs confirmed that for N1 amplitude there was a significant effect of Condition, $F(3, 16) = 4.65, p < .05$.

Post hoc Tukey tests revealed that only in the synchronous condition, N1 amplitude in the AV modality was significantly smaller ($0.8 \mu\text{V}$) than in the A-only modality ($p < .001$). N1 latency was not significantly affected by modality ($F < 1$) and did not differ between conditions, $F(3, 16) = 1.76, p = .20$. Overall bimodal P2 was reduced in amplitude ($0.7 \mu\text{V}$), $F(1, 18) = 11.15, p < .01$, and latency was shortened (6 msec), $F(1, 18) = 12.31, p < .01$. These amplitude and latency effects were not affected by Condition, $F(3, 16) = 2.39, p = .11$ and $F(3, 16) = 0.20, p = .89$, respectively. N2 amplitude at electrode FCz to bimodal presentations was $0.7 \mu\text{V}$ larger than for A-only presentations, $F(1, 18) = 18.65, p < .001$, whereas there was no effect of Condition, $F(3, 16) = 2.63, p < .09$. The effect of modality on N2 latency (peak N2 latency facilitation of 6 msec

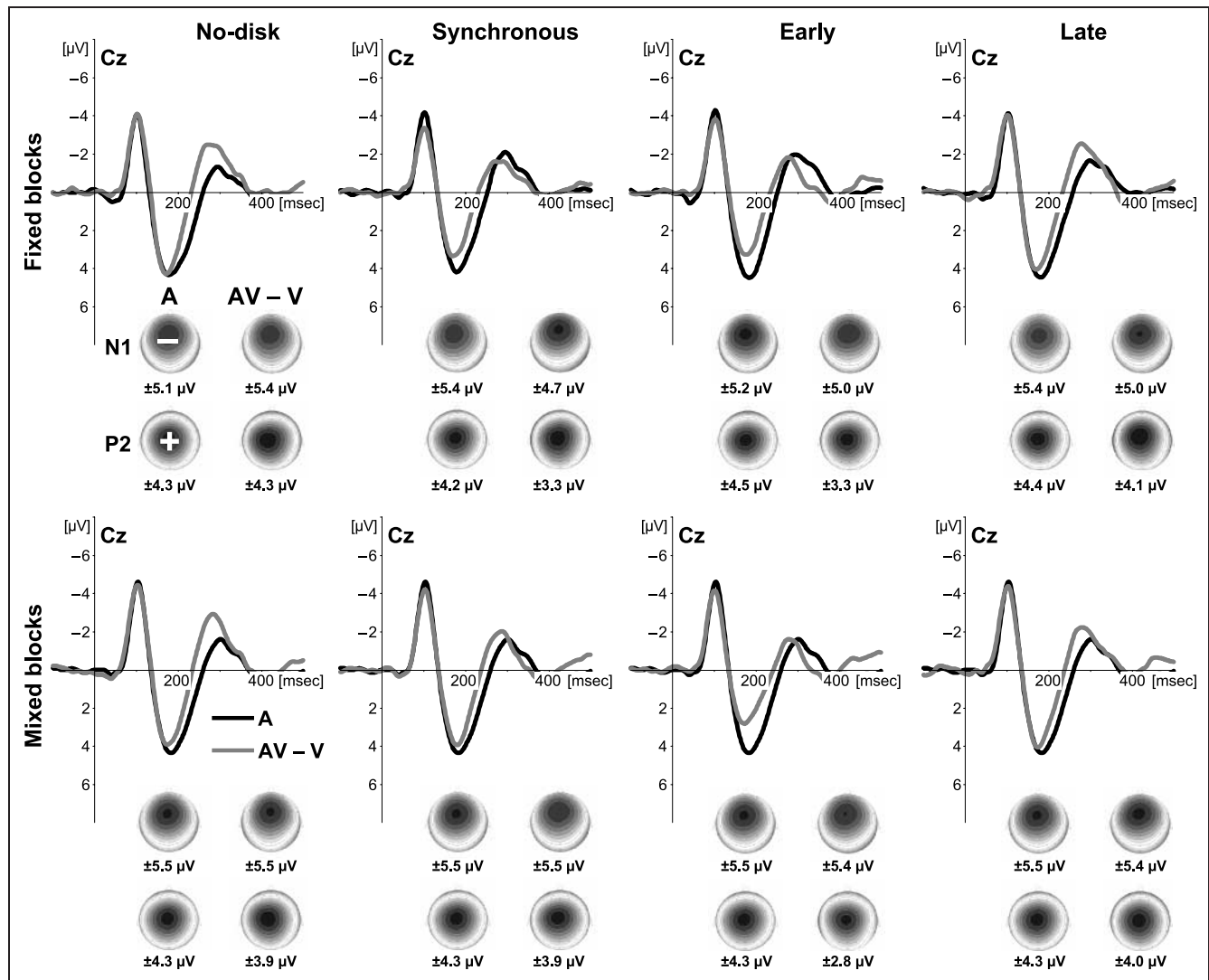


