MEG demonstrates a supra-additive response to facial and vocal emotion in the right superior temporal sulcus

Cindy C. Hagan[a,b,1], Will Woods[b], Sam Johnson[b], Andrew J. Calder[c], Gary G. R. Green[a,b], and Andrew W. Young[a,b]

[a]Department of Psychology, University of York, Heslington YO10 5DD, United Kingdom; [b]York Neuroimaging Centre, University of York, The Biocentre, York Science Park, Heslington YO10 5DG, United Kingdom; and [c]MRC Cognition and Brain Sciences Unit, Cambridge CB2 7EF, United Kingdom

An influential neural model of face perception suggests that the posterior superior temporal sulcus (STS) is sensitive to those aspects of faces that produce transient visual changes, including facial expression. Other researchers note that recognition of expression involves multiple sensory modalities and suggest that the STS also may respond to crossmodal facial signals that change transiently. Indeed, many studies of audiovisual (AV) speech perception show STS involvement in AV speech integration. Here we examine whether these findings extend to AV emotion. We used magnetoencephalography to measure the neural responses of participants as they viewed and heard emotionally congruent fear and minimally congruent neutral face and voice stimuli. We demonstrate significant supra-additive responses (i.e., where AV > [unimodal auditory + unimodal visual]) in the posterior STS within the first 250 ms for emotionally congruent AV stimuli. These findings show a role for the STS in processing crossmodal emotive signals.

audio-visual emotion | emotional faces | emotional voices | fear | gamma

Some aspects of faces, such as identity, are relatively fixed whereas other aspects, such as facial expression, can change from moment to moment. A highly influential neural model of face perception suggests that the aspects of faces that change visibly (e.g., eye gaze, lip movement, facial expression) are processed by a neural pathway leading to the posterior superior temporal sulcus (STS) (1, 2). However, both lip movement and facial expression are generally accompanied by changing auditory signals. Indeed, in daily life emotion commonly is conveyed through multiple signals involving the face, voice, and body. Although originating from different sensory modalities, these transient emotive signals are perceived and integrated within our social interactions with seemingly minimal effort. It therefore is possible that recognition of emotional expression is an intrinsically crossmodal process and that the sensitivity of the posterior STS to the changeable aspects of faces may extend to voices, as the posterior STS could serve to integrate emotional signals arising from faces and voices (3). During emotion perception, the role of the posterior STS may extend beyond the visual function of perceiving facial emotion to serve a wider purpose related to the crossmodal integration of facial and vocal signals.

The STS is a prime candidate for the integration of visual and auditory emotive signals. The STS lies between primary auditory and visual cortices, and studies of audiovisual (AV) speech integration have demonstrated supra-additive responses in the STS to congruent speech stimuli (4, 5). Supra-additivity is a stringent criterion for multisensory integrative regions and is based on the known electrophysiological behavior of signal integration (6, 7). Supra-additivity occurs when the observed multisensory effect exceeds the sum of the unisensory components, i.e., when AV > [unimodal auditory (A) + unimodal visual (V)]. The purpose of the present study was to examine whether the supra-additivity observed by Calvert et al. (4) can be extended to AV nonverbal cues of emotion. We use supra-additivity to see if, akin to the findings observed in AV speech, a supra-additive response in STS is observed during the presentation of AV nonverbal cues of emotion.

Calvert et al. (4) have shown supra-additive responses in the STS to AV speech, but the time course over which these supra-additive responses occur remains unknown. We therefore used magnetoencephalography (MEG) to provide information about the neural time course, frequency content, and location of crossmodal emotion integration. Magnetic fields are not susceptible to spatial smearing as they pass through the scalp, and MEG therefore provides better spatial resolution than either EEG or event-related potential (ERP) techniques, with equal temporal resolution. We identified crossmodal regions in MEG using an MEG equivalent of the supra-additivity criterion proposed by Calvert et al. (4). The supra-additivity criterion has been criticized in functional MRI (fMRI) research because the transformation of neuronal responses into the blood oxygen level-dependent signal measured by fMRI is inherently nonlinear (8). MEG evades this criticism by measuring a direct index of neuronal activity (i.e., magnetic fields) rather than the signal dependent on the blood oxygen level. We understand that some have criticized the supra-additivity criterion as too stringent when measuring mixed neuronal populations (9), but it is precisely this stringency that appealed to us. We therefore chose to investigate the neural time course of supra-additive responses in the posterior STS.

