Brain Responses to Auditory and Visual Stimulus Offset: Shared Representations of Temporal Edges

Marcus Herdener,^{1,2,*†} Christoph Lehmann,^{1†} Fabrizio Esposito,³
 Francesco di Salle,⁴ Andrea Federspiel,¹ Dominik R. Bach,¹
 Klaus Scheffler,⁵ and Erich Seifritz¹

¹University Hospital of Psychiatry Bern, Bern, Switzerland

²Forensic-psychiatric Service, University of Bern, Bern, Switzerland

³Department of Neurological Sciences, University of Naples "Federico II", Naples, Italy

⁴Department of Neuroscience, University of Pisa, Pisa, Italy

⁵Department of Medical Radiology, MR-Physics, University of Basel, Basel, Switzerland

Abstract: Edges are crucial for the formation of coherent objects from sequential sensory inputs within a single modality. Moreover, temporally coincident boundaries of perceptual objects across different sensory modalities facilitate crossmodal integration. Here, we used functional magnetic resonance imaging in order to examine the neural basis of temporal edge detection across modalities. Onsets of sensory inputs are not only related to the detection of an edge but also to the processing of novel sensory inputs. Thus, we used transitions from input to rest (offsets) as convenient stimuli for studying the neural underpinnings of visual and acoustic edge detection per se. We found, besides modality-specific patterns, shared visual and auditory offset-related activity in the superior temporal sulcus and insula of the right hemisphere. Our data suggest that right hemispheric regions known to be involved in multisensory processing are crucial for detection of edges in the temporal domain across both visual and auditory modalities. This operation is likely to facilitate cross-modal object feature binding based on temporal coincidence. *Hum Brain Mapp* 30:725–733, 2009. © 2008 Wiley-Liss, Inc.

Key words: BOLD; functional imaging; crossmodal integration; edge detection; object formation

INTRODUCTION

When dealing with natural scenes, sensory systems have to process an often tremendous flow of information. Early

processing of sensory inputs produces elements that require grouping, which follows principles described by Gestalt psychologists [Wertheimer, 1938]. Grouping, e.g. by similarity or temporal proximity, subserves figure-ground segregation and has been suggested for object formation in the visual and auditory domain [Kubovy and Van Valkenburg, 2001]. Edges, or object boundaries, are particularly important for figure-ground segregation that implies that some parts of the environment are perceived to go together whereas others do not. Kubovy and Van Valkenburg [2001] propose that in the visual domain space and time are indispensable attributes for object formation, while frequency and time are crucial for auditory object formation. The importance of edges for vision is well-known since the seminal work of Hubel and Wiesel [1962] and the implications of edges for the segregation of auditory events have been described more recently [Fishbach et al., 2001, 2003].

[†]These two authors contributed equally to the present study.

Contract grant sponsor: SNF grant; Contract grant number: PP00B-103012.

*Correspondence to: Marcus Herdener, M.D., University Hospital of Psychiatry Bern, Bolligenstr. 111, 3000 Bern 60, Switzerland. E-mail: herdener@puk.unibe.ch

Received for publication 10 August 2007; Revised 16 December 2007; Accepted 18 December 2007

DOI: 10.1002/hbm.20539

Published online 11 February 2008 in Wiley InterScience (www.interscience.wiley.com).

However, object perception is generally not carried out within a single modality [Gibson, 1966, 1979, 1982; Stein and Meredith, 1993]. The possession of multiple sensory systems and integration of sensory information across modalities provides substantial advantages. It enhances the ability to detect, locate, and discriminate external stimuli [for review see Stein and Meredith, 1993]. For example, reaction times to congruent inputs from more than one modality are significantly shorter than those to unimodal stimuli [Frens et al., 1995; Hughes et al., 1994; Molholm et al., 2002]. Such crossmodal advantages depend on the detection of some point of commonality between the different sensory inputs. Two major features, proximity in space and time, seem to determine whether different modal cues will be perceived as arising from the same event or object, and will be integrated for perceptual gain [Radeau, 1994].

