Multisensory events in our natural environment unfold at multiple temporal scales over extended periods of time. This functional magnetic resonance imaging study investigated whether the brain uses transient (onset, offset) or sustained temporal codes to effectively integrate incoming visual and auditory signals within the cortical hierarchy. Subjects were presented with 1) velocity-modulated radial motion, 2) amplitude-modulated sound, or 3) an in phase combination of both in blocks of variable durations to dissociate transient and sustained blood oxygen level-dependent responses. Audiovisual interactions emerged primarily for transient onset and offset responses highlighting the importance of rapid stimulus transitions for multisensory integration. Strikingly, audiovisual interactions for onset and offset transients were dissociable at the functional and anatomical level. Low-level sensory areas integrated audiovisual inputs at stimulus onset in a superadditive fashion to enhance stimulus salience. In contrast, higher order association areas showed subadditive integration profiles at stimulus offset possibly reflecting the formation of higher order representations. In conclusion, multisensory integration emerges at multiple levels of the cortical hierarchy using different temporal codes and integration profiles. From a methodological perspective, these results highlight the limitations of conventional event related or block designs that cannot characterize these rich dynamics of audiovisual integration.

Keywords: amplitude modulation, crossmodal, fMRI, multisensory, sustained, transient

Introduction

In our natural multisensory environment, events evolve at multiple temporal scales over extended periods of time. To predict the incoming sensory signals, the human brain may have adapted to the temporal structure of its environment. More specifically, the cortical hierarchy is thought to re-capitulate the temporal hierarchy of the environment with low-level areas encoding fast and high-level areas slow dynamic changes (Kiebel et al. 2008). Indeed, manipulating the temporal scales of the sensory inputs revealed a hierarchy of temporal receptive fields along the visual and auditory processing streams (Griffiths et al. 1998; Boemio et al. 2005; Hasson et al. 2008; Overath et al. 2008). In the visual domain, low-level visual areas maintained high response reliability despite severe disruptions of a video’s temporal structure, while response reliability in high-level areas decreased for piecewise scrambled videos (Hasson et al. 2008). In the auditory domain, the rate of noise bursts determined the neural activity and the associated shape of the blood oxygen level-dependent (BOLD) response (Harms and Melcher 2002; Harms et al. 2005) (for related findings using amplitude-modulated noise, see Giraud et al. 2000). For instance, in Heschl’s gyrus (HG), a sustained response for trains of noise bursts at low rate changed into transient responses for high rates possibly mediating subjects’ perceptual change from individually resolved noise bursts into a fused percept of a continuous stream (Harms and Melcher 2002). Importantly, the waveform of the BOLD response for a given rate varied across the cortical hierarchy. As in the visual system, low-level areas with putatively smaller temporal receptive fields showed a sustained response reflecting the series of individual bursts, while higher order areas exhibited more transient responses that were limited to train onsets and offsets (Seifritz et al. 2002; Harms et al. 2005) (for related findings in the visual domain, see Uludag 2008). In particular, transient offset responses have been implicated in edge detection in the visual and auditory modalities (Downar et al. 2000; Herdener et al. 2009). Collectively, these studies suggest that the brain may use multiple temporal response codes at different levels of the cortical hierarchy to analyze the temporal pattern of incoming visual and auditory signals (e.g., a sustained response component coding continuous sensory stimulation; transient onset and offset response components coding only the temporal edges of that sensory stimulation).