The Haxby model (1, 2) suggests that the posterior STS responds to the changeable aspects of faces, and we do not dispute this suggestion. Rather, we ask whether the changeable facial signals analyzed by the STS could be crossmodal or are exclusively visual. We therefore used static facial photographs to minimize responses in the posterior STS that otherwise could be attributed to the transient facial changes made during emotional expression. The powerful nature of emotion integration is evident from studies showing that participants integrate emotional faces and voices despite a lack of temporal contiguity between the 2 sensory streams (10) and despite instructions to ignore 1 sensory modality and focus exclusively on the other (Experiments 2 and 3 in ref. 10). We paired static photographs of fearful faces with nonverbal vocal cues of fear (i.e., screams) to remain consistent with the methods used in previous behavioral and neuroimaging studies of crossmodal emotion integration (10–13) and to minimize possible laterality effects associated with AV speech (4, 5).
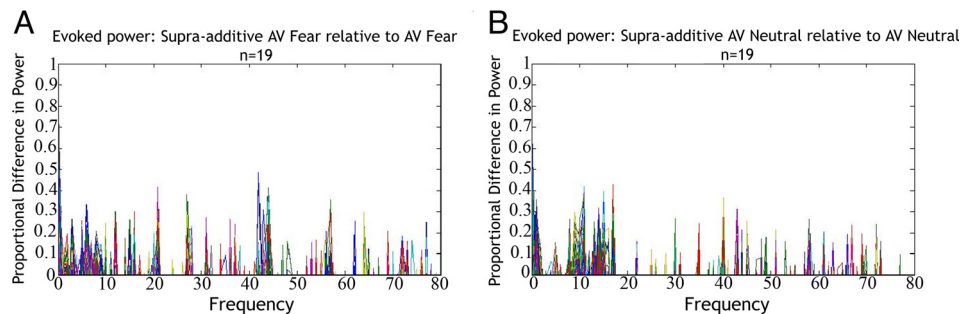

Fig. 1. Proportion of supra-additive difference in evoked power within each frequency bin (i.e., supra-additive AV/AV) at the group level for (*A*) fear and (*B*) neutral AV conditions. Colored lines represent the evoked response captured by a single sensor generated at the group level.

We used fearful expressions to examine the emotion circuit specified by Haxby and colleagues (1, 2) and because fear is an emotion with a clear evolutionary basis (14) that often is examined in neuroimaging paradigms (15, 16). We also presented neutral faces with minimally congruent neutral nonverbal vocal signals (i.e., polite coughs) to determine the extent to which integration mechanisms are engaged for facial–vocal pairings irrespective of congruence, because it has been suggested that facial and vocal integration is a mandatory process (10, 17). To test whether a supra-additive response would be observed in the STS during AV presentation of emotion, we presented participants with unimodal faces (V condition) and nonverbal vocal expressions (A condition). We also were interested in the time course and frequency content of the STS response, because previous studies demonstrate rapid response enhancements to AV emotion (13, 18, 19, 20), and recent reviews implicate the gamma frequency band (i.e., 30–80 Hz) in crossmodal integration (20).

Results

Using MEG, we continuously recorded 19 participants during facial, vocal, and facial + vocal trials. There were 96 trials in each of 6 conditions (AFear, VFear, AVFear, ANeutral, VNeutral, and AVNeutral), with each trial consisting of a fixation cross, a 700-ms stimulus (A, V, or AV), and an intertrial interval. Trials were randomized across participants.

MEG data were analyzed in 2 stages; first at the level of the sensors that surround the participant's head during the initial capturing of the data and then at the level of the brain sources that perpetuate the signals captured by the sensors surrounding the participant's head.

Sensor Level. We first analyzed the data at the sensor level to assess and identify the frequency content underlying supra-additive activity. Sensor data were analyzed for evoked activity in each condition for each participant and then for the group of participants. Evoked activity retains the phase-dependent information of the brain response and is important for identifying responses that are phase-locked to the onset of a presented stimulus. Only differences observed at the group level are reported here.

Evoked analysis was performed upon the group of participants for all 6 conditions using a 1,000-ms time window (300 ms before and 700 ms after stimulus onset). The averaged data across participants for every condition were subjected to a fast Fourier transform to examine the frequency content of the MEG signal. The data then were squared to ascertain the evoked power present in each condition for all 248 channels. A notch filter and Hanning window were applied to remove 50 Hz and 60 Hz components (U.K. and U.S. mains electricity) and to minimize spectral leakage, respectively.