As time is an indispensable attribute for objects in both vision and audition, we hypothesized that common neural underpinnings form the basis for the detection of temporal edges across modalities. We considered on and offsets of sensory inputs to be particularly suitable to study the neural underpinnings of edge detection as they contrast a specific sensory input against no physical input. Onset responses to auditory and visual inputs have been studied extensively [Harms and Melcher, 2002; Konishi et al., 2001; Pantev et al., 1991; Seifritz et al., 2002]. However, transient responses to stimulus onset include both temporal edge-related activation and activity related to new afferent input processing. In contrast, activity elicited by cessation of a stimulus, which is not confounded by additional sensory inputs, reflects edge detection per se [Herdener et al., 2007].

Therefore, we used functional magnetic resonance imaging (fMRI) and examined offset responses to auditory and visual stimulus patterns to test for shared neural substrates of temporal acoustic and visual edge detection, which might be involved in the integration of object features across modalities.

MATERIALS AND METHODS

Subjects

Seven healthy volunteers (all males, age range 28–37 years) were recruited from an academic environment. None of the subjects reported a history of major medical, neurological, or psychiatric disorders or of psychotropic medication. All subjects gave written informed consent before participation.

Stimulation

Auditory stimulation

Pure sine tones with 1 kHz frequency and constant intensity of 95 dB sound pressure level were amplitude-

modulated with a square wave envelope (smoothed using exponential onset and offset ramps of 10 ms duration) and four different sound pulse repetition rates (0.5, 1, 2, and 4 Hz) with a constant sound pulse length of 100 ms and four resulting interpulse intervals (1.85, 0.87, 0.39, 0.14 s). Each single acoustic stimulation block lasted 58.4 s (corresponding to 30 fMRI volumes). With the four repetition rates used for amplitude modulation, each stimulation block either consisted of 30 (0.5 Hz), 60 (1 Hz), 120 (2 Hz), or 240 (4 Hz) sound pulses of length 100 ms. The noise reduction by the headphones (Commander XG, Resonance Technology, Northridge, CA) of approximately 30 dB and the spectral composition of the continuous scanner noise [Seifritz et al., 2006] enabled a clear perception of all experimental stimuli.

Visual stimulation

A red light-emitting-diode (LED) grid (10 × 10 LEDs) has been driven by a signal described by the same input parameters as those used for acoustic stimulation (light pulse duration, 100 ms; four repetition rates, 0.5, 1, 2, and 4 Hz). The LED grid was fixated at the head-coil of the scanner directly in front of the face of the subjects.

None of the subjects reported pronounced unpleasantness related to either auditory or visual stimulus presentation in interviews conducted after acceptance trials of sensory stimulation before each scanning session and after the experiment.

The experiment was split up into two sessions carried out on different days. In the first session, subjects were exposed to auditory stimulation only, whereas in the second session visual stimuli were presented. Each session consisted of four consecutive fMRI runs of 375 volumes each. Each run contained two repetitions of each of the four experimental stimulation types (i.e. two stimulation blocks of each of the four different pulserates per run), randomly intermixed and separated by nine (one at the beginning, one at the end) silent intervals of 29.2 s (Fig. 1A).

The duration of each of the two experimental sessions was 48 min and 40 s (that is four times 375 volumes per run). Subjects were instructed to concentrate on the auditory signals and to fixate the LED grid, respectively. They were not asked to carry out any output tasks.

Image Acquisition

The fMRI data were acquired on a 1.5-T standard clinical MRI scanner (Sonata, Siemens, Erlangen, Germany) equipped with a circularly polarized radio frequency head coil. The subjects' head was stabilized with foam pads to minimize movement during the experiment. A T1-weighted high-resolution data set covering the whole brain was collected for each subject with a three-dimensional magnetization-prepared rapid acquisition gradient echo sequence with 1.2 (1 × 1 × 1.2) mm³ cubic voxels. The functional volumes were positioned parallel to the lateral

Figure 1.