Figure 4. ERPs time locked to AV onset at electrode Cz, for synchronous (AV stimulus occurred when two moving disks hit the visual stimulus), early (AV stimulus occurred 240 msec before the moving disks would hit the visual stimulus), late (AV stimulus occurred 240 msec after the moving disks hit the visual stimulus), and no-disk (no visual anticipatory movement) conditions. The scalp topographies are displayed for auditory-only and AV - V N1 and P2. The range of the voltage maps in microvolts is displayed below each map. At the top, the ERPs for fixed blocks (temporal AV asynchrony was fixed within a block of trials) are shown, and at the bottom, the ERPs for mixed blocks (temporal AV asynchrony varied from trial to trial) are shown.

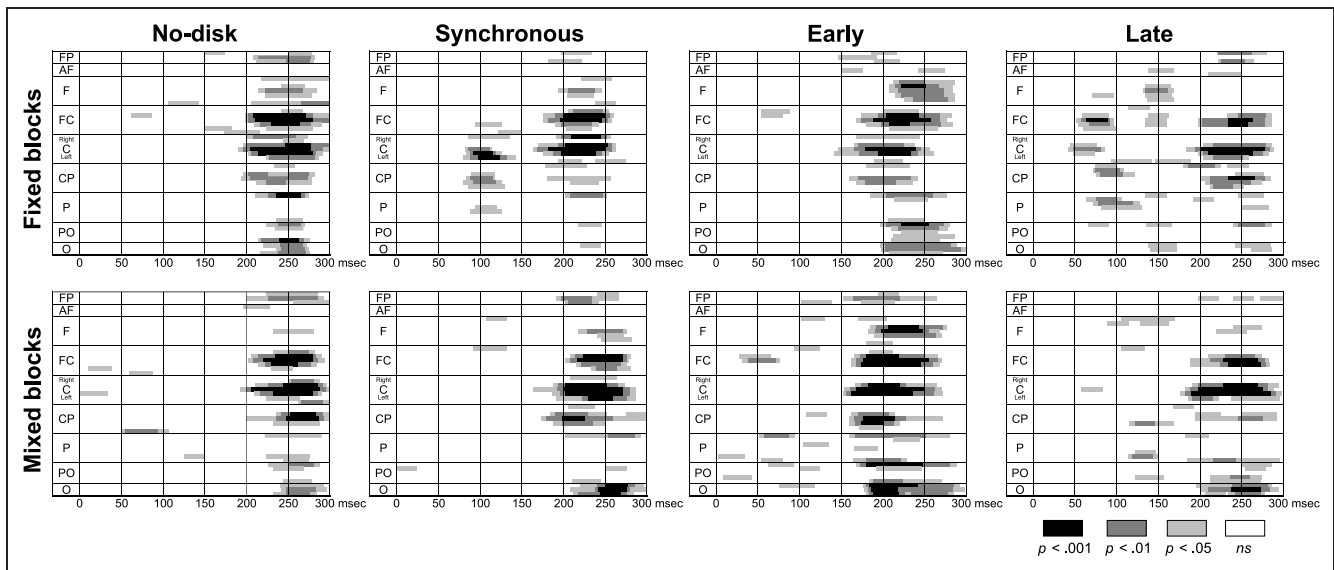


Figure 5. Pointwise t tests on the difference wave $AV - (A + V)$ evaluating AV interactions at every electrode in a 1- to 300-msec poststimulus window for the synchronous, early, late, and no-disk conditions. At the top, the results for fixed blocks are shown, and at the bottom, the results for mixed blocks are shown.

for bimodal stimuli) approached significant levels, $F(1, 18) = 4.24, p = .05$. N2 latency was unaffected by Condition ($F < 1$).