The means and standard deviations of the evoked power within each condition were calculated. The mean evoked power spectra of both unimodal conditions were summed for all 248 sensor channels. This summed power then was subtracted from the mean evoked power spectrum of the respective AV condition to analyze the presence of supra-additivity in both neutral and fear AV conditions at the group level in accordance with Calvert's criterion (4). Although the theta frequency band elicited the largest supra-additive response, this result may have been caused by the disproportionate amount of power observed in the lower-frequency bands. To compare the relative contributions of each frequency bin to the supra-additive response, as opposed to the frequency bin delivering the largest supra-additive signal, the supra-additive differences in power then were divided by the power spectra of the respective AV condition. This procedure allowed us to examine the proportion of supra-additivity present in each frequency bin relative to the evoked power of the AV condition. Supra-additive responses were observed broadband up through 80 Hz (Fig. 1). We therefore chose first to apply a broadband filter (3–80 Hz) to our data at the source level, before filtering within individual bands of frequency.

Brain Source Level. Statistical parametric maps were generated from the supra-additive comparisons [i.e., AV − (A+V); see *Materials and Methods* for details about the procedure] to localize the neural source of supra-additive increases and decreases in power observed broadband (3–80 Hz). Beamforming techniques provide rich frequency information, so we repeated the beamforming procedure with spectral filters for frequency bands commonly used in clinical studies (theta = 4–8 Hz; alpha = 8–13 Hz; beta = 13–30 Hz; gamma = 30–80 Hz) to examine supra-additive increases and decreases in power within each frequency band (see *SI Text*). Statistical parametric maps were generated first at the level of the individual and then at the group level in Montreal Neurological Institute (MNI) (standard brain) space.

Our analysis was focused on brain areas of interest identified through previous studies of AV speech and emotion integration. Because we were interested primarily in supra-additive responses in the STS, and because the supra-additivity criterion is a stringent means by which to identify crossmodal neural activity, we chose not to adjust our criterion for significance further to account for family-wise error. Parametric t-tests were used to identify significant supra-additive increases and decreases in power for fear and neutral conditions within overlapping 500-ms time windows across the entire duration of the stimulus. Motivated by our a priori examination of supra-additivity, we detail findings representing supra-additive increases in power [AV > (A+V)] in the following sections; however, supra-additive decreases in power [AV < (A+V)] also were observed broadband and within different frequency bands. These data are presented in Table 1 and in the SI.

$$AVFear > (AFear + VFear)$$

Table 1. Coordinates in MNI space and associated peak t-scores showing the significant differences (1-tailed) in power observed broadband (3–80 Hz) for the main effects of audio-visual fear minus (auditory fear + visual fear) and audio-visual neutral minus (auditory neutral + visual neutral)

Brain regions	BA	P-value	T-scores	Coordinates x	y	z
Broadband AVFear versus (AFear + VFear) 0–500 ms						
L anterior cingulate cortex, L superior frontal gyrus	10	<0.0005	4.18	−10	24	18
R inferior parietal lobule, R superior temporal gyrus, R superior temporal sulcus, R insula	40/22	<0.0005	4.09	70	−46	24
R postcentral gyrus, R thalamus	3	<0.01	2.74	40	−16	24
R middle temporal gyrus	21	<0.01	2.65	70	−10	−16
L precentral gyrus	6	<0.01	2.44	−40	−6	38
R superior frontal gyrus, R medial frontal gyrus	8	<0.025	2.33	14	44	44
R precentral gyrus, R thalamus, R caudate	6	<0.05	2.02	54	−10	54
R middle grontal gyrus	10	<0.05	1.82	30	70	14
L postcentral gyrus	1	<0.05	1.75	−66	−20	38
L middle temporal gyrus	21	<0.05	−2.02	−68	−26	−16
Broadband AVNeutral versus (ANeutral + VNeutral) 0–500 ms						
R precuneus	7	<0.01	−2.72	24	−56	54
R inferior parietal lobule	40	<0.01	−2.69	34	−46	48
R inferior parietal lobule	40	<0.01	−2.66	40	−46	58
L thalamus	-	<0.025	−2.13	−14	−38	16
L precuneus	7	<0.05	−2.00	−20	−50	48
L thalamus	-	<0.05	−1.90	−6	−20	16
R parahippocampal gyrus, R hippocampus	-	<0.05	−1.88	40	−10	−22
L supramarginal gyrus	40	<0.05	−1.74	−50	−46	34
L inferior parietal lobule	40	<0.05	−1.74	−56	−36	34
L inferior parietal lobule	40	<0.05	−1.73	−46	−36	28

Positive t-scores reflect significant increases in power; negative t-scores reflect significant decreases in power.
BA = Brodmann area; L = left; R = right.