Temporal patterns are known as prominent cues for audiovisual integration in subcortical, sensory-specific, and higher order association areas (Calvert 2001; Amedi et al. 2005; Foxe and Schroeder 2005; Ghazanfar and Schroeder 2006; Driver and Noesselt 2008; Stein and Stanford 2008). Even though multisensory signals do not have to be precisely physically synchronous, they have to co-occur within a certain temporal window in order to be integrated (Wallace et al. 1996; Miller and D’Esposito 2005; Vatakis and Spence 2006; van Wassenhove et al. 2007; Lewis and Noppeney, forthcoming; Maier et al. forthcoming). Additional cues for audiovisual integration can be gained from the temporal covariation of continuous auditory and visual signals (Chandrasekaran et al. 2009). This raises the question whether multisensory interactions emerge primarily in transient or sustained temporal response codes. To our knowledge, previous studies have not been able to dissociate the contributions of transient onset or offset and sustained responses to multisensory integration, since they either conformed to rapid-event related or conventional block designs. The current functional magnetic resonance imaging (fMRI) study optimized the experimental design to characterize the roles of onset, offset, and sustained response components in audiovisual integration. Subjects were presented with 1) random dot kinematograms of radial motion, 2) auditory noise, or 3) the combination of both in blocks of variable durations. Both the velocity of the motion and the amplitude of the sound were modulated at 0.1 Hz to provide continuous synchrony cues for audiovisual integration. Modeling transient, sustained, and modulatory response components within one general linear model allowed us to address the following 3 issues: First, we identified superadditive and subadditive audiovisual interactions for the onset, offset, and...
sustained response components. We hypothesized that multi-sensory interactions would primarily emerge for the transient response code, since they reflect prominent temporal cues essential for audiovisual integration. Second, it is well established that sensory signals can be combined in superadditive, additive, or subadditive fashion and that signal intensity plays an important role in determining the multisensory integration profile (Perrault et al. 2005; Stanford et al. 2005; Werner and Noppeney 2009). The current study investigated whether the same sensory signals might be combined in superadditive or subadditive interactions for the different response components. Third, guided by previous findings in the unisensory domains (Seifritz et al. 2002; Uludag 2008), we investigated whether audiovisual interactions may be spatially segregated for different response components. A spatial segregation for audiovisual interactions of the various response components may indicate regional variation in temporal receptive field sizes, which in turn may inform us about a region's functional specialization.

Materials and Methods

Subjects
Seventeen right-handed subjects (8 females; mean age: 26.5 years; standard deviation: 2.6) with no history of neurological or psychiatric illness participated in the study after giving informed consent. All subjects had normal or corrected-to-normal vision and reported normal hearing. The study was approved by the human research ethics committee of the medical faculty at the University of Tübingen.

Stimuli
The visual stimulus was a random dot kinematogram of expanding radial motion at 100% coherence presented with a fixation cross (0.2° visual angle) on a black background within a rectangular window of 16° × 22° visual angle (modified from the Cogent Graphics sample library; http://www.vislab.ucl.ac.uk/). The dot density decreased exponentially with eccentricity, uniformly in all directions. On each frame (60 Hz refresh rate), each dot was displaced outward at a small visual angle with the dot velocity increasing exponentially with eccentricity (see: supplementary video). In addition, the dot velocities were amplitude modulated at 0.1 Hz within a range of 2-8°/second. Figure 1A shows the velocity profile of the 150 dots in the display as a function of time at a given eccentricity (Fig. 1C). Stimulus blocks and events were pseudorandomly presented in 6 experimental sessions (~10 min) with each block and event type (i.e., V10s, A10s, AV10s, V20s, A20s, AV20s, V30s, A30s, AV30s) being presented at least once within a single session. The order of the scanning sessions was counterbalanced across subjects.

To control subjects' attentional level to a certain degree, they were instructed to fixate and respond as accurately and fast as possible to targets. Targets (300-ms duration) were simple visual (i.e., color change of the fixation cross), auditory (i.e., complex tone), or audiovisual (i.e., color change of fixation cross and complex tone) events that occurred after 2.5, 5, or 7.5 s in visual, auditory, or audiovisual blocks of 10-s durations. Approximately 20% of all blocks contained a target. Target blocks were pseudorandomly inserted and were modeled separately as effects of no interest in the general linear model.

Experimental Set Up

Visual and auditory stimuli were presented using Cogent (John Romaya, Vision Lab, University College London, UK; http://www.vislab.ucl.ac.uk/), running under Matlab 7.0 (MathWorks Inc.) on a Windows (Microsoft Inc.) PC. Visual stimuli were back projected onto a Plexiglas screen using a LCD projector (JVC Ltd.) visible to the subject through a mirror mounted on the magnetic resonance (MR) head coil. Auditory stimuli were presented monophonically using MR-compatible headphones (MR Confon GmbH). Subjects performed a behavioral task using a MR-compatible custom-built button device connected to the stimulus computer.

MRI
A 3-T Siemens Magnetom TrioTim System (Siemens) was used to acquire both T1-weighted anatomical images (176 sagittal slices, time repetition [TR] = 1900 ms, time echo [TE] = 2.26 ms, time to inversion = 900 ms, flip angle = 90°, field of view [FOV] = 256 × 224 mm, image matrix = 256 × 224, voxel size = 1 × 1 × 1 mm³) and T2T-weighted axial echo-planar images with BOLD contrast (gradient-echo echo-planar imaging, Cartesian k-space sampling, TR = 3080 ms, TE = 40 ms, flip angle = 90°, FOV = 192 × 192 mm, image matrix = 64 × 64, 38 slices acquired sequentially in ascending direction, 3.0 × 3.0 × 2.6 mm³ voxels, interslice gap = 0.4 mm). To increase the effective sampling rate, the stimulus onset asynchronies were not simple multiples of the TR, jittering event onset relative to scan onset (Josephs et al. 1997; Miezien et al. 2000). There were 6 sessions with a total of 206 volume images per session. The first 3 volumes were discarded to allow for T1 equilibration effects. Subjects were allowed a 15-min break before the fourth session outside of the scanner. The high-resolution anatomical image volume was acquired at the end of the experiment.

