frontiers in
INTEGRATIVE NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 31 May 2012
doi: 10.3389/fnint.2012.00026

Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events

Jeroen J. Stekelenburg* and Jean Vroomen

Department of Cognitive Neuropsychology, Tilburg University, Tilburg, Netherlands

Edited by:
Hermann J. Mueller, University of Munich, Germany

Reviewed by:
Daniel Senkowski, Charité, University Medicine, Germany
Edmund C. Lalor, Trinity College Dublin, Ireland

****Correspondence:***
Jeroen J. Stekelenburg, Department of Medical Psychology and Neuropsychology, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, Netherlands.
e-mail: j.j.stekelenburg@uvt.nl

In many natural audiovisual events (e.g., a clap of the two hands), the visual signal precedes the sound and thus allows observers to predict *when*, *where*, and *which* sound will occur. Previous studies have reported that there are distinct neural correlates of temporal (when) versus phonetic/semantic (which) content on audiovisual integration. Here we examined the effect of visual prediction of auditory location (where) in audiovisual biological motion stimuli by varying the spatial congruency between the auditory and visual parts. Visual stimuli were presented centrally, whereas auditory stimuli were presented either centrally or at 90° azimuth. Typical sub-additive amplitude reductions (AV − V < A) were found for the auditory N1 and P2 for spatially congruent and incongruent conditions. The new finding is that this N1 suppression was greater for the spatially congruent stimuli. A very early audiovisual interaction was also found at 40–60 ms (P50) in the spatially congruent condition, while no effect of congruency was found on the suppression of the P2. This indicates that visual prediction of auditory location can be coded very early in auditory processing.

Keywords: audiovisual integration, spatial congruity, visual prediction

INTRODUCTION

Many ecological settings are multisensory in nature with a causal relationship between the involved unisensory modalities, as in the case of hearing and seeing someone speak (Winkler et al., 2009). Quite often, visual information also leads the auditory information as in the case where two objects collide, or in the case of audiovisual speech where lip movements precedes actual phonation for up to several hundredths of milliseconds (Klucharev et al., 2003; van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2007). This visual anticipatory information allows observers to predict several aspects of the upcoming auditory signal, like its timing and content. Several electrophysiological markers underlying this predictive information have been found in studies aimed at tracking the time course of audiovisual speech integration. These report that the auditory-evoked N1 and P2 components of the event-related brain potential (ERP) are attenuated and speeded up when the auditory signal (monosyllabic syllables) is accompanied by concordant visual speech input (Klucharev et al., 2003; Besle et al., 2004; van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2007; Arnal et al., 2009). These sub-additive interactions are not only found in speech, but also in other naturalistic and artificial non-speech events provided that the visual information precedes and predicts sound onset as in the case of a clap of the two hands (Stekelenburg and Vroomen, 2007) or a collision of two disks (Vroomen and Stekelenburg, 2010). Of equal importance, there is no N1-suppression when there is no visual anticipatory information about sound onset as in the case in a video recording of a saw that suddenly moves (Stekelenburg and Vroomen, 2007; Vroomen and Stekelenburg, 2010). The functional interpretation of the N1-suppression may be related to a reduction of signal uncertainty, dampened sensation of loudness, or lowered computational demands for auditory brain areas.

This hypothesis fits with results from the early 1970s where in motor-sensory research it was found that the auditory N1 is dampened by self-generated sounds (Schafer and Marcus, 1973; McCarthy and Donchin, 1976; Martikainen et al., 2005) if compared to sounds replayed to the participant. Similar effects of motor prediction were found for the visually evoked N1 (Gentsch and Schütz-Bosbach, 2011; Gentsch et al., 2012). This motor-induced effect on the N1 has been attributed to reduced temporal uncertainty induced by a forward model that predicts and inhibits the sensory consequences of one's own actions (Schafer and Marcus, 1973). Also, work on unimodal auditory processing indicates that the auditory N1 can be attenuated by expectations of time (Lange, 2009, 2010).