Broadband (3–80 Hz). Clusters of significant supra-additive increases in power for the fear condition were observed in broadband analyses. One cluster extended from the right posterior to anterior STS [Brodmann area (BA) 22] with activation spreading to encompass the right insula (Fig. 2) and peaking in the right inferior parietal lobule (BA 40). Other regions of significant supra-additive increases in power observed broadband are reported in Table 1.

$$AVNeutral > (ANeutral + VNeutral)$$

Broadband (3–80 Hz). No significant clusters of activity representing significant supra-additive increases in power for the neutral condition were observed in broadband analyses (Table 1).

Discussion

The aim of this study was to examine the presence of supra-additive responses in the STS during conditions of crossmodal presentation of emotion. Here we demonstrate a supra-additive response for crossmodal emotion stimuli using electrophysiological techniques and show evoked supra-additive broadband activity at the sensor level for emotionally congruent crossmodal (facial and vocal) stimuli. Our source-space analyses demonstrated that supra-additive broadband activation was observed in the STS within the first 250 ms of AVFear stimulus presentation. These data therefore confirm that the contribution of the STS to the perception of facial emotion is related to its wider role in crossmodal integration, consistent with the suggestion put forward by Calder and Young (3).

| 0–500ms | 50–550ms | 100–600ms | 150–650ms | 200–700ms | 250–750ms | 300–800ms |
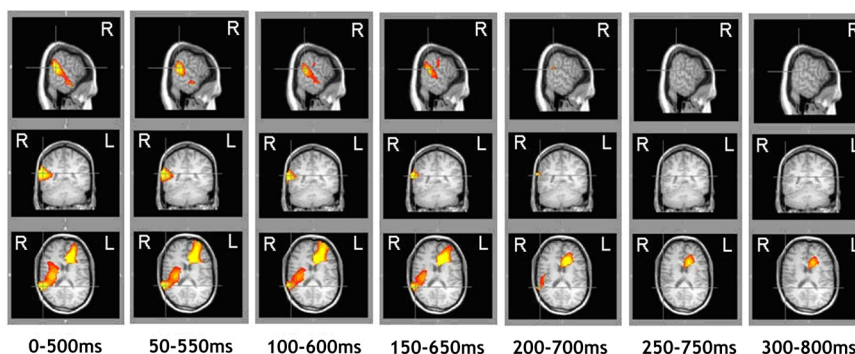
Fig. 2. Supra-additive increase in power observed broadband across the entire duration of the AVFear stimulus (i.e., from 0 to 800 ms, stimulus onset through stimulus offset). Each column depicts the supra-additive response observed in a 500-ms time window beginning at 0 ms, with time windows of subsequent columns increasing in increments of 50 ms up to 300 ms. Crosshairs are placed at the MNI coordinate in the right STG (60, −46, 18) with activation thresholded from $P < 0.05$ to $P < 0.005$. T-values observed at the coordinate are as follows: 0–500 ms = 3.43; 50–550 ms = 3.19; 100–600 ms = 2.64; 150–650 ms = 2.79; 200–700 ms = 2.28; 250–750 ms = 0; 300–800 ms = 0.

Although our results are in line with previous fMRI and EEG studies of AV emotion integration (8, 13, 21), the neural response to crossmodal cues of fear may be a special example of crossmodal emotion integration under potentially dangerous situations. Our neutral minimally congruent condition elicited subthreshold (i.e., not significantly supra-additive) broadband activation in regions adjacent to the right STS. Broadband responses in the STS occurred only for voxels showing decreases in power to visual stimuli and increases in power to auditory stimuli when compared independently with the fixation cross baseline. When broadband activity was broken down into the frequency bands typically used in clinical studies, supra-additive gamma activity (30–80 Hz) contributed most to the STS broadband response; supra-additive gamma activation occurred in the right STS and temporal regions for both fear and neutral conditions (see SI).