The topography of N1 and P2 did not significantly differ between A and AV for all conditions ($ps > .14$).

As depicted in Figure 3, source analysis on A-only N1 and N1 suppression to bimodal presentations in the synchronous condition revealed similar (but slightly more bilateral) neural generators in the auditory cortex as found in Experiment 1.

Running t test analysis (Figure 5) confirmed that when asynchrony was fixed, auditory N1 was modulated by the moving disks in the synchronous condition in an 80- to 120-msec window, whereas for the no-disk and early conditions, no interactions were found at this latency. For the late condition, subadditive N1 interactions were found somewhat more posteriorly at the central-parietal electrodes. Figure 5 further shows that late (superadditive) interactions were found for all conditions at N2 latency (approximately 200–300 msec).

Mixed Blocks

When asynchronies varied from trial to trial, N1 amplitude and latency did not differ significantly between conditions, $F(3, 16) = 1.91, p = .17$ and $F(3, 16) = 2.95, p = .07$, respectively. There was no overall modulation of N1 amplitude and latency, $F(1, 18) = 1.85, p = .19$ and $F(1, 18) = 1.42, p = .25$, respectively. P2 amplitude differed significantly between conditions, $F(3, 16) = 11.32, p < .001$. Post hoc test showed that only for the early ($p < .001$) and synchronous ($p < .05$) conditions, P2 of the AV condition was smaller than for A-only. P2 latency was shorter for bimodal stimuli than for A-only stimuli, $F(1, 18) = 12.15, p <$

.01. This effect did not differ significantly between conditions, $F(3, 16) = 2.64, p = .08$. N2 amplitude to bimodal presentations was $0.6 \mu\text{V}$ larger than for A-only presentations, $F(1, 18) = 5.89, p < .05$. N2 amplitude did not differ between conditions, $F(3, 16) = 1.86, p = .18$. N2 latency was shortened for 11 msec in the bimodal condition, $F(1, 18) = 12.06, p < .01$, which did not differ between conditions, $F(3, 16) = 1.71, p = .21$.

The topography of N1 for the randomized-condition blocks did not significantly differ between A-only and $AV - V$ for all conditions ($ps > .11$). P2 topography for the early, synchronous, and no-disk conditions did not differ between bimodal and A-only presentations. A Modality \times Electrode interaction, $F(48, 864) = 4.48, p < .05$, was found for the late condition, indicating a more posterior P2 distribution for bimodal presentations.

Running t test analysis (Figure 5) confirmed that there were no reliable AV interactions at auditory N1 latency (80–120 msec). Figure 5 further shows similar late superadditive interactions as found in the fixed blocks.

Discussion

The main result of Experiment 2 is that N1 suppression to bimodal stimuli was found in the synchronous condition in the fixed blocks, whereas there was no N1 suppression in any condition in the mixed blocks. Given that the fixation task is identical for both mixed and fixed blocks, it is unlikely that the N1 suppression can be ascribed to differences in visual attention between the synchronous and the no-disk condition. We further observed a substantial reduction of P2 amplitude to bimodal stimuli in the early condition for both fixed and mixed

blocks. Late interactions were found for each condition in both fixed and mixed blocks, which suggests that they occur independently of the predictive value of the visual stimulus.

GENERAL DISCUSSION

The present study demonstrates that the amplitude of the auditory-evoked N1 is reduced when the temporal occurrence of an artificial AV event—the deformation of a rectangle that is synchronized with a tone—is made predictable by moving disks that touch the rectangle. The deformation of the rectangle itself (without moving disks) induced no reduction of the auditory N1 (see also Fort et al., 2002a, 2002b; Giard & Peronnet, 1999). Furthermore, the N1 reduction was abolished when the asynchrony between the AV event and the collision of moving disks varied from trial to trial, thus demonstrating that visual information has to precede and reliably predict the onset of an AV event.

Conditions under Which N1 Suppression Occurs

Others have argued before that the suppression of the auditory N1 induced by AV presentations is exclusively related to the integration of AV speech because no N1 suppression was found with artificial AV combinations like pure tones combined with geometrical shapes (Fort et al., 2002b; Giard & Peronnet, 1999) or spoken and written forms (Raij et al., 2000). Also in monkey studies, it has been found that the primate auditory cortex integrates facial and vocal signals through enhancement and suppression of local field potentials, whereby the majority of these multisensory responses was specific to face/voice integration (Ghazanfar, Maier, Hoffman, & Logothetis, 2005). These comparisons, though, left unexplained what the unique properties of AV speech are that cause the effect. It might be that faces evoke face-specific neurons that drive integration (possibly via the STS), but it might also be the ecological validity of speech, the meaningful relationship between audition and vision, the fact that visual speech provides phonetically relevant information, or the dominance of the auditory modality in AV speech (Barraclough, Xiao, Baker, Oram, & Perrett, 2005).

