

# Audiovisual Non-Verbal Dynamic Faces Elicit Converging fMRI and ERP Responses

Julie Brefczynski-Lewis · Svenja Lowitzsch ·  
Michael Parsons · Susan Lemieux · Aina Puce

Received: 2 February 2009 / Accepted: 31 March 2009 / Published online: 23 April 2009  
© The Author(s) 2009. This article is published with open access at Springerlink.com

**Abstract** In an everyday social interaction we automatically integrate another's facial movements and vocalizations, be they linguistic or otherwise. This requires audiovisual integration of a continual barrage of sensory input—a phenomenon previously well-studied with human audiovisual speech, but not with non-verbal vocalizations. Using both fMRI and ERPs, we assessed neural activity to viewing and listening to an animated female face producing non-verbal, human vocalizations (i.e. coughing, sneezing) under audio-only (AUD), visual-only (VIS) and audiovisual (AV) stimulus conditions, alternating with Rest (R). Underadditive effects occurred in regions dominant for sensory processing, which showed AV activation greater than the

dominant modality alone. Right posterior temporal and parietal regions showed an AV maximum in which AV activation was greater than either modality alone, but not greater than the sum of the unisensory conditions. Other frontal and parietal regions showed Common-activation in which AV activation was the same as one or both unisensory conditions. ERP data showed an early superadditive effect ( $AV > AUD + VIS$ , no rest), mid-range underadditive effects for auditory N140 and face-sensitive N170, and late AV maximum and common-activation effects. Based on convergence between fMRI and ERP data, we propose a mechanism where a multisensory stimulus may be signaled or facilitated as early as 60 ms and facilitated in sensory-specific regions by increasing processing speed (at N170) and efficiency (decreasing amplitude in auditory and face-sensitive cortical activation and ERPs). Finally, higher-order processes are also altered, but in a more complex fashion.

This article is published as part of the Special Issue on Multisensory Integration.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10548-009-0093-6) contains supplementary material, which is available to authorized users.

J. Brefczynski-Lewis · S. Lowitzsch · M. Parsons ·  
S. Lemieux · A. Puce  
Center for Advanced Imaging & Department of Radiology, West  
Virginia University, Morgantown, WV, USA

M. Parsons · A. Puce  
Center for Advanced Imaging & Departments of Behavioral  
Medicine & Psychiatry, West Virginia University, Morgantown,  
WV, USA

S. Lowitzsch  
Center for Advanced Imaging & Department of Mathematics,  
West Virginia University, Morgantown, WV, USA

A. Puce  
Center for Advanced Imaging & Department of Neurobiology/  
Anatomy, West Virginia University, Morgantown, WV, USA

M. Parsons  
Psychology & Psychiatry, Cleveland Clinic, Cleveland, OH,  
USA

S. Lemieux  
Social and Life Sciences Imaging Center, The Pennsylvania  
State University, Hershey, PA, USA

S. Lowitzsch  
Max-Planck Forschungsgruppe, Institut für Optik, Informatik  
und Photonik, Universität Erlangen-Nürnberg, Erlangen,  
Nuremberg, Germany

J. Brefczynski-Lewis · A. Puce (✉)  
Department of Psychological & Brain Sciences, Indiana  
University, 1101 E 10th St, Bloomington, IN 47405, USA  
e-mail: ainapuce@indiana.edu

**Keywords** Facial motion · STS · IPL · IFG · Social cognition · Multisensory integration

## Introduction

Everyday social interactions involve the integration of auditory and visual information from speech and non-verbal social cues. These latter cues are often underemphasized in humans, as most attention tends to focus on the spoken word (Campbell et al. 2001; Capek et al. 2008; Frith and Frith 2007; Kawashima et al. 1999; Macaluso et al. 2004; MacSweeney et al. 2000). Humans generate many non-verbal vocalizations that are accompanied by readily identifiable stereotypical facial gestures (Howell 1900). Non-verbal vocalizations likely engage higher-order processing, and can be overlooked, misused, or misinterpreted by those with social cognition disorders (Golarai et al. 2006; Luyster et al. 2008; Sarfati et al. 1997; Troisi et al. 1998). Non-verbal vocalizations can be communicative as one may purposely vocalize or exaggerate non-verbal cues to send a message, such as burp to signal the deliciousness of a meal, or one may purposely suppress a sign or yawn to conceal dissatisfaction or boredom. Social and other advantages may thus come from the ability to interpret information about the mental, emotional, or homeostatic state of individuals as conveyed through multisensory non-verbal cues. This is supported by studies that show greater activation to human non-verbal stimuli versus other non-human categories in multiple regions including STS, frontal parietal regions, and insula (Fecteau et al. 2007; Lewis et al. 2008).

In a normal context, the accurate interpretation of socially related non-verbal information requires appropriate integration of multisensory input, usually visual and auditory information, which can change based on incoming information quality. In a noisy situation like a crowded bar, one observes lip and face movements more than in a quiet setting, as the visual information can effectively amplify the audio by up to 11 dB (MacLeod and Summerfield 1987). Behavioral studies of both speech and non-speech stimuli indicate that multiple (congruent) stimulus modalities lead to improved processing, with both shorter reaction times and increased accuracy compared to either modality alone (Grant and Walden 1996; Miller 1982; Sumbly and Pollack 1954).

These behavioral advantages for multisensory stimuli manifest as differences in timing, amount and type of brain activity compared to unisensory stimuli. However, studies have revealed conflicting results including both facilitation,

in the form of faster and *decreased* brain responses (for fMRI Martuzzi et al. 2007; Wright et al. 2003; for ERPs Besle et al. 2004; van Wassenhove et al. 2005), and enhancement, or *increased* activation, for multisensory versus unisensory stimuli (Hubbard et al. 2008; Kayser et al. 2007). The reasons for these differences in multisensory effects are not understood, although some studies suggest that they may be related to factors such as congruency (Puce et al. 2007; Saint-Amour et al. 2007), whether one modality predicts the other (Ghazanfar et al. 2005; Stekelenburg and Vroomen 2007), or neuronal population properties (Laurienti et al. 2005; Stevenson et al. 2007). Even more complex results have been seen for higher-order regions, with effects (in speech-related studies) seen in posterior superior temporal sulcus (pSTS), inferior parietal lobule (IPL), and inferior frontal cortex (IFC) (Calvert et al. 2001; Kawashima et al. 1999). In the current study we were particularly interested in multisensory effects in pSTS due to its postulated role in social related processes (Redcay 2008), and links to different visual, auditory, and motor processes (Beauchamp et al. 2004).

To investigate multisensory effects related to human non-verbal vocalizations and accompanying facial movements, we studied neural responses elicited to an animated synthetic female face producing various non-verbal vocalizations (i.e. coughing, yawning), using both functional magnetic resonance imaging (fMRI) and event-related potentials (ERPs). We presented stimuli under three conditions. In the audiovisual (AV) condition participants saw the animated face and heard congruent human vocalizations. In the visual (VIS) condition, only the animated face was seen, whereas in the auditory (AUD) condition only the vocalizations were heard. Randomized blocked presentations of AV, VIS and AUD conditions were alternated with rest (R). Two participant groups ( $n = 10$  for fMRI,  $n = 13$  for ERPs) responded to infrequent unisensory targets (animated face blinking, or uttering “mmm” without a visual change to the face). Our hypothesis predicted that sensory-specific regions specialized for a given unisensory condition, would show facilitated processing (faster times to peak and reduced amplitudes) in the presence of a multisensory stimulus. Specifically, for the fMRI experiment, we predicted a reduced BOLD signal for the AV versus either unimodal condition in sensory regions. For the ERP experiment, we predicted reduced amplitudes and faster latencies for early ERP components. In addition, we predicted that higher-order regions, especially right pSTS, would show greater AV activation (versus unisensory conditions) due to specialization in multisensory and/or social processes.

