

# *Inverse Effectiveness and Multisensory Interactions in Visual Event-Related Potentials with Audiovisual Speech*

**Ryan A. Stevenson, Maxim Bushmakin, Sunah Kim, Mark T. Wallace, Aina Puce & Thomas W. James**

## **Brain Topography**

A Journal of Cerebral Function and Dynamics

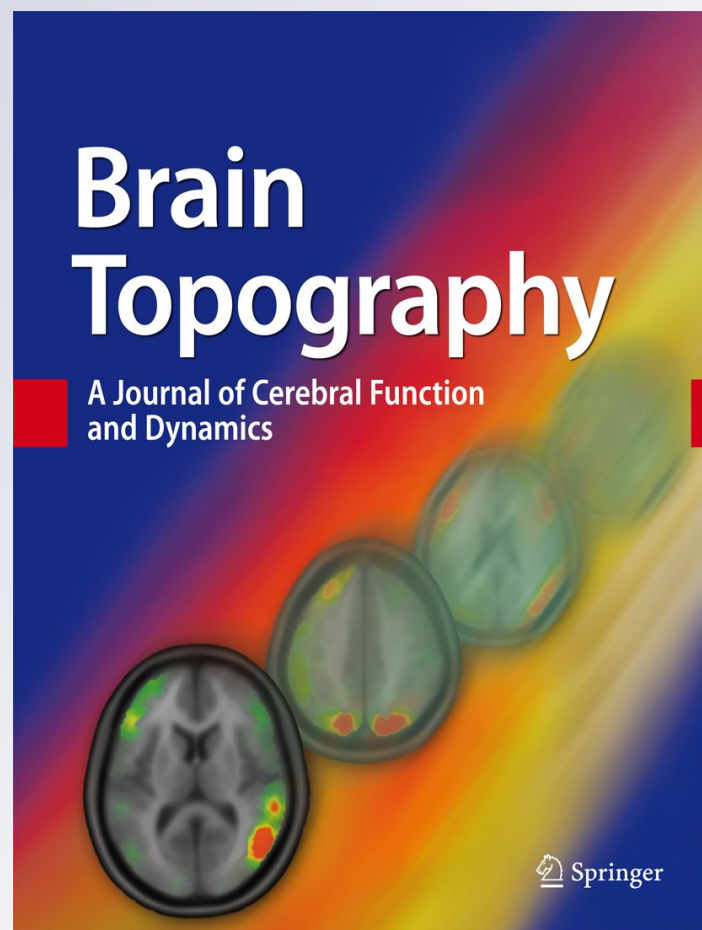
ISSN 0896-0267

Volume 25

Number 3

Brain Topogr (2012) 25:308-326

DOI 10.1007/s10548-012-0220-7



**Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.**

# Inverse Effectiveness and Multisensory Interactions in Visual Event-Related Potentials with Audiovisual Speech

Ryan A. Stevenson · Maxim Bushmakin ·  
Sunah Kim · Mark T. Wallace · Aina Puce ·  
Thomas W. James

Received: 16 August 2011 / Accepted: 5 February 2012 / Published online: 25 February 2012  
© Springer Science+Business Media, LLC 2012

**Abstract** In recent years, it has become evident that neural responses previously considered to be unisensory can be modulated by sensory input from other modalities. In this regard, visual neural activity elicited to viewing a face is strongly influenced by concurrent incoming auditory information, particularly speech. Here, we applied an additive-factors paradigm aimed at quantifying the impact that auditory speech has on visual event-related potentials (ERPs) elicited to visual speech. These multisensory interactions were measured across parametrically varied stimulus salience, quantified in terms of signal to noise, to provide novel insights into the neural mechanisms of audiovisual speech perception. First, we measured a monotonic increase of the amplitude of the visual P1-N1-P2 ERP complex during a spoken-word recognition task with increases in stimulus salience. ERP component

amplitudes varied directly with stimulus salience for visual, audiovisual, and summed unisensory recordings. Second, we measured changes in multisensory gain across salience levels. During audiovisual speech, the P1 and P1-N1 components exhibited less multisensory gain relative to the summed unisensory components with reduced salience, while N1-P2 amplitude exhibited greater multisensory gain as salience was reduced, consistent with the principle of inverse effectiveness. The amplitude interactions were correlated with behavioral measures of multisensory gain across salience levels as measured by response times, suggesting that change in multisensory gain associated with unisensory salience modulations reflects an increased efficiency of visual speech processing.

**Keywords** Multisensory integration · Inverse effectiveness · ERPs · P1-N1-P2 · N170 · Speech perception · Face perception

R. A. Stevenson · M. Bushmakin · A. Puce · T. W. James  
Department of Psychological and Brain Sciences,  
Indiana University, Bloomington, IN, USA

R. A. Stevenson · S. Kim · A. Puce · T. W. James  
Program in Neuroscience, Indiana University,  
Bloomington, IN, USA

R. A. Stevenson (✉) · M. T. Wallace  
Department of Hearing and Speech Sciences, Vanderbilt  
University Medical Center, MRB 3, Room 7110,  
Nashville, TN 37232, USA  
e-mail: ryan.andrew.stevenson@gmail.com

M. Bushmakin · S. Kim · A. Puce · T. W. James  
Cognitive Science Program, Indiana University,  
Bloomington, IN, USA

S. Kim  
Vision Science Program, School of Optometry,  
University of California, Berkeley, CA, USA

## Introduction

In our everyday interactions with others, face-to-face communication relies on evaluating the incoming audiovisual speech of the other individual. Speech perception is an inherently multisensory process. This can be demonstrated behaviorally by evaluating the effect of visual speech, meaning the oral and facial articulations associated with the auditory utterance, on the ability of a listener to perceive an utterance correctly. Visual speech increases the ability of a listener to perceive an utterance correctly (Sumby and Pollack 1954), increases the speed at which a listener perceives an utterance (Soto-Faraco et al. 2004), and can even change the perception of the utterance (McGurk and MacDonald 1976). These modulations in

speech perception found with audiovisual speech relative to auditory- and visual-only speech are driven by interactions of visual and auditory neural processing mechanisms. Audiovisual interactions at the individual-neuron level have been characterized most thoroughly in the superior colliculus of animal models, a midbrain structure involved in orienting responses to auditory, visual, and somatosensory stimuli (Meredith and Stein 1983, 1986; Wallace et al. 1992, 1993, 1996, 1998, 1996). At the level of the single neuron, neural responses exhibit interactions that mirror behavioral interactions (Stein et al. 1988). Similar to what has been shown in both behavioral accuracy (Lovelace et al. 2003; Nelson et al. 1998; Stein and Wallace 1996; Wilkinson et al. 1996) and response time (RT; Diederich and Colonius 2004; Hershenson 1962), the most commonly found interaction is multisensory gain, where the response of an individual cell with audiovisual stimulation exceeds the response predicted by independent unisensory stimulations (in the case of action potentials measured from a single neuron, the maximum response to a unisensory presentation; Meredith and Stein 1983; Stein et al. 1988).

Multisensory gain in both neuronal responses and behavior has been consistently found to depend on low-level stimulus factors, such as stimulus efficacy, spatial congruency, and temporal coincidence (for review, see Stein and Meredith 1993). Specifically, lower stimulus efficacy (usually controlled through stimulus salience quantified in terms of signal-to-noise ratio) produces greater performance benefits when stimuli are presented in both sensory modalities, an interaction known as inverse effectiveness (Meredith and Stein 1986; Stein and Meredith 1993). Inverse effectiveness is also found in the spike counts of individual neurons, with low-salience stimuli producing greater multisensory gain than high-salience stimuli (Meredith and Stein 1983; Perrault et al. 2003; Stanford et al. 2005; Stein et al. 2009; Wallace et al. 1996, 1998).