Our study paired emotionally congruent static photographs of fearful faces with temporally fluctuating (i.e., asynchronous) fearful sounds and observed supra-additive broadband activation in the right STS. In line with the view of Haxby and colleagues (1, 2) that the STS is involved in analyzing transiently changing facial signals, research suggests that the posterior STS is involved in the perception of both overt (22) and implied (23, 24) biological motion. This suggestion is supported by two recent fMRI studies of crossmodal emotion integration; one study showed significant enhancements in the left posterior STS for temporally asynchronous emotional faces and voices (8), and the other study showed that activation in the bilateral STG correlated positively with behavioral improvements in the identification of temporally synchronous emotional facial and vocal displays (21). Although it was not tested explicitly in the present paradigm, our findings are consistent with the notion that the simultaneous presentation of a congruent vocal sound may be sufficient to imply motion to a static face; further studies that address this question directly are warranted.

We also compared the time course of the fear supra-additive response with the time courses of the responses observed during the individual unimodal conditions. The response to the unimodal auditory stimulus occurred within the first 150 ms, whereas the response to the unimodal visual stimulus occurred within the first 300 ms. A recent study in monkeys has shown that facial and vocal integration is mediated through interactions between the STS and the auditory cortex (25). The supra-additive response in the human posterior STS therefore could arise through interactions with the human auditory cortex, similar to the crossmodal integration process observed in monkeys. Given that crossmodal integration occurs during the first 250 ms, it is possible that the auditory response facilitates the visual response during crossmodal processing.

Our finding that supra-additivity occurs within the first 250 ms of crossmodal emotion presentation demonstrates the rapidity with which facial and vocal fearful emotion is integrated. Similar results were reported in an ERP study that showed AV emotion elicits significant enhancements in the auditory N1 component as early as 110 ms poststimulus when congruent AV signals are compared independently with incongruent AV signals (13). Indeed, our broadband results achieved the highest levels of significance within the first 100 ms of stimulus presentation. Some suggest that the rapid timescale of crossmodal emotion integration reflects the mandatory nature of the integration of facial and vocal stimuli (18). The mandatory nature of facial and vocal emotion integration is evident in the behavioral studies of crossmodal emotion integration from which the paradigm used here was adopted (10). That integration can be achieved under unnatural crossmodal conditions, such as static faces and temporally fluctu-

ating voices, has been suggested speak to the powerful nature of the effect (10).

MEG studies are limited by the requirement that a large number of trials must be presented for each condition; this requirement correspondingly limits the number of conditions in any given experiment. From the current experiment, it is not possible to discern whether the results presented here are specific to the emotion of fear or whether these results extend to other high-priority emotions, such as anger. It also is possible that a supra-additive response in the STS will be observed for all crossmodally presented emotions, including nonpriority emotions such as happiness and sadness. A final possibility is that supra-additive responses in the STS are elicited only when the nonverbal crossmodal signals perceived are sufficiently congruent. Further crossmodal studies therefore are required to establish the merit of any one or all three of these possibilities. All three possibilities suggest the involvement of the right STS in some aspect of the crossmodal integration of nonverbal facial and vocal signals.

In conclusion, we demonstrate that congruent facial and vocal signals of emotion elicit supra-additive broadband activation in the right posterior STS within the first 250 ms of crossmodal stimulus presentation. Subthreshold supra-additive activity was observed in posterior temporal regions when minimally congruent facial and vocal cues were presented. For both fear and neutral conditions, supra-additive activity in the gamma frequency band—a band previously implicated in sensory integration—was observed to correspond best with the supra-additive broadband response observed. These findings therefore demonstrate a role for the right posterior STS in the integration of facial and vocal signals of emotion and are consistent with the integration hypothesis of the STS proposed by Calder and Young (3). We interpret these findings as suggesting a wider role for the posterior STS in face perception, a role that involves the integration of rapidly changing signals from faces and voices. The most parsimonious conclusion from the present data is that the posterior STS is sensitive to nonverbal expressions arising from faces and voices. Our hope is that these data provide a critical insight into the nature of the role played by the posterior STS as it relates to face perception.

Materials and Methods

Participants. We recruited 28 healthy participants from the University of York. We excluded 9 participants because of scanner problems, excessive head movement, or the presence of electrical noise in the data, leaving a total of 19 participants [10 male; mean age. = 24.44 (SD 4.23) years, range = 19.22–33.41 years] for inclusion in the final data analysis. All participants were right-handed, with normal hearing and with normal or corrected-to-normal vision. Participants were without history of neurological injury and were offered a small stipend for study participation. Ethical approval was granted jointly by the York University Department of Psychology and the York Neuroimaging Centre.