## Materials and Methods

### Participants

For the fMRI study 10 right-handed healthy males participated (ages: 24–37 years, mean 28 years). For the ERP study, there were 13 right-handed participants (18 originally collected, 5 excluded, for the 13 included participants: 7 males, ages: 19–43 years, mean 29 years). All participants had either normal or corrected-to-normal vision and gave informed consent in a study approved by the Institutional Review Board for the Protection of Human Participants at West Virginia University.

### Stimuli and Task

Participants viewed  $4 \times 4$  degree videos of a synthetic female face producing facial movements and vocalizations. Stimuli were seven non-speech vocalizations with accompanying face movements consisting of a cough, sneeze, burp, yawn, laugh, sigh and whistle. Two infrequently presented unisensory target stimuli, a blink (visual) and an uttered ‘mmm’ (auditory), made participants focus on visual and auditory sensory input equally. In the Audiovisual condition, there was a 33 ms (or 1 video frame) delay between the peak movement (i.e. fully opened mouth) and sound. Animations were based on filmed real life movements associated with the seven non-verbal vocalizations of three different actors.

Stimulus type was pseudorandomly ordered within 20 s stimulus blocks consisting of 10 trials each of combined Audiovisual stimulation (AV), Auditory stimulation only (AUD), and Visual stimulation only (VIS) (Fig. 1). In the AV participants saw the face making the facial movements and heard the associated non-speech vocalizations. In VIS, participants observed the face making movements without

hearing the vocalizations. In AUD, participants viewed a neutral colored plain background (RGB = 140, 132, 127) and heard the vocalizations. The absence of the face for the AUD condition prevented an ‘incongruent’ stimulus (face still but vocalization present), but made an event-related design difficult due to onset effects. Visual motion duration and sound duration was 600 and 567 ms respectively, for all non-target stimulus types. Participants maintained their gaze on an ever-present green fixation cross and pressed a single response button when either of one of the two specified unisensory targets were seen or heard. Behavioral responses were monitored to ensure attentional alertness. Minor variations in timing occurred for the fMRI versus ERP paradigms.

### Data Acquisition

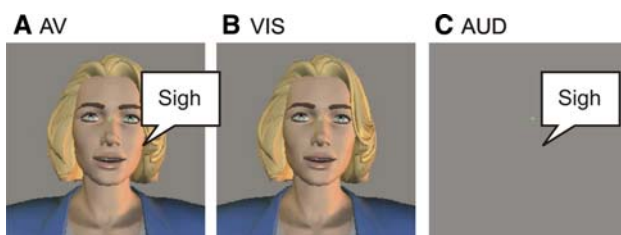
#### Functional MRI Study

Data were acquired on a 3 Tesla GE Horizon LX MRI scanner and quadrature birdcage headcoil. We used a 14 slice split-sagittal acquisition (Puce et al. 2003), where 7 sagittal slices (3 mm thickness + 1 mm gap) were taken in each hemisphere to maximally visualize the cortex of the STS and STG (see Supplementary Fig. 1). A series of 125 gradient echo echoplanar volumes were acquired over each of the three, 4 min 10 s stimulation periods (after before and after rest period removal, total = 375 volumes) using the following parameters: TE = 35, TR = 2000,  $\alpha = 70^\circ$ , NEX = 1, BW = 125, FOV = 24 mm, matrix =  $128 \times 128$  (in-plane resolution of 1.875 mm), slice thickness = 3 mm, gap = 1 mm. In the Talarach  $x$  plane, sagittal slice coverage was from  $x = -34$  to  $-67$ , and  $x = 34$  to 67.

A T1-weighted whole brain volume which was acquired as a high-resolution spoiled gradient-recalled acquisition in a steady state (SPGR) (voxel size =  $1.2 \times 0.9375 \times 0.9375$  mm; FOV = 240; matrix =  $256 \times 256$ ; 124 slices).

#### EEG/ERP Study

Participants were seated comfortably in an armchair in a dimly lit room with a white noise generator. A continuous 128-channel recording of 124 channels of scalp EEG (QuikCap, Compumedics Neuroscan, El Paso, TX, USA) and 4 channels of horizontal and vertical electrooculograph (EOG) was taken using Neuroscan 4.3 software (Compumedics, Neuroscan, El Paso, TX, USA). Data were sampled at 250 Hz/channel and bandpass filtered from 0.1 to 100 Hz and amplified with a gain of 5,000. A reference consisted of two electrodes placed either side of the nose or on the cheek close to the nose. The midline frontal ground



**Fig. 1** Example of a stimulus still frame depicted at the middle of an animation. **a** In the AV condition, the face is present and the non-verbal vocalization accompanies the visual stimulus. Here a yawn is depicted and the open mouth and narrowing eyes can be clearly seen. **b** In the VIS condition only the moving face is present. **c** In the AUD condition only the vocalizations are heard. For all conditions a green fixation cross was located in same position on the screen throughout scans (between the eyes when the face present in the AV and VIS conditions)

electrode was sited on the electrode cap itself. Electrode impedances were kept below 10 k $\Omega$ .

## Data Analysis

### *Functional MRI*

Data reconstruction was implemented via Analysis of Functional Neural Images (AFNI), version 2.31 software (Cox 1996). Data processing steps included offline image reconstruction in conjunction with smoothing in Fourier space via a Fermi window (full width at half maximum = 1 voxel), correction for differences in slice-timing, and 6-parameter rigid-body motion correction. The motion estimates over the course of the scan for translation (inferior–superior, right–left, and anterior–posterior) and rotation (yaw, pitch, roll) parameter estimates were used as covariates in further analyses.

Each image time series was spatially registered to the volume closest in time to the high-resolution structural scan both within-plane and then in all three planes using an iterative linear least squares method, to reduce the effects of head motion. AUD, VIS, and AV blocks were analyzed with a least-squares general linear model (GLM) fit that modeled each activation block and head motion parameters. Each regressor consisted of an ideal hemodynamic response function for the specified block type, obtained by convolving the event time file (across 3 concatenated imaging runs) with a  $\gamma$ -variate function. The beta-weights resulting from the GLM analysis were converted to percent signal change using the mean overall baseline and spatially smoothed using a 4 mm Gaussian filter. These percentage signal change maps were transformed into standardized Talaraich space.

A voxel-by-voxel parametric two-tailed  $t$  test was used on the percent signal change maps for a group comparison of each condition (versus rest) separately, plus AV versus AUD + VIS.  $P$ -value correction for multiple comparisons was based on a combination of threshold cutoff and cluster extent using 3dmerge (AFNI). Minimal cluster size was calculated using Monte Carlo simulation program AlphaSim (AFNI). For a masked AFNI image, AlphaSim ran 1,000 iterations, with a radius connectivity of 4.1 (since slice thickness + gap was 4 mm) and image defined Gaussian filters with FWHM determined with 3dFWHM. The minimal cluster size to avoid false cluster detection was 57 voxels for  $P < 0.05$ , 16 voxels for  $P < 0.01$ , 11 voxels for  $P < 0.005$ , 6 voxels for  $P < 0.001$ . Alpha maps were overlaid on inflated PALS atlas cortical model brains (Van Essen 2005; Van Essen et al. 2001).

Regions of interest (ROIs) were based on significant activation from the analyses above. The average time course of the MR BOLD response in select ROIs was

generated using the AFNI 3dDeconvolve program with the *iresp* option. The average time courses for each condition (AUD, VIS, AV, ApV) were averaged within each ROI and normalized across datasets. For a given hemisphere, we took voxels showing significant activation from that hemisphere plus its mirror opposite correlate (using 3dLRflip in AFNI), such that each ROI had equivalent right and left hemisphere volumes.