From this literature, it also appears that predictions about the informational content are processed in a later stage of auditory processing. In multisensory studies it has been found that the N1-suppression is not affected by whether the auditory and visual information are congruent or incongruent (e.g., hearing /ba/ while lipreading /fu/), but AV integration of informational content (whether phonetic or semantic) affects the auditory P2 component as it is modulated by stimulus congruency in both speech (Klucharev et al., 2003) and

non-speech stimuli (Stekelenburg and Vroomen, 2007). This suggests the existence of two functionally distinct integrative mechanisms with different time-courses (see also Arnal et al., 2009). Klucharev et al. (2003) hypothesized that the early effects at N1 reflect AV interactions in the processing of general features shared by the acoustic and visual stimulus such as coincidence in time, and—at this stage untested—spatial location.

The current study was set-up to further explore the time-course and functional significance of visual predictive coding on auditory processing. One hitherto unexplored aspect is that, besides prediction of time and content, visually anticipatory information can also predict the likely *location* of the auditory signal because the origin of a sound usually corresponds with the location of the visual signal. Here, we thus examined whether spatial congruency between auditory and visual anticipatory information affects the auditory-evoked potentials N1, P2, or other components. For spatially congruent events, the location of the auditory and visual stimulus were aligned in the center, while for the incongruent condition there was a large separation of 90° between the auditory and visual stimulus. This large separation effectively prevented a ventriloquist effect (i.e., vision capturing the apparent sound location) to occur (Colin et al., 2001), and the reported effects were, therefore, devoid of neural correlates associated with ventriloquism (Bonath et al., 2007). We expected that if predictive coding entails prediction of sound location (over and above timing), then more suppression should be found when the locations of the auditory and visual stimulus were congruent rather than incongruent because a "confirmed" prediction lowers computational demands. Alternatively, if the brain does not use visual anticipatory information about sound location, then no effect of audiovisual spatial congruency should be observed.

METHODS

PARTICIPANTS

Twenty-two (19 woman, mean age 18.5, SD 1.1) healthy participants took part in the experiment. All were students from Tilburg University who reported normal hearing and normal or corrected-to-normal vision. All of them were naive to the purpose of the study. They received course credits for their participation. Informed consent was obtained from all participants. This study was conducted in accordance with the Declaration of Helsinki and was approved by the local Ethics Committee of Tilburg University.

STIMULI AND PROCEDURE

The experiment took place in a dimly lit and sound attenuated room. Visual stimuli were presented on a 19-inch monitor positioned at eye-level, at 70 cm from the participant's head. The sounds emanated from either of two speakers. One speaker was located directly below the monitor, the other one was located on the left side of the participant, perpendicular to the left ear, at the same height and distance as the central speaker. Four audiovisual stimuli were used that in previous studies induced reliable N1-suppression (Stekelenburg and Vroomen, 2007). Two stimuli were the syllables /bi/ and /fu/ pronounced by a Dutch female speaker whose entire face was visible on the screen. The other stimuli were natural human actions, i.e., a clap of two hands and a tap of a spoon against a cup. For each stimulus category three exemplars were recorded so that there were 12 unique recordings in total. The videos were presented at a rate of 25 frames/s with an auditory sample rate of 44.1 kHz. The size of the video frames subtended 14° horizontal and 12° vertical visual angle. Sound level was measured with a sound-level meter with the microphone pointing toward the auditory source. Peak intensity was 65 dB(A) for the central and lateral speakers. The duration of the auditory sample was 306–325 ms for /bi/, 594–624 ms for /fu/, 292–305 ms for the spoon tapping on a cup, and 103–107 ms for the clapping hands. Average duration of the video was 3 s, including a 200-ms fade-in and fade-out, and a still image (200–500 ms) at the start (**Figure 1**). A blank screen of 500–1000 ms followed each trial. The inter-stimulus interval (from auditory onset) was on average 3.7 s. The time from the start of the articulatory movements until voice onset was, on average, 160 ms for /bi/ and 200 ms for /fu/. The time from the start of the movements of the arm(s) until sound onset in the non-speech stimuli was 280 ms for the clapping hands and 320 ms for the tapping spoon.