Data Analysis
The data were analyzed with statistical parametric mapping (using SPM5 software; the Wellcome Department of Imaging Neuroscience, London; http://www.fil.ion.ucl.ac.uk/spm) (Friston et al. 1995). Scans from each subject were realigned using the first as a reference, spatially normalized into Montreal Neurological Institute (MNI) standard space (Evans et al. 1992), resampled to 3 × 3 × 3 mm³ voxels and spatially smoothed with a Gaussian kernel of 8-mm full-width at half-maximum. The time series of all voxels were high-pass filtered to 1/128 Hz, and a correction for temporal autocorrelation (AR1) was applied. The visual, auditory, and audiovisual brief 1-s events of the fMRI experiment were modeled as an event-related unit impulse (i.e., delta function) logged to the event onset. The 10 s, 20 s, and 30 s visual, auditory, and audiovisual blocks were modeled as a linear combination of 5 response components: 1) onset, 2) offset, 3) offset modulation, 4) sustained, and 5) sinusoidal modulation (Fig. 2). The onset (ONSET) and offset (OFFSET) responses of the activation blocks were modeled by event-related unit impulses logged to block onset and block offset. The offset impulses were parametrically modulated by the duration of the preceding block (=offset modulation, OFFMOD). The sustained (SUSTAINED) response was modeled as a series of 20 (10 s block), 40 (20 s block), and 60 (30 s block) delta functions equally spaced.

Downloaded from https://academic.oup.com/cercor/article-abstract/21/4/920/289665 by guest on 31 March 2019
between block onset and block offset. The impulse series of the sustained response were parametrically modulated at 0.1 Hz according to the sinusoidal auditory amplitude and visual velocity modulation (=sinusoidal intensity modulation [SIM]).

The motivation for using these 5 response components is as follows (see also Introduction):
1. To dissociate multisensory integration processes that emerge for transient responses due to rapid stimulus changes and for sustained responses due to the continuous audiovisual stimulation, we modeled ONSET/OFFSET and SUSTAINED responses separately.
2. To investigate whether multisensory integration processes differ for stimulus ONSETs that necessitate rapid orienting responses and OFFSETs signaling that attention can be disengaged, these 2 components were modeled separately.
3. In addition to the rapid transitions at stimulus ONSET and OFFSET, the sinusoidal intensity variation provides slow temporal cues that may facilitate audiovisual binding. The role of these slow temporal cues is evaluated by modeling the sinusoidal modulation of auditory and visual stimulus intensity.
4. In the unisensory domain, it is well known that the OFFSET responses depend on the length of the preceding stimulation period possibly as a result of perceptual learning/adaptation (Harms and Melcher 2002). To accommodate these effects, our model parametrically modulated the magnitude of the OFFSET response by the length of the preceding stimulation period.

The regressors encoding the visual, auditory, and audiovisual 1 s events and the 10 s, 20 s, and 30 s blocks (with 5 regressors) were entered into the design matrix after convolving each event-related unit impulse with a canonical hemodynamic response function. In addition, the statistical model included the 3 different (visual, auditory, audiovisual) target types and their corresponding target blocks as regressors of no interest (i.e., blocks containing targets were modeled separately). Realignment parameters were included as nuisance covariates to account for residual motion artifacts.
Figure 2. The BOLD response for the 10 s, 20 s, and 30 s stimuli (horizontal lines indicate the length of the stimulus blocks) was modeled as a linear combination of 5 response components: onset (ONSET) and offset (OFFSET) transients, modulation of the offset response by length of the preceding stimulation block (OFFMOD), sustained response (SUSTAINED), and 0.1 Hz sinusoidal modulation of the sustained response (SIM). The 5 predictors are shown after convolution with the canonical hemodynamic response function.

Regressors specific effects for each subject were estimated according to the general linear model and passed to a second-level analysis as contrasts. This involved creating contrast images of each effect relative to fixation baseline (i.e., brief 1 s stimuli, onset, offset, and sustained response components separately for the visual, auditory, and audiovisual conditions, as averaged over sessions) at the first level. Furthermore, superadditive (i.e., \((AV-rest) - (V-rest) + (A-rest))\) and subadditive (i.e., \((V-rest) + (A-rest) - (AV-rest))\) interaction contrasts (Calvert 2001; Calvert and Lewis 2004) were formed separately for the brief 1 s events and the onset, offset, and sustained response components. These contrast images were entered into independent second-level one-sample \(t\)-tests. Inferences were made at the second level to allow for a random effects analysis and inferences at the population level (Friston et al. 1999).

