Multisensory processing of naturalistic objects in motion: A high-density electrical mapping and source estimation study

Daniel Senkowski, Dave Saint-Amour, Simon P. Kelly, and John J. Foxe

Introduction

Imagine a drop hitting a water surface. Effortlessly, our senses merge the visual percept of the drop with the sound of the splash. The gain in speed and accuracy afforded by the multisensory integration of inputs to the different sensory modalities allows us to rapidly process the features of naturalistic objects in motion. The study of multisensory processing of naturalistic objects in motion is therefore crucial to our understanding of brain processing in everyday life (De Gelder and Bertelson, 2003).

Traditionally, integrative multisensory processing was thought to occur exclusively at higher-order cortical convergence zones, while processing within lower-level sensory cortices was considered to be specific to a single modality. Recent electrophysiological and functional magnetic resonance imaging (fMRI) studies, however, have demonstrated that inputs from multiple sensory modalities can be integrated at the earliest stages of cortical processing, in regions that were originally thought to be modality specific (Calvert, 2001; Fort et al., 2002; Foxe et al., 2000, 2002; Foxe and Schroeder, 2005; Giard and Peronnet, 1999; Molholm et al., 2002; Schroeder et al., 2003; Schroeder and Foxe, 2002, 2005; Senkowski et al., 2005, 2006; Talsma and Woldorff, 2005). In most of these studies simple non-naturalistic stimuli, such as sinusoidal tones and static visual gratings were used to examine multisensory processing. The multiple sensory inputs that continuously enter our sensory systems in naturalistic situations, however, are often more complex. Thus far, it is unknown at which stages in the processing hierarchy multisensory inputs of naturalistic objects in motion are integrated.

Predictions about when and where naturalistic audiovisual objects in motion are integrated in the brain can be derived from findings relating to unisensory object and motion processing within the visual and auditory modalities. In a recent event-related potentials (ERP) study on auditory discriminative object processing comparing sounds of living versus man-made objects, Murray et al.
reported the earliest effects of auditory object discrimination about 100 ms after stimulus onset in right temporal and left inferior frontal cortex. In addition, fMRI studies have revealed an involvement of the superior temporal (STG) and middle temporal gyrus (MTG) in the processing of environmental sounds (Lewis et al., 2004, 2005; Thierry et al., 2003). Research on object recognition in the visual modality, on the other hand, has shown that static naturalistic pictures can be processed on a semantic level within the first 150 ms after stimulus presentation (Bacon-Mace et al., 2005; Thorpe et al., 1996; VanRullen and Thorpe, 2001; Doniger et al., 2000, 2001). FMRI studies, in which auditory and visual motion processing was directly compared, have shown conjoint activations in response to auditory and visual motion stimuli in areas involving lateral parietal cortex, lateral frontal cortex, anterior midline, anterior insular cortex, intraparietal sulcus, and ventral premotor cortex (Lewis et al., 2000; Bremmer et al., 2001). In line with other studies (Beauchamp et al., 2002; Ferber et al., 2003; Grossmann and Blanke, 2002), visual motion processing was also associated with activations in medial temporal gyrus (including area MT), superior temporal sulcus, and lateral occipital complex (LOC).

The processing of naturalistic multisensory objects in motion has been examined in recent MRI studies by Beauchamp and colleagues (Beauchamp et al., 2004a,b; Beauchamp, 2005). These studies have revealed a network of cortical areas for naturalistic multisensory stimulus integration including LOC, STG, and MTG. Due to the low temporal resolution of fMRI, however, these studies did not provide information about the temporal dynamics underlying naturalistic multisensory object processing. In addition, no non-naturalistic motion control stimuli were used in these recent fMRI studies. For instance, Beauchamp (2004b) compared the interactions for the processing of semantically congruent multisensory stimuli (e.g., a moving hammer and a bang sound) with the interactions for the processing of semantically incongruent multisensory stimuli (e.g., a moving hammer and a ring sound). Thus, the degree to which the integration effects previously reported by Beauchamp and colleagues are specific to naturalistic stimulation remains to be elucidated.