Schematic illustration of stimulation, data acquisition and BOLD-signal modeling. **(A)** Sound and light pulses were presented at four randomly intermixed repetition rates in stimulation blocks of 58.4 s duration (corresponding to 30 functional volumes), followed by 29.2 s without sensory inputs (corresponding to 15 functional volumes; for details see text). For data acquisition we used continuous-sound fMRI. By emitting continuous rather than pulsed gradient sound (see Fig. 1 in [Seifritz et al., 2006] for detailed illustration of sound envelopes), this novel image acquisition

method (i) enhances the signal to noise ratio in auditory fMRI and (ii) avoids interferences between temporal patterns of experimental stimuli and those related to (conventional) gradient sounds. **(B)** Shown are the modeled BOLD response components to acoustic and visual stimulation blocks (onset-, sustained-, and offset-response) and the linear summation of these three components (full model), which entered the general model analyses (for details see text in Materials and Methods section).

sulcus and consisted of a gradient-recalled echo-planar imaging sequence with 16 image slices having a thickness of 5 mm and a volume repetition time of 1.947 s (field of view, 180^2 mm²; matrix, 64^2 pixels (resulting voxel-size: 2.8125 mm $\times$ 2.8125 mm $\times$ 5 mm); echo time, 61 ms; flip angle, 90°; bandwidth, 1,280 Hz/pixel; slice acquisition time, 122 ms).

To maximize the signal-to-noise ratio in auditory fMRI with preserved time resolution, we used continuous-sound fMRI [Seifritz et al., 2006]. This image acquisition method emits continuous rather than pulsed gradient sound. On comparing the conventional echo-planar imaging sequences, the sequence reduces resting state blood oxygen level-dependent (BOLD) baseline in auditory cortex and enhances BOLD amplitude during simple sound stimuli. Specifically, the use of continuous sound fMRI minimized the presumed saturation in the auditory core region related to stimulation by the scanner's gradient sound bursts. In the visual domain, no differential BOLD responses to light stimuli in the visual cortex have been

observed comparing continuous-sound and conventional fMRI (see also Fig. 4 in [Seifritz et al., 2006]).

Image Processing and Statistical Analysis

The first two volumes in each run were discarded to allow for T1 equilibration effects. Image time-series analysis was performed using BrainVoyager QX 1.7 (Brain Innovation, Maastricht, The Netherlands). The time-series were corrected for slice acquisition time through sinc interpolation, realigned with their corresponding T1 volumes, warped into standard space [Talairach and Tournoux, 1988], resampled into 3 mm isotropic voxels, motion-corrected using Levenberg-Marquarts's least square fit for six spatial parameters, highpass-filtered for removal of low-frequency drifts, corrected voxelwise for linear drifts, and spatially smoothed using a 6-mm full-width at half-maximum Gaussian kernel. Prior to multisubject statistical analyses, the image time-series were normalized to have a mean signal value of 100, to reduce intersession- and

Figure 2.

Heteromodal regions involved in temporal edge detection. A conjunction of visually (yellow) and auditory (blue) evoked offset responses reveals heteromodal zones (green), which are sensitive to temporal edges from various modalities. These right-hemispheric heteromodal areas (STS, Talairach coordinates $x/y/z$,

52/−44/17; insula, Talairach coordinates, $x/y/z$ mm, 36/20/6; $q < 0.001$ FDR) are located between zones, which show a more modality-specific sensitivity for temporal edge detection. Activation maps are projected on the averaged anatomy of the individual subjects.

intersubject-variance caused by the variability in the receiver sensitivity. Through this normalization the unit of individual values fluctuating around that mean is considered percent signal change.

In a first step, we defined the cortical areas showing both visual and auditory significant offset response evoked by the end of the preceding stimulation as regions-of-interest (ROI) for further investigation. This was done by applying fixed effects general linear model analyses [Friston et al., 1995] with condition-specific stimulus boxcar functions, convolved with a γ -kernel [Boynton et al., 1996] to model the hemodynamic response behavior. The following predictors entered the design matrix: two modalities (acoustic, visual stimulation), a set of four repetition rates (0.5, 1, 2, and 4 Hz), and three BOLD response components (onset, sustained, offset). This resulted in a total of 24 different predictors ($2 \times 4 \times 3$). The boxcar model of onset and offset responses (subsequently convolved with the gamma kernel), had a length of two volumes, the onset starting with the first volume of stimulation and the offset starting with the first volume after stimulation offset. For the sustained response, a boxcar starting with the second volume and ending with the last volume of stimulation has been defined (Fig. 1B). This way, we modeled the responses to

the on and offsets marking the edges of sensory stimulation blocks as single events, which are independent from sustained responses following ongoing sensory stimulation [Fox et al., 2005; Harms et al., 2005; Harms and Melcher, 2002; Herdener et al., 2007; Seifritz et al., 2002].