In contrast with this vocal-specific view, we observed striking similarities between the neural correlates of AV integration of speech, ecologically valid nonspeech events (i.e., clapping hands) containing visual anticipatory information (Stekelenburg & Vroomen, 2007), and artificial events (deformation of a geometrical shape) containing visual anticipatory information. Clearly, then, early AV interactions in the auditory cortex are not speech specific but crucially depend on anticipatory information whether present in speech, humanlike nonspeech, or, as demonstrated here, completely artificial events.

From this perspective, one can ask why others (Senkowski et al., 2007) did not find decrement of auditory potentials for natural AV objects containing visual anticipatory motion (a drop falling and hitting a water surface). A reason for the absence of auditory suppression in this study might be that the interval between the visual motion onset and the sound was too short (~100 msec). Typically, attenuation of auditory N1 is obtained with considerably longer intervals (Besle, Fort, Delpuech, et al., 2004). Moreover, N1 suppression becomes stronger if the visual interval increases (Stekelenburg & Vroomen, 2007). Likewise, in monkey studies, suppressed responses in auditory cortex were primarily found if the interval between initiation of mouth movements and voice-onset times was long (Ghazanfar et al., 2005). This thus suggests that a minimum of a head start of visual information is needed for the modulation of auditory processing (see also Aoyama, Endo, Honda, & Takeda, 2006).

It is also of interest to note that when the asynchrony between the disks and the AV event was fixed in a block of trials, there was no N1 reduction if the AV event came 240 msec before the disks touched the rectangle, despite the fact that participants might have in this case predicted when the stimulus was to occur because the moving disks gave an estimate (although impoverished if compared with the synchronous condition) about the onset of the AV event. This finding is noteworthy because it has been shown that perception of AV synchrony is to some extent malleable. For example, when participants are exposed to a fixed AV delay, they adapt their criterion of what constitutes AV synchrony provided that the AV asynchrony is relatively small (<~100 msec; Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vroomen et al., 2004). At large asynchronies, though, there is no adaptation, and stimuli do not combine into a multisensory percept. The current results confirm this finding for the early condition because at the large asynchrony used here (240 msec), it seems unlikely that there is adaptation to AV asynchrony, which explains why there was also no N1 suppression. Nevertheless, when the AV event in the fixed blocks was lagging, N1 suppression (at the central-parietal electrodes) was observed. This may indicate an asymmetric adaptation to AV synchrony. This asymmetry may correspond with the asymmetric temporal window of integration of AV stimuli, which holds that AV asynchrony tolerance is larger when the visual signal leads rather than when it lags (van Wassenhove, Grant, & Poeppel, 2007).

When discussing the conditions under which N1 suppression occurs, one could ask whether the visual task as used in the present study (detection of a change in fixation in catch trials) had an effect on the observed AV interactions. For example, would similar results be obtained if participants were engaged in an auditory rather than visual task? There are several arguments why task-related effects will only be secondary. First, others have observed that depression of auditory N1 (in case of AV speech) is also obtained when attention is focused on

the auditory modality (Besle, Fort, Delpuech, et al., 2004) rather than the visual one, as used in the current study and that of Stekelenburg and Vroomen (2007). Furthermore, van Wassenhove et al. (2005) manipulated the attended modality (focusing on either the visual or the auditory modality) and found no effect of this manipulation. Finally, the finding that in Experiment 2 N1 suppression was found in fixed blocks but not in mixed blocks cannot be explained by visual attention because the same task and identical stimuli were used in both conditions. The N1 suppression is thus unlikely to result from attention being directed toward the visual modality.