### *Event-Related Potentials*

ERP analysis was performed using Neuroscan 4.3 Software (Neurosoft, Inc., Sterling, VA, USA). EEG data were first segmented into 1500 ms epochs with 100 ms pre-stimulus baseline based on the event markers which identified each trial type. The EEG data of the target trials were not included in subsequent analyses (similar to the fMRI study). Epochs containing artifact registering greater than  $\pm 100$   $\mu$ V, due primarily to eye blinks, or electromyographic activity due to face or head motion, were excluded from subsequent analyses. We excluded data from five of the eighteen participants (three participants had technically suboptimal studies due to excessive eye blinks/muscle activity in their EEG data, and two participants were deemed to be overly familiar with the stimuli and showed low alertness levels during the study), leaving thirteen participants in the final ERP analyses.

The zero time point was the start of the audio, visual or simultaneous audiovisual stimulus. Individual epochs were normalized relative to a 100 ms prestimulus baseline, and linear trend was calculated and removed across the entire epoch, based on the prestimulus baseline. Stimulus types for each condition (AUD, VIS, AV) were averaged across all 6 runs. Each participant's averaged ERP data were then digitally smoothed with a zero phase-shift low pass filter (cut-off 30 Hz, 6 dB/octave).

Group averages were constructed and the averaged ERPs were scrutinized to identify ERP peaks and troughs. P100, N170, P250, and P500 ERP components were identified in the group average waveforms. Area under the curve (AUC) ranges were also selected for certain broader peaks. Latency ranges (windows) were selected for each grand average ERP peak or trough, and an automated peak picking routine was then run on the averaged ERP data of each individual subject. Area under the curve (AUC) measures were also taken for selected ERP components. Each subject's ERP waveforms and ERP peak amplitude and latency measures and AUCs were exported as sets of ASCII files.

Topographic voltage maps were created from the grand average ERP data at peak and trough timepoints to examine the regional distribution of ERP activity. Data from multiple sensors showing similar ERP behavior were then

averaged as noted in the results section, with location of sensor markers determined by averaging Polhemus digitizer locations.

Data at eye channels were also displayed in order to determine whether ERP signals may have been influenced by systematic, but subtle, eye movements. The signal excursion for the artifact free data in the eye channels was small (on the order of  $\mu\text{V}$ ) and therefore did not appear to be due to actual eye movements which typically generate signals on the order of mV. In addition, the lower horizontal EOG channels did not show an equal and opposite negativity, suggesting that the positivity in the upper vertical EOG channel, located on the forehead, was likely due to frontal brain activity and not to eye related activity per se.

### Statistical Analysis of ERP Data

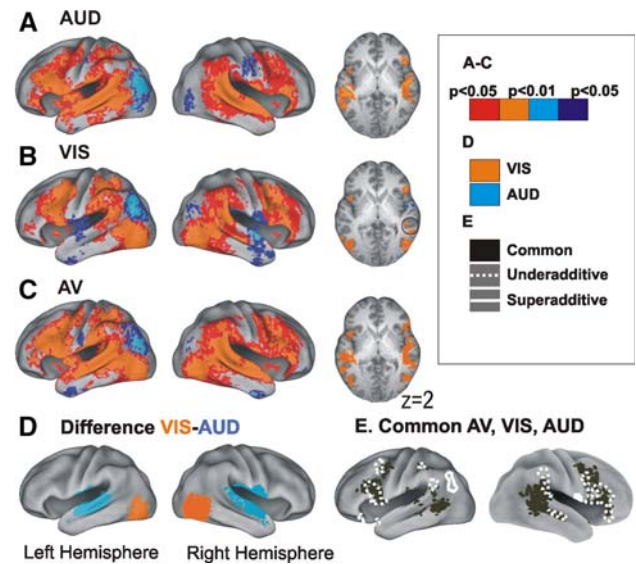
Student's *t* tests, one-way (Condition) and two-way (Condition by Hemisphere) ANOVAs of peak amplitudes, latencies, and AUC for particular ERP components were analyzed using SPSS V15. In order to objectively determine the timepoints for AUC measures, we calculated timepoint by timepoint values for the *t* test difference for AV versus AUD plus VIS. Thus, we set as time regions for AUC, periods of sustained (20 ms, 5 timepoints) significant differences ( $t > 1.67$  for  $n = 60$  epochs) between the multisensory and sum of the unisensory conditions. We used similar calculations to measure the time elapsed, after which no significant peaks occurred, an effective Return to Baseline (RTB). To determine the RTB we first calculated the *t* value versus zero for each point on the waveform. The RTB was defined as the end of the last significant peak of sustained (20 ms, 5 timepoints) significance.

## Results

### fMRI Data

Ten participants completed the non-verbal unisensory target detection task in a 3T MRI scanner. A split sagittal slice acquisition optimized sampling of temporal cortex, but excluded medial regions as well as more medial aspects of frontal parietal cortex, fusiform and early visual cortex.

All three conditions produced robust activation in sensory and higher-order cognitive regions. AUD and AV conditions produced additional and extensive activation of mid- to anterior STS and mid-insula (Fig. 2a, c), whereas VIS and AV conditions produced activation of lateral occipital and posterior middle temporal gyrus (LO/pMTG) and lateral fusiform gyrus (Fig. 2b, c). Brain regions



**Fig. 2** Group fMRI activation maps for each stimulus condition, AUD (a) VIS (b) and AV (c) versus REST. Warm colors represent net positive BOLD signal, cool colors represent net negative BOLD signal. d Difference maps for VIS versus AUD. Regions more active in the VIS condition are represented by warm colors ( $P < 0.01$  corrected). e Common activation maps (black) for AV, VIS and AUD ( $P < 0.001$  corrected). Overlaid regions show mathematical superadditivity (solid white lines) and underadditivity (dashed white lines)  $P < 0.05$  corrected)

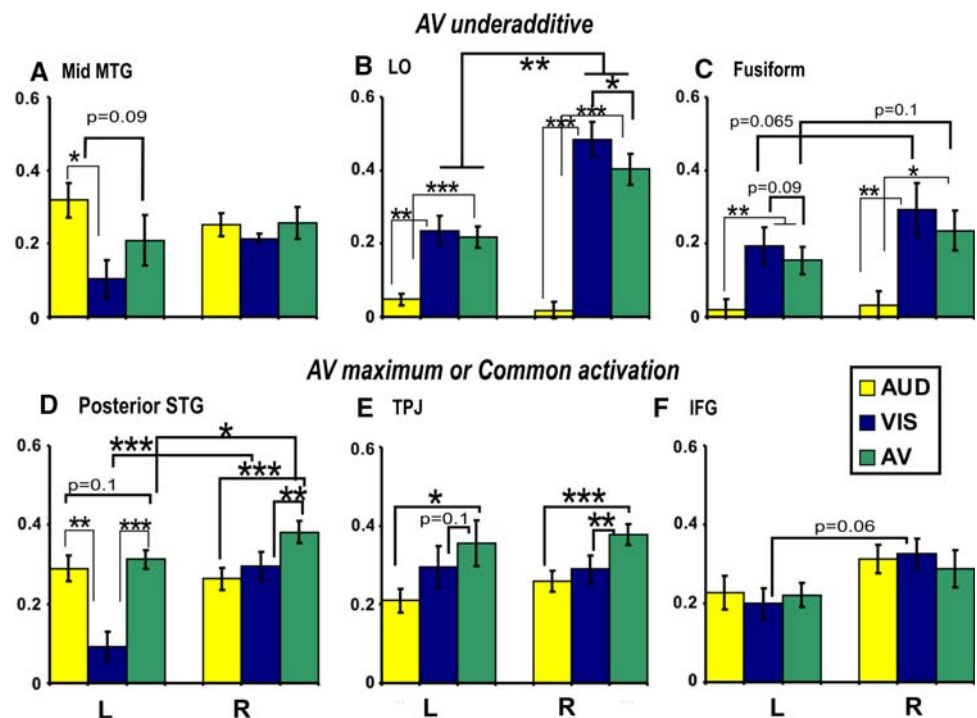
showed multisensory relationships that fell into four main categories (Fig. 3):

- (1) *superadditive*, defined as audiovisual greater than the sum of auditory alone and visual alone i.e.  $AV > ApV$ ;
- (2) *underadditive*, defined as audiovisual less than the sum of auditory and visual alone, and audiovisual less than the dominant sensory modality e.g.  $AV < ApV$  and  $AV < AUD$  or  $VIS$ ;
- (3) *AV maximum*, defined as audiovisual greater than either unisensory condition along ( $AV > VIS$  and  $AV > AUD$ , and  $VIS > 0$ ,  $AUD > 0$ ).
- (4) *Common activation*, defined as AV activation equal to one or both conditions ( $AV = AUD$  and/or  $VIS$ ). Note that for both AV maximum and Common activation, audiovisual would be less than the sum of the unisensory conditions ( $AV < ApV$ ).