Multisensory interactions have also been documented in non-invasive, population-level neural responses of humans. Specifically, electroencephalographic (EEG) and magnetoencephalographic (MEG) studies have provided evidence for a wide range of multisensory interactions. Recent findings have utilized the temporal precision of averaged evoked electrical and magnetic neural responses to show that processes thought to be solely unisensory may in fact be multisensory at least in a limited sense. These include but are not limited to visual evoked potentials to faces (Brefczynski-Lewis et al. 2009; Joassin et al. 2004; Klucharev et al. 2003; Latinus et al. 2010; Puce et al. 2007) and objects (Fort et al. 2002a; Giard and Peronnet 1999; Molholm et al. 2002, 2004). The visual evoked potential is typically composed of a triphasic complex referred to as P1-N1-P2, is elicited with a wide range of visual stimuli

(Allison et al. 1999; Bentin et al. 1996), is commonly recorded as a posterior occipito-temporal scalp potential, and has been related to invasive studies of neural activity in striate and extrastriate visual cortices, lateral occipito-temporal regions (Allison et al. 1999), middle temporal gyrus (Joassin et al. 2004), fusiform gyrus (FG) and inferior temporal cortex (ITC; Allison et al. 1994a, b; Joassin et al. 2004).

While the visual evoked potential, as the name implies, has been described as a response to visual stimuli, recent evidence has suggested that many regions that were once thought to be solely unisensory receive input from other sensory modalities (Calvert et al. 1997; Haxby et al. 1994; Kawashima et al. 1995; Laurienti et al. 2002; Macaluso et al. 2000; Martuzzi et al. 2007; Molholm et al. 2002; Watkins et al. 2006, 2007). For example, direct cortical projections between primary auditory and visual cortices have now been documented in the primate brain (Cappe et al. 2009; Clavagnier et al. 2004; Falchier et al. 2002, 2010; Musacchia and Schroeder 2009; Rockland and Ojima 2003; Smiley and Falchier 2009). In addition to these anatomical connections, the P1-N1 complex, often referred to as the P100 and N170, has shown modulations with the addition of an auditory stimulus. For example, face and voice pairings have shown a subadditive response (Brefczynski-Lewis et al. 2009; Joassin et al. 2004; Klucharev et al. 2003; van Wassenhove et al. 2005), and a smaller N170 when auditory and visual stimuli are incongruent (Puce et al. 2007; Stekelenburg and Vroomen 2007; for a discussion of subadditive effects, see Vroomen and Stekelenburg 2010). Also, a recent study recording from posterior occipital electrodes typically associated with unisensory visual responses identified an early (40–60 ms) interaction with simple, non-speech stimuli at low-intensity but not high-intensity, where responses became more subadditive with lower stimulus intensities (Senkowski et al. 2011). The directionality of voltage waveforms exhibiting these effects cannot be disambiguated (Murray et al. 2008), however, the presence of interactions across stimulus-intensity levels itself provides evidence for integration. While the impact that visual speech has on auditory comprehension and accuracy has been rather well documented, much less is known about the converse relationship—specifically the impact that auditory speech has on the processing of visual speech tokens and faces. Consequently, the current study seeks to examine this issue.

In the current study we investigate the impact that stimulus salience of audiovisual speech has on the visual-evoked potential. To do this, we employ an additive-factors experimental design that is novel to the methods of EEG and MEG. This paradigm measures changes in interactions between two cognitive processes by parametrically varying an experimental added factor, in this case salience. The

additive-factors paradigm was originally used in behavioral studies using RT measures to study interactions in cognitive processes (Sternberg 1969, 1998, 2001) and has since been adapted to fMRI studies both theoretically (Sartori and Umiltà 2000) and empirically (James and Stevenson 2012; James et al. 2009; Kim and James 2010; Kim et al. 2011; Stevenson et al. 2009, 2010). The paradigm has also successfully been applied to measuring interactions between sensory modalities, where changes in the audio-visual response across multiple levels of a given stimulus factor was compared to the multifactorial change in the unisensory response to demonstrate the principle of inverse effectiveness (James and Stevenson 2012; James et al. 2009; Kim and James 2010; Kim et al. 2011; Stevenson et al. 2009, 2010). Here, our additive-factors design focused on stimulus salience, with an emphasis on interactions between salience levels and multisensory gain in the P1-N1-P2.

## Methods and Materials

### Methods Overview

The experiment consisted of two phases, a psychophysical phase, and a subsequent EEG phase. During the first phase, participants identified auditory spoken words or visual spoken words at a range of signal-to-noise ratio (SNR) levels. Accuracies from each individual subject were used to determine the SNR level associated with three behavioral accuracy levels for each sensory modality. In the second phase, the same participants were presented with audio-only, visual-only, and audiovisual presentations at their three individual SNR levels while EEG recordings, accuracy, and RT were measured.

### Participants

Participants included 15 right-handed, native-English speakers (5 female, mean age = 25 years, range = 18–29, SD = 2.7). All participants reported normal or corrected-to-normal visual acuity, normal hearing, and no previous neurological impairments. The experimental protocol was approved by the Indiana University Institutional Review Board and Human Subjects Committee and all participants provided written consent.

### Stimulus Materials

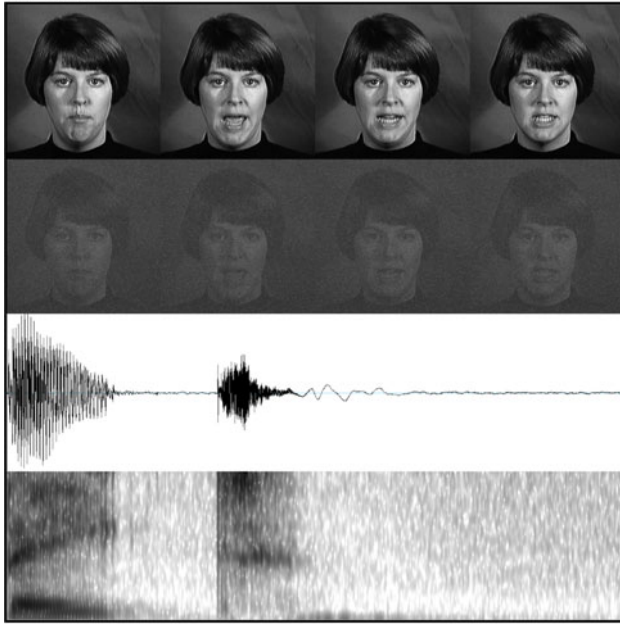
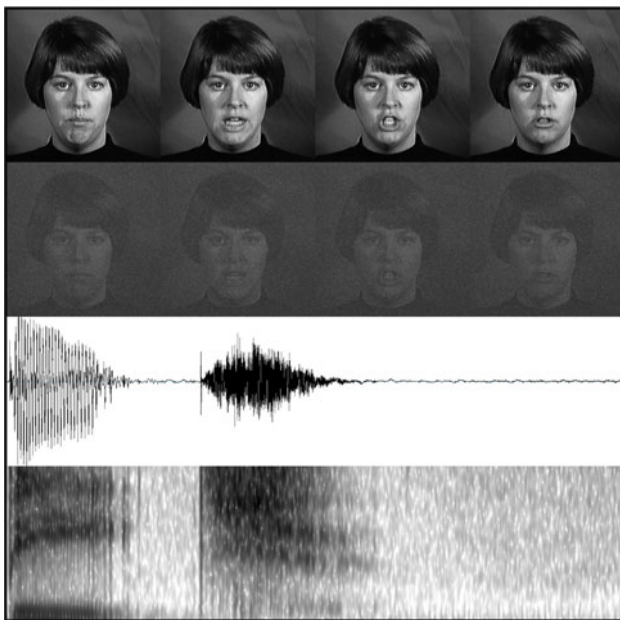
Stimuli included dynamic, audiovisual (AV) recordings of a female speaker saying two nouns, “bait” and “beach.” Stimuli were selected from a previously published stimulus set, The Hoosier Audiovisual Multi-Talker Database

(Sheffert et al. 1996), and have been previously used successfully in studies of multisensory perception (Stevenson et al. 2009, 2010, 2011; Stevenson and James 2009). All stimuli were spoken by speaker F1 in the database. We selected monosyllabic words that had high levels of accuracy on both visual-only and audio-only recognition (Lachs and Hernandez 1998) and resided in low-density lexical neighborhoods (Luce and Pisoni 1998; Sheffert et al. 1996). The auditory amplitudes of each utterance were equated as measured by root-mean-square contrast. Also, chosen stimuli all began with the bilabial voiced stop /b/, and stimulus onset began with auditory onset, ensuring that no pre-auditory articulatory movement could be seen visually prior to auditory onset. Total stimulus presentation time lasted 1 s. See Fig. 1 for auditory and visual representations of stimuli.