There were five experimental conditions; Ac (audio from the center, no video), Al (audio from lateral, no video), Vc (video from central, no audio), AcVc (audio and video from central), and AlVc (audio from lateral, video from center). For each condition, a total of 72 experimental trials were presented, separately for each of the four stimuli across 12 blocks, amounting to a total of 1440 trials. Trial order was randomized. To ensure that participants were looking at the video during stimulus presentation, they had to detect, by key press, the occasional occurrence of catch trials (8% on top of the total number of experimental trials). Catch trials occurred equally likely in all conditions. Catch trials contained a superimposed small white spot—either between the lips and nose for the speech stimulus, or at collision site for

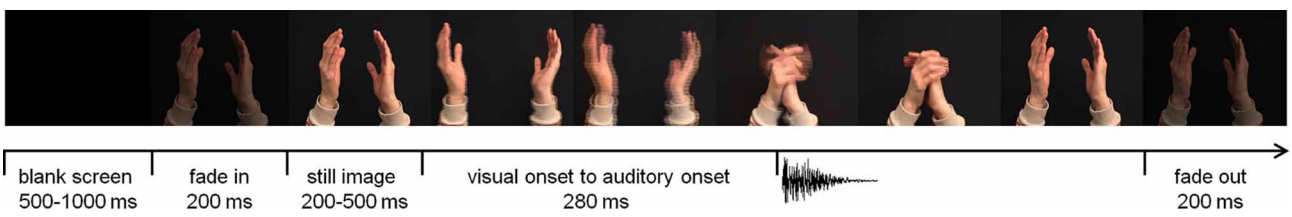

FIGURE 1 | Time-course of an audiovisual trial (hand clap). The visual onset to auditory onset differed per stimulus type (hand clap 280 ms; tapping spoon 320 ms; /bi/ 160 ms; /fu/ 200 ms).

the hands or at the site where the spoon hit the cup—for 120 ms. The appearance of the spot varied quasi-randomly within 300 ms before or after the onset of the sound. In the Ac and Al conditions the spot was presented on a dark screen at about the same position and at the same time as in the other conditions.

EEG RECORDING AND ANALYSIS

The electroencephalogram (EEG) was recorded at a sample 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 the extended International 10–20 system. Two additional electrodes served as reference (Common Mode Sense [CMS] active electrode) and ground (Driven Right Leg [DRL] passive electrode). EEG was referenced offline to an average of left and right mastoids and band-pass filtered (0.5–30 Hz, 24 dB/octave). The raw data were segmented into epochs of 800 ms, including a 100-ms prestimulus baseline. ERPs were time-locked to the sound onset in the AV and A conditions, and to the corresponding time stamp in the V condition. After EOG (Gratton et al., 1983), epochs with an amplitude change exceeding $\pm 120\,\mu V$ at any EEG channel were rejected. ERPs of the non-catch trials were averaged per condition (Ac, Al, Vc, AcVc, and AlVc) across all stimuli. As in previous studies, multisensory interactions were examined by comparing ERPs evoked by A stimuli with the corresponding AV minus V (AV − V) ERPs (Besle et al., 2004; Stekelenburg and Vroomen, 2007; Arnal et al., 2009; Vroomen and Stekelenburg, 2010). The additive model (A = AV − V) assumes that the neural activity evoked by the 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 et al., 1995). If the bimodal response differs (supra-additive or sub-additive) from the sum of the two unimodal responses, this is attributed to the interaction between the two modalities (Giard and Peronnet, 1999; Molholm et al., 2002; Klucharev et al., 2003; Besle et al., 2004; Teder-Sälejärvi et al., 2005; Stekelenburg and Vroomen, 2007; Vroomen and Stekelenburg, 2010). Critical comparisons in the current study were between the AV interactions of which visual and auditory signals originated from the same location and AV interactions of which visual and auditory signals originated from different locations. We, therefore, calculated congruent (AcVc − Vc) and incongruent (AlVc − Vc) difference waves and compared them to the corresponding A conditions (i.e., Ac and Al, respectively). The auditory N1 and P2 had a central maximum, and analyses were, therefore, conducted at nine central electrodes surrounding Cz. The peak amplitude of N1 was scored in a window of 70–150 ms. The peak amplitude of P2 was scored in a window of 120–250 ms. To test possible differences in AV interactions of congruent and incongruent sound locations, N1 and P2 scores of the congruent and incongruent difference waves were subtracted from the corresponding A conditions; (Ac − [AcVc − Vc]) and (Al − [AlVc − Vc]). These difference scores were submitted to a repeated measures MANOVA with as within-subjects variables Congruency (AV locations congruent versus incongruent) and Electrode (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2).