For the conjunction of auditory and visual offset regions (offset ROIs), a statistical map for each of the two modalities was created, based on the contrast offset versus sustained response (pooled across the four different pulse rates). The statistical parametric maps were thresholded at $q < 0.001$ false discovery rate (FDR), as described by Genovese et al. [2002]. Conjunction was defined as the intersection of the two statistical maps, thresholded at the specified rate [Nichols et al., 2005]. Only regions showing an offset response in both modalities at the threshold chosen then entered a detailed ROI analysis. This consisted of a set of four one-factorial analysis of variance (ANOVA) of the peak offset response, defined as the averaged BOLD signal value three volumes (5.84 s) after stimulus offset.

RESULTS

First, modality-specific offset-related activity was defined by contrasting BOLD activity elicited at the cessation of

TABLE I. Regions showing offset responses after either visual or auditory stimulation

Brain regions	BA	Side	No of voxels	Talairach coordinates (center of gravity; x, y, z)
Brain regions showing visual offset response				
Temporal lobe, superior temporal gyrus (extending to superior temporal sulcus)	22/13	Right	2239	51, -43, 16
Sublobar, insula/frontal lobe, inferior frontal gyrus	45/13	Right	667	34, 24, 6
Parietal lobe, precuneus	7	Right	3,562	4, -42, 44
Temporal lobe, superior temporal gyrus	39	Right	179	47, -61, 18
Sublobar, claustrum/insula	—	Left	169	-37, -11, -2
Sublobar, claustrum/insula	—	Right	742	37, -3, -6
Brain regions showing auditory offset response				
Temporal lobe, superior temporal gyrus (extending to superior temporal sulcus)	13/40	Right	1,392	53, -44, 20
Sublobar, insula	13	Right	585	36, 17, 6
Frontal lobe, paracentral lobule	31	Right	810	2, -32, 44
Parietal lobe, inferior parietal lobule	40	Right	799	50, -40, 40
Occipital lobe, cuneus	18	Right	2,680	7, -75, 28
Frontal lobe, middle frontal gyrus	9	Left	326	-39, 39, 35
Occipital lobe, precuneus	31	Left	761	-15, -73, 28
Sublobar, lentiform nucleus	—	Left	120	-24, 9, 13,

Areas showing BOLD activity elicited by offsets of either visual or auditory stimulation blocks. Clusters are thresholded at a voxel-level threshold of $q < 0.001$ FDR and a minimal cluster size of 135 voxels. Brain regions and Brodmann Areas (BA) were labeled using the Talairach Daemon software (<http://ric.uthscsa.edu>).

stimulus blocks versus sustained responses in order to disentangle offset-related activity from responses related to processing of afferent inputs. The conjunction of the resulting auditory and visual offset response maps [Nichols et al., 2005] yielded two ROIs, both of which were located in the right hemisphere (see Fig. 2): One in the posterior part of the right superior temporal sulcus (STS, Talairach coordinates $x/y/z$, 52/-44/17 mm; cluster size, 794 mm³), and the other in the right insula (Talairach coordinates, $x/y/z$, 36/20/6 mm; cluster size, 68 mm³). These regions, both responding significantly to visual and auditory stimulus offsets, are interposed between areas showing a more modality-specific response pattern extending to the superior temporal gyrus (auditory offset) and the medial temporal gyrus (visual offset), respectively (Fig. 2; all clusters of unimodal offset detection are reported in Table I). Note that the lateralization of heteromodal brain regions as revealed by the conjunctional analysis is not due to a lateralization in these areas only within one modality, since modality-specific acoustic and visual offset responses in STS and insula both show an asymmetry in favor of the right hemisphere.

The event-related average of the offset BOLD signal time-course (baseline computed as the eight last volumes of the epoch preceding the stimulus offset, that is sustained BOLD activity related to preceding sustained stimulation) is shown for the right STS ROI and the right insula ROI in Figure 3.

Four one factorial ANOVAs of the peak offset response revealed no significant difference for the four sound and light stimulus repetition rates (auditory offset in right STS, $P = 0.724$; visual offset in right STS, $P = 0.219$; auditory

offset in right insula, $P = 0.231$; visual offset in right insula, $P = 0.181$).