All stimuli in both phases of the study were presented using MATLAB software (MATHWORKS Inc., Natick, MA) with the Psychophysics Toolbox (Brainard 1997; Pelli 1997). Visual stimuli were presented on a 48 inch Samsung® Plasma Screen television positioned three meters from the participant. Visual stimulus dimensions were 400 × 400 pixels and hence subtended 16 × 16° of visual angle. Auditory stimuli were presented through two Cyber Acoustics® speakers placed directly below the screen at a distance of 3 m from the participant.

### Pre-EEG Psychophysics Protocol

Participants first completed a psychophysical protocol using the method of constant stimuli (Laming and Laming 1992) to identify each individual’s behavioral thresholds in the auditory and visual sensory modalities. Both auditory and visual unisensory were presented at fifteen SNR levels. Auditory and visual stimuli were overlaid with dynamic white noise at a set level (level of auditory noise = 50 dB<sub>A</sub>), with the auditory signal adjusted through variations in the root mean square (RMS) and visual signal adjusted through variations in contrast by changing the luminance of each pixel towards or away from the mean luminance on a frame-by-frame basis. Participants were presented with 20 trials at each SNR level, in both sensory modalities, for a total of 300 auditory and 300 visual trials. The orders of both sensory modality and SNR level were randomized. Each trial began with either an auditory- or visual-only presentation of a single spoken word, followed by a screen reminding participants of their two response options (bait or beach). Participants completed a two-alternative forced-choice (2AFC) task identifying the word that they heard or saw via button press with either the left or right index fingers. The subsequent trial automatically began after a brief, randomly jittered pause following the participant’s response, and breaks were offered after every 50 trials.

**A Bait stimuli****B Beach stimuli**

**Fig. 1** Stimuli. Graphical representations for each of the stimuli used. Each *panel* displays full-contrast frames from the visual stimuli, reduced-contrast, noisy frames of the visual stimuli, auditory waveforms, and auditory spectrograms of the utterances

Participants were instructed to respond as quickly and accurately as possible.

Using the accuracies recorded in the psychophysical behavioral session, a psychometric function was fit to each individual participant's data, and each participant's individual threshold levels were extracted at which the participant responded with accuracy rates of 97.5, 85, and

72.5% independently for both auditory and visual presentations. The three SNR levels at each of the three thresholds were then used in the subsequent EEG phase of the experiment for that individual participant. The choice to use salience levels associated these accuracy levels, although not ideal for analysis of multisensory gain in accuracy, was necessary as EEG data were only analyzed for correct trials. Mean signal dB levels of auditory stimuli at 97.5, 85, and 72.5% accuracy were 34, 30.9, and 28.6, respectively. Mean Michelson Contrasts for visual stimuli at 97.5, 85, and 72.5% accuracy were 0.83, 0.77, and 0.70, respectively.

**EEG Protocol**

Within 48 h of the behavioral psychophysics phase, participants returned for the EEG phase of the experiment. Stimuli used in the EEG phase included auditory-only, visual-only, and audiovisual trials at each of the three SNR levels associated with 97.5, 85, and 72.5% accuracy for the individual participant. Audiovisual trials consisted of the simultaneous presentation of the auditory and visual stimuli at their respective unisensory SNR levels, for example, an 85% audiovisual presentation included the presentation of a visual stimulus at the visual SNR at the 85%-accuracy threshold and an auditory stimulus at the auditory 85% threshold.

The nine experimental conditions consisted of auditory high, medium, and low ( $A_H$ ,  $A_M$ , and  $A_L$ , respectively), visual high, medium, and low ( $V_H$ ,  $V_M$ , and  $V_L$ , respectively), and multisensory high, medium, and low ( $AV_H$ ,  $AV_M$ , and  $AV_L$ , respectively). A total of 50 trials were presented in each condition, for a total of 450 stimulus presentations over the entire EEG recording session. Each trial was identical to those previously described in the psychophysics phase except that following a participant's response, a fixation cross was presented, and the participant pressed the space bar to continue on to the next trial. Time intervals between button-presses to start the trial and stimulus onset included a 1 s pause followed by an additional randomly jittered pause between 1 and 500 ms, avoiding slow anticipatory potentials (Teder-Salejarvi et al. 2002). The task was identical to the psychophysics phase, and breaks were again offered every 50 trials. Behavioral accuracies were calculated interactively during the experiment, and after the randomized 450 trials, participants were presented with an additional group of trials equal in number and condition to the trials with which they responded incorrectly until each condition had 50 correct trials (referred to henceforth as remedial trials). Identical to the first phase, participants responded as quickly and accurately as possible.

## EEG Measurement and Analysis

A continuous scalp EEG recording was performed using a 256-channel HydroCel™ Geodesic Sensor Net (Electrical Geodesics Inc., Eugene OR, USA; EGI). Data were sampled at 250 Hz per channel, bandpass filtered from 0.1 to 100 Hz, and amplified with a gain of 5000 through a Net Amps 300™ Geodesic amplifier (EGI), and were recorded relative to a vertex reference electrode. Continuously recorded EEG data were stored for subsequent offline analysis using NetStation V 4.4 software (EGI). Electrode impedances were kept below 70 kΩ consistent with the manufacturer's recommendations.

Posthoc analysis of EEG data consisted of the following steps: first, data were digitally lowpass filtered at 0.1–40 Hz using an infinite impulse response filter. The continuous EEG data were then segmented into 1,100 ms epochs containing each delivered stimulus, with a 100 ms pre-stimulus baseline and 1,000 ms period after stimulus onset. Only correct trials were included, with incorrect trials excluded from further analysis. Epochs containing either eye blinks or electromyographic activity were excluded from analysis using a semi-automated routine which excluded epochs with extreme amplitude values exceeding  $\pm 50 \mu\text{V}$  with additional visual inspection. Subsequently, bad channels (where voltage changes exceeding  $\pm 150 \mu\text{V}$ ) were identified and removed and replaced with a Net Station spherical-spline interpolation of surrounding electrode recordings. The EEG recordings from one participant had persistent artifacts on multiple channels and hence the data from this participant were excluded from subsequent analyses. The mean 100 ms pre-stimulus baseline amplitude was measured, and each epoch was baseline corrected to a mean zero amplitude level of 100 ms pre-stimulus baseline. All epochs for each condition ( $A_H$ ,  $A_M$ ,  $A_L$ ,  $V_H$ ,  $V_M$ ,  $V_L$ ,  $AV_H$ ,  $AV_M$ ,  $AV_L$ ) were then averaged for each participant to be further analyzed on an individual subject basis. For visualization purposes, individual participant's averaged ERP waveforms were then combined into a grand averaged ERP as a function of condition. ERP data were then re-referenced relative to a common average reference.