RESULTS

Participants detected 99% of the catch trials, indicating that they indeed watched the monitor. **Figures 2** and **3** show that for both auditory locations, AV interactions were associated with N1 and P2 suppression. As reported before (Stekelenburg and Vroomen, 2007), the video substantially reduced the amplitude of the auditory N1, $[F_{(1,\ 21)} = 30.54, p < 0.001]$. Most importantly, this intersensory effect was influenced by spatial Congruency, $[F_{(1,\ 21)} = 5.53, p < 0.05]$, indicating that the N1 reduction was greater for the spatially congruent AcVc condition (a $1.7\,\mu V$ reduction) than for the spatially incongruent AlVc condition (a $1.1\,\mu V$ reduction). This congruency effect was not affected by Electrode ($F < 1$) (**Figure 4**).

We also tested for each sound location separately the intersensory effect on N1 amplitude with the variables Modality (A versus AV − V) and Electrode in a repeated measure MANOVA. For both congruent and incongruent conditions, N1 suppression was significant, $[F_{(1,\ 21)} = 29.99, p < 0.001, F_{(1,\ 21)} = 17.83, p < 0.001]$, respectively. To further delineate whether the difference in N1 suppression should be attributed to differences in A-only versus AV − V, we separately tested A-only and AV − V between the two locations. The N1 of the AcVc − Vc condition was smaller than AlVc − Vc condition, $[F_{(1,\ 21)} = 14.66, p < 0.01]$, but there was no difference in the N1 between Ac and Al, $[F_{(1,\ 21)} = 1.39, p = 0.25]$. This further suggests that the effect of location on N1-suppression was due to differences in AV integration, and not to differences in Ac versus Al *per se*. The same MANOVA as for the N1 amplitude on the latency scores showed that N1 latency was not affected by stimulus modality ($F < 1$), nor was there an effect of Congruency ($F < 1$) or a Congruency × Electrode interaction, $[F_{(1,\ 21)} = 1.02, p = 0.47]$.

The same MANOVA on the P2 showed that the P2 was also reduced in amplitude ($1.9\,\mu V$) and speeded up (7 ms) alike in the bimodal conditions AcVc and AlVc, $[F_{(1,\ 21)} = 23.03, p < 0.001]$ and $[F_{(1,\ 21)} = 5.98,\ p < 0.05]$, respectively. Importantly, the intersensory effects on the P2 amplitude and P2 latency were *not* affected by Congruency, $[F_{(1,\ 21)} = 1.07, p = 0.31$ and $F < 1]$, respectively. There were also no Congruency × Electrode interactions for P2 amplitude (**Figure 4**) and P2 latency, $[F_{(1,\ 21)} = 1.16, p = 0.38$ and $F_{(1,\ 21)} = 1.33, p = 0.31]$.

Figure 2 also suggests that there was an early effect of spatial congruency at the P50 component. The P50 was scored by calculating mean activity in a 40–60 ms window and showed a maximum at the fronto-central electrodes. The difference scores (Ac − [AcVc − Vc]) and (Al − [AlVc − Vc] were submitted to a repeated measures MANOVA with the within-subjects variables Congruency (AV locations congruent versus incongruent) and Electrode (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2). There was a Congruency × Electrode interaction, $[F_{(8,\ 14)} = 2.90, p < 0.05]$. Simple effect test, examining the effect of congruency at each electrode, showed that this effect was localized mainly at electrode Cz ($p < 0.05$). Separate tests for the congruent and incongruent conditions showed that at Cz significant AV interactions were found for congruent presentations, $[t_{(21)} = 2.62, p < 0.05]$, but not for incongruent presentations ($t < 1$).