In summary, the data demonstrate that temporal edges, represented by the offset of auditory and visual stimulation blocks, elicit BOLD activity in common heteromodal clusters of right hemispheric STS and insula.

DISCUSSION

Our results demonstrate that the posterior part of STS and insula of the right hemisphere are involved in the processing of temporal edges of both acoustic and visual sensory streams presented at various rates. These heteromodal areas demarking the boundaries of putative sensory objects are interposed in between regions showing modality-specific preferences for offset responses (see Fig. 2).

These findings are consistent with recent animal work demonstrating a conglomeration of multisensory neurons in transitional multisensory zones located mainly at the border of modality-specific domains, which have been suggested not only to play roles specific to the two represented modalities but also to be involved in the brain's ability to integrate information from multiple senses [Wallace et al., 2004]. In humans, a patchy organization within multisensory areas like STS has been revealed recently by high-resolution fMRI-studies, suggesting that intervening multisensory patches are located among unimodal clusters [Beauchamp et al., 2004a]. Accordingly, a functional architecture of multisensory regions has been suggested, in which information from various modalities is brought into close anatomical proximity, facilitating the integration in

Figure 3.

Event-related average of BOLD offset responses in multimodal regions after auditory or visual stimulation blocks. The cessation of auditory (left column) and visual (right column) stimulation is represented at timepoint 0, that is, the increase of BOLD responses after stimulus offset is presented relative to a baseline of BOLD signal levels related to preceding sustained stimulation.

Offset responses in areas showing heteromodal properties (right STS; right insula; see also regions labeled in green in Fig. 2) were pooled for the four different stimulus repetition rates (0.5, 1, 2, and 4 Hz) since a set of 4 one factorial ANOVAs did not reveal any significant difference between repetition rates. Error bars represent standard errors of the mean.

the intervening cortex. In addition, future investigations using special imaging techniques at very high spatial resolutions might thus have the potential to further characterize an even more fine-grained patchy organization [see for example Beauchamp et al., 2004a] within the heteromodal zones located between the unimodal areas crucial for offset detection as identified in our study.

Projections to the likely homologue of human STS from auditory and visual association cortices have been identified anatomically in monkeys [Seltzer et al., 1996]. Because of its spatial location between higher-level auditory and visual cortices, STS is ideally situated to link various auditory and visual object features like stimulus on- and offset or other congruencies [Calvert et al., 2001].

Most neuroimaging studies in humans that have demonstrated multisensory properties of STS used complex auditory and visual objects like pictures or sound recordings of

tools or living objects [Beauchamp et al., 2004b] or linguistic material [Calvert et al., 2000; Raji et al., 2000] as stimuli. However, it seems unlikely that STS is specialized only in the processing of these classes of complex objects.

Our data suggest that STS is also crucially involved in recognizing basic object features like temporal edges. We tentatively hypothesize that this information is subsequently used for the integration of temporal coincident sensory inputs from different modalities. Because of the co-responsiveness of the heteromodal cluster in right-hemispheric STS (located between modality-specific regions) to temporal edges at the offset acoustic and visual inputs, and the crucial role of this region for integration of audio-visual inputs [for review see e.g. Amedi et al., 2005], we suggest that the sensitivity of these areas to supramodal temporal characteristics of sensory objects might facilitate crossmodal binding of other sensory features based on

temporal coincidence. However, it should be noted that the design of our study does not provide direct evidence for multisensory integration of auditory and visual inputs since it does not allow to test for crossmodal interaction effects [cf. Calvert and Thesen, 2004; Foxe et al., 2002; Lehmann et al., 2006; Molholm et al., 2002], and that future studies are needed to further evaluate the importance of temporal edge detection within STS for multisensory integration.

The suggested role of STS for detection of common supramodal temporal features of sensory inputs, however, is in accordance with previous work, which by the use of positron-emission-tomography showed that synchronous versus asynchronous audiovisual speech yielded increased activity in STS [Macaluso et al., 2004].

The lateralization of heteromodal (and modality-specific) stimulus offset detection of nonlinguistic stimuli to STS and insula of the right hemisphere found in our study is in accordance with previous work demonstrating that auditory [Herdener et al., 2007] and visual stimulus offset [Fox et al., 2005] is preferably processed in right-hemispheric non-primary cortices. In addition, the analysis of temporal structure of nonlinguistic auditory inputs seems to be processed preferentially in the right hemisphere [Kircher et al., 2004; Mustovic et al., 2003]. In contrast, studies investigating activation related to the temporal synchronicity of linguistic multisensory signals found a predominantly left-hemispheric activity in STS [Calvert et al., 2000; Macaluso et al., 2004].