Voltage maps were plotted from the grand average ERP data to examine the time courses of the regional distributions of ERP activity. A subset of nine electrodes each located over the left and right lateral occipito-temporal scalp was selected a priori based on previous reports (e.g., Bentin et al. 1996; Brefczynski-Lewis et al. 2009; Giard and Peronnet 1999) and confirmed to have local amplitude maxima as seen on voltage maps, showing peak activity during the P1-N1-P2 response. ERP waveforms from each bilateral nine electrode cluster were extracted from both the grand average ERP data set as well as from each individual

participant's averaged ERP as a function of condition. Average peak amplitudes and peak latencies across each hemispheric cluster as well as peak-to-peak amplitudes and latencies were then calculated for unisensory (A and V), multisensory (AV), and summed unisensory ( $A + V$ ) ERP peaks and troughs using two-way, paired sample t-tests, as further described in the results section.

## Additive Factors Analysis

In analysis of these data, we used an additive-factors metric to directly test for change in multisensory interactions across SNR levels. When a multisensory interaction is expressed as

$$AV - (A + V) \neq 0,$$

a change in multisensory gain across levels would thus be expressed as:

$$AV_H - (A_H + V_H) \neq AV_L - (A_L + V_L),$$

which can be reorganized as:

$$AV_H - AV_L \neq (A_H - A_L) + (V_H - V_L).$$

Throughout the analysis we present our data in this form, comparing changes in responses to multisensory stimuli across salience levels to changes responses to unisensory stimuli, which tests for changes in multisensory interactions across these salience levels.

## Results

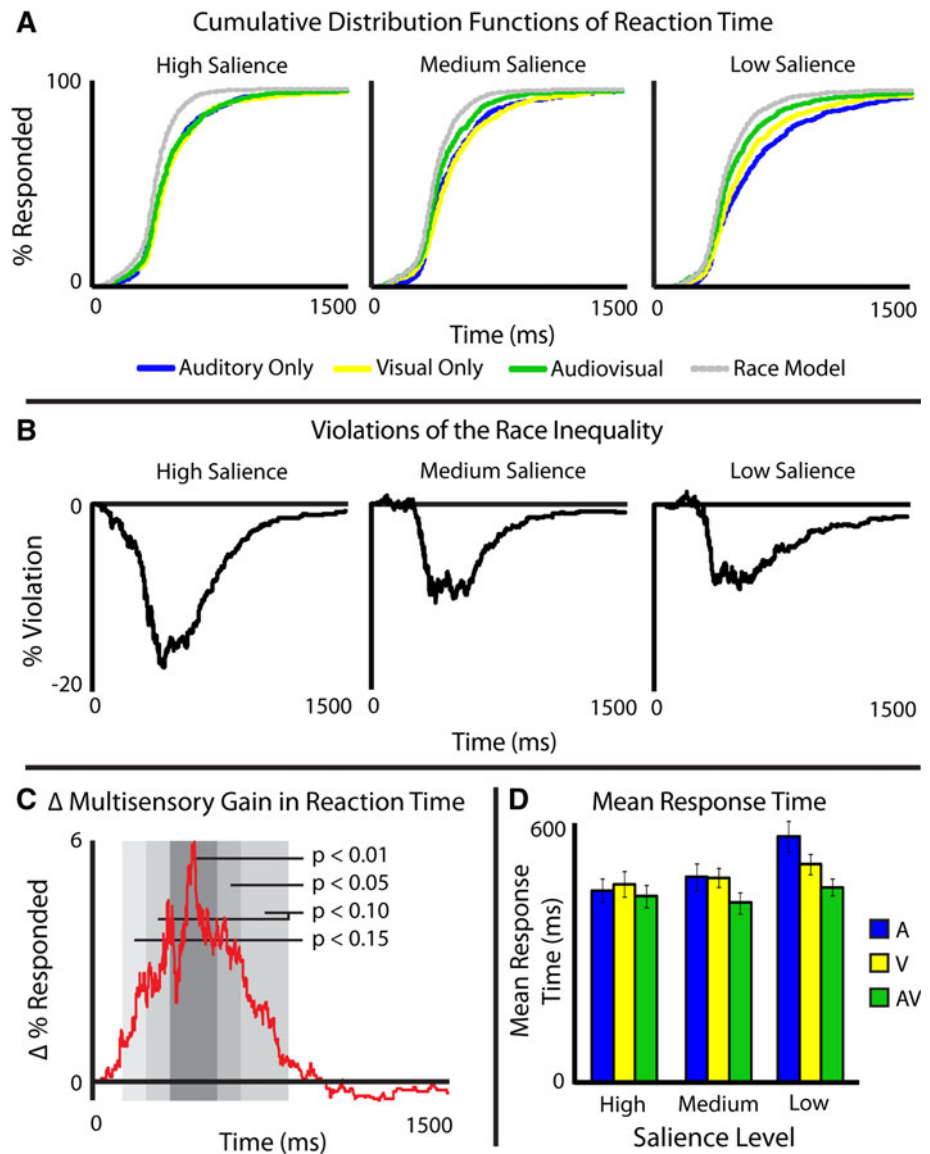
### Behavioral Data Recorded During the EEG Phase of the Experiment

#### Response Time Data

Participant's RTs were measured for all correct, randomized trials during the EEG recording session. Incorrect trials and remedial trials (which consisted primarily of unisensory, low-SNR trials, and were thus not randomized), were excluded from RT analysis. Mean RTs (Fig. 2d) and cumulative distribution functions (CDFs) with 1 ms intervals were calculated for each individual, for each sensory modality, at each SNR level (Fig. 2a). Predicted non-interactive multisensory CDFs were calculated with a parallel, non-interactive model, or race model (Raab 1962), that accounts for statistical facilitation of two independent processes. The predicted CDFs will be referred to as  $pAV_H$ ,  $pAV_M$ , and  $pAV_L$  for the predicted multisensory high-, medium-, and low-SNR levels, respectively.

Each individual's CDFs tested for salience effects were analyzed using the A, V, and AV individual CDFs of RT. Kolmogorov–Smirnov statistical tests were run comparing

**Fig. 2** Response time data. Cumulative distribution functions at each salience level and each modality can be seen as well as race model calculations in *Panel A*, and differences between the AV presentations and race models can be seen in *Panel B*. Change in gain across salience levels can be seen in *Panel C*. *Panel D* shows mean response times for each modality at each salience level



individual's  $A_H$  to  $A_L$ ,  $V_H$  to  $V_L$ , and  $AV_H$  to  $AV_L$ . In the auditory and visual comparison, all 15 subjects showed ordered salience effects (i.e., high salience was faster than medium was faster than low,  $p < 0.001$ ), while 13 of 15 ( $p = 0.0074$ ) showed ordered salience effects in the AV comparison. Given these consistent findings, group CDFs were calculated for each condition as the arithmetic mean of all individuals CDFs (Fig. 2a). Salience effects were then analyzed using the A, V, and AV group CDFs of RT. Kolmogorov–Smirnov tests were run on the group data comparing  $A_H$  to  $A_L$  ( $KS = 0.20$ ,  $p < 3.98e^{-11}$ ),  $V_H$  to  $V_L$  ( $KS = 0.13$ ,  $p < 7.00e^{-5}$ ), and  $AV_H$  to  $AV_L$  ( $KS = 0.7$ ,  $p < 0.06$ ). Unisensory RTs thus showed large salience effects, with slower responses being seen as stimulus salience decreased. Multisensory RTs showed a marginally significant salience effect in the same direction, but this effect was markedly smaller than the unisensory effect.