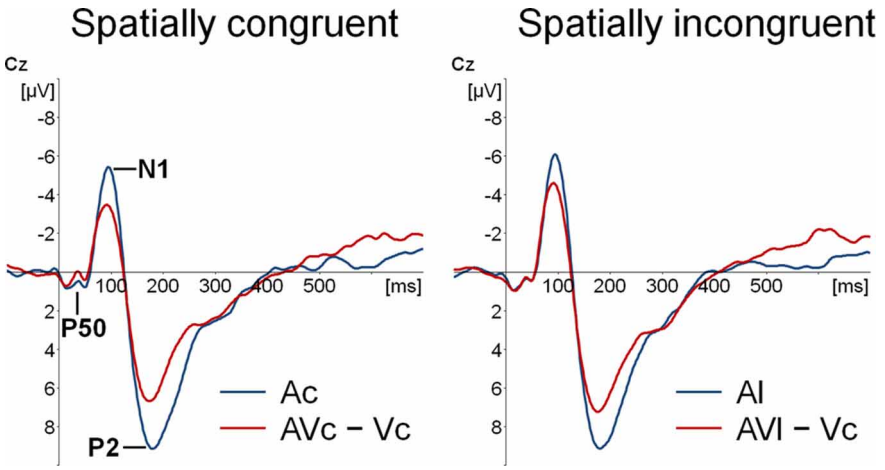

FIGURE 2 | Event-related potentials (ERPs) time-locked to auditory onset at electrode Cz. Separately for spatially congruent and incongruent audiovisual presentations the auditory-only (Ac and Al) ERP and the audiovisual minus visual-only difference wave (AVc − Vc and AVl − Vc) are displayed.

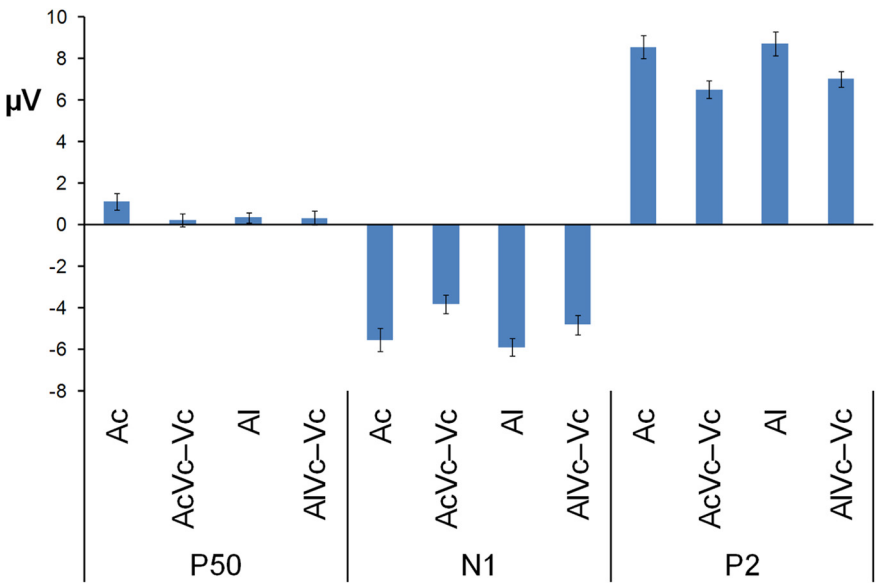

FIGURE 3 | Mean voltage in μV of P50, N1, and P2 averaged across the electrodes used in the MANOVA for auditory-only stimuli presented centrally and laterally (Ac and Al) and audiovisual difference waves for spatially congruent and incongruent AV presentations (AVc − Vc and AVl − Vc). The bars indicate one standard error of mean.

DISCUSSION

Our results support theoretical models that assume that the brain uses distinct sources of information to predict subsequent sensory inputs. More specifically, our results replicate the by now well-established finding that suppression of auditory N1 and P2 constitutes the neural consequence of an interaction of audiovisual stimuli containing anticipatory visual motion (Besle et al., 2004; van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2007; Arnal et al., 2009; Vroomen and Stekelenburg, 2010). One small difference with a previous study (Stekelenburg and Vroomen, 2007) that used the same stimuli was that both N1 and P2 peaked earlier in the bimodal condition whereas in the current study latency facilitation was limited to P2. It may be inferred that the latency facilitation of N1 is less robust than the suppression of the N1 amplitude. This is further supported by a study using audiovisual speech also showing a reduction in N1 amplitude, but not in N1 latency (Besle et al., 2004). The new finding here is that the N1 suppression was greater for spatially congruent than incongruent AV stimuli. We hypothesized that a visual signal that is naturally leading the auditory signal would allow observers to predict not only the onset and content, but also the location of the sound. As demonstrated before, temporal prediction is predominantly reflected in N1-suppression because it only occurs when anticipatory visual movements reliably predict sound onset,