Each effect was tested in 2 search volumes. The first search volume included all voxels, that is, the whole brain (54,005 voxels). In order to increase the sensitivity of the analysis with respect to the superior temporal sulcus (STS) as an a priori candidate region for audiovisual object recognition (Calvert 2001; Amedi et al. 2005; Werner and Noppeney 2009), the second search volume for the random effects analysis was limited to the subset of 3884 voxels (at a spatial resolution of 3 x 3 x 3 mm\(^3\)) that were located within the middle temporal gyri (MTG) and superior temporal gyri (STG) bilaterally within MNI space as defined by the Automated Anatomical Labeling library (Tzourio-Mazoyer et al. 2002) using the MarsBar (http://marsbar.sourceforge.net/) toolbox (Brett et al. 2002). Unless otherwise stated, we report activations at \(P < 0.05\), corrected at the cluster level for the search volume (i.e., the entire brain or the STS as indicated in the results tables) using an auxiliary uncorrected height threshold of \(P < 0.001\) (Friston et al. 1994).

Results of the random effects analysis were superimposed onto the SPM-rendered standard brain and a subjects' T\(_1\)-weighted normalized brain, using the MRIcron software (http://www.sph.sc.edu/comd/rorden/mricron.html).

Another In-depth General Linear Model Analyses

Analysis of the Amplitude of Sinusoidal Modulation

The parameter estimate pertaining to the sinusoidal modulation in our primary analysis is influenced by both, the phase and the magnitude of the sinusoidal modulation at 0.1 Hz in regional activity time courses during the visual, auditory, and audiovisual blocks. To dissociate magnitude from phase information, a second analysis modeled the modulation of the sustained response with 2 sinusoidal parametric modulators, with the second modulator \((SIM + 90°)\) being phase shifted by 90° with respect to the first (SIM) one (Fig. S34). This enabled us to compute unambiguously the magnitude of the modulation according to:

\[
M = \sqrt{(\text{SIM} + 90°)^2 + \text{SIM}^2}
\]

Since differences in phase across regions and conditions can arise from numerous reasons (e.g., variation of vasculature and associated hemodynamic response function, neural coding of velocity vs. acceleration), we report only fMRI results pertaining to differences in response magnitude.

Finite Impulse Response Model for Time Course Analysis

The regional time courses were visualized, that is, plotted using a finite impulse response model that modeled 1 s, 10 s, 20 s, and 30 s visual, auditory, and audiovisual stimulus periods and the 3 target conditions by 12 (1 s events), 15 (10 s blocks), 18 (20 s blocks), and 21 (30 s blocks) contiguous boxcar functions of peristimulus time. Each boxcar was set to 3.08 s duration. Each of these time bins (separately for visual, auditory, and audiovisual conditions and for block durations) were used to form a regressor in the design matrix with realignment parameters included as nuisance covariates to account for residual motion artifacts. Regressor-specific effects for each subject were estimated according to the general linear model (Friston et al. 1995) and passed to a second-level analysis as contrast images, that is, the parameter estimate of each time bin and each condition averaged across sessions. These contrast images were entered into second-level analyses of variance (ANOVA). To obtain BOLD signal change (%) over time (across subject’s mean ± standard error of the mean [SEM]), the individual parameter estimates for each time bin were weighted by the scaling of their basis functions separately for 1 s, 10 s, 20 s, and 30 s stimuli (under visual, auditory, and audiovisual stimulation).

Behavioral Data

During the experiment, subjects were engaged in a target detection task to render the experiment comparable with previous passive viewing/listening paradigms while maintaining some degree of attentional control. All subjects (mean ± SEM) achieved ceiling performance (i.e., 100% detection accuracy) for targets from all sensory modalities with visual (522 ± 18 ms), auditory (458 ± 19 ms), and audiovisual (411 ± 17 ms) reaction times (RTs). A one-way repeated measurement ANOVA of the targets’ RTs (with the factor sensory input: auditory, visual, audiovisual) identified a significant main effect of sensory input \((F_{4,425} = 58.3; P < 0.001)\) after Greenhouse-Geisser correction. Post hoc comparisons (Bonferroni corrected, \(P < 0.05\)) revealed significant multisensory RT facilitations, that is, audiovisual responses were faster than auditory and visual responses. Furthermore, auditory RTs were significantly shorter than visual RTs. These results confirm that subjects attended to visual, auditory, or both sensory inputs during the experiment.