Experimental Stimuli. *Visual stimuli.* Fearful and neutral facial expressions of 2 male and 2 female actors (JJ, WF, MF, SW) from the Ekman and Friesen Facial Affect series (26) were used. Expressions were selected based on similarities in facial action unit scores and high recognition ratings for fearful and neutral expressions (26). One of the female Ekman fear faces (MF) was caricatured by 50% to produce a more discernible fearful face without affecting the facial action unit score. The hair was removed from each face, and the faces were presented in greyscale to minimize contrast differences between stimuli.
Auditory stimuli. We used 4 nonverbal vocal expressions of fear (i.e., screams) and neutral vocal sounds (i.e., polite coughs). Sound clips were obtained from professional sound-effects internet sites or were created using acting volunteers from the Department of Psychology at the University of York and the Cognition and Brain Sciences Unit, Cambridge.
Audio-visual stimuli. Each face was paired with a fear or neutral vocal sound to make expression-matched pairings of faces and voices. The face–voice pairings were gender congruent, and each emotive sound was paired consistently with a single actor.

Design and Procedure. *Experimental paradigm.* Participants were asked to attend to the voice and the face of each actor while trying to maintain central fixation during each trial. A task-irrelevant response was included whereby participants were asked to identify and report when the letter 'B' or 'R' appeared at the center of the screen after a select number of trials (96 in total, pseudorandomly chosen) by using 2 buttons on an ergonomic response box. This subsidiary task was included to ensure that participants remained attentive to, and centrally fixated upon, each stimulus and provided a way to monitor performance.

Experimental trials. All trials began with a black fixation cross, 3 × 3 cm in size, presented for 500 ms in the center of the screen against a solid gray background. Next a visual, auditory, or AV stimulus appeared in the center of the screen for 700 ms. Immediately following stimulus offset, a solid gray screen appeared for 1,300 ms.

There were 96 presentations of each condition: 192 unimodal voice trials (96 fear, 96 neutral); 192 unimodal face trials (96 fear, 96 neutral); and bimodal face–voice trials in which the face was congruent with the simultaneously presented vocal expression (192 bimodal trials; 96 fear congruent, 96 neutral congruent). All stimuli were presented using Presentation 2005 software (Neurobehavioral Systems). Trials were randomized across participants to minimize habituation effects. The experiment was presented in 3 runs of equal duration lasting ≈10 min each. To minimize fatigue, participants were allowed short breaks of a few seconds between runs.

Response trials. We included 96 additional trials in which a response was requested from the participant. On these occasions, either the letter 'B' or the letter 'R' appeared at the center of the screen directly after stimulus offset. The letter remained on the screen for 250 ms and was followed by a solid gray screen 1,050 ms in duration directly after letter offset. Response trials were followed by a ''dummy'' trial (16 from each condition, pseudorandomly chosen and counterbalanced across conditions). Dummy trials were discarded from the overall analysis of the data because of potential motor response contamination.

Data Acquisition. *Magnetoencephalography.* MEG data were acquired at the York Neuroimaging Centre using a 248-channel Magnes 3600 whole-scalp recording system (4-D Neuroimaging) with superconducting quantum interference device-based first-order magnetometer sensors. A Polhemus stylus digitizer (Polhemus Isotrak) was used to digitize each participant's head, nose, and eye orbit shapes before data acquisition to facilitate accurate co-registration with MRI data. Coils were placed in front of the left and right ears and at three equally spaced locations across the forehead to monitor head position prior to and following data acquisition. Data from four participants with head movement values of 0.75 cm or greater at two or more coils were excluded from both sensor and source-space analyses.

Participants were seated during the experiment. Magnetic brain activity was digitized continuously across the three runs. Images were projected onto a screen at a viewing distance of ≈70 cm and subtending a viewing angle of 8° for faces and 0.3° for letters. Faces were presented in small size (5 × 9 cm) to help minimize participant eye saccades. During response trials, the letters displayed were 5 mm × 1 cm in size to ensure that central fixation was maintained throughout each stimulus presentation. Auditory stimuli were presented at a comfortably audible level via Etymotic Research ER30 earphones. Participants were monitored throughout the scan using a video camera situated in a magnetic shielded room.

For all participants data were online filtered with a direct current (DC) filter and were sampled at a rate of 678.17 Hz (bandwidth 200 Hz).