Two regions showed mathematical *superadditivity*, the right insula/frontal operculum and left angular gyrus. However, this resulted from negative activation versus baseline in one or in all three conditions, and neither region showed significant positive activation for the AV condition (solid white outlines, Fig. 2d).

Several regions showed significant or near-significant *underadditive* effects including MTG, LO/pMTG, and lateral fusiform gyri, with the AV condition showing decreased activation compared with either unisensory

**Fig. 3** Group fMRI data ROI analyses: underadditive BOLD responses. Histograms depict relative fMRI percent signal change for all three conditions in *underadditive* ROIs (a–c) *AV maximum* (d–e) and *common-activation* ROIs (f). Asterisks indicate pairwise *t* test significance: \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$



condition or the sum of the unisensory conditions, ApV (Table 1A). The AUD-preferred region, left mid-MTG, showed a trend ( $P < 0.1$ ) of  $AV < AUD$ . Similarly, for the VIS-preferred regions, LO and fusiform gyrus, there was a significant difference and trend, respectively, of  $AV < VIS$ . LO and fusiform also showed a right-hemisphere bias (Table 1A).

*AV maximum and Common activation* effects were seen in frontal, parietal and temporal regions (Black overlay at  $P < 0.001$ , Fig. 2d). The pSTS, TPJ, IFG, and DLPFC all showed strong condition effects for  $AV < ApV$  (Table 1B). Portions of these regions were also revealed in a voxelwise *t* test of AV versus ApV (dashed white lines in Fig. 2d). The pSTS showed a significant hemisphere effect (right > left), with the VIS condition showing the strongest lateralization ( $t = 4.98$ ,  $P < 0.005$ ). *AV maximum* activation was seen in several of these regions, including right posterior pSTG ( $P < 0.01$  versus VIS,  $P < 0.001$  versus AUD), and TPJ ( $P < 0.01$ , versus VIS,  $P < 0.05$  versus AUD). IFG and DLPFC showed *common activation*, with the AUD condition showing the least activation in DLPFC.

The STS and IFG regions, in addition to showing at least a trend towards hemisphere effects for amplitude of activation (Table 1A), also showed a greater number of active voxels in the right versus left hemisphere (Right STS: 134 versus Left STS: 0; Right IFG: 1871 mm<sup>3</sup> versus Left IFG: 157 mm<sup>3</sup>), thus showing right hemisphere dominance in both magnitude and extent of activation.

A separate group of 13 participants participated in the ERP version of the experiment. Since the neutral, eyes

forward face was present as a baseline for the duration of VIS and AV blocks (except during movements), the zero time point for ERP measurements is at the onset of the facial movement and/or simultaneous vocalization (for AUD, onset of vocalization). In general, ERP waveforms revealed modality-specific early components with characteristic topographies and morphologies, in addition to a late positivity (Fig. 4). Stimuli containing auditory stimulation (AUD and AV conditions) showed the typical auditory N140 with an amplitude maximum at midline central electrodes. Stimuli containing visual stimulation (VIS and AV conditions) showed the typical face-specific N170 at bilateral temporo-occipital electrodes, showing delayed latencies typical of dynamic visual stimuli. In addition to showing both typical auditory and face-related components, the AV condition also elicited a unique early positivity in left parieto-occipital electrodes. All three conditions showed a diffuse late positivity which lasted up to 1500 ms, and varied in amplitude between conditions. A *t* test of AV versus ApV was used to search for regions of potential multisensory effects. Again, effects were grouped into the four categories of *superadditive*, *underadditive*, *AV maximum* and *common-activated*.

#### *Superadditivity at 60–148 ms*

At this relatively early post-stimulus time range, the AV condition elicited an early positivity in bilateral temporo-occipital electrodes that was not seen in either unisensory condition (Figs. 4, 5a). We performed a timepoint-by-

**Table 1** Summary of significant and trend effects resulting from 2-way ANOVA comparisons for (A) fMRI and (B) ERPs

	Amplitude (AV vs. ApV)	Amplitude (single conditions)	
<b>(A) fMRI results</b>			
Facilitated underadditivity			
mid-MTG	Condition: AV < ApV	$F(2,18) = 17.60^{**}$	Condition: AV < AUD $P = 0.1$ n.s.
LO/pMTG	Condition: AV < ApV	$F(2,18) = 17.25^{**}$	Condition: AV < VIS $F(2,18) = 6.24^*$
	Hemisphere: right > left	$F(2,18) = 24.46^{**}$	Hemisphere: right > left $F(2,18) = 28.97^{***}$
Fusiform	Condition: AV < ApV	$F(2,18) = 6.81^*$	Condition: AV < VIS $P = 0.14$ n.s.
	Hemisphere: right > Left	$P = 0.065$ n.s.	Hemisphere: right > left $P = 0.062$
Common activation			
pSTS	Condition: AV < ApV	$F(2,18) = 52.2^{***}$	$t$ test Rt AV > AUD*** ANOVA n.s.
			AV > VIS**
			Lf: AV > AUD***
			–
	Hemisphere: right > Left	$F(2,18) = 16.5^{**}$	–
	Condition × hemisphere	$F(2,18) = 13.1^{**}$	n.s.
	Condition: AV < ApV	$F(2,18) = 11.0^{**}$	n.s.
TPJ			$t$ test Rt: AV > AUD*** ANOVA n.s.
			AV > VIS**
			Lf: AV > AUD*
			AV > VIS $P = 0.1$ n.s.
			–
	Hemisphere: right = left	n.s.	n.s.
	Condition: AV < ApV	$F(2,18) = 31.4^{***}$	n.s.
IFG			n.s.
	Hemisphere: right > left	$P = 0.17$ n.s.	n.s.
	Condition × hemisphere	$F(2,18) = 6.4^*$	n.s.
	Condition: AV < ApV	$F(2,18) = 20.2^{**}$	n.s.
DLPFC			$F(3,18) = 120.2^{***}$
			Condition: AUD < AV or VIS
<b>(B) ERP results</b>			
Superadditive			
60–148 ms	Condition: AV > ApV	$F(2,24) = 11.98^{**}$	Condition: AV < AUD/VIS $F(3,36) = 4.30^*$
	Condition × hemisphere	$F(2,24) = 16.67^{**}$	–
Facilitated underadditivity			
N140	Condition: AV < ApV	$F(2,24) = 10.86^{**}$	Condition: AV < AUD $P = 0.13$
N170	Condition: AV < ApV	$F(2,24) = 21.61^{**}$	Condition: AV < VIS $F(2,24) = 7.07^*$
	Hemisphere: right > left	$F(2,24) = 5.28^*$	Condition: AV < VIS $F(2,24) = 5.71^*$
	Hemisphere: left > right	$F(2,22) = 5.57^*$	Condition: VIS > AUD or AV $F(3,36) = 45.07^{**}$
AV maximum or common 230–304 ms (12 samples)			–

Table 1 continued

	Amplitude (AV vs. ApV)	Amplitude (single conditions)	Latency
400–600 ms (11 samples)	Condition: AV < ApV $F(2,20) = 25.43^{**}$	Condition: AUD < VIS or AV $F(3,33) = 8.49^{**}$ $t$ test Lf: AV > AUD <sup>***</sup> AV > VIS $P = 0.08$ n.s. Rt: AV > AUD <sup>***</sup>	Condition: VIS > AUD or AV $F(3,33) = 22.20^{***}$

The middle columns show the results of AV versus ApV, while the right columns show the effects for AV versus VIS and AV versus AUD. B has an additional set of columns for ERP latency effects. Asterisks indicate  $P$ -value of  $F$  statistic: \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ , \*\*\*\*  $P < 0.00001$ . All effects were tested, but non-significant interactions are listed as n.s. or are omitted. In several cases, noteworthy  $t$  tests are listed

timepoint  $t$  test analysis to determine the time range showing significant differences between AV and AUD plus VIS (i.e. AV versus ApV), and performed an AUC analysis for this time range (60–148 ms). A two-way ANOVA revealed superadditivity, with main effects of Hemisphere [ $F(2,24) = 16.67$ ,  $P < 0.01$ ].