In the present ERP study, we examined the neuronal dynamics underlying the integration of short video clips of naturalistic multisensory objects in motion (i.e., a naturalistic water splash) using high-density electrical mapping and local autoregressive averages (LAURA) distributed source estimation (Grave de Peralta et al., 2001, 2004). The specificity of integration effects to naturalistic object processing was investigated by comparing multisensory interactions for naturalistic stimuli with those for abstract but to some extent comparable non-naturalistic control stimuli. Using the high temporal resolution of EEG, this study aimed to provide detailed information about the timing of multisensory integration of naturalistic objects. In addition, the application of the LAURA distributed linear inverse solution enabled us to examine the neuronal structures associated with the integration of naturalistic objects in motion. In line with previous studies on multisensory integrative processing of basic auditory and visual information (Giard and Peronnet, 1999; Molholm et al., 2002), we expected to find the earliest interactions for naturalistic objects in motion in early sensory areas. Since frontal cortical areas are also involved in early sensory integrative processing (e.g., Molholm et al., 2002), we also expected to observe an early involvement of frontal areas. Finally, based on the studies by Beauchamp and colleagues, we also predicted that effects of multisensory processing would be found in STG and MTG.

### Methods

#### Participants

Twelve neurologically normal paid volunteers participated in the study. One participant was excluded from the analysis on the basis of extensive eye movement artifacts. The remaining 11 participants (all right-handed, mean age 26.7±5.9, 6 females) reported normal hearing and had normal or corrected-to-normal vision. The Institutional Review Board of the Nathan Kline Institute for Psychiatric Research approved the experimental procedures, and each subject provided written informed consent.

#### Procedure

Participants were presented with a random stream of naturalistic and non-naturalistic video clips, and static “target” stimuli. Each stimulus class (i.e., naturalistic and non-naturalistic) consisted of unisensory-auditory (A), unisensory-visual (V), and multisensory AV stimuli. The subject’s task was to silently count target stimuli and to verbally report the total number of targets to the experimenter after each block. Target probability was 15% with A, V and AV equally likely.

In addition, about 1247 “no-stimulus” events (~30%) were intermixed in the stream of naturalistic, non-naturalistic, and target stimuli. No-stimulus events, which have been used previously in faster-rate fMRI studies (e.g., Eimer et al., 2002; Senkowski et al., 2005), are “blank” trials containing no stimulation, randomly distributed in the trial sequence as if they were real stimulus events.

Busse and Woldorff (2003) have recently shown that for a presentation rate of about 30%, these no-stimulus events do not elicit physiological responses by themselves. The selective average of these responses represents the average overlap of adjacent trials. As demonstrated by Talsma and Woldorff (2005), subtracting the time-locked averages of experimental trials from no-stimulus events can effectively eliminate overlapping activity from adjacent trials and/or anticipatory activity, which are contaminating factors when unisensory responses are summed and compared with multisensory stimuli (see e.g., Teder-Sälejärvi et al., 2002). Each subject underwent 35 experimental blocks, each containing 84 trials (naturalistic, non-naturalistic, or targets) and 36 no-stim events. All stimuli were presented with a randomly varying inter-stimulus-interval (ISI; measured from the offset of one trial to the onset of the next) between 500 and 800 ms (mean 650 ms). The duration of each experimental block was about 2 min.

#### Stimuli

A naturalistic water splash clip, recorded with a RedLake MotionScope camera at frame rates of 250–1000 per second, was used with permission of the North Carolina School of Science and Mathematics. Since the splash clip was originally filmed without sound, the sound of a naturalistic splash was separately recorded and digitized at 44.1 kHz with 16-bit resolution and dubbed over the clip for the naturalistic multisensory stimulation. The
naturalistic and non-naturalistic clips are provided in the online supplementary material. Fig. 1 illustrates the naturalistic and non-naturalistic multisensory clips, which consisted of 14 frames of 40 ms duration, leading to a total clip duration of 560 ms. For the non-naturalistic control stimuli, each of the 14 naturalistic frames was filtered using a Gaussian blur filter (filter radius = 40 pixels), as implemented in Adobe® Photoshop® 6.01. The filter blurs an image while keeping the luminance level constant. Thus, the non-naturalistic control clip consisted of a visual motion pattern, which was comparable with the splash motion in terms of luminance level.

The auditory stimuli had a duration of 180 ms and a maximum sound-pressure-level (SPL) of 66 dB. To create the non-naturalistic auditory stimulus, the naturalistic stimulus waveform was convolved with a brown noise signal using Cool Edit Pro 2.0 (Syntrillium Software). Thus, the non-naturalistic auditory stimuli here consisted of brown noise modeled by the sound of the naturalistic water splash. This has the effect of balancing the global energy level in naturalistic and non-naturalistic auditory stimuli. The Fast Fourier Transform (FFT) (Hamming window, segment length = 2048) was applied to ensure the similarity of amplitude spectra for both stimulus classes. Naturalistic and non-naturalistic multisensory AV clips consisted of a combination of visual clips and auditory stimuli. The visual stimulus onset preceded the sound onset by 100 ms for all multisensory stimuli (Fig. 1).