Late AV Interactions

The current results show that the auditory P2 is functionally dissociated from the N1 and thus likely reflecting different processes. When the AV asynchrony varied from trial to trial, P2 amplitude in the early condition was greatly suppressed whereas N1 amplitude remained unaffected. A similar dissociation between AV N1 and P2 suppression was reported earlier when incongruent AV pairings were used. For example, P2 but not N1 suppression was stronger when visual /bi/ was combined with incongruent auditory /fu/ rather than congruent /bi/ (Stekelenburg & Vroomen, 2007). This may suggest that P2 suppression is enhanced whenever the incoming AV signals do not match (temporal, semantic, or phonetic) from what can be expected from the leading visual signal. The observation that P2 amplitude is suppressed by both visual prediction and AV incongruence may become understandable if one considers that the P2 is not an unitary response but that there are many different neural generators active within a 150- to 250-msec window (Crowley & Colrain, 2004). Suppression of P2 by stimulus incongruence thus likely results from other generators than those underlying visual temporal prediction of N1.

There were also superadditive AV interactions in a 200- to 300-msec window. Similar AV interactions at this latency have been reported before (Senkowski et al., 2007; Teder-Sälejärvi et al., 2002). Given that these late interactions were found for artificial and natural stimuli and whether the leading visual stimulus was predictive for the temporal occurrence or not, it appears that at this stage, AV interactions occur independently of the nature of the AV events. These superadditive effects might represent an interaction in auditory association cortex or in polysensory cortex of the superior temporal plane (Teder-Sälejärvi et al., 2002).

The Underlying Neural Network of N1 Suppression

Our findings also speak to the issue about the underlying neural mechanisms. Several studies suggested that

AV speech interactions are mediated by speech motor regions and brain areas comprising the “mirror neuron system” (Broca’s area, premotor cortex, and anterior insula; Ojanen et al., 2005; Skipper et al., 2005; Callan et al., 2003). Assuming that the mirror neuron system is also active during perception and imitation of meaningful goal-directed hand movements (Koski et al., 2002; Grezes et al., 1999; Iacoboni et al., 1999), it seemed conceivable that the subadditive AV interactions obtained for humanlike actions like speaking faces or clapping hands (Besle, Fort, Delpuech, et al., 2004) could have been mediated by the mirror neuron system as well. However, given that similar AV suppressive effects were observed with highly artificial stimuli, it seems unlikely that the mirror neuron system is critically involved. Is the mechanism underlying the N1 suppression to artificial AV stimuli the same as the one found with AV speech? At this stage, there is no reason to postulate different mechanisms. The principal factor, in speech and non-speech alike, is the reliability of the visual signal of predicting sound occurrence. The more reliable, the bigger the N1 reduction.

There were no large or consistent topographical differences between the auditory-only and suppressed AV N1, which suggests that vision modulates the neural generators of auditory N1 itself (Adler et al., 1982). Source analysis indeed suggested that the suppression of N1 amplitude was generated in the auditory cortex at the same location as of the auditory N1. These results closely resemble those of an fMRI study in which a light announced the presentation of a sound (Lehmann et al., 2006). In that study, neural activity in the primary auditory cortex (wherein the main generators of auditory N1 are located; Näätänen & Picton, 1987) evoked by an artificial sound was suppressed when it was preceded by a visual stimulus (LED). Current data indicate that the sound paired with the leading visual stimulus did not excite different neuronal populations than sound alone but instead resulted in deactivation of auditory cortex.

The interaction effects in auditory cortex might also reflect feedback inputs from the STS where unisensory signals of multisensory objects are initially integrated (Calvert et al., 1999). However, this feedback interpretation from STS has been challenged by an MEG study in which it was demonstrated that interactions in the auditory cortex (150–200 msec) preceded activation in the STS region (250–600 msec; Möttönen, Schurmann, & Sams, 2004). In addition, an ERP study demonstrated that visual speech input may affect auditory-evoked responses via subcortical (brainstem) structures (Musacchia, Sams, Nicol, & Kraus, 2006). These extremely early AV interactions at the level of the brainstem (~11 msec) may only become understandable if one realizes that the visual input in AV speech precedes the auditory signal by tens, if not hundreds, of milliseconds (Munhall & Tohkura, 1998), which, as demonstrated here, might be crucial for this effect to be obtained.