Neuroimaging Data

To orient the reader, we first briefly characterize the neural systems underlying the ONSET, OFFSET, and SUSTAINED
response components (of the 10 s, 20 s, and 30 s blocks) and the 1 s events separately for the visual and auditory conditions relative to fixation (for details, see Supplementary materials). Second, we report superadditive (AV > V + A) and subadditive (V + A > AV) interaction contrasts separately for the ONSET, OFFSET, and SUSTAINED components and the brief 1 s events as indices for multisensory integration (corrected for multiple comparisons within the entire brain or the STS (STG/MTG) search volume). The response profile in each significant region is then characterized (at an uncorrected statistical threshold) according to its response (relative to baseline) separately for the ONSET, OFFSET, SUSTAINED, SIM, and OFFMOD components. These additional statistical comparisons serve only as in-depth data characterization and are not presented as valid statistical comparisons in their own right (note: they are not independent from the interaction contrasts).

The Visual and Auditory Neural Systems for Onset, Offset, and Sustained Responses and Brief 1 s Events

In the visual system, onset, offset, and sustained responses were found in mostly nonoverlapping areas with onset and sustained responses located primarily at lower and offset responses at higher levels of the sensory processing hierarchy. The extensive neural systems for the brief 1 s visual events encompassed both onset and offset visual systems (Fig. 3; see also: Supplementary results, Tables S1–S3, Fig. S1).

In the auditory system, onset, offset, and sustained responses were found in partly nonoverlapping areas with sustained responses located primarily at lower and offset responses at higher hierarchical levels, while onset responses existed at both lower and higher hierarchical levels of auditory processing. Finally, in the visual system, the neural systems for the brief 1 s auditory events included the onset and offset auditory systems (Fig. 3; see also: Supplementary results, Tables S4 and S5).

These results reveal commonalities in the functional organization of the visual and auditory systems. Onset, offset, and sustained responses were anatomically segregated to a certain degree. While offset responses were limited to higher order association areas, sustained and onset responses extended into low level and even primary sensory areas. As a consequence, a conjunction analysis across sensory modalities revealed a common system only for the offset responses and the brief 1 s events in the posterior and middle STS (Fig. 6C,D; see also: Supplementary Table S6). In contrast, onset or sustained systems did not generalize across sensory modalities. Both in the visual and auditory domains, the neural systems identified for the brief 1 s events encompassed onset and offset systems. This demonstrates that a conventional event-related design cannot further dissociate onset and offset systems based on their functional and temporal properties.

Audiovisual Interactions for Onset Transients

At stimulus onset, superadditive interactions (i.e., an audiovisual response that is greater than the sum of 2 unisensory responses) were detected predominantly in visual and auditory onset systems, including the bilateral calcarine sulcus (CaS), fusiform gyrus (FFG), HG, and STG (Fig. 4A,B and Table 1). Interestingly, in the visual cortex, audiovisual interactions in the onset and offset system were observed only in regions representing the peripheral visual field (see: Supplementary Fig. S2) (Wandell et al. 2007) that have previously been shown to receive direct anatomical projections from the core and parabelt areas of the auditory cortex (Falchier et al. 2002).

In these regions, the superadditive effects were mediated by 2 complementary mechanisms: A mutual unisensory inhibition and a multisensory coexcitation. First, signals of the nonpreferred sensory modalities induced deactivations in sensory cortices. For instance, auditory signals deactivated visual cortex, while visual signals led to deactivations in auditory areas. Second, when presented concurrently, signals from the nonpreferred sensory modality amplified the response to inputs from the preferred sensory modality generating significant multisensory enhancements relative to the maximal unisensory responses in CaS and FFG and nonsignificant enhancements in HG and STG (Fig. 4C).

For completeness, no subadditive interactions were observed for onset responses.

Audiovisual Interactions for Offset Transients

No superadditive interactions were detected for offset responses (even at $P < 0.05$ uncorrected). However, subadditive offset responses were observed in the visual offset system, including the left anterior intraparietal sulcus (aIPS), the right posterior MTG overlapping with hMT+/V5+, and the right posterior superior temporal sulcus (pSTS) (Fig. 5A,B and Table 1). Yet, the specific

![Figure 3](https://example.com/figure3.jpg)

**Figure 3.** Activations for (A) transient response components, that is, onset (orange), offset (cyan), overlap onset/offset (green), (C) sustained response components, and (B) the brief (1s) stimuli for visual (V, top), auditory (A, middle), and audiovisual (AV, bottom) stimulations are rendered on template brains. Height threshold: $P < 0.001$, uncorrected. Extent threshold: >40 voxels.