Target stimuli consisted of a Gabor patch (unisensory-V targets), a 1000 Hz sinusoidal tone (unisensory-A targets) and the Gabor patch presented together with the tone (multisensory AV targets). The Gabor patch was a Gaussian-windowed sinusoidal checkerboard (spatial frequency = 1 cycle per degree; Gaussian standard deviation = 1.5; contrast = 90%) having the same size and mean luminance (43 cd/m²) as the naturalistic and non-naturalistic visual stimuli. Target tones were presented at 66 dB SPL, with 5 ms linear rise and fall times. Visual, auditory, and audiovisual targets had a duration of 100 ms.

The auditory stimuli (naturalistic, non-naturalistic, and targets) were presented through two speakers placed to the left and to the right of the monitor. Visual stimuli were centrally presented and participants were instructed to fixate a central fixation cross that was visible during the ISI. All visual stimuli (naturalistic, non-naturalistic, and targets) subtended an angle of 9.8° vertically and 8.9° horizontally.

**Data acquisition and event-related potentials**

High-density EEG recordings were acquired using a BioSemi Active-Two electrode system with 4 EOG and 164 scalp electrodes. Data were band-pass filtered from 0.05 to 100 Hz during recording at a sample rate of 512 Hz and off-line re-referenced to average reference. To further reduce high frequency noise, a 55 Hz low-pass filter was applied prior to the data analysis. The data were epoched from −200 ms before to 500 ms after visual onset. Baselines were computed from a −100 ms to 0 ms time interval before visual onset and subtracted for each trial. For artifact rejection, trials were automatically excluded from averaging if the standard deviation within a moving 200 ms time window exceeded 30 μV in any one of the channels in the averaging epoch and channels with extreme high and/or low frequency artifacts throughout the entire recording were inter-
polated using a model of the amplitude topography at the unit sphere surface using all channels without high and low frequency artifacts (Perrin et al., 1989, 1990). After artifact rejection, an average of 303 trials per condition was accepted into the analysis for each subject. The time locked average of the no-stimulus events (see above) was subtracted from the ERPs of naturalistic and non-naturalistic stimuli for each subject individually in order to eliminate overlapping activity from adjacent trials and/or anticipatory activity. An illustration of grand-average ERPs for naturalistic stimuli before and after no-stimulus subtraction is provided in the supplement material (Supplement Figure 1).

**Data analysis**

To verify that participants attended to the stimuli throughout the experiment, the number of targets verbally reported by the participants after each experimental run was compared with the
actual number of targets. Since the only purpose of the target stimuli was to control for the participants attention, the ERPs of these stimuli were not analyzed.

In Fig. 2 unisensory-auditory and unisensory-visual ERPs are shown with naturalistic and non-naturalistic conditions superimposed. It can be seen that the naturalistic and non-naturalistic ERP traces are similar in morphology and in the latencies of prominent components. Fig. 3 shows the ERPs to naturalistic and non-naturalistic multisensory-audiovisual stimuli. Of particular interest was the “rhythmic” character of the ERP pattern, with multiple cycles apparent for both naturalistic and non-naturalistic multisensory stimuli.

In line with previous studies (e.g., Calvert, 2001; Fort et al., 2002; Giard and Peronnet, 1999; Molholm et al., 2002; Murray et al., 2005a; Saint-Amour et al., 2007; Talsma et al., 2007), multisensory interactions in ERPs were assessed by directly comparing the ERPs to multisensory stimuli with the linear summation of the responses to the respective unisensory constituents (ERP AV vs. (ERP A+ERP V)). The statistical analysis of multisensory interactions focused on the three distinct main response intervals shown as shaded areas in Fig. 3: 120–140 ms, 210–250 ms, and 300–350 ms after sound onset. The statistical analysis was carried out on integrated potential measurements averaged across electrodes and time interval within each of three regions of interest (ROI) that were selected based on the topographical response pattern (Figs. 4 and 5). Two ROIs over lateral occipital scalp regions, each consisting of 6 channels, and one medial-central ROI that consisted of 5 electrodes were selected (Fig. 4). Fig. 5 shows the topographical maps of naturalistic and non-naturalistic ERPs for the three distinct response intervals. The figure illustrates a stable activity pattern over lateral occipital scalp across the three analysis intervals and suggests differences in integration effects between naturalistic and non-naturalistic stimuli. Analyses to this point suggest, to some extent, comparable processing mechanisms for naturalistic and non-naturalistic stimuli, which is supported by (1) the similar activity traces shown in Figs. 2–4, and (2) the almost identical topographical distribution shown in Fig. 5.