MRI. Standard structural MRI scans were obtained for co-registration with MEG. Images on a 3-T scanner (HD Excite; General Electric) using a whole-head coil (8-channel high T-resolution brain array). The scanner has a 3-T 60-cm magnet. To maximize magnetic field homogeneity, an automatic shim was applied before scanning. Using an IR-prepared fast spoiled gradient recalled pulse sequence (repetition time = 6.6 ms, echo time = 2.8 ms, flip angle = 20°, and an inversion time of 450 ms), we imaged 176 1-mm-thick 3-D sagittal slices parallel to midline structures covering the whole brain. The field of view was 290 × 290 mm, and the matrix size was 256 × 256, giving an in-plane spatial resolution of 1.13 mm.

Localizer and calibration scans were performed before performing a high-resolution T1 volume with voxel dimensions of 1 × 1.13 × 1.13 mm. For better elimination of distortion and improved co-registration of MRI and MEG data, 3-D gradient warping corrections and edge-enhancement filters were applied.

Analysis of Imaging Data. *Sensor level analysis.* MEG data were cleaned of artifacts and then were analyzed on an epoch-by-epoch (e.g., trial-by-trial)

basis in relation to the sensors surrounding the participant's head. Faulty sensors and epochs containing swallow, eye saccade, blink, and electrical noise artifacts within a 1,200-ms time window (500 ms before stimulus onset and 700 ms after stimulus onset) were identified. We excluded 2 participants with greater than 30% of unusable trials from further analyses (≈200 or more epochs were rejected). Before analyses were performed at the sensor level, artifact-contaminated epochs were removed, and faulty sensors were zeroed. Differences in overall DC level among sensors were removed.

Source level analysis. MEG data were co-registered with each participant's structural MRI scan first by matching the surface of the head digitization maps to the surface of the structural MRI (27). A spatial filter (''beamformer'') technique was applied to localize sources related to each task (28). A 1,300-ms time window (500 ms prestimulus plus 800 ms through stimulus duration) was used to create the covariance matrix that determined the spatial filter properties of the beamformer. For each participant, spatial filers were obtained for each isolated point on a grid covering the volume of the brain. The spatial filters estimate the current source contribution of a single point within the brain grid volume, independent of all other points on the grid (29). The power at each grid point for specified frequency bands was calculated for every epoch of all conditions and represented by a neuromagnetic activity index (NAI). This procedure was performed to provide maps of activity corresponding to each trial for every condition. Statistical parametric contrasts of these maps were created between conditions and were used to localize sources of activity.

The spatial and temporal dynamics of the broadband response were analyzed using a beamformer, which captures both evoked and induced activity (i.e., total power). The broadband response was designated as the neuromagnetic activity between 3 and 80 Hz and was selected based upon the results observed at the sensor level. Frequencies lower than 3 Hz were not included in our broadband analyses because insufficient numbers of waveform cycles were represented in the 500-ms time window of our beamforming analyses. To minimize the contribution of task-irrelevant oscillatory activity, the NAI of a ''passive'' time window was subtracted from the NAI of an ''active'' time window to produce an ''active minus passive'' contrast for all conditions. The passive period was designated as neuromagnetic activity occurring during the 500-ms prestimulus period of each stimulus. The active period was designated as the neuromagnetic activity occurring during each stimulus presentation across a moving 500-ms time window. As such, active periods were either from 0–500 ms, 50–550 ms, 100–600 ms, 150–650 ms, 200–700 ms, 250–750 ms, or 300–800 ms after stimulus onset. A 600-ms temporal ''buffer'' was applied to both passive and active time windows to eliminate edge effects when filtering. Epochs from unimodal A and V contrasts were paired according to their respective temporal order of presentation and then were summed. This procedure generated A+V epochs and provided an estimate of variance across epochs. Parametric statistics were used to assess the presence of supra-additive differences in power between the generated A+V epochs and the respective AV epochs.

Masking procedure. An additional criterion for integration sites was used to minimize supra-additive responses observed in unisensory regions [consistent with Calvert's method (4)], such that activations in supra-additive areas were required to respond, at least minimally, during both unimodal conditions. Beamforming approaches reveal both increases and decreases in power, yielding four possible alternatives that equally satisfy the criterion of minimal activity in both A and V conditions. Voxels displaying supra-additivity could produce (*i*) increases in power in both unimodal conditions such that t-values greater than 0 are observed in both A and V conditions, (*ii*) decreases in power in both unimodal conditions such that t-values less than 0 are observed in both A and V conditions, (*iii*) increases in power in the A condition with t-values greater than 0 and decreases in power in the V condition with t-values less than 0, or (*iv*) increases in power in the V condition with t-values greater than 0 and decreases in power in the A condition with t-values less than 0.