RT data were also analyzed to assess multisensory gain. Multisensory gain was computed by subtracting the CDFs from the non-interactive race models (Raab 1962) predicted from the measured multisensory CDFs at each SNR level (e.g.,  $AV_H - pAV_H$ ; Fig. 2a, black). On an individual basis, 12 of 15 individuals showed violations of the race model at high SNR ( $p = 0.0352$ ), 14 of 15 at medium SNR ( $p = 0.001$ ), and 13 of 15 at low SNR ( $p = 0.0074$ ), demonstrating that this effect was not driven by the results of a single or a few participants, but rather was consistent across participants. Given the consistency across subjects, a race model and violations of the race model were then calculated for the group CDFs (Fig. 2b). In the averaged group data, violations of the race model (Miller 1982) were observed at both the medium- and low-SNR levels (for 104 and 115 ms, respectively), but not at the high-SNR levels.

To measure an interaction between multisensory gain and stimulus salience in RTs, we used a difference-of-differences calculation (James et al. 2012; Stevenson et al. 2009) that compares the changes in multisensory gain across the three levels of SNR. With multisensory gain calculated as the predicted multisensory response CDF subtracted from the measured multisensory response CDF (e.g.,  $AV_H - pAV_H$ ), change in multisensory gain was calculated as a difference of differences (Fig. 2c), or  $[(AV_L - pAV_L) - (AV_H - pAV_H)]$ . The order in which the subtraction is performed dictates the sign of the measured change in multisensory gain. Given the order of subtraction (low – high) to assess gains, a resulting positive distribution infers inverse effectiveness. This difference-of-differences distribution in group RTs were divided into 100 ms bins and the area under the curve within each bin was measured, with bins statistically greater than zero indicating inverse effectiveness (Fig. 2c). As predicted by the principle of inverse effectiveness, the level of multisensory gain increased parametrically as stimuli became less effective.

#### Accuracy Data

Participants' accuracies were measured for all randomized trials (Fig. 3). Remedial trials (which consisted primarily of unisensory, low-SNR trials, and were thus not randomized), were excluded from accuracy analysis. Mean unisensory accuracies parametrically decreased with decreases in SNR. Mean multisensory accuracies also parametrically decreased as SNR decreased, and in all cases significantly exceeded the maximum unisensory response in a pair-wise *t*-test ( $p < 0.01$ , 0.02, and 0.0005 for high, medium, and low SNR, respectively). Predicted accuracies with multisensory stimuli based on unisensory accuracies, assuming no interaction, were also calculated as the sum of probabilities that the individual correctly identified the unisensory stimulus minus the probability that the individual

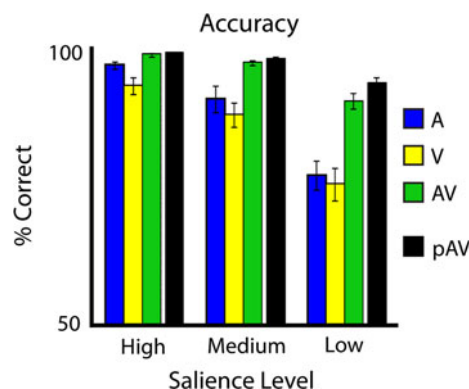
identified both  $[p(A) + p(V) - p(A)p(V)]$ . The predicted multisensory accuracies also parametrically decreased with decreases in SNR. Measured and predicted multisensory accuracies did not show significant differences in the high- or medium-SNR levels in a paired *t*-test ( $p = 0.33$  and 0.11, respectively), but did show a significant difference in the low-SNR level, where the measured multisensory accuracy was *less* than the predicted accuracy ( $p = 0.04$ ). Given the high accuracy level found in the unisensory conditions, the predicted accuracies, particularly for the high and medium salience levels, were at ceiling, precluding any analysis of a change in multisensory gain across stimulus salience levels. The choice to use salience levels associated with high accuracy levels was necessary as EEG data were only analyzed for correct trials.

#### ERP Data Overview

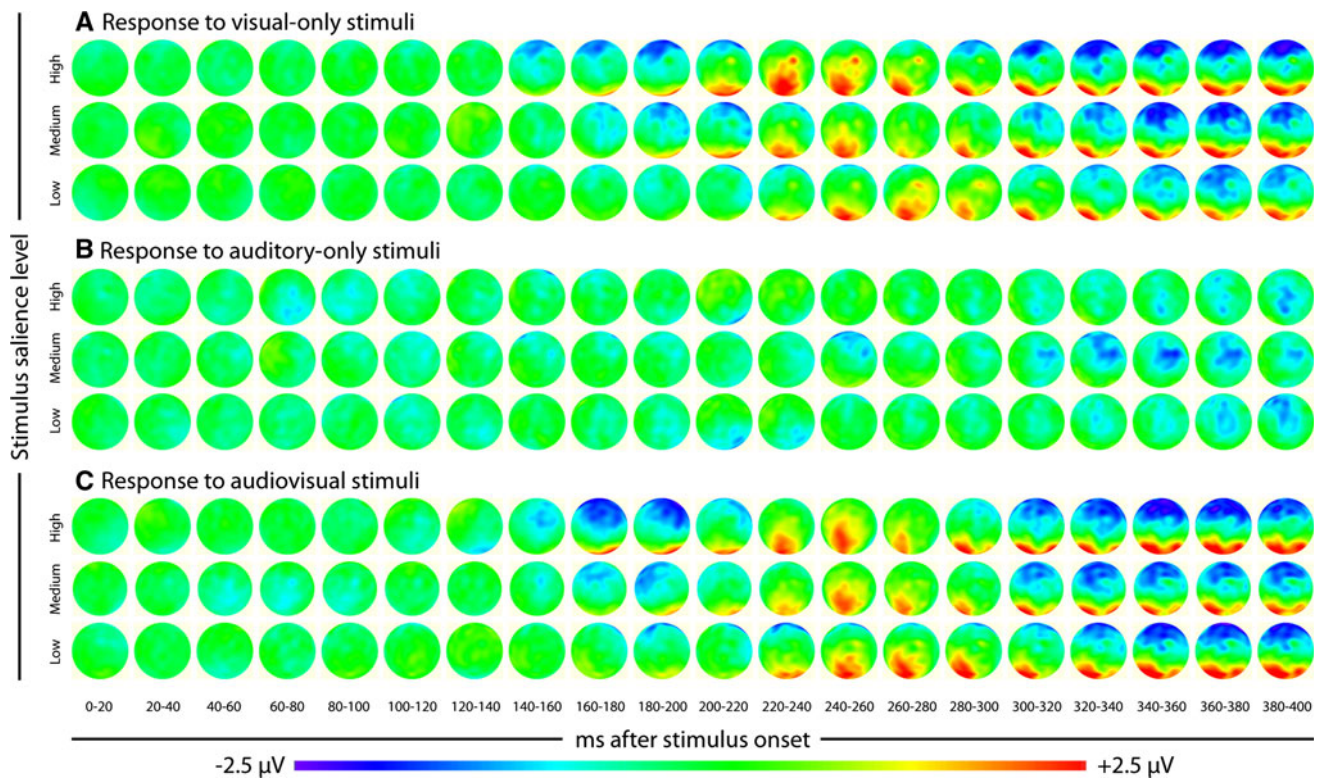
The main ERP analysis focused on a posterior scalp triphasic ERP complex extracted from a priori selected electrodes that previous reports have identified as responding maximally to visual stimuli, a finding confirmed in our data as illustrated topographically in Fig. 4. This tri-phasic complex consists of a positive–negative–positive series of components, which we will refer to as P1–N1–P2 (see Fig. 5 for group data). Analysis focused on three effects on the amplitudes and latencies of individually extracted P1–N1–P2 complexes, an effect of stimulus salience within sensory modality, multisensory gain, and a sensory modality by stimulus salience interaction. Stimulus-salience effects were measured in peak-to-peak amplitudes and latencies of ERP components in each sensory modality. Multisensory gains were measured as differences from the summed unisensory response at each stimulus salience level. Finally, a more sophisticated analysis designed to measure multisensory interactions across an added stimulus factor, in this case stimulus salience, was also conducted, comparing the differences of multisensory and summed unisensory response.