To establish the presence of multisensory interactions, three levels of analysis were applied: The first level of the analysis was conducted to render a full description of the spatio-temporal properties of the AV interactions for naturalistic and non-naturalistic stimuli. The multisensory stimuli were compared with the linear summation of the responses to the respective unisensory constituents using point-wise running t-tests (two-tailed) for each scalp electrode. Comparable with previous studies (e.g., Molholm et al., 2002), an AV interaction was defined as at least 11 consecutive data points meeting a 0.05 alpha criterion (11 data points = 21.5 ms at a 512 Hz digitization rate) (for details of the use of running t-tests see e.g., Guthrie and Buchwald, 1991). In the second level of analysis repeated measures analyses of variance (ANOVA) were performed separately for the three main response intervals described above using the within-subject factors stimulus class (naturalistic and non-naturalistic), stimulus type (AV and A+V), and ROI (medial-central, left-occipital, right-occipital). In the third level of analysis, follow up ANOVAs using the factor stimulus type (AV and A+V) were performed separately for the three ROIs and separately for naturalistic and non-naturalistic stimuli when effects of the factor stimulus class were found significant in the second level of analysis.

Electrical source estimation

To examine the sources underlying multisensory integration effects obtained at the scalp, the LAURA distributed source estimation approach was applied using Cartool software (http://brainmapping.unige.ch/Cartool.php). LAURA is based on a linearly distributed inverse solution, “...where the relationship between the brain activity at one point and its neighbors is expressed in terms of a local autoregressive estimator with coefficients depending upon a power of the distance from the point.” (Grave de Peralta et al., 2001, p. 533). Technical details about the LAURA regularization strategy have been explicitly described by Grave de Peralta et al. (2001, Appendix B). The
sources of the LAURA approach were fitted into a realistic head model with a solution space of 4024 nodes, where voxels are restricted to the gray matter of the Montreal Neurological Institute’s (MNI’s) average brain divided into a regular grid with 6 mm spacing. A particular advantage of the LAURA approach in comparison to the dipole source analysis approach is that it is capable of dealing with multiple simultaneously active sources of a priori unknown location and makes no assumptions regarding the number or location of active sources (Grave de Peralta et al., 2004). Recent studies have successfully applied the LAURA approach for estimating the sources of auditory and visual ERPs (e.g., Murray et al., 2004, 2006; Sehatpour et al., 2006). For instance, Sehatpour et al. (2006) compared visual responses in lateral–occipital cortex during a perceptual closure task in an fMRI and an EEG study investigating the same subjects. The authors reported very similar regions of activation in the fMRI study and in the LAURA analysis of ERP data, providing evidence for the reliability of the LAURA approach. However, it is important to emphasize that the spatial resolution of this analysis approach is much lower than the spatial resolution of fMRI. For this reason, the results of the LAURA analysis here will be reported descriptively on a more global cortical perspective. The LAURA approach was applied for each subject and each stimulus condition separately and subsequently averaged across subjects for each condition. To estimate the Brodmann areas which best fit the LAURA source activity pattern, the location of each activity pattern was compared with the location of the Brodmann areas in Talairach space. To test for multisensory integration effects, within subject $t$-tests were performed on a node-by-node basis comparing the mean LAURA estimated activity of A+V vs. AV responses averaged across each of the pre-defined time intervals. Thereby, the mean of the LAURA inverse solution was estimated for each subject and each condition separately for the ERP averages across each of the three time intervals. Specifically, the LAURA source estimation was computed for each sample point before averaging across all sample points of a time interval in source space.

Results

Behavioral data for target stimuli

Participants correctly counted the number of targets in 81% of all experimental blocks. In 15% of blocks the participants miscounted the number of targets by one (e.g., counted 9 or 11 when there were actually 10 targets). In only 4% of all blocks (pooled for all subjects) participants miscounted the number of targets by more than one. These findings demonstrate that participants reliably attended to the stimuli during the experimental blocks.

Point-wise running $t$-tests for scalp electrodes

Fig. 6 shows the results of the point-wise running $t$-test for 160 scalp electrodes for naturalistic and non-naturalistic stimuli. The time intervals that were selected for the ROI analysis are highlighted in gray. There was no clear pattern of integration effects before 120 ms after sound onset. For the 120–140 ms time interval, however, significant effects were found at occipital and some frontal electrode sites for naturalistic stimuli, and that all effects are reported relative to sound onset.