924 Transient and Sustained Responses in Audiovisual Integration  ·  Werner and Noppeney
interaction patterns differed across these regions. While the offset responses for audiovisual signals were comparable with unisensory inputs in posterior STS, they were suppressed relative to unisensory stimuli in anterior IPS. Furthermore, the 3 regions differed in sensory selectivity as expressed in the onset and sustained responses. Surprisingly, the sensory selectivity even within a single region differed for onset and sustained responses suggesting that higher visual and association areas may contain multiple neuronal populations that differ in temporal and sensory preference. For instance, the anterior IPS showed auditory onset but visual sustained responses. In contrast, posterior STS showed auditory selectivity for onset responses but a rather amodal response profile for sustained and offset responses. Furthermore, the offset responses in STS linearly increased with preceding block length particularly for audiovisual stimulation suggesting that posterior STS adapts (or fine-tunes) more strongly to the slow temporal features of the combined auditory and visual intensity profiles leading to a pronounced rebound from adaptation at stimulus offset (Fig. 5C).

Audiovisual Interactions for Sustained Responses
No significant audiovisual interactions (i.e., neither subadditive nor superadditive) were observed for sustained brain activity. This absence of nonlinearities for the sustained response...
component is probably a result of sensory selectivity. The sustained response component was strictly sensory selective in the majority of the areas, that is, it was elicited for only one sensory input, with the response to bisensory stimulation then being statistically indistinguishable from the response to the preferred unisensory stimulation. Hence, the audiovisual response was a simple sum of the auditory and visual responses.

**Audiovisual Effects for the SIM**

The SIM of the sustained response was significantly stronger for audiovisual than unisensory inputs suggesting that both auditory and visual signals influence the response dynamics in posterior STS (Fig. 5C). More specifically, the parameter estimate for the SIM was more negative in the audiovisual condition. This indicates a stronger influence of audiovisual signals on the response dynamics compared to the unisensory conditions. The difference in the SIM parameter estimates across conditions may reflect a change in magnitude or a phase shift of the response to the modulated stimulus.

### Table 1

Audiovisual interaction effects for onsets, offsets, and 1 s stimuli

<table>
<thead>
<tr>
<th>Regions</th>
<th>MNI coordinates</th>
<th>z-score (peak)</th>
<th>P value (cluster)</th>
<th>V_xext</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. superior temporal gyrus (middle)</td>
<td>-48 -12 -3</td>
<td>4.80</td>
<td>0.039</td>
<td>46</td>
</tr>
<tr>
<td>L. insula (posterior)</td>
<td>-39 -18 3</td>
<td>3.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. Heschl's gyrus</td>
<td>-48 -21 6</td>
<td>3.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. Heschl's gyrus</td>
<td>39 -27 15</td>
<td>4.13</td>
<td>0.000</td>
<td>139</td>
</tr>
<tr>
<td>R. parietal operculum</td>
<td>60 -15 12</td>
<td>3.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. insula</td>
<td>45 0 -6</td>
<td>3.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. Heschl's gyrus</td>
<td>48 -12 3</td>
<td>3.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. Heschl's gyrus</td>
<td>51 -18 3</td>
<td>3.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. cuneus</td>
<td>-15 -81 21</td>
<td>4.16</td>
<td>0.000</td>
<td>503</td>
</tr>
<tr>
<td>L. lingual gyrus</td>
<td>-12 -87 30</td>
<td>3.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. cuneus</td>
<td>15 -87 36</td>
<td>3.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. cuneus</td>
<td>3 -81 18</td>
<td>3.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. fusiform gyrus</td>
<td>-33 -60 -12</td>
<td>3.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. calcarine sulcus</td>
<td>-12 -81 0</td>
<td>3.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. fusiform gyrus</td>
<td>30 -54 -15</td>
<td>3.92</td>
<td>0.036</td>
<td>47</td>
</tr>
<tr>
<td>R. lingual gyrus</td>
<td>21 -60 -3</td>
<td>3.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subadditive audiovisual interactions: onset</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. intraparietal sulcus (anterior)</td>
<td>-36 -51 45</td>
<td>4.23</td>
<td>0.020</td>
<td>51</td>
</tr>
<tr>
<td>R. human middle temporal region (hMT+)</td>
<td>-42 -45 39</td>
<td>3.64</td>
<td>0.021*</td>
<td>25</td>
</tr>
<tr>
<td>R. superior temporal sulcus (posterior)</td>
<td>57 -69 6</td>
<td>3.60</td>
<td>0.029*</td>
<td>22</td>
</tr>
<tr>
<td>R. superior temporal sulcus</td>
<td>60 -63 3</td>
<td>3.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. human middle temporal region</td>
<td>66 -51 12</td>
<td>3.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. superior temporal sulcus (posterior)</td>
<td>60 -45 15</td>
<td>3.41</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Discussion