Functional Interpretation of N1 Suppression

What is the functional interpretation of the N1 reduction? It might be reasoned that the leading visual signal increases alertness because visual anticipatory information serves as a warning signal (a cue) that directs attention to the auditory channel. As alerting took place only in the bimodal condition, the AV interaction effect on the auditory N1 would then, in essence, reflect the difference between attended versus unattended auditory information. Such an attentional account, however, is unlikely to account for the present findings because directing attention to the auditory modality generally results in an amplitude increase rather than decrease of ERP components in the time window of auditory N1 (Besle, Fort, Delpuech, et al., 2004). Conversely, as suggested by others (Oray, Lu, & Dawson, 2002), the leading visual signal may have captured involuntary attention thereby gating the input to the primary auditory cortex. N1 suppression in the synchronous condition in Experiment 1 would then be the result of attention to the auditory modality being distracted away by the collision of the disks. However, Experiment 2 refutes this hypothesis. Although the moving disks and the visual collision in Experiment 2 could have attracted attention in both early, synchronous, and late conditions, auditory N1 was only attenuated when the AV event could be accurately predicted (i.e., in the synchronous and late conditions when temporal asynchrony was fixed). These results therefore rule out the possibility that involuntary attention to the moving visual stimuli caused the N1 suppression. Rather, they corroborate the notion that the N1 suppression is the result of reduced temporal uncertainty induced by a leading visual signal.

Interestingly, the idea that the amplitude of auditory N1 is attenuated by reduced temporal uncertainty has been propagated since the early 1970s. In motor-sensory research, it has been demonstrated that auditory-evoked potentials can be modulated by whether a sound is induced by self-initiation or not. For example, auditory N1 is smaller to self-generated tones (Martikainen, Kaneko, & Hari, 2005; McCarthy & Donchin, 1976; Schafer & Marcus, 1973) than to the same sounds replayed to the subject, an effect attributed to reduced temporal uncertainty (Schafer & Marcus, 1973). Motor-to-auditory inhibition has been localized in the auditory cortex (Martikainen et al., 2005) and linked to a forward model in which modulation of auditory cortical response to self-generated actions allows immediate distinction of self and externally generated auditory stimuli (Heinks-Maldonado, Mathalon, Gray, & Ford, 2005). The forward model predicts and inhibits the sensory consequences of one's own actions so that more processing capacity can be allocated to external stimuli (Blakemore & Decety, 2001; Blakemore, Wolpert, & Frith, 2000). A similar predictive forward model may also be at work for the observed visual-to-auditory inhibition in the current study and others (Besle, Fort, Delpuech, et al., 2004), where sensory consequences of visual information are being pre-

dicted (van Wassenhove et al., 2005). Available evidence suggests that temporal information and not informational content is transmitted in the forward model because auditory N1 suppression is affected by the temporal relationship and not by the phonetic or semantic AV congruence (Stekelenburg & Vroomen, 2007). For future research, it will be of interest to further explore the specificity of auditory suppression because an answer on this question will have important consequences for the functional interpretation of this phenomenon.

Conclusion

Our results demonstrate that the neural correlates underlying AV integration are not specific to humanlike actions because they are found with artificial stimuli as well. The suppression of the auditory-evoked N1 is induced by synchronized visual information that reliably predicts sound onset. These properties are inherent of AV speech but also many other nonspeech events, whether biological or not.

Acknowledgments

The Cartool software (<http://brainmapping.unige.ch/Cartool.htm>) was programmed by Denis Brunet, from the Functional Brain Mapping Laboratory, Geneva, Switzerland, and was supported by the Center for Biomedical Imaging (CIBM) of Geneva and Lausanne.

Reprint requests should be sent to Jean Vroomen, Department of Psychology, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands, or via e-mail: J.Vroomen@uvt.nl.

REFERENCES

- Adler, L. E., Pachtman, E., Franks, R. D., Pecevich, M., Waldo, M. C., & Freedman, R. (1982). Neurophysiological evidence for a defect in neuronal mechanisms involved in sensory gating in schizophrenia. *Biological Psychiatry*, *17*, 639–654.
- Aoyama, A., Endo, H., Honda, S., & Takeda, T. (2006). Modulation of early auditory processing by visually based sound prediction. *Brain Research*, *1068*, 194–204.
- Barracough, N. E., Xiao, D., Baker, C. I., Oram, M. W., & Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*, *17*, 377–391.
- Barth, D. S., Goldberg, N., Brett, B., & Di, S. (1995). The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex. *Brain Research*, *678*, 177–190.
- 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.
- Besle, J., Fort, A., & Giard, M. H. (2004). Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cognitive Processing*, *5*, 189–192.
- Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*, 561–567.

- Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *NeuroReport*, *11*, R11–R16.
- Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *NeuroReport*, *14*, 2213–2218.
- Callan, D. E., Jones, J. A., Munhall, K., Kroos, C., Callan, A. M., & Vatikiotis-Bateson, E. (2004). Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *Journal of Cognitive Neuroscience*, *16*, 805–816.
- Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport*, *10*, 2619–2623.
- Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk–MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology*, *113*, 495–506.
- Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, *115*, 732–744.
- Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002a). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex*, *12*, 1031–1039.
- Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002b). Early auditory–visual interactions in human cortex during nonredundant target identification. *Brain Research, Cognitive Brain Research*, *14*, 20–30.
- 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.
- Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484.
- Grave de Peralta Menendez, R., Gonzalez Andino, S. L., Lantz, G., Michel, C. M., & Landis, T. (2001). Noninvasive localization of electromagnetic epileptic activity: I. Method descriptions and simulations. *Brain Topography*, *14*, 131–137.
- Grezes, J., Costes, N., & Decety, J. (1999). The effects of learning and intention on the neural network involved in the perception of meaningless actions. *Brain*, *122*, 1875–1887.
- Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology*, *28*, 240–244.
- Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, *42*, 180–190.
- Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*, 2526–2528.
- Klucharev, V., Möttönen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Brain Research, Cognitive Brain Research*, *18*, 65–75.
- Koski, L., Wohlschläger, A., Bekkering, H., Woods, R. P., Dubeau, M. C., Mazziotta, J. C., et al. (2002). Modulation of motor and premotor activity during imitation of target-directed actions. *Cerebral Cortex*, *12*, 847–855.
- Lehmann, C., Herdener, M., Esposito, F., Hubl, D., di Salle, F., Scheffler, K., et al. (2006). Differential patterns of multisensory interactions in core and belt areas of human auditory cortex. *Neuroimage*, *31*, 294–300.
- Martikainen, M. H., Kaneko, K., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex*, *15*, 299–302.
- McCarthy, G., & Donchin, E. (1976). The effects of temporal and event uncertainty in determining the waveforms of the auditory event related potential (ERP). *Psychophysiology*, *13*, 581–590.
- McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203–208.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.
- Meylan, R. V., & Murray, M. M. (2007). Auditory–visual multisensory interactions attenuate subsequent visual responses in humans. *Neuroimage*, *35*, 244–254.
- Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave de Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology*, *115*, 2195–2222.
- Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *Journal of Neuroscience*, *27*, 4120–4131.
- Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual–auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex*, *14*, 452–465.
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Brain Research, Cognitive Brain Research*, *14*, 115–128.
- Möttönen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Brain Research, Cognitive Brain Research*, *13*, 417–425.
- Möttönen, R., Schürmann, M., & Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: A magnetoencephalographic study. *Neuroscience Letters*, *363*, 112–115.
- Munhall, K. G., & Tohkura, Y. (1998). Audiovisual gating and the time course of speech perception. *Journal of the Acoustical Society of America*, *104*, 530–539.
- Musacchia, G., Sams, M., Nicol, T., & Kraus, N. (2006). Seeing speech affects acoustic information processing in the human brainstem. *Experimental Brain Research*, *168*, 1–10.
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*, 375–425.
- Ojanen, V., Möttönen, R., Pekkola, J., Jääskeläinen, I. P., Joensuu, R., Autti, T., et al. (2005). Processing of audiovisual speech in Broca's area. *Neuroimage*, *25*, 333–338.
- Oray, S., Lu, Z. L., & Dawson, M. E. (2002). Modification of sudden onset auditory ERP by involuntary attention to visual stimuli. *International Journal of Psychophysiology*, *43*, 213–224.
- Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron*, *28*, 617–625.

- Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., et al. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, *127*, 141–145.
- Schafer, E. W., & Marcus, M. M. (1973). Self-stimulation alters human sensory brain responses. *Science*, *181*, 175–177.
- Senkowski, D., Saint-Amour, D., Kelly, S. P., & Foxe, J. J. (2007). Multisensory processing of naturalistic objects in motion: A high-density electrical mapping and source estimation study. *Neuroimage*, *36*, 877–888.
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: Motor cortical activation during speech perception. *Neuroimage*, *25*, 76–89.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge MA: MIT Press.
- Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, *19*, 1964–1973.
- Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, *17*, 1098–1114.
- Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Brain Research, Cognitive Brain Research*, *14*, 106–114.
- 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.
- Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Research, Cognitive Brain Research*, *22*, 32–35.