---

1 In describing the results, we would remind the reader that onset of the visual clip preceded the sound onset by 100 ms for all multisensory stimuli,
whereas no consistent pattern of integration effects was observed for non-naturalistic stimuli. For the two longer latency time intervals (i.e., 210–250 ms and 300–350 ms), significant integration effects were observed at multiple locations across the scalp for both stimulus classes. Multisensory integration effects were examined in further detail in the ROI and LAURA analysis described below.

Multisensory interactions 120–140 ms after auditory onset

The three-way ANOVA for the 120–140 ms interval revealed a significant main effect of ROI ($F(2,20)=57.81, p<0.0001$) due to positive occipital and negative medial-central amplitudes. In addition, interactions between the factors stimulus class and stimulus type ($F(1,10)=8.14, p<0.02$), and stimulus type and ROI ($F(2,20)=57.81, p<0.02$) were observed. There was also a three-way interaction between the factors stimulus class, stimulus type, and ROI ($F(2,20)=3.83, p<0.04$), suggesting different patterns of integration effects for naturalistic and non-naturalistic stimuli. Follow-up ANOVAs were performed to test for multisensory interactions at each of the three ROIs, separately. For naturalistic stimuli a significant effect of stimulus type was found at the right ($F(1,10)=7.02, p<0.03$) and left occipital ROIs ($F(1,10)=7.04, p<0.03$). Occipital amplitudes for naturalistic stimuli were larger for A+V responses (average left/right=$8.69 \mu V$) compared to multisensory AV responses (average left/right=$6.82 \mu V$) (i.e., a sub-additive effect; see Figs. 4 and 5). For the non-naturalistic control stimuli, no integration effects were found, suggesting that there were no multisensory interactions in the 120–140 ms time interval for non-naturalistic stimuli. Due to the absence of significant interactions on the scalp level for non-naturalistic stimuli, the LAURA distributed linear inverse solution for the 120–140 ms interval were computed for naturalistic stimuli only.

Fig. 7 shows the LAURA distributed source estimations for naturalistic stimuli for AV and A+V responses. A dominant activity pattern in the lateral occipital complex (LOC) was found for AV and A+V responses. The voxel-by-voxel analysis revealed significant interactions in the comparison of A+V vs. AV LAURA source estimations in occipital cortex, right temporal lobe, right insular, and prefrontal regions including the right medial frontal gyrus (MFG).

Multisensory interactions 210–250 ms after auditory onset

The three-way ANOVA revealed a significant main effect of stimulus class ($F(1,10)=11.53, p<0.01$) and of ROI ($F(1,10)=29.98, p<0.0001$). The effect of stimulus class was due to higher amplitudes for naturalistic (3.96 $\mu V$) compared to non-naturalistic stimulation (2.48 $\mu V$), while the effect of ROI was due to different polarities for medial-central (negative) and occipital (positive) ERP amplitudes (Fig. 4). In addition, significant interactions between stimulus class and ROI ($F(2,20)=8.61, p<0.002$) and between stimulus type and ROI ($F(2,20)=8.99, p<0.002$) were found. The three-way interaction between the factors stimulus class, stimulus type, and ROI approached the significance level ($F(2,20)=2.80, p<0.08$), indicating a trend toward different patterns of integration effects for naturalistic and non-naturalistic stimulation. The follow-up ANOVAs for naturalistic stimuli revealed a significant effect of stimulus type at the right-occipital ROI ($F(1,10)=9.53, p<0.02$).
due to larger amplitudes for multisensory AV (7.36 μV) compared to A+V responses (6.7 μV), i.e., a super-additive effect. In addition, a significant effect of stimulus type was observed for the medial-central ROI ($F(1,10)=14.21, p<0.004$) that was explained by more negative amplitudes for multisensory (−2.66 μV) compared to A+V stimuli (−1.65 μV). The follow-up ANOVAs for non-naturalistic stimuli revealed a significant effect for the medial-central ROI ($F(1,10)=4.99, p<0.04$), again due to more negative amplitudes to multisensory AV (−1.49 μV) compared to A+V stimuli (−0.97 μV). The effect of stimulus type for the right and left occipital ROIs approached the significance level ($p<0.06$ and $p<0.09$, respectively).