#### Trend Towards Underadditivity for N140

AUD and AV, but not VIS, conditions elicited a central negativity at 144.7 ms, typical for auditory stimuli (Fig. 6a). Peak amplitude analysis revealed a trend for AUD > AV (Table 1B), but found no significant super- or underadditivity.

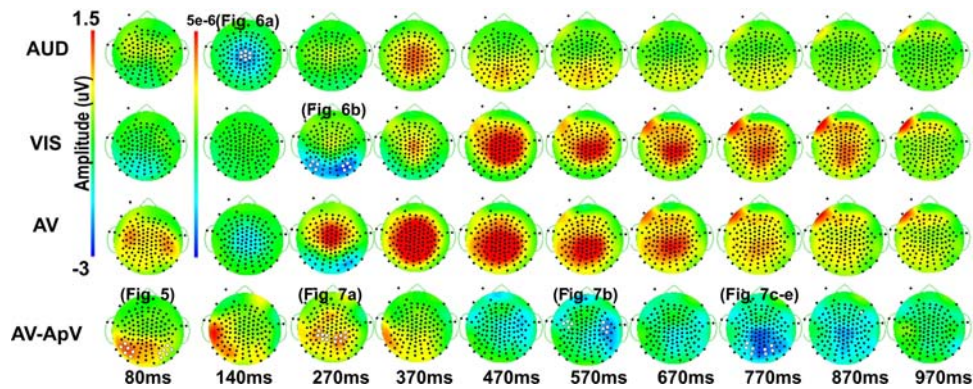
#### Underadditivity for N170

All three conditions produced a negativity at an average 276.7 ms, with preference for conditions including visual stimuli (Fig. 6b). The waveform was characteristic of the N170 which is elicited by dynamic faces, peaking at temporo-occipital electrodes, with a right hemisphere bias (Fig. 5c, white dots). As the stimulus was dynamic, the N170 was considerably delayed relative to the 170 ms typical for the presentation of static face stimuli (see Puce et al. 2000). Face movement began at 0 ms and was generally identifiable as a particular “vocalization” by 33 ms. Although the AV and VIS N170s were larger, the AUD condition also elicited a negativity that had a similar timecourse at these electrodes (274.8 ms). The N170 was underadditive, such that AV < ApV, and showed condition and hemisphere main effects (Table 1B). Peak analysis revealed reduced amplitude and decreased latency for the AV versus VIS condition, as well as a right hemisphere bias for amplitude (Table 1B).

#### AV Maximum and Common Activation for Late Positivities

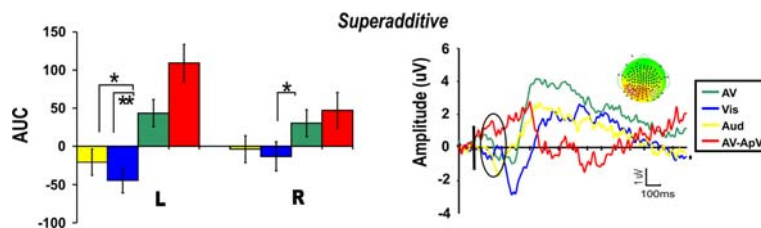
All three conditions elicited widely distributed late positivities (Fig. 7). A strict AV versus ApV  $t$  test revealed an underadditivity at bilateral temporo-parietal electrodes at 230–304 ms (timepoint-by-timepoint  $t$  test, Fig. 7a). Even though the peaks were broad, we used a semi-automated peak analysis with verification of peaks in individual subjects data. We wanted to examine the data for latency differences, and more strictly apply multisensory criteria (at the time of the AV and AUD peaks, the VIS peak has yet to appear). The temporo-parietal peak occurred at 240 and 244 ms respectively for the AUD and AV conditions, but was delayed at 328 ms for the VIS condition. This peak was considered a *common activation*, as analysis, using the homologous peaks for all three conditions, did not show





**Fig. 4** Group average topographic ERP maps as a function of condition and time. Topographic maps are depicted at post-stimulus time points of 80, 140, 270 ms, and then every 100 ms for all three stimulus conditions (*top three rows*). The *bottom row* shows

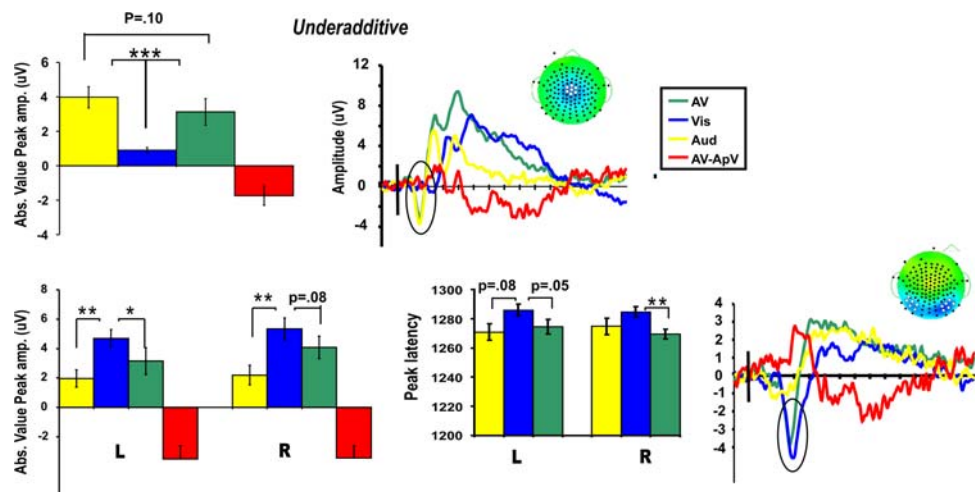
difference maps for the AV condition minus the sum of the unisensory conditions (ApV). *Red-yellow* shows positive ERP activity, *blue-aqua* show negative ERP activity. Topographic maps showing timepoints analyzed in subsequent figures are labeled



**Fig. 5** Superadditivity in group ERP averaged data at bilateral temporo-occipital electrodes (from 60 to 148 ms). Histograms for area under the curve (AUC) analysis for time range 60–148 ms. Topographic map shows 12 sampled electrodes (*white dots*).