The present study investigated the role of onset, offset, and sustained responses in multisensory integration. First, we demonstrate that audiovisual interactions emerge primarily for transient (rather than sustained) responses at stimulus onset and offset highlighting the importance of rapid stimulus transitions for multisensory integration. Indeed, rapid transitions in environmental states provide important synchrony cues informing the brain about which sensory signals are generated by the same event and should hence be integrated (Stein and Meredith 1993). Second, audiovisual interactions for onset and offset transients were doubly dissociable, both at the functional and anatomical level. For onset responses, they were superadditive and located in low-level sensory areas possibly mediating an increase in stimulus salience. For offset responses, they were subadditive and located in higher order sensory or classical multisensory integration areas possibly reflecting the

---

Note: P value, corrected at cluster level for the whole brain (54,005 voxels). V_xext, voxel extent = number of voxels in cluster.

*the STS (STG, MTG) search volume (3884 voxels).
formation of higher order representations. Thus, localization of the audiovisual interactions of the onset and offset responses converged with the more basic anatomical segregation into a "low-level" onset and "higher level" offset neural systems.

**Integration in Sensory-Specific Regions**

In line with accumulating evidence for multisensory interactions in sensory-specific brain areas (Schroeder and Foxe 2002; van Atteveldt et al. 2004; Ghazanfar et al. 2005; Lehmann et al. 2006; Watkins et al. 2006; Bonath et al. 2007; Martuzzi et al. 2007; Meienbrock et al. 2007; Werner and Noppeney 2010), superadditive audiovisual interactions at stimulus onset occurred selectively in primary and low-level visual and auditory areas. In all of those areas, unisensory preferred and non-preferred inputs elicited mirror reversed response profiles for onset and offset responses (Laurienti et al. 2002). For instance, in primary auditory areas, auditory input elicited activation at stimulus onset and deactivation at stimulus offset, while visual input evoked deactivation at stimulus onset and activation at stimulus offset (and vice versa for primary visual areas). During unisensory stimulation, activations in primary auditory and visual areas were thus characterized by a seesaw relationship. Importantly, while inputs from the nonpreferred sensory modality induced a deactivation when presented alone, they amplified the response to a concurrently presented input from the preferred sensory modality leading to superadditive interactions. In other words, competitive or inhibitory interactions between sensory cortices for unisensory stimulation turned into cooperative or excitatory interactions for multisensory stimulation. These superadditive interactions were elicited only by the rapid transitions at stimulus onset suggesting that they crucially depend on the temporal co-occurrence of multiple sensory inputs. Neurophysiological studies in nonhuman primates have shown that multisensory integration processes in primary sensory cortices are governed by tight temporal constraints possibly via mechanisms of phase resetting (Lakatos et al. 2007; Kayser et al. 2008). Furthermore, functional imaging studies in humans have also suggested their
Transient and Sustained Responses in Audiovisual Integration

Auditory inputs may withdraw attentional resources from visual processing (and vice versa) but boost temporally co-occurring visual (or auditory) stimuli. Even though we cannot fully exclude potential attentional top-down effects (Somers et al. 1999; Petkov et al. 2004; Busse et al. 2005; Fairhall and Macaluso 2009), previous studies demonstrated superadditive audiovisual interactions in primary auditory cortex irrespective of task context (Werner and Noppeney 2010) or an individual’s cognitive state (Kayser et al. 2008). Furthermore, superadditive interactions were observed at stimulus onset in low-level sensory regions in the absence of interactions in higher order association areas. This is in line with previous electroencephalography and intracranial recordings in humans showing audiovisual integration effects in low-level sensory areas before activity in higher order association areas and hence feedback influences may emerge (Molholm et al. 2004; Mishra et al. 2007; Besle et al. 2008). Collectively, these results point to audiovisual integration mechanisms that might (at least in part) be mediated via direct connections between sensory regions or feedforward thalamocortical projections (Schroeder et al. 2003) rather than feedback from higher order areas (Driver and Noesselt 2008; Musacchia and Schroeder 2009).