The primary motivation was to minimize supra-additive responses in unisensory areas. The response in the primary visual cortex was observed as a decrease in power for the V condition, whereas the response in the primary auditory cortex was observed as an increase in power for the A condition when compared independently with the fixation cross baseline. Therefore mask alternative 3 (detailed above) was adopted and beamformed to localize voxels meeting the two criteria for A and V conditions. When both criteria were not met, the voxel's contribution to the overall calculation of statistical power was removed.

1. Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. *Trends in Cognitive Sciences* 4:223–233.

2. Haxby JV, Hoffman EA, Gobbini MI (2002) Human neural systems for face recognition and social communication. *Biol Psychiatry* 51:59–67.

3. Calder AJ, Young AW (2005) Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience* 6:641–651.
4. 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.
5. 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.
6. Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. *Science* 389–391.
7. Stein BE, Jiang W, Stanford TR (2004) in *The Handbook of Multisensory Processes*, ed. Calvert G, Spence C, Stein BE (MIT Press, Cambridge, MA) pp. 243–264.
8. Ethofer T, Pourtois G, Wildgruber D (2006) Investigating audiovisual integration of emotional signals in the human brain. *Prog Brain Res*, 156:345–361.
9. Beauchamp MS (2005) See me, hear me, touch me: Multisensory integration in lateral occipitaltemporal cortex. *Curr Opin Neurobiol* 15:145–153.
10. De Gelder B, Vroomen J (2000) The perception of emotions by ear and by eye. *Cognition and Emotion* 14:289–311.
11. Dolan RJ, Morris JS, de Gelder B (2001) Crossmodal binding of fear in voice and face. *Proc Natl Acad Sci USA* 98:10006–10010.
12. Pourtois G, de Gelder B, Bol A, Crommelinck M (2005) Perception of facial expressions and voices and of their combination in the human brain. *Cortex* 41:49–59.
13. Pourtois G, de Gelder B, Vroomen J, Rossion B, Crommelinck M (2000) The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport* 11:1329–1333.
14. Darwin, C (1998) *The Expression of the Emotions in Man and Animals*, ed. Ekman, P. (HarperCollinsPublishers, London).
15. Morris JS, Scott SK, Dolan RJ (1999) Saying it with feeling: Neural responses to emotional vocalizations. *Neuropsychologia* 37:1155–1163.
16. Morris JS, et al. (1996) A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 383:812–815.
17. De Gelder B, Pourtois G, Vroomen J, Bachoud-Levi AC (2000) Covert processing of faces in prosopagnosia is restricted to facial expressions: Evidence from cross-modal bias. *Brain Cognit* 44:425–444.
18. De Gelder B, Böcker KB, Tuomainen J, Hensen M, Vroomen J (1999) The combined perception of emotion from voice and face: Early interaction revealed by human electric brain responses. *Neurosci Lett* 260:133–136.
19. Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *J Cognit Neurosci* 11:473–490.
20. Senkowski D, Schneider T, Foxe J, Engel A (2008) Crossmodal binding through neural coherence: Implications for multisensory processing. *Trends in Neurosciences* 31:401–409.
21. 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.
22. Frith CD, Frith U (1999) Interacting minds—a biological basis. *Science* 286:1692–1695.
23. Kourtzi Z, Kanwisher N (2000) Activation of human MT/MST by static images with implied motion. *J Neurosci* 12:48–55.
24. Senior C, et al. (2000) The functional neuroanatomy of implicit-motion perception or representational momentum. *Curr Biol* 10:16–22.
25. Ghazanfar AA, Chandrasekaran C, Logothetis NK (2008) Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *J Neurosci* 28:4457–4469.
26. Young A, Perrett D, Calder A, Sprengelmeyer R, Ekman P (2002) Facial expressions of emotion: Stimuli and tests (FEEST). (Thames Valley Test Company, Bury St. Edmunds).
27. Kozinska D, Carducci F, Nowinski K (2001) Automatic alignment of EEG/MEG and MRI data sets. *Clinical Neurophysiology* 112:1553–1561.
28. Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Bio-Medical Engineering* 44:867–880.
29. Huang MX, et al. (2004) Commonalities and differences among vectorized beamformers in electromagnetic source imaging. *Brain Topography* 16:139–158.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

NEUROSCIENCE