The LAURA distributed source estimation for naturalistic and non-naturalistic AV and A+V responses revealed a clear activity pattern in LOC (Figs. 7 and 8). For naturalistic stimuli (Fig. 7, middle right panel), multisensory interactions were observed in regions including occipital cortex, left MFG and parts of left inferior frontal gyrus (IFG), right temporal lobe, right lateral precentral gyrus, left medial temporal gyrus (MTG), and parts of left superior temporal gyrus (STG). For non-naturalistic stimuli, robust multisensory interactions were observed in a network of areas including left MFG, left inferior parietal cortex, left MTG and STG, and right lateral precentral cortex (Fig. 8, upper right panel).

**Multisensory interactions 300–350 ms after auditory onset**

The three-way ANOVA for the 300–350 ms interval revealed a significant main effect of stimulus class ($F(1,10)=19.39, p<0.002$) and ROI ($F(2,20)=57.86, p<0.001$) (Figs. 4 and 5). The ROI amplitudes were larger for naturalistic (3.02 μV) compared to non-naturalistic stimuli (1.42 μV). The ROI effect was again due to negative medial-central (−2.65 μV) and positive occipital amplitudes (average left/right=4.66 μV). In addition, significant interactions between stimulus type and stimulus class ($F(1,10)=12.75, p<0.006$), stimulus type and ROI ($F(2,20)=11.08, p<0.0007$), stimulus class and ROI ($F(2,20)=5.04, p<0.02$), and a three-way interaction between stimulus class, stimulus type, and ROI ($F(2,20)=4.41, p<0.03$) were found. The follow-up ANOVAs for naturalistic stimuli revealed significant effects at the medial-central ROI ($F(1,10)=5.63, p<0.04$) and the right and left occipital ROIs ($F(1,10)=7.15, p<0.03$ and $F(1,10)=5.95, p<0.04$, respectively). The amplitudes at medial-central ROI electrodes were more negative for A+V (−3.4 μV) compared to AV (−2.72 μV) stimuli (sub-additive effect). At occipital ROIs larger amplitudes were observed for A+V (average left/right 6.39 μV) compared to AV (average left/right 5.75 μV) stimuli (sub-additive effect). For non-naturalistic stimuli, a significant effect was observed at the medial-central ROI ($F(1,10)=5.44, p<0.05$) due to more negative amplitudes for A+V (−2.47 μV) compared to AV (−2.02 μV) stimuli (sub-additive). The ANOVAs for the occipital ROIs did not reveal significant integration effects.

As for the earlier time intervals, the LAURA distributed source estimations for naturalistic and non-naturalistic AV and A+V stimuli were dominated by a clear activity pattern in LOC (Figs. 7 and 8). For naturalistic stimuli robust multisensory interactions were observed in right temporal lobe, frontal areas including the right MFG, and right precentral gyrus (including posterior MTG) (Fig. 7, lower right panel). Multisensory interactions for non-naturalistic stimuli were localized to right temporal lobe (including posterior MTG) and MFG. In addition, there were scattered interactions in precentral, occipital, and parietal cortices for non-naturalistic stimuli.

**Discussion**

We observed robust visual responses over lateral occipital scalp regions that were source-localized to lateral–occipital complex (LOC) for naturalistic and abstract non-naturalistic multisensory motion stimuli during the entire analysis interval.
This clearly supports the notion that the LOC is an important structure for the processing of audiovisual objects in motion (Beauchamp, 2005). LOC is also known to play an important role in multisensory object processing (Amari et al., 2001, 2002, 2005; Molholm et al., 2004; Murray et al., 2005b). However, although there were robust activations in LOC for naturalistic and non-naturalistic stimuli, there was no clear pattern of integration effects in this region, suggesting that auditory and visual inputs of multisensory objects in motion, as presented here, are not primarily integrated in LOC.

Multisensory integration effects for naturalistic and non-naturalistic stimuli were found in a widespread network of audiovisual convergence zones; beginning 120–140 ms after sound onset for naturalistic stimuli but somewhat later (210–250 ms) for non-naturalistic stimuli. The pattern of integration effects, while similar in some respects, also differed considerably between naturalistic and non-naturalistic stimuli. We discuss these findings in more detail below.

Multiple stages of multisensory integration for naturalistic objects in motion

The fast rate at which semantic visual object recognition is carried out has been demonstrated by Thorpe and colleagues (Bacon-Mace et al., 2005; Thorpe et al., 1996), who showed that even complex visual images presented in an unpredictable manner can be processed within the first 150 ms after stimulus presentation. For specific kinds of visual discrimination processes, like face recognition and face gender discrimination, effects have been described as early as 50 ms poststimulus (Seek et al., 1997; Mouchetant-Rostaing et al., 2000). In the present paradigm where there is only one ‘real’ object that is ever presented, it is likely that a near-complete object representation (i.e., of the splash) has been established in visual cortex by the time of sound onset, or very shortly thereafter (since the visual onset preceded the sound onset by 100 ms).