#### Stimulus Salience Effects on Amplitude

Salience effects on the amplitudes of the ERP components of the P1–N1–P2 complex from bilateral lateral occipito-temporal montages were measured for each individual using  $3 \times 2$  repeated measures ANOVAs (salience level  $\times$  hemisphere), with amplitudes, and detailed statistics from salience-level comparisons reported in Table 1 and a graphical representation in Fig. 6. P1 was measured relative to the pre-stimulus baseline, while N1 and P2 were quantified using peak-to-peak amplitude changes (i.e., P1–N1, N1–P2) due to significant salience effects in P1, which may have impacted these later ERP components.

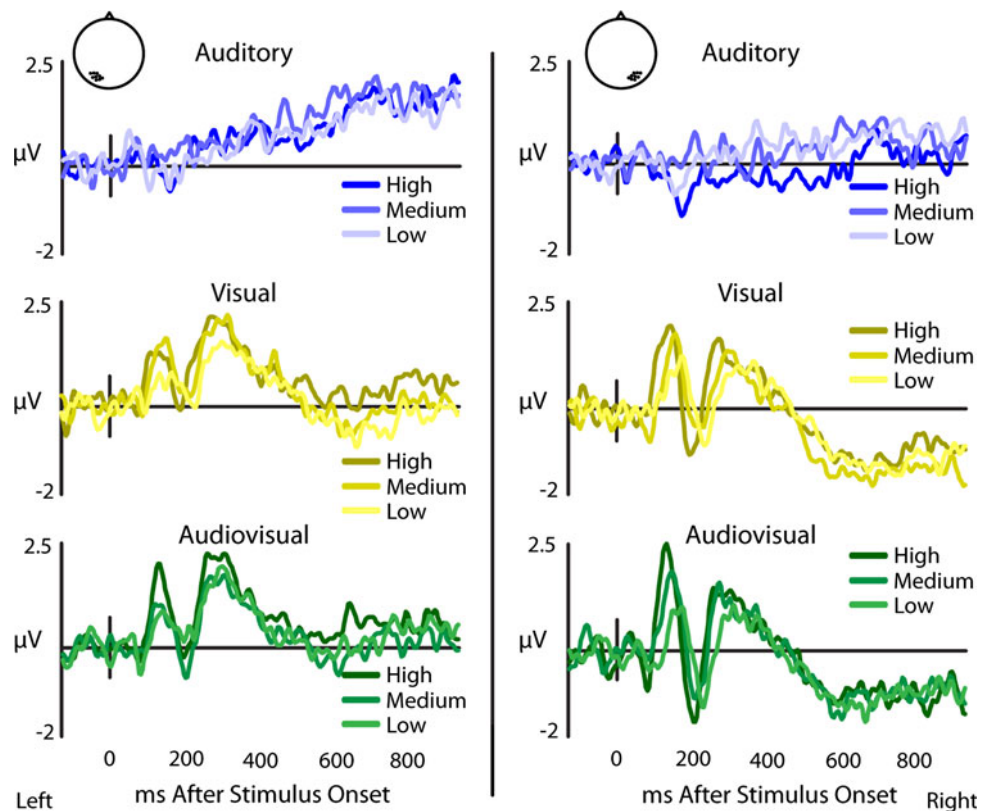


**Fig. 3** Accuracy data. Accuracies for each modality at each salience level, as well as the predicted accuracy for audiovisual trials based on unisensory accuracies



**Fig. 4** Topographic maps of ERP responses. Group average topographic maps of responses to visual-only (a), auditory-only (b), and audiovisual stimuli for the whole scalp

**Fig. 5** Effects of stimulus salience on posterior temporal recordings. Group-averaged ERPs from a nine electrode cluster overlying the posterior temporal scalp in *left* and *right* hemispheres with visual (yellow) and audiovisual (green) presentations show variable P1-N1-P2 responses relative to stimulus salience, while auditory (blue) presentations evoke no P1-N1-P2 response (color figure online)



**Table 1** Mean P1, P1-N1, and N1P2 amplitudes in  $\mu\text{V}$  across subjects as a function of saliency level (high, medium, low)

Hemi	Modality	ERP	High	Medium	Low	$F_{(2,39)}$	$p$
Right	V	P1	2.98 (0.35)	2.96 (0.33)	2.32 (0.48)	2.47	<0.11
		P1-N1	-4.62 (0.71)	-4.40 (0.63)	-3.59 (0.51)	2.15	<0.14
		N1-P2	3.97 (0.71)	3.30 (0.69)	2.78 (0.53)	1.78	NS
	AV	P1	3.28 (0.41)	2.77 (0.43)	1.84 (0.42)	10.66	<4.17e <sup>-4</sup>
		P1-N1	-5.58 (0.67)	-4.31 (0.63)	-3.73 (0.53)	13.79	<8.28e <sup>-5</sup>
		N1-P2	4.32 (0.67)	3.60 (0.57)	3.71 (0.46)	1.07	NS
	A + V	P1	2.48 (0.50)	2.90 (0.58)	2.52 (0.43)	0.46	NS
		P1-N1	-5.09 (0.71)	-4.34 (0.58)	-4.43 (0.51)	1.17	NS
		N1-P2	5.59 (0.67)	4.03 (0.60)	2.88 (0.48)	10.85	<3.75e <sup>-4</sup>
Left	V	P1	2.34 (0.41)	2.45 (0.48)	1.86 (0.40)	2.11	<0.15
		P1-N1	-3.31 (0.50)	-3.32 (0.41)	-2.85 (0.37)	1.43	NS
		N1-P2	3.53 (0.54)	3.21 (0.42)	3.06 (0.37)	0.87	NS
	AV	P1	2.69 (0.47)	1.75 (0.39)	1.28 (0.42)	11.18	<3.14e <sup>-4</sup>
		P1-N1	-3.95 (0.56)	-2.89 (0.47)	-2.17 (0.34)	10.81	<3.83e <sup>-4</sup>
		N1-P2	3.51 (0.44)	2.95 (0.40)	3.10 (0.40)	0.96	NS
	A + V	P1	2.46 (0.38)	3.18 (0.59)	2.05 (0.52)	2.63	<0.09
		P1-N1	-3.89 (0.49)	-3.88 (0.50)	-3.38 (0.28)	0.55	NS
		N1-P2	4.87 (0.62)	3.17 (0.41)	2.48 (0.41)	7.12	<0.004

Values in parentheses represent standard deviations of the group mean

Hemi hemisphere, V visual, AV audiovisual, A + V = summed unisensory response amplitudes;  $F$  = results of a 1-factor, repeated-measures ANOVA across the three saliency levels

Audiovisual P1 responses showed significant amplitude changes across saliency levels. With decreases in stimulus saliency, P1 amplitudes became less positive ( $F_{(2,78)} = 15.51$ ,  $p < 3.68e^{-5}$ ) and P1-N1 amplitudes became less negative ( $F_{(2,78)} = 15.47$ ,  $p < 3.75e^{-5}$ ) in both hemispheres. No significant saliency effects were seen in the audiovisual N1-P2 ( $F_{(2,78)} = 1.38$ ). Summed unisensory responses showed significant no significant saliency effects in the P1 ( $F_{(2,78)} = 1.82$ ) or P1-N1 ( $F_{(2,78)} = 0.73$ ), but did show a significant bilateral saliency effect in the N1-P2 ( $F_{(2,78)} = 14.36$ ,  $p < 6.28e^{-5}$ ) which became less positive as saliency decreased. Visual responses showed slight amplitude decreases in the P1-N1-P2 complex bilaterally with less salient stimuli, with marginally significant reduction in the P1 ( $F_{(2,78)} = 2.90$ ,  $p < 0.08$ ), P1-N1 ( $F_{(2,78)} = 2.33$ ,  $p < 0.12$ ), and N1-P2 ( $F_{(2,78)} = 2.05$ ,  $p < 0.15$ ) amplitudes with lower saliency stimuli. As a general synopsis of these results, as stimulus saliency decreased, the amplitudes of the individual components of the P1-N1-P2 complex declined.