Integration in Higher Order Association Areas

Subadditive interactions at stimulus offset were observed in posterior MTG (possibly overlapping with hMT+/V5+), intraparietal sulcus (IPS), and STS that have previously been implicated in multisensory integration (Calvert 2001; Macaluso et al. 2003; Wright et al. 2003; Beauchamp et al. 2004; Barraclough et al. 2005; Saito et al. 2005; Schlack et al. 2005; Hein et al. 2007; Noesselt et al. 2007; Ghazanfar et al. 2008; Noppeney et al. 2008; Dahl et al. 2009; Sadaghiani et al. 2009; Stevenson and James 2009; van Atteveldt et al. 2009; Werner and Noppeney 2009). The STS and IPS are ideally suited to integrate inputs from multiple sensory modalities based on their extensive bidirectional anatomical connectivity to visual and auditory areas (Neal et al. 1990; Seltzer and Pandya 1994; Falchier et al. 2002; Rockland and Ojima 2003). While these regions were sensory selective in their onset responses, their offset responses were observed for auditory and visual signals leading to subadditive interactions. More specifically, the IPS showed suppressive interactions (i.e., the bisensory response is smaller than both unisensory responses), that is, a response profile that cannot be explained by response nonlinearities of the BOLD response but suggests audiovisual interactions at the neuronal level. In contrast, the STS showed an amodal response profile with comparable responses to auditory, visual, and audiovisual stimulation (though we acknowledge the saturation of the BOLD response as an alternative explanation). Furthermore, the STS sustained responses exhibited a stronger modulation at 0.1 Hz for audiovisual than unisensory stimulations suggesting that slow continuous signals from both sensory modalities drive the temporal dynamics in STS. This complex response profile for onset, offset, and sustained components suggests that higher order association regions may contain multiple neuronal populations differing in terms of temporal selectivity and sensory preference. From the perspective of predictive coding, the “offset” neuronal populations may form higher order amodal representations (Werner and Noppeney 2009, 2010) abstracted from the sensory inputs with the amodal rebound of activity in STS reflecting a prediction error signal induced by the abrupt change in stimulus structure.
at block offset (Rao and Ballard 1999; Garrido et al. 2009; Friston 2010; Noppeney et al. 2010). In line with this predictive coding (or alternative adaptation) framework, the offset response has been shown to depend indeed on the temporal structure (e.g., rate, regularity, and length) (Harms and Melcher 2002; Herdener et al. 2007) or block length (this study) of the preceding stimulation period. However, clearly future studies are needed to support this hypothesis.

**Methodological Considerations: Limitations of Conventional Event-Related or Block Designs for Identification of Multisensory Interactions**

Importantly, our experimental design also included brief 1 s stimuli equivalent to a conventional event-related design. Not surprisingly, the 1 s stimuli were not able to reveal the complex temporal profile of audiovisual interactions; yet, more importantly, they even failed to reveal audiovisual integration processes in low-level sensory areas and identified only the subadditive interactions in the posterior STS/STG. The parameter estimate plots showed that this "null finding" was not just a result of less efficient estimation but rather due to the inherent temporal smoothing of the BOLD response that caused the opposing effects at stimulus onset and offset to be canceled (e.g., in CaS and HG). From a methodological perspective, these findings suggest that superadditive audiovisual interactions in low-level sensory areas at stimulus onset may be missed in conventional event-related or block designs. They may only emerge in more complex experimental designs that enable the dissociation of onset and offset responses.

**Conclusions**

In conclusion, multisensory integration emerges at multiple levels of the cortical hierarchy, including primary and higher order association areas. Low-level sensory areas integrate audiovisual inputs at stimulus onset in a superadditive fashion to enhance stimulus salience and enable an initial scene analysis. In contrast, higher order association areas show subadditive interactions at stimulus offset possibly reflecting the formation of higher order representations. This functional and anatomical segregation suggests that super- and subadditive integration profiles may serve different purposes in multisensory integration. The predominance of audiovisual interactions for transient response components highlights the importance of rapid transitions in environmental states for multisensory integration. From a methodological perspective, our results highlight the limitations of conventional event related and block designs that may fail to detect superadditive interactions at stimulus onset because of the inherent smoothing of the BOLD response that precludes the dissociation of onset and offset responses.

**Funding**

Max Planck Society.

**Supplementary Material**

Supplementary material can be found at: http://www.ercor.oxfordjournals.org/

**Notes**

We thank Kamil Uludag, Marcus Herdener, and Massimiliano Di Luca for very helpful advice and useful discussions. Conflict of Interest: None declared.

**References**


Transient and Sustained Responses in Audiovisual Integration