Waveform for 6 averaged right temporal occipital electrodes shows unique peak for AV (*black circle*). Asterisks indicate pairwise *t* test significance

**Fig. 6** Underadditivity in group averaged ERP data in relatively early post-stimulus timeranges. Histograms for peak amplitude for **a** N140 and both peak amplitude and latency for **b** N170. Averaged ERP waveforms for sampled electrode sites (*white dots* on topographic maps), appear at the right of the histograms. Asterisks indicate pairwise *t* test significance. *P*-values listed for non-significant trends

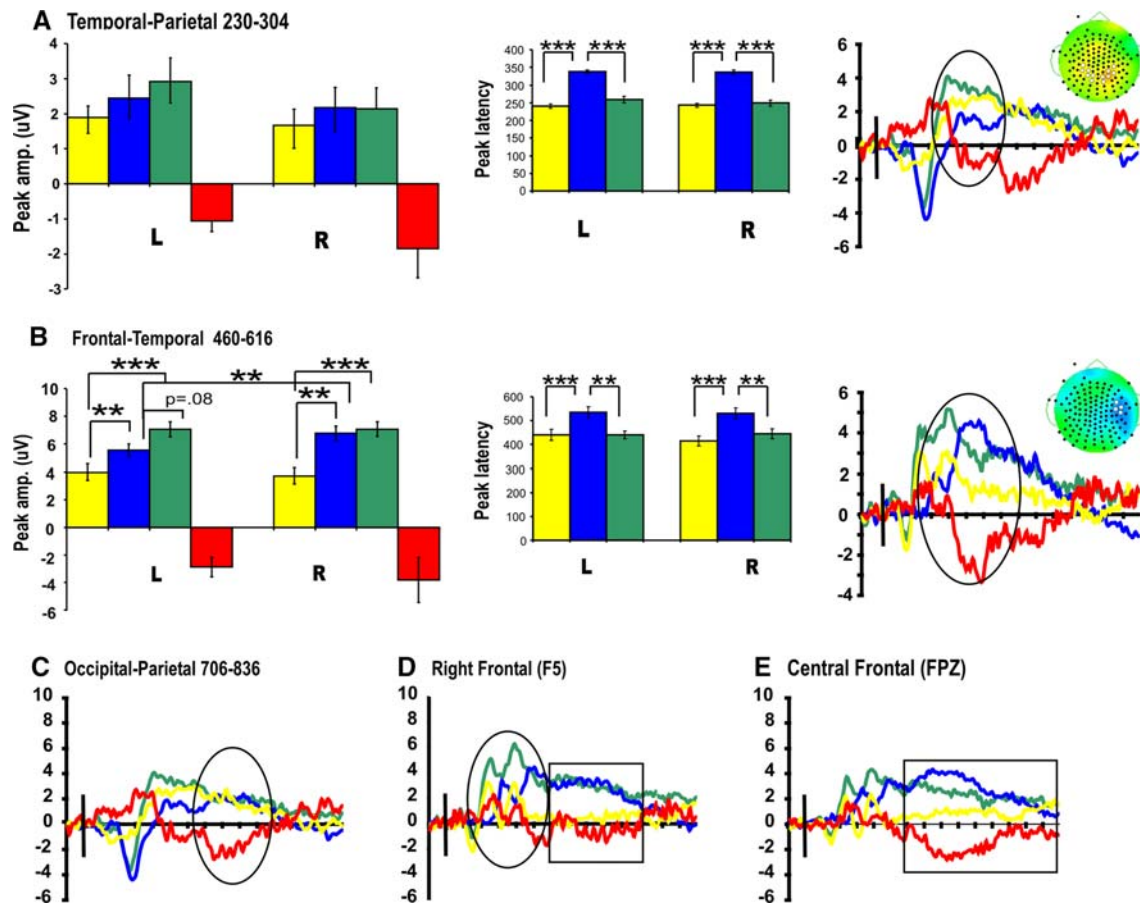


any significant amplitude differences between conditions (Table 1B).

An underadditivity was seen at right fronto-temporal electrodes at 460–616 ms (timepoint-by-timepoint *t* test, Fig. 7b). Latency of this broad peak was also greatest for the VIS condition (492 ms) and was similar across the AUD and AV conditions (384 and 400 ms, respectively). Analysis of homologous peaks revealed a Condition effect,

with the AUD condition showing the smallest amplitude (Table 1B). This peak was considered an AV *maximum*, as the AV condition was significantly greater than either unisensory condition in the right hemisalp, and a similar trend was seen in the mirror opposite electrodes (Fig. 7b).

*Common activation* was seen at several other electrode sites, include occipital and frontal electrodes. Occipital electrodes showed equivalent sustained activation for all



**Fig. 7** Later ERPs histograms and waveforms. Charts showing peak amplitude and latency analysis for Common activation **a** 12 temporoparietal electrodes (6 in each hemisphere) in the post-stimulus timerange, 230–304 ms and AV maximum, **b** 8 frontal-temporal electrodes in the timerange 460–616 ms. Rightmost panel in each row

shows ERP waveforms for each condition along with AV-ApV topographic maps (sampled electrodes are white dots). Asterisks indicate pairwise *t* test significance. Very late timeranges show common activation in occipital and frontal electrode sites (**c–e**)

three conditions in the 700–800 ms range (Fig. 7c). In contrast, for frontal electrodes, F5 and FPZ, late sustained activation was only seen for AV and VIS conditions and not the AUD condition (rectangles, Fig. 7d, e). Note at F5 the AV maximum peaks at the 250 ms and 450 ms ranges (circle, Fig. 7d).

### Results Summary

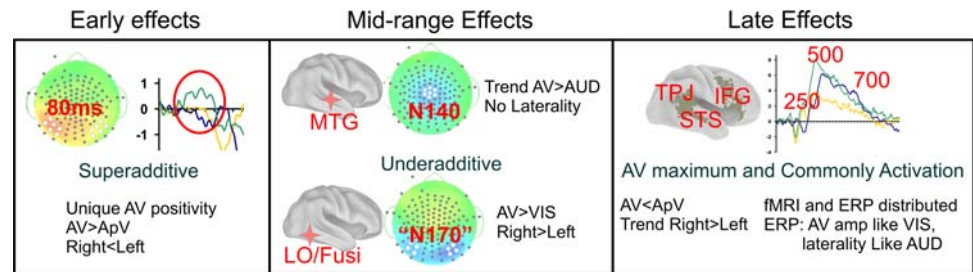
The ERP data showed a unique early positivity for the AV condition starting at 60 ms. In sum, however, there was convergence of ERP and fMRI data. Both the N140 (generated in the superior temporal plane) (Giard et al. 1994; Godey et al. 2001; Ponton et al. 2002), and the auditory STS showed non-significant trends towards underadditivity. For VIS related processing, AV activation was significantly reduced (smaller amplitude ERPs and BOLD responses) compared to the preferred unisensory stimuli (VIS). Finally, multiple “higher-order” cortical regions, as well as late time ranges typically associated with more

higher-order processes showed significant or near significant trends of AV maximum activation (AV greater than either condition alone). Other regions/timeranges showed common activation in which AV and one or both conditions showed similar degrees of activation. Several of these regions and timeranges showed a right hemisphere bias.

### Discussion

Using an animated synthetic face and associated real human non-verbal vocalizations we elicited reliable fMRI activation to unisensory and multisensory stimulation in an imaging study designed to optimally image the STS/STG in its entirety, while still including face-sensitive and auditory regions in lateral sensory cortex. In a second group of subjects we elicited reproducible and consistent ERPs to the relatively long durations of the facial motion and associated vocalizations. Subjects were asked to detect unisensory target stimuli (a blink and an “mmm” sound) so

**Fig. 8** Summary of main fMRI and ERP findings in terms of time of occurrence relative to stimulus onset and type of multisensory phenomenon



that we could study audiovisual integration without a bias to a particular sensory modality. We discuss our results, summarized in Fig. 8, in the context of other audiovisual non-biological and speech integration studies, and in more general terms of social cognition.

#### Multisensory Effects in Sensory-Related Processes

Multisensory effects were seen in early regions (fMRI data) and ERP components which supported our hypotheses predicting facilitation effects. Interestingly, we also observed a unique early AV ERP component. This AV positivity peaking around 60–80 ms is similar to that seen in recent audiovisual integration studies (Giard and Perronet 1999; Shams et al. 2005). Somewhat surprisingly, these studies, which used *non-biological* stimuli, showed early AV integration effects whose laterality was opposite to ours. Giard and Perronet (1999) proposed that their early ERP response may stem from the recruitment of specific multisensory cells in or near striate cortex, where bisensory cells have been seen observed in animals (Fishman and Michael 1973; Morrell 1972). Due to our slice selection in our fMRI study, we could not confirm whether multisensory effects occurred in early visual cortex, however, such early effects in humans have been seen in other studies (Martuzzi et al. 2007).