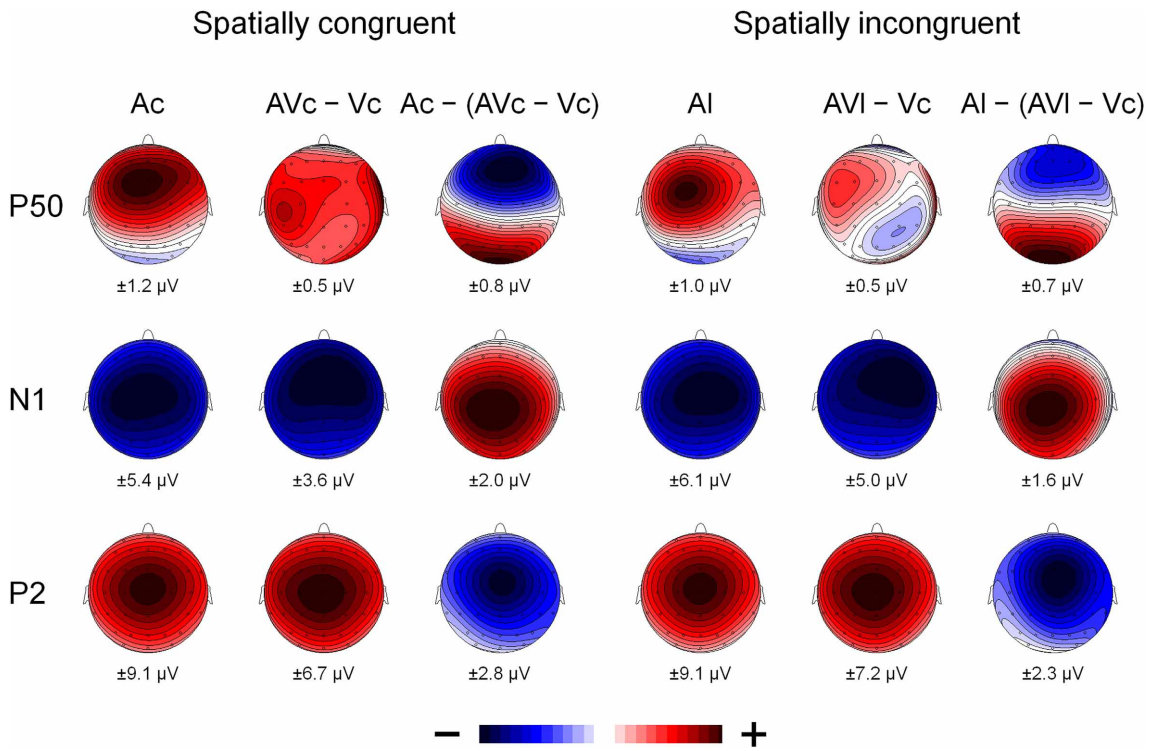

FIGURE 4 | The scalp topography of P50, N1, and P2 for the A-only ERPs (Ac and Al), the AV difference waves (AVc − Vc and AVl − Vc) and the multisensory interactions represented by the A-only ERP minus the AV difference wave (Ac − [AVc − Vc] and Al − [AVl − Vc]) for spatially congruent and incongruent presentations. The range of the voltage maps in μV is displayed below each map.

while it is abolished when vision does not predict sound onset (Stekelenburg and Vroomen, 2007; Vroomen and Stekelenburg, 2010). In a similar vein, it thus appears that the N1 is also sensitive to spatial prediction, given that the visually induced auditory N1-suppression was reduced when the auditory location did not match the predicted location. It thus seems likely that the N1 suppression reflects a process in which both the temporal onset and the location of the sound is predicted on the basis of the leading visual signal.

The spatially congruent AV stimuli also induced early integration effects at 40–60 ms at the central sites, while no such early integration was found for spatially discordant AV stimuli. Similar early AV interactions have been demonstrated in studies on AV integration with more basic artificial stimuli (Giard and Peronnet, 1999; Molholm et al., 2002; Talsma et al., 2007; Sperdin et al., 2009). This suggests that spatial congruity is a necessary condition for these early interactions.