The involvement of the right insular cortex in the processing of intersensory temporal synchrony has been shown in a positron-emission-tomography study where the subjects were asked to judge temporal asynchronies between subsequent visual and auditory inputs [Bushara et al., 2001]. However, this function has not been traditionally assigned to the human insula, which has previously been suggested to be involved in autonomic, visceral, somatosensory, vestibular, smell, taste, and language processing [Augustine, 1996; Mesulam and Mufson, 1984; Penfield and Faulk, 1955]. Similarly, functional neuroimaging studies showed activation of the insula under a wide variety of behavioral conditions including pain perception, anticipatory anxiety, reflex conditioning, associative learning, and association of auditory and visual information but did not specifically investigate temporal aspects of multisensory stimulation as the basis for insula activity [Büchel et al., 1998; Calvert et al., 1997; Fischer et al., 1998; Ploghaus et al., 1999].

Although it cannot be fully excluded that the insular activation found in our experiments is accordingly related to subjectively experienced slight unpleasantness of auditory and visual stimulation, this explanation seems unlikely, since none of the subjects reported pronounced unpleasantness related to sensory stimuli. Moreover, painful nerve stimulation has been shown to induce tonic BOLD responses in right insula without transient peaks at on- and offset, in contrast to similar nonpainful stimula-

tion, which results mainly in transient rather than sustained insular responses at the onset of stimulus blocks [Downar et al., 2003]. Here, accordingly, we found a transient (rather than sustained) increase of BOLD activity related to the offset of sensory stimulation is not associated with a pronounced unpleasantness.

Animal studies suggest that the insular cortex receives auditory and visual inputs via multiple pathways from primary sensory cortices, multisensory regions and, moreover, via subcortical projections from thalamic nuclei [Graybiel, 1973; Guldin and Markowitsch, 1984; Hicks et al., 1988]. Moreover, an anatomical connectivity between insula and STS has been suggested by studies using diffusion tensor imaging to map white matter fascicles in the human brain [Catani et al., 2002]. Therefore, it is possible that the processing of temporal edges in STS and the insular region is organized in a parallel and/or serial fashion, with activity in heteromodal regions of STS probably driven by inputs from primary sensory regions and activity in the insula related to inputs from subcortical and/or multimodal association areas like STS. However, because of the sluggishness of BOLD responses and the comparatively poor temporal resolution it was not possible to draw conclusions about the exact temporal order of activation, which should be further investigated using electrophysiological techniques.

The potential role of STS and insula, which show a co-responsiveness to auditory and visual edges, in mediating audiovisual psychological phenomena like the McGurk effect, whereby perceived speech sounds can be influenced by seen lip movements [McGurk and MacDonald, 1976], should be addressed in future investigations, since this effect is affected by temporal synchrony of the audio-visual inputs (although a temporal window of about 200 ms for integration of cross-modal information has been demonstrated behaviorally [van Wassenhove et al., 2006]). In addition, impact of temporal proximity of heteromodal stimuli on (spatial) perception has been demonstrated behaviorally for nonspeech audiovisual inputs [Slutsky and Recanzone, 2001], and their integration might also rely on brain areas involved in temporal edge detection.

In conclusion, our data suggest that regions in the STS and the insula of the right hemisphere are crucially involved in the demarcation of the temporal edges represented by the offset of acoustic and visual stimulus trains. Besides modality-specific edge-related activity, a conjunction analysis revealed overlapping multisensory zones located amongst modality-specific regions and being susceptible for both auditory and visual edge detection. We, therefore, tentatively hypothesize that these regions are capable to combine modality-specific features arising from the same object based on the synchrony of their temporal occurrence into a complex object representation, since STS and insula are known to be crucially involved in the multisensory processing of complex audiovisual objects [Amedi et al., 2005; Beauchamp et al., 2004a,b; Raji et al., 2000]. They receive inputs from unimodal cortices as well as

from subcortical structures [Graybiel, 1973; Guldin and Markowitsch, 1984; Hicks et al., 1988; Seltzer et al., 1996] and are thus ideally located for the integration of auditory and visual information.