Aside from the unique AV ERP signal, AUD and VIS-related sensory-related regions and ERP signals showed multisensory effects characteristic of facilitation, as predicted by our hypothesis. Significant effects were seen in VIS-related regions (LO) and ERP components (N170), and trends in the same direction were seen in fMRI activated AUD-related regions and ERP components (mid-MTG and N140). The strongest case of fMRI and ERP convergence was at mid-level visual processes, characterized by a right hemisphere bias and decrease in both amplitude (fMRI and ERP) and speed (ERP) of AV versus VIS.

The auditory trend towards facilitation was consistent with the role of the centrally located N140 in multisensory integration (Besle et al. 2004; Puce et al. 2007; van Wassenhove et al. 2005). A study by Puce et al. (2007) showed the largest N140s were elicited when a dynamic human

face (relative to house and primate face stimuli) was paired with incongruous sounds, suggesting that the context provided by a conspecific (human) face influences associated auditory processing. Additionally, when congruous sounds were presented, the N140 was largest to both human and primate faces when paired with species-appropriate vocalizations relative to a house stimulus whose front door opened with a creaking door sound (Puce et al. 2007). However, unlike previous ERP studies using speech stimuli (Besle et al. 2004; van Wassenhove et al. 2005), our results here did not reach significance, perhaps due to differences in timing of facial movements relative to vocalizations. In our paradigm, face movement and audio were simultaneous, although there was a natural delay in the movement peaking for our non-verbal stimuli (e.g. fully open mouth, upturned eyes in the sigh in Fig. 1), which is opposite to speech stimuli. Ghazanfar et al. (Ghazanfar et al. 2005) also have shown that timing plays a critical role in multisensory effects in an experiment in which monkeys were presented conspecific coos and grunts along with images of primate faces, however, in this experiment static faces of primates were utilized. Multisensory neurons in the auditory core and belt regions showed more enhancement when the face versus audio delay was less than 100 ms, and facilitation when the delay was greater than 200 ms. The variance of these studies based on timing underscores the importance of subtle audio versus visual onset time differences in multisensory processing.

Reduced activation in the AV condition could be due to various causes, including less energy demands brought about by facilitated processing. Alternatively, the relative decrease in amplitude in the multisensory relative to the unisensory conditions may be due a smaller population of neurons with exclusively multisensory versus unisensory preferences (Beauchamp 2005; Laurienti et al. 2005). Alternatively, the distribution of resources available to process these stimuli might be limited over multiple sensory cortical regions. We favor the increased efficiency explanation in the light of the behavioral facilitation effects observed in many studies (Bolognini et al. 2005; Gondan et al. 2005; Grant and Walden 1996; Miller 1982; Sumbly and Pollack 1954), although all three explanations are plausible and cannot be differentiated in the current

dataset. Multisensory optimization may also take the form of synchronization of the phase of oscillatory stimuli (such as gamma band activity) (Engel et al. 2001; Schroeder et al. 2008; Senkowski et al. 2008), which can produce important behavioral sequelae (Schroeder et al. 2008). Future studies quantifying ERP amplitudes and latencies, oscillatory activity and behavior in a combined manner might better clarify the underlying nature of these processes.

#### *Higher-Order Underadditive Effects*

We saw significant activation by all three conditions in putative higher-order cognitive processes (based on latencies (ERP data) and origins (fMRI data); Doeller et al. 2003; van Herten et al. 2006; Vuilleumier and Pourtois 2007). Here a clear case for convergence is more difficult to make since later ERPs are typically diffusely distributed, making source localization challenging, as they can potentially come from multiple sources (Siedenberg et al. 1996; Soltani and Knight 2000). However in our study, later ERPs, and activation in higher-order brain regions as revealed by fMRI showed spatially distributed responses and a combination of *AV maximum* and *Common activation* responses. From the fMRI side, this network of frontal, temporal and parietal regions has been implicated in other studies as playing an important role in multisensory perception (for review see Ghazanfar and Schroeder 2006), as well as for understanding speech and socially related stimuli (Calvert et al. 2001; Moll and de Oliveira-Souza 2007). For the ERPs, there were multiple distributed late peaks that showed an AV response with properties from both unisensory conditions (AV elicited larger amplitudes like the VIS condition and faster latencies like the AUD condition).

It is much more difficult to attribute higher-order activation as being specifically related to multisensory processing, as these regions did not show superadditivity, perhaps due to ceiling effects from these robust stimuli (Stevenson et al. 2007). In addition, higher order processes can be non-specific and can be very sensitive to other factors such as attention. The target stimuli were always unisensory, and therefore there were two possible unisensory targets in the AV blocks. It is possible that in the AV and VIS conditions responding to the corresponding unisensory target stimulus may have resulted in potentially greater stimulus-driven attentional effects. Having said that, the STS, IFG, and TPJ have shown potential multisensory behavior in other studies where the task requirements did not involve such contingencies (Calvert et al. 2001; Kawashima et al. 1999). Additionally, the blocked-event design in this study could conceivably have produced some refractoriness effects in the data, and in other studies (Calvert et al. 2001; Kawashima et al. 1999), albeit

unlikely. Ideally, an event-related design would circumvent these kinds of issues.

Notably, all three conditions in our study activated right pSTS, a region previously shown to be important in social-related multisensory processing (Redcay 2008). Right pSTS along with right TPJ, were the only regions in which there was maximum activation in the AV condition. Further evidence of pSTS importance in both multisensory and social processing come from prior research on a possible pSTS homologue in monkeys, the Anterior Superior Temporal Polysensory Area (STPa), which responds to visual biological motion, faces, and head and body view and direction (Jellema et al. 2000; Oram and Perrett 1996), and projects to higher order cognitive and emotional processing regions such as the amygdala and prefrontal regions (Oram and Perrett 1996).

In addition, we saw right-hemisphere dominated effects in STS and frontal regions, which is typically not seen in most speech related studies (Campbell et al. 2001; Capek et al. 2008; Hubbard et al. 2008; Kawashima et al. 1999; Macaluso et al. 2004; MacSweeney et al. 2000; Skipper et al. 2005). However, left lateralized activation has been less strongly observed in multisensory studies of simple speech, syllables, and emotional prosody (Kreifelts et al. 2007; Olson et al. 2002; Wright et al. 2003). It is possible that higher-order regions in both hemispheres have multisensory properties and are recruited based on verbal versus non-verbal relevance.

#### Summary

Both imaging modalities produced datasets that were very complex, yet there was a surprising degree of convergence between the ERP and fMRI data (Fig. 8). Underadditivity dominated the multisensory effects in earlier regions as supported by the significant (VIS areas) and trend towards (AUD) smaller and faster responses for AV versus unisensory stimuli. These data, along with previous behavioral studies, suggest that early or mid-sensory regions may be optimized to process multisensory stimuli, if information from multiple modalities is available (Foxy and Schroeder 2005). Multiple “higher-order” cortical regions, as well as late ERP activity typically associated with more higher-order processes showed underadditive effects driven by common activation for all conditions of non-verbal human stimuli, with a dominance of the AV condition in temporal regions. In particular the unique right pSTS effects confirm the important role of pSTS in social cognition, and again show the tendency toward right lateralization for social-related stimuli.