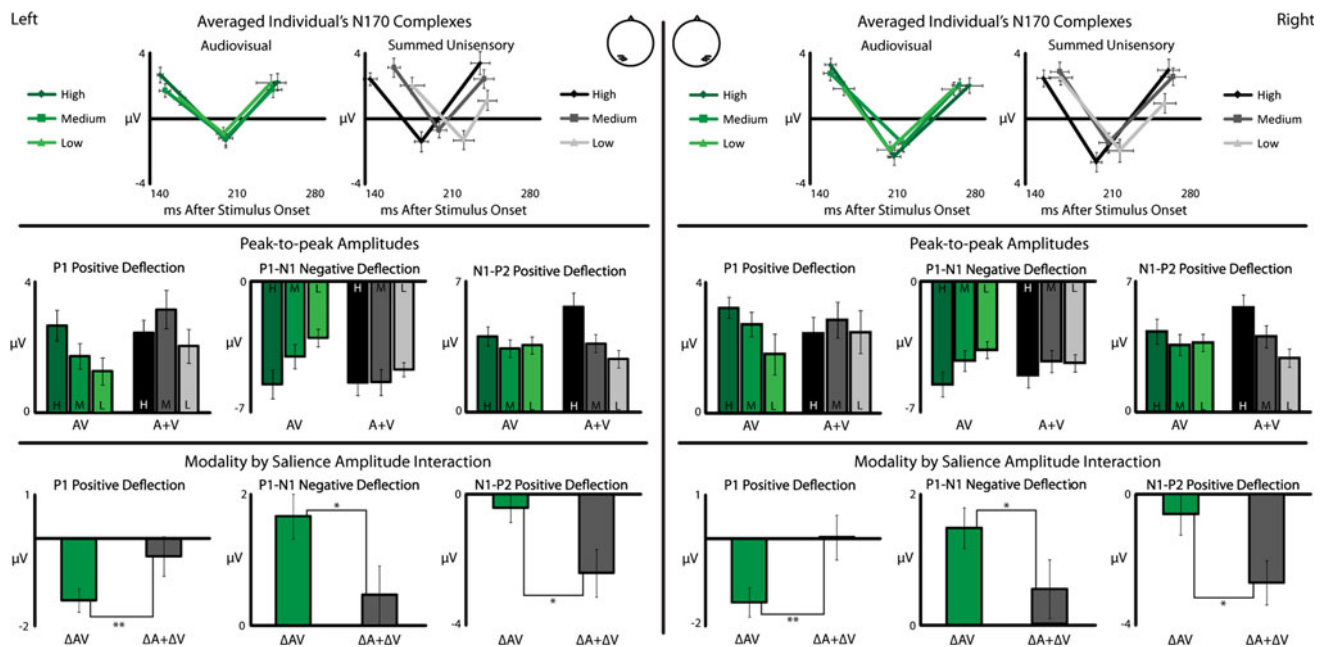
#### Stimulus Saliency Effects on Latency

Saliency effects on the latencies of the components of the P1-N1-P2 complex were measured relative to stimulus onset with visual, audiovisual, and summed unisensory responses for bilateral lateral occipito-temporal montages

(Fig. 6) using  $3 \times 2$  repeated measures ANOVAs (saliency level  $\times$  hemisphere), with exact latencies, and detailed statistics from pairwise comparisons between high- and low-saliency conditions ( $F$  values, and  $p$  values) reported in Table 2). Relative to stimulus onset, visual presentations showed earlier responses with increasing stimulus saliency. This was observed bilaterally for P1 ( $F_{(2,78)} = 12.92$ ,  $p < 1.27e^{-4}$ ), N1 ( $F_{(2,78)} = 8.59$ ,  $p < 0.002$ ), and P2 ( $F_{(2,78)} = 6.93$ ,  $p < 0.004$ ). Audiovisual presentations showed bilateral changes in latencies only in the P1 component ( $F_{(2,78)} = 5.47$ ,  $p < 0.02$ ) and only in the left hemisphere that were earlier with more salient stimuli bilaterally. No saliency effect on N1 ( $F_{(2,78)} = 0.40$ ) or P2 ( $F_{(2,78)} = 0.50$ ) latency was observed. With higher saliency stimuli, summed unisensory presentations exhibited earlier latencies in the P1 significantly ( $F_{(2,78)} = 8.08$ ,  $p < 0.002$ ) in the left hemisphere and the N1 component was earlier ( $F_{(2,78)} = 16.90$ ,  $p < 1.99e^{-5}$ ), significantly in the left and marginally significantly in the right hemisphere. The P2 showed no latency effect ( $F_{(2,78)} = 0.08$ ).

#### Multisensory Interactions as Indexed by Amplitude Changes

Interactions between auditory and visual processing measured in ERP amplitudes were assessed by comparing multisensory responses with the sum of the constituent



**Fig. 6** Changes in multisensory interactions across salience levels. Individually extracted P1-N1-P2 responses were averaged and are shown in the top row with audiovisual (green) and summed unisensory (gray) presentations. Amplitudes of the P1-N2 and N1-

P2 are shown for each salience level, with darker colors representing more salient presentations. Changes of amplitudes across salience levels are shown in the bottom row (color figure online)

**Table 2** Mean P1-N1-P2 latencies in ms across subjects as a function of saliency level (high, medium, low)

Hemi	Modality	ERP	High	Medium	Low	$F_{(2,39)}$	$p$
Right	V	P1	147 (5.7)	165 (6.0)	182 (6.3)	18.86	$<8.71e^{-6}$
		N1	191 (8.1)	216 (6.9)	215 (12.0)	3.51	$<0.05$
		P2	247 (10.5)	263 (7.4)	278 (11.4)	2.89	$<0.08$
	AV	P1	153 (5.0)	152 (5.9)	164 (9.4)	1.34	NS
		N1	209 (5.5)	216 (4.7)	208 (10.6)	0.47	NS
		P2	276 (11.3)	267 (6.6)	262 (10.9)	0.53	NS
	A + V	P1	151 (5.9)	165 (8.3)	166 (10.4)	1.10	NS
		N1	198 (4.6)	210 (8.1)	221 (10.9)	2.80	$<0.08$
		P2	262 (9.6)	266 (11.2)	258 (9.9)	0.14	NS
Left	V	P1	149 (6.6)	159 (6.3)	170 (7.5)	4.23	$<0.03$
		N1	192 (6.7)	210 (7.7)	218 (7.2)	12.31	$<1.74e^{-4}$
		P2	250 (8.1)	273 (6.7)	284 (8.8)	7.83	$<0.003$
	AV	P1	144 (5.2)	148 (5.0)	161 (5.44)	4.55	$<0.03$
		N1	202 (6.0)	202 (6.0)	200 (6.5)	0.10	NS
		P2	248 (8.8)	244 (8.1)	242 (13.2)	0.08	NS
	A + V	P1	139 (6.3)	161 (5.6)	177 (10.9)	10.89	$<3.66e^{-4}$
		N1	185 (6.1)	201 (7.0)	223 (9.1)	13.37	$<1.02e^{-4}$
		P2	238 (8.8)	241 (9.0)	244 (8.4)	0.018	NS

Values in parentheses represent standard deviations of the group mean

Hemi hemisphere, V visual, AV audiovisual, A + V = summed unisensory response amplitudes;  $F$  = results of a 1-factor, repeated-measures ANOVA across the three salience levels

unisensory responses. Additionally, an additive-factors analysis was performed. This additional analysis has the benefit of excluding spurious findings due to common activations in auditory and visual ERPs (see “Appendix”).