REFERENCES

- Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005): Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res* 166:559–571.
- Augustine JR (1996): Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Res Brain Res Rev* 22:229–244.
- Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004a) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat Neurosci* 7:1190–1192.
- Beauchamp MS, Lee KE, Argall BD, Martin A (2004b) Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41:809–823.
- Boynton GM, Engel SA, Glover GH, Heeger DJ (1996): Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16:4207–4221.
- Büchel C, Morris J, Dolan RJ, Friston KJ (1998): Brain systems mediating aversive conditioning: An event-related fMRI study. *Neuron* 20:947–957.
- Bushara KO, Grafman J, Hallett M (2001): Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci* 21:300–304.
- Calvert GA, Thesen T (2004): Multisensory integration: Methodological approaches and emerging principles in the human brain. *J Physiol (Paris)* 98:191–205.
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997): Activation of auditory cortex during silent lipreading. *Science* 276:593–596.
- Calvert GA, Campbell R, Brammer MJ (2000): Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649–657.
- Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001): Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14:427–438.
- Catani M, Howard RJ, Pajevic S, Jones DK (2002): Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage* 17:77–94.
- Downar J, Mikulis DJ, Davis KD (2003): Neural correlates of the prolonged salience of painful stimulation. *Neuroimage* 20:1540–1551.
- Fischer H, Andersson JL, Furmark T, Fredrikson M (1998): Brain correlates of an unexpected panic attack: A human positron emission tomographic study. *Neurosci Lett* 251:137–140.
- Fishbach A, Nelken I, Yeshurun Y (2001): Auditory edge detection: A neural model for physiological and psychoacoustical responses to amplitude transients. *J Neurophysiol* 85:2303–2323.
- Fishbach A, Yeshurun Y, Nelken I (2003): Neural model for physiological responses to frequency and amplitude transitions uncovers topographical order in the auditory cortex. *J Neurophysiol* 90:3663–3678.
- Fox MD, Snyder AZ, Barch DM, Gusnard DA, Raichle ME (2005): Transient BOLD responses at block transitions. *Neuroimage* 28:956–966.
- Foxe JJ, Wylie GR, Martinez A, Schroeder CE, Javitt DC, Guilfoyle D, Ritter W, Murray MM (2002): Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *J Neurophysiol* 88:540–543.
- Frens MA, Van Opstal AJ, Van der Willigen RF (1995): Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept Psychophys* 57:802–816.
- Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, Turner R (1995): Analysis of fMRI time-series revisited. *Neuroimage* 2:45–53.
- Genovese CR, Lazar NA, Nichols T (2002): Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15:870–878.
- Gibson JJ (1966): *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton Mifflin.
- Gibson JJ (1979): *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum.
- Gibson JJ (1982): What is involved in surface perception. In: Beck J, editor. *Organization and Representation in Perception*. Hillsdale, NJ: Lawrence Erlbaum. pp 151–157.
- Graybiel AM (1973): The thalamo-cortical projection of the so-called posterior nuclear group: a study with anterograde degeneration methods in the cat. *Brain Res* 49:229–244.
- Guldin WO, Markowitsch HJ (1984): Cortical and thalamic afferent connections of the insular and adjacent cortex of the cat. *J Comp Neurol* 229:393–418.
- Harms MP, Guinan JJ Jr, Sigalovsky IS, Melcher JR (2005): Short-term sound temporal envelope characteristics determine multi-second time patterns of activity in human auditory cortex as shown by fMRI. *J Neurophysiol* 93:210–222.
- Harms MP, Melcher JR (2002): Sound repetition rate in the human auditory pathway: Representations in the waveshape and amplitude of fMRI activation. *J Neurophysiol* 88:1433–1450.
- Herdener M, Esposito F, Di Salle F, Lehmann C, Bach DR, Scheffler K, Seifritz E (2007): BOLD correlates of edge detection in human auditory cortex. *Neuroimage* 36:194–201.
- Hicks TP, Benedek G, Thurlow GA (1988): Modality specificity of neuronal responses within the cat's insula. *J Neurophysiol* 60:422–437.
- Hubel DH, Wiesel TN (1962): Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160:106–154.
- Hughes HC, Reuter-Lorenz PA, Nozawa G, Fendrich R (1994): Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. *J Exp Psychol Hum Percept Perform* 20:131–153.
- Kircher TT, Rapp A, Grodd W, Buchkremer G, Weiskopf N, Lutzenberger W, Ackermann H, Mathiak K (2004): Mismatch negativity responses in schizophrenia: A combined fMRI and whole-head MEG study. *Am J Psychiatry* 161:294–304.
- Konishi S, Donaldson DI, Buckner RL (2001): Transient activation during block transition. *Neuroimage* 13:364–374.
- Kubovy M, Van Valkenburg D (2001): Auditory and visual objects. *Cognition* 80:97–126.
- Lehmann C, Herdener M, Esposito F, Hubl D, di Salle F, Scheffler K, Bach DR, Federspiel A, Kretz R, Dierks T, Seifritz E (2006): Differential patterns of multisensory interactions in core and belt areas of human auditory cortex. *Neuroimage* 31:294–300.