**Acknowledgements** We thank Mr. Kamal Shaikh for transforming each subject’s high-resolution SPGR to Talairach coordinates in

AFNI, Dr. Jim Thompson for discussions regarding image analysis, and Ms. Mary Pettit for administrative assistance. This study was funded by NIH Grant R01 NS049436, and the Radiology Department at West Virginia University.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Beauchamp MS (2005) Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics* 3:93–113
- Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat Neurosci* 7:1190–1192
- Besle J, Fort A, Delpuech C, Giard MH (2004) Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J NeuroSci* 20:2225–2234
- Bolognini N, Frassinetti F, Serino A, Ladavas E (2005) “Acoustical vision” of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Exp Brain Res* 160:273–282
- 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
- Campbell R, MacSweeney M, Surguladze S, Calvert G, McGuire P, Suckling J, Brammer MJ, David AS (2001) Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Brain Res Cogn Brain Res* 12:233–243
- Capek CM, Macsweeney M, Woll B, Waters D, McGuire PK, David AS, Brammer MJ, Campbell R (2008) Cortical circuits for silent speechreading in deaf and hearing people. *Neuropsychologia* 46:1233–1241
- Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173
- Doeller CF, Opitz B, Mecklinger A, Krick C, Reith W, Schroger E (2003) Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *Neuroimage* 20:1270–1282
- Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci* 2:704–716
- Fecteau S, Belin P, Joanne Y, Armony JL (2007) Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage* 36:480–487
- Fishman MC, Michael P (1973) Integration of auditory information in the cat’s visual cortex. *Vision Res* 13:1415–1419
- Foxe JJ, Schroeder CE (2005) The case for feedforward multisensory convergence during early cortical processing. *NeuroReport* 16:419–423
- Frith CD, Frith U (2007) Social cognition in humans. *Curr Biol* 17:R724–R732
- Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? *Trends Cogn Sci* 10:278–285
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25:5004–5012
- Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11:473–490
- Giard MH, Perrin F, Echallier JF, Thevenet M, Froment JC, Pernier J (1994) Dissociation of temporal and frontal components in the human auditory N1 wave: a scalp current density and dipole model analysis. *Electroencephalogr Clin Neurophysiol* 92:238–252
- Godey B, Schwartz D, de Graaf JB, Chauvel P, Liegeois-Chauvel C (2001) Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. *Clin Neurophysiol* 112:1850–1859
- Golarai G, Grill-Spector K, Reiss AL (2006) Autism and the development of face processing. *Clin Neurosci Res* 6:145–160
- Gondan M, Niederhaus B, Rosler F, Roder B (2005) Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. *Percept Psychophys* 67:713–726
- Grant KW, Walden BE (1996) Evaluating the articulation index for auditory-visual consonant recognition. *J Acoust Soc Am* 100:2415–2424
- Howell W (1900) *An American Text-book of Physiology*, 2nd edn. W.B. Saunders & Company, p 454. Digitized Oct 16 2007
- Hubbard AL, Wilson SM, Callan DE, Dapretto M (2008) Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Hum Brain Mapp* 30:1028–1037
- Jellema T, Baker CI, Wicker B, Perrett DI (2000) Neural representation for the perception of the intentionality of actions. *Brain Cogn* 44:280–302
- Kawashima R, Imaizumi S, Mori K, Okada K, Goto R, Kiritani S, Ogawa A, Fukuda H (1999) Selective visual and auditory attention toward utterances—a PET study. *Neuroimage* 10:209–215
- Kayser C, Petkov CI, Augath M, Logothetis NK (2007) Functional imaging reveals visual modulation of specific fields in auditory cortex. *J Neurosci* 27:1824–1835
- Kreifelts B, Ethofer T, Grodd W, Erb M, Wildgruber D (2007) Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage* 37:1445–1456
- Laurienti PJ, Perrault TJ, Stanford TR, Wallace MT, Stein BE (2005) On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Exp Brain Res* 166:289–297
- Lewis J, Puce A, Engel L (2008) Cortical processing of human vs. non-human categories of action sounds. *Neuroimage 14th meeting: human brain mapping conference, #694 Abstr#1229*
- Luyster RJ, Kadlec MB, Carter A, Tager-Flusberg H (2008) Language assessment and development in toddlers with autism spectrum disorders. *J Autism Dev Disord* 38:1426–1438
- 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
- MacLeod A, Summerfield Q (1987) Quantifying the contribution of vision to speech perception in noise. *Br J Audiol* 21:131–141
- MacSweeney M, Amaro E, Calvert GA, Campbell R, David AS, McGuire P, Williams SC, Woll B, Brammer MJ (2000) Silent speechreading in the absence of scanner noise: an event-related fMRI study. *NeuroReport* 11:1729–1733
- Martuzzi R, Murray MM, Michel CM, Thiran JP, Maeder PP, Clarke S, Meuli RA (2007) Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cereb Cortex* 17:1672–1679
- Miller J (1982) Divided attention: evidence for coactivation with redundant signals. *Cogn Psychol* 14:247–279
- Moll J, de Oliveira-Souza R (2007) Moral judgments, emotions and the utilitarian brain. *Trends Cogn Sci* 11:319–321
- Morrell F (1972) Visual system’s view of acoustic space. *Nature* 238:44–46

- Olson IR, Gatenby JC, Gore JC (2002) A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Brain Res Cogn Brain Res* 14:129–138
- Oram MW, Perrett DI (1996) Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J Neurophysiol* 76:109–129
- Ponton C, Eggermont JJ, Khosla D, Kwong B, Don M (2002) Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. *Clin Neurophysiol* 113:407–420
- Puce A, Smith A, Allison T (2000) ERPs evoked by viewing facial movements. *Cogn Neuropsychol* 17:221–239
- Puce A, Syngienotis A, Thompson JC, Abbott DF, Wheaton KJ, Castiello U (2003) The human temporal lobe integrates facial form and motion: evidence from fMRI and ERP studies. *Neuroimage* 19:861–869
- Puce A, Epling JA, Thompson JC, Carrick OK (2007) Neural responses elicited to face motion and vocalization pairings. *Neuropsychologia* 45:93–106
- Redcay E (2008) The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neurosci Biobehav Rev* 32:123–142
- Saint-Amour D, De Sanctis P, Molholm S, Ritter W, Foxe JJ (2007) Seeing voices: high-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia* 45:587–597
- Sarfati Y, Hardy-Bayle MC, Besche C, Widlocher D (1997) Attribution of intentions to others in people with schizophrenia: a non-verbal exploration with comic strips. *Schizophr Res* 25:199–209
- Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. *Trends Cogn Sci* 12:106–113
- Senkowski D, Schneider TR, Foxe JJ, Engel AK (2008) Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci* 31:401–409
- Shams L, Iwaki S, Chawla A, Bhattacharya J (2005) Early modulation of visual cortex by sound: an MEG study. *Neurosci Lett* 378:76–81
- Siedenberg R, Goodin DS, Aminoff MJ, Rowley HA, Roberts TP (1996) Comparison of late components in simultaneously recorded event-related electrical potentials and event-related magnetic fields. *Electroencephalogr Clin Neurophysiol* 99:191–197
- Skipper JJ, Nusbaum HC, Small SL (2005) Listening to talking faces: motor cortical activation during speech perception. *Neuroimage* 25:76–89
- Soltani M, Knight RT (2000) Neural origins of the P300. *Crit Rev Neurobiol* 14:199–224
- Stekelenburg JJ, Vroomen J (2007) Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci* 19:1964–1973
- Stevenson RA, Geoghegan ML, James TW (2007) Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Exp Brain Res* 179:85–95
- Sumby W, Pollack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212–215
- Troisi A, Spalletta G, Pasini A (1998) Non-verbal behaviour deficits in schizophrenia: an ethological study of drug-free patients. *Acta Psychiatr Scand* 97:109–115
- Van Essen DC (2005) A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *Neuroimage* 28:635–662
- Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH (2001) An integrated software suite for surface-based analyses of cerebral cortex. *J Am Med Inform Assoc* 8:443–459
- van Herten M, Chwilla DJ, Kolk HH (2006) When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *J Cogn Neurosci* 18:1181–1197
- van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci USA* 102:1181–1186
- Vuilleumier P, Pourtois G (2007) Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging. *Neuropsychologia* 45:174–194
- Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G (2003) Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb Cortex* 13:1034–1043