Whereas spatial congruency affected AV interactions at N1, no such effect was found at the P2 component. This is in line with a study demonstrating that same- and different AV location pairings showed both similar and different AV interactions (Teder-Sälejärvi et al., 2005). The dissociation between N1 and P2 effects also verifies the hypothesis of a study (Klucharev et al., 2003) stating that the AV interactions at N1 reflect interactions in the processing of general features shared by the acoustic and visual stimulus, specifically spatial and temporal correspondence, while later interactions at P2 latency reflect interactions at phonetic, semantic, or associative level (Stekelenburg and Vroomen, 2007). The distinction between these two qualitatively different integration mechanisms with different underlying time-courses is supported by a MEG/fMRI study (Arnal et al., 2009). The latter study proposed that two distinct neural routes are involved in the audiovisual integration of speech. These authors conjectured that predictive visual information affects auditory perception via a fast direct visual to auditory pathway which conveys physical visual but no phonological characteristics. After the visual-to-auditory predictive mechanism a secondary feedback signal is followed via STS, which signals the error (if present) between visual prediction and auditory input. Because visual predictive information about auditory location affects early (P50, N1) potentials, it seems reasonable to maintain that within this dual route model, AV integration of location is realized via a fast direct route.

One can also ask to which extent the present results are modulated by the effects of attention on multisensory processing. The task of participants was to detect visual catch trials in the center of fixation. This implies that in the congruent condition, auditory stimuli were presented at the attended location (the center), whereas in the incongruent condition they were presented at an unattended location. Could it be, then, that differences in "spatial attention" rather than "accuracy of sensory prediction" underlie the present results of spatial congruency. Indeed, it has been argued that attention can modulate the neural correlates of multisensory integration (Talsma et al., 2010). Typically,

in these studies (Senkowski et al., 2005; Talsma and Woldorff, 2005) attention is manipulated by presenting auditory, visual, and audiovisual stimuli randomly to two lateral spatial positions and instructing participants to focus their attention at only one of these locations during a block of trials. When stimuli are presented at the attended location, multisensory (AV) stimuli elicit *larger* ERP waveforms (N1, P2) than the sum of the visual and auditory (A + V) parts alone, whereas at the unattended location, the difference between the AV and A + V is smaller. Note that this result is exactly the opposite pattern what was found here, because we obtained *smaller* ERPs, not larger, if sounds were presented at the audiovisual congruent (i.e., attended) location. In addition, the interaction of attention with multisensory integration was associated with enhanced late fronto-centrally distributed potentials (Busse et al., 2005; Talsma and Woldorff, 2005), whereas in the current study there was no hint of late congruency effects. We conjecture that the critical difference is that we used stimuli with visual predictive information that preceded sound onset, whereas these other studies used *synchronized* AV stimuli, thus without visual anticipatory information. Future studies might try to further disentangle the effects of spatial attention and sensory prediction on multisensory integration. One could, for example, envisage a study in which visual stimuli with predictive information are presented at fixation or far from fixation, while sounds are presented from audiovisual congruent or incongruent locations. On the attentional account, distance from fixation should matter, while on the sensory prediction account it is the spatial congruency between the auditory and visual information that matters.

The here reported effects of spatial congruity differ in several aspects (timing and location over the scalp) from earlier studies on spatial location. One study (Teder-Sälejärvi et al., 2005) found ERP interactions that differed according to spatial congruity which included a phase and amplitude modulation of visual-evoked activity localized to the ventral occipito-temporal cortex at 100–400 ms, and an amplitude modulation of activity localized to the superior temporal region at 260–280 ms. Another study (Gondan et al., 2005) also found effects of spatial congruity as ERPs to spatially congruent and spatially incongruent bimodal stimuli started to differ over the parietal cortex around 160 ms after stimulus onset. We conjecture, though, that a critical difference is that both studies used *synchronized* AV stimuli, thus without visual anticipatory information. Other potentially relevant differences are that we used natural rather than artificial stimuli (flashes and beeps), and we used a larger degree of separation between auditory and visual stimuli [90° in our study versus 40° in Gondan et al. (2005) and 60° in Teder-Sälejärvi et al. (2005)].

Our study is also relevant for the question as to whether N1-suppression to audiovisual presentations is evoked by factors other than visual prediction. Initial studies on visually induced suppression of auditory N1 (e.g., Besle et al., 2004; van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2007) cannot rule out that visual anticipatory movement might have summoned involuntary transient attention to the visual modality, thereby depleting attentional resources of the auditory modality (Pilling, 2009). This depletion of auditory resources might then be reflected in a suppression of the auditory N1. However, if this kind of non-spatial depletion of auditory resources were the sole determinant of the N1-supression, one would expect no differential effect of spatial congruity on N1-suppression because it should be identical for both congruent and incongruent locations. The current results, therefore, refute a depletion account of N1-suppression.