- Macaluso E, George N, Dolan R, Spence C, Driver J (2004): Spatial and temporal factors during processing of audiovisual speech: A PET study. *Neuroimage* 21:725–732.
- McGurk H, MacDonald J (1976): Hearing lips and seeing voices. *Nature* 264:746–748.
- Mesulam M, Mufson E (1984): Association and auditory cortices. In: Peters AJE, editor. *Cerebral Cortex*. New York: Plenum. pp 179–224.
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (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.
- Mustovic H, Scheffler K, Di Salle F, Esposito F, Neuhoﬀ JG, Hennig J, Seifritz E (2003): Temporal integration of sequential auditory events: silent period in sound pattern activates human planum temporale. *Neuroimage* 20:429–434.
- Nichols T, Brett M, Andersson J, Wager T, Poline JB (2005): Valid conjunction inference with the minimum statistic. *Neuroimage* 25:653–660.
- Pantev C, Makeig S, Hoke M, Galambos R, Hampson S, Gallen C (1991): Human auditory evoked gamma-band magnetic fields. *Proc Natl Acad Sci USA* 88:8996–9000.
- Penfield W, Faulk ME Jr (1955): The insula; further observations on its function. *Brain* 78:445–470.
- Ploghaus A, Tracey I, Gati JS, Clare S, Menon RS, Matthews PM, Rawlins JN (1999): Dissociating pain from its anticipation in the human brain. *Science* 284:1979–1981.
- Radeau M (1994): Auditory-visual spatial interaction and modularity. *Curr Psychol Cogn* 13:3–51.
- Raij T, Uutela K, Hari R (2000): Audiovisual integration of letters in the human brain. *Neuron* 28:617–625.
- Seifritz E, Esposito F, Hennel F, Mustovic H, Neuhoﬀ JG, Bilecen D, Tedeschi G, Scheffler K, Di Salle F (2002): Spatiotemporal pattern of neural processing in the human auditory cortex. *Science* 297:1706–1708.
- Seifritz E, Di Salle F, Esposito F, Herdener M, Neuhoﬀ JG, Scheffler K (2006): Enhancing BOLD response in the auditory system by neurophysiologically tuned fMRI sequence. *Neuroimage* 29:1013–1022.
- Seltzer B, Cola MG, Gutierrez C, Massee M, Weldon C, Cusick CG (1996): Overlapping and nonoverlapping cortical projections to cortex of the superior temporal sulcus in the rhesus monkey: Double anterograde tracer studies. *J Comp Neurol* 370:173–190.
- Slutsky DA, Recanzone GH (2001): Temporal and spatial dependency of the ventriloquism effect. *Neuroreport* 12:7–10.
- Stein BE, Meredith MA (1993): *The Merging of the Senses*. Cambridge, MA: MIT Press.
- Talairach J, Tournoux P (1988): *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging*. Stuttgart: Thieme.
- van Wassenhove V, Grant KW, Poeppel D (2007): Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45:598–607.
- Wallace MT, Ramachandran R, Stein BE (2004): A revised view of sensory cortical parcellation. *Proc Natl Acad Sci USA* 101:2167–2172.
- Wertheimer M. 1938. *Laws of organization in perceptual forms*. In: Ellis W, editor. *A Source Book of Gestalt Psychology*. London: Routledge and Kegan Paul. pp 71–78.