The earliest effects of multisensory processing of naturalistic objects in motion were observed 120 to 140 ms after the onset of the splash sound in a network of areas including occipital cortex, right lateral temporal lobe, insular, and right prefrontal cortex (including right MFG). The integration effects in early visual and auditory sensory areas are in line with previous studies on multisensory processing of basic audiovisual inputs (e.g., Giard and Peronnet, 1999; Molholm et al., 2002; Schroeder et al., 2003) and with findings from fMRI studies using naturalistic stimuli (Beauchamp et al., 2004b). A larger activation pattern for a multisensory compared to a unisensory motion discrimination task in insular cortex has been previously reported by Lewis et al. (2000). In addition, Bushara et al. (2001) have demonstrated that insular cortex plays an important role in the detection of auditory–visual stimulus onset asynchronies. The authors hypothesized that insular cortex mediates temporally defined multisensory audiovisual interactions. This assumption would fit with the effects in insular cortex for the naturalistic clips here where the visual onset preceded the auditory onset by 100 ms. The effects in right MFG are in line with the finding of interactions over right fronto-temporal scalp for the processing of audiovisual objects (Giard and Peronnet, 1999) and with the observation of right fronto-central interactions for basic audiovisual inputs (Molholm et al., 2002). The convergence of these findings suggests a key role for right frontal cortex in multisensory audiovisual processing.

Multisensory interactions in the 210–250 ms interval were observed in a network of areas including occipital cortex, right lateral temporal lobule, left MFG and IFG, right precentral gyrus, left MTG, and parts of left STG. Since it has been shown that discriminative auditory object processing directly involves left prefrontal areas (Murray et al., 2006), the multisensory interactions in left precentral cortex in the present study could be an indicator of an integration effect on the processing of the naturalistic sound. The integration effects in right precentral areas are in agreement with recent fMRI studies comparing motion processing in the auditory and visual modalities (Bremmer et al., 2001; Lewis et al., 2000), whereas the observation of interactions in left MTG and STG fits with the Beauchamp et al. (2004a,b) studies on naturalistic audiovisual object integration. As such, the present results extend evidence for an important role played by MTG, STG, and precentral areas in multisensory motion processing.

Multisensory integration effects for naturalistic stimuli in the 300–350 ms time interval were observed in left MFG, right lateral precentral gyrus, and right lateral temporal lobule, which were areas that also showed interactions in the 210–250 ms time interval. Interestingly, additional interactions were found in the posterior part of right MTG, which is in line with recent findings on the processing of naturalistic environmental sounds (Murray et al., 2006; Thierry et al., 2003). We therefore suggest that the integration effects in posterior MTG are likely to reflect higher-order integrative processing of the naturalistic splash sound.

Multisensory integration effects for naturalistic vs. non-naturalistic stimuli

The pattern of integration effects, while similar in some respects, differed considerably between naturalistic and non-naturalistic stimuli. First, integration effects in the 120–140 ms time interval were observed for naturalistic but not for non-naturalistic stimuli. Second, there were robust integration effects in inferior parietal cortex (IPC) for non-naturalistic stimuli but not for naturalistic stimuli.

The absence of multisensory integration effects in the 120–140 ms interval for non-naturalistic stimuli might be related to the fact that these stimuli are not encoded semantically. It is possible that the visual object in the naturalistic stimulation condition, where the splash clip is presented 100 ms prior to the splash sound onset, is semantically encoded by the time of sound onset (Bacon-Mace et al., 2005; Mouchetant-Rostaing et al., 2000; Thorpe et al., 1996). In contrast, the non-naturalistic stimuli, which are not identifiable as objects, do not have a semantically meaningful representation. That semantic processes are involved in the early integration effects for naturalistic stimuli is supported by the fact that the multisensory interactions for naturalistic stimuli were found in a network of areas that has been previously associated with semantic object processing. However, we cannot rule out the possibility that differences in physical stimulus properties between the two stimulus classes might have also contributed the absence of early integration effects for non-naturalistic stimuli. Although both stimulus classes were identical with respect to luminance, naturalistic visual motion stimuli contained higher spatial frequencies compared to non-naturalistic stimuli. This may provide one explanation for the larger P1 amplitudes and subsequent differences on later ERP components between naturalistic and non-naturalistic clips (Fig. 2). In addition, although the acoustic
stimuli were balanced for their overall spectral frequency and the non-naturalistic stimulus was created by directly convolving the naturalistic stimulus waveform with a brown noise signal, there were still some differences in the spectral energy across time between the two stimulus classes. Future studies are necessary to investigate the influence of the spectral acoustic energy and visual spatial frequency and on naturalistic and non-naturalistic audio-visual stimulus processing.