In summary, we found that the auditory-evoked N1 and P2 were suppressed when accompanied by their corresponding visual signals. These sub-additive AV interactions have previously been attributed to visual prediction of auditory onset (Stekelenburg and Vroomen, 2007; Vroomen and Stekelenburg, 2010). The crucial finding here is that spatial congruity between A and V also affected AV interactions at early ERP components: for spatially incongruent pairings, no AV interactions at P50 and less N1-suppression was found than for spatially congruent pairings, whereas suppression of P2 remained unaffected by spatial congruency. This suggests that visuo-spatial and visuo-temporal information have different time-courses in AV integration: spatial prediction has earlier effects on auditory processing (P50, N1) than temporal prediction (N1, P2).

REFERENCES

Arnal, L. H., Morillon, B., Kell, C. A., and Giraud, A. L. (2009). Dual neural routing of visual facilitation in speech processing. *J. Neurosci.* 29, 13445–13453.

Barth, D. S., Goldberg, N., Brett, B., and Di, S. (1995). The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex. *Brain Res.* 678, 177–190.

Besle, J., Fort, A., Delpuech, C., and Giard, M. H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur. J. Neurosci.* 20, 2225–2234.

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H. J., and Hillyard, S. A. (2007). Neural basis of the ventriloquist illusion. *Curr. Biol.* 17, 1697–1703.

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., and Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18751–18756.

Colin, C., Radeau, M., Deltenre, P., and Morais, J. (2001). Rules of intersensory integration in spatial scene analysis and speechreading. *Psychol. Belg.* 41, 131–144.

Gentsch, A., Kathmann, N., and Schütz-Bosbach, S. (2012). Reliability of sensory predictions determines the experience of self-agency. *Behav. Brain Res.* 228, 415–422.

Gentsch, A., and Schütz-Bosbach, S. (2011). I did It: unconscious expectation of sensory consequences modulates the experience of self-agency and its functional signature. *J. Cogn. Neurosci.* 23, 3817–3828.

Giard, M. H., and Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J. Cogn. Neurosci.* 11, 473–490.

Gondan, M., Niederhaus, B., Rösler, F., and Röder, B. (2005). Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. *Percept. Psychophys.* 67, 713–726.

Gratton, G., Coles, M. G., and Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* 55, 468–484.

Klucharev, V., Möttönen, R., and Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Brain Res. Cogn. Brain Res.* 18, 65–75.

Lange, K. (2009). Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain Cogn.* 69, 127–137.

Lange, K. (2010). Can a regular context induce temporal orienting to a target sound? *Int. J. Psychophysiol.* 78, 231–238.

Martikainen, M. H., Kaneko, K., and Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cereb. Cortex* 15, 299–302.

McCarthy, G., and 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.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., and Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res. Cogn. Brain Res.* 14, 115–128.

Pilling, M. (2009). Auditory event-related potentials (ERPs) in audiovisual speech perception. *J. Speech Lang. Hear. R.* 52, 1073–1081.

Schafer, E. W., and Marcus, M. M. (1973). Self-stimulation alters human sensory brain responses. *Science* 181, 175–177.

Senkowski, D., Talsma, D., Herrmann, C. S., and Woldorff, M. G. (2005). Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. *Exp. Brain Res.* 166, 411–426.

Sperdin, H. F., Cappe, C., Foxe, J. J., and Murray, M. M. (2009). Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Front. Integr. Neurosci.* 3:2. doi: 10.3389/neuro.07.002.2009

Stekelenburg, J. J., and Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cogn. Neurosci.* 19, 1964–1973.

Talsma, D., Doty, T. J., and Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb. Cortex* 17, 679–690.

Talsma, D., Senkowski, D., Soto-Faraco, S., and Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* 14, 400–410.

Talsma, D., and Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J. Cogn. Neurosci.* 17, 1098–1114.

Teder-Sälejärvi, W. A., Di Russo, F., McDonald, J. J., and Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *J. Cogn. Neurosci.* 17, 1396–1409.

van Wassenhove, V., Grant, K. W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1181–1186.

Vroomen, J., and Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *J. Cogn. Neurosci.* 22, 1583–1596.

Winkler, I., Denham, S. L., and Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci.* 13, 532–540.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 March 2012; accepted: 14 May 2012; published online: 31 May 2012.

Citation: Stekelenburg JJ and Vroomen J (2012) Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. Front. Integr. Neurosci. ***6****:26. doi: 10.3389/fnint.2012.00026*