Of special interest was the observation of multisensory integration effects in inferior parietal cortex (IPC), which were observed exclusively for non-naturalistic stimuli. An increase in IPC activity has been previously reported for simultaneously presented but semantically unrelated audio-visual motion stimuli (i.e., visual motion dots in combination with auditory inputs that elicit the perception of a moving sound) (Lewis et al., 2000). In addition, it has been demonstrated that IPC plays an important role in spatial audio-visual integrative processing (Sestieri et al., 2006), suggesting that IPC supports integrative processing between auditory and visual inputs with spatially coherent motion. Given these previous findings, it would have to be considered somewhat surprising that the naturalistic motion stimuli used in the present study do not appear to have substantially engaged this region. One possible explanation might lie in the extent of processing that is necessary for semantically related versus semantically unrelated stimuli. That is, perhaps the additional IPC activity for unrelated auditory and visual inputs lies in the extended search for relatedness, whereas semantically meaningful stimuli are processed relatively automatically by this region and further semantic processing occurs downstream in higher-order processing regions of the frontal lobes. Clearly this account is speculative and will require further research where the extent of semantic relatedness between sensory inputs is systematically manipulated.

Another interesting finding in this study was that all effects in the LAURA source estimation analysis were sub-additive, i.e., the linear summation of A+V responses was larger than multisensory AV responses. At first glance, this observation seems to contradict the results obtained from the scalp recorded data, where some effects also appeared to be super-additive (e.g., AV responses were larger than A+V responses in the 210–250 ms time interval). However, the relative strength between conditions of activity obtained from a single scalp position or a scalp ROI does not necessarily directly correspond to the LAURA distributed source activity patterns, since the latter are derived from the relative activity distribution across the whole scalp.

Finally, one could ask whether the observed effects here might also be reflected in oscillatory responses. We and others have recently reported relationships between multisensory processing and oscillatory responses for basic auditory-visual stimulation (Sakowitz et al., 2005; Senkowski et al., 2006, 2007a,b). As illustrated in Fig. 3, the phases of activity appeared regularly-spaced about every 100 ms, which corresponds to the alpha frequency range (about 10 Hz). In an exploratory re-analysis of the here presented data, we have recently investigated possible multisensory integration effects in the alpha band (Kelly et al., 2006). In this re-analysis, we did not find evidence for multisensory processing in oscillatory alpha activity for naturalistic or non-naturalistic objects in motion. Future studies are necessary to examine the relationships between naturalistic multisensory processing and oscillatory responses in more detail.

Conclusion

This study focused on the neuronal dynamics of multisensory processing of naturalistic objects in motion. Multiple-stages of integration effects beginning at 120–140 ms after sound onset in occipital cortex, right temporal lobule, insular, and MFG were found for naturalistic objects in motion, suggesting that these regions play an important role during the first stage of integrative naturalistic object processing. Integration effects at longer latencies were observed in a widespread cortical network involving temporal, occipital, and frontal areas, possibly reflecting higher-order multisensory processing. The earliest integration effects for non-naturalistic stimuli were found 210–250 ms after sound onset in higher cortical areas including IPC. As such, there were clear differences in the cortical networks activated by multisensory motion stimuli as a consequence of the semantic relatedness of the separate sensory inputs.

Acknowledgments

This work was supported by grants from the U.S. National Institute of Mental Health (MH65350) and the National Institute of Aging (AG22696) to Dr. J.J. Foxe. We would like to express our sincere appreciation to Marina Shpaner and Jennifer Montesi for their technical assistance and to two anonymous reviewers for their helpful comments on this article. The Cartool software (http://brainmapping.unige.ch/Cartool.php) has been programmed by Denis Brunet, from the Functional Brain Mapping Laboratory, Geneva, Switzerland, and is supported by the Center for Biomedical Imaging (CIBI)of Geneva and Lausanne. We are also very grateful to Martin Winters from the North Carolina School of Science and Mathematics for his permission to use the splash clips for this experiment. The original clip and further clips can be found at: http://courses.ncssm.edu/hsi/splashes/video/video.htm.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2007.01.053.

References