A comparison of audiovisual and summed unisensory P1 peak amplitudes revealed differences only at the medium salience level and only in the left hemisphere ( $t_{(13)} = 2.62$ ,  $p < 0.03$ ). For P1-N1 amplitude, no difference was found

between audiovisual and summed unisensory presentations at the high salience level (Right;  $t_{(13)} = 1.42$ , Left;  $t_{(13)} = 0.77$ ). For the medium salience level, a significant difference was seen in the left ( $t_{(13)} = 2.50$ ,  $p < 0.03$ ) but not right ( $t_{(13)} = 0.77$ ) hemisphere. Finally, for low salience stimuli, a significant differences in the left ( $t_{(13)} = 3.50$ ,  $p < 0.004$ ) and marginally significant difference in the right ( $t_{(13)} = 2.03$ ,  $p < 0.06$ ) hemisphere was found. In each of these cases of significant differences from the summed unisensory prediction, the actual multisensory responses were subadditive relative to the summed unisensory responses. Overall, this analysis demonstrated that as stimulus salience was decreased, the multisensory responses of the P1-N1 decreased more than predicted based on the summed unisensory response. For the N1-P2 amplitude, in the right hemisphere, significant differences were seen in the high ( $t = 3.60$ ,  $p < 0.03$ ) and low ( $t = 2.30$ ,  $p < 0.04$ ) salience levels, but not the medium ( $t_{(13)} = 1.71$ ). In the left hemisphere, the only significant difference was seen at the high-salience level ( $t_{(13)} = 2.41$ ,  $p < 0.04$ ) and not the medium ( $t_{(13)} = 1.15$ ) or low ( $t_{(13)} = 1.62$ ) levels. Hence, for N1-P2 amplitudes, the multisensory response becomes less subadditive as salience decreases in the right hemisphere, with no trend in the left.

While the differences seen through comparing the actual audiovisual response to the summed unisensory responses can shed light on interactions across salience levels, the interaction between sensory modality and stimulus salience was directly assessed through an additive-factors analysis. The difference in peak-to-peak amplitudes in the P1-N1-P2 complex (i.e., P1-N1 and N1-P2) across salience level was calculated for each individual for both audiovisual and summed unisensory responses (Fig. 6, middle panel). As with RT, we measured differences between changes in unisensory and multisensory responses across salience levels. Differences between audiovisual (e.g.,  $AV_H - AV_L$ , or  $\Delta AV$ ) and summed unisensory differences [e.g.,  $(A_H - A_L) + (V_H - V_L)$ , or  $(\Delta A + \Delta V)$ ] across the added factor of stimulus salience are indicative of a multisensory interaction (Fig. 6, bottom panel). Furthermore, these measures are independent of common activation (see “Appendix”).

For P1 amplitude, a significant interaction was found between the multisensory and summed unisensory responses in both the left ( $t_{(13)} = 3.14$ ,  $p < 0.008$ ) and right ( $t_{(13)} = 3.15$ ,  $p = 0.008$ ) hemispheres. Here, multisensory gain *decreased* at lower salience levels. In P1-N1 amplitude, a similar interaction was found between the multisensory and summed unisensory responses in both the left ( $t_{(13)} = 2.61$ ,  $p < 0.03$ ) and right ( $t_{(13)} = 2.28$ ,  $p < 0.04$ ) hemispheres, with the multisensory changes larger than the changes predicted by the summed unisensory responses. That is, the multisensory gain *decreased* at lower salience levels. An interaction was also found in the N1-P2 component in both

the left ( $t_{(13)} = 2.5$ ,  $p < 0.03$ ) and right ( $t_{(13)} = 2.82$ ,  $p < 0.02$ ) hemispheres, however this interactions showed an *increase* in multisensory gain at lower stimulus salience level, consistent with classic inverse effectiveness.

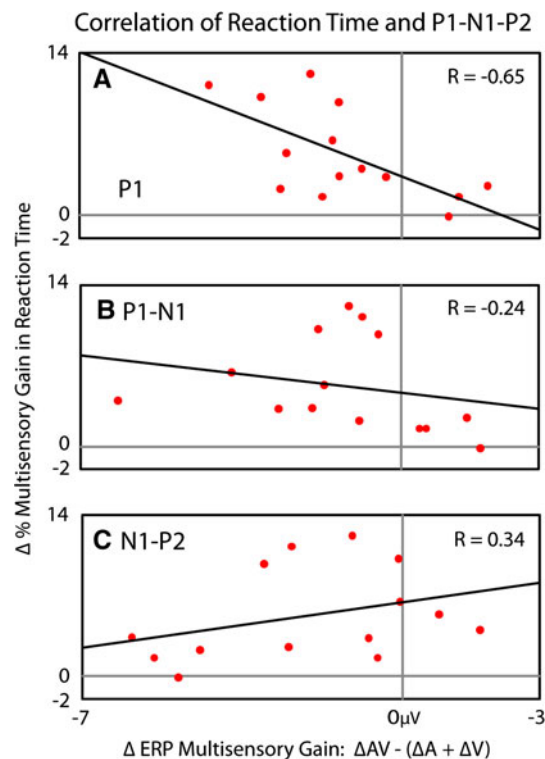
These two analyses of multisensory interactions of P1-N1-P2 amplitudes confirm that visual ERPs are modulated by auditory stimuli, and also show that these multisensory interactions change with the stimulus salience. For P1 and P1-N1, as stimulus salience is reduced, the multisensory gain (relative to the prediction of the summed unisensory responses) is also reduced. In contrast, the bilateral N1-P2 shows increased multisensory gain as stimulus salience is reduced. Taking these findings into account, it would appear that the main drivers of the multisensory gain effect might be P1 and P2, with respect to decreasing and increasing multisensory gain, respectively.

#### *Relationship of RT to Multisensory Interactions as Indexed by ERP*

As highlighted above, multisensory interactions across salience levels were found for both RTs and ERPs. In particular, RTs exhibited inverse effectiveness, particularly early in participant's CDFs (Fig. 2c), and ERPs exhibited a multisensory interaction across salience levels in the P1, P1-N1, and N1-P2 (Fig. 6). To determine if there was a relationship between multisensory interactions at these two levels of analysis, correlations between the behavioral and neural changes in multisensory gain across salience level were calculated (Fig. 7). Each of the components in each individual's P1-N1-P2 responses from the previously described a priori selected lateral occipito-temporal electrodes was thus compared to that individual's change in multisensory gain in RT across stimulus levels. Degree of inverse effectiveness in RTs was strongly correlated with the multisensory interaction in the P1 ( $R = -0.65$ ) and weakly correlated with the multisensory interaction in the P1-N1 ( $R = -0.24$ ). More specifically, a greater degree of inverse effectiveness in RTs was associated with a greater *reduction* in the amplitude of the interaction as indexed in the individual's P1 and P1-N1 deflections. In contrast, levels of inverse effectiveness in RTs were moderately positively correlated with the multisensory interaction in the N1-P2 ( $R = 0.34$ ). Thus, a greater degree of inverse effectiveness in RTs was associated with a greater *increase* in the amplitude of the interaction seen in the individual's N1-P2 deflections.

#### *Multisensory Interactions in Electrode Clusters Selected Post hoc*

In addition to the multisensory interactions across stimulus salience level that were measured a priori within the visual



**Fig. 7** Correlation between multisensory gain in response time and P1-N1-P2 amplitudes across salience levels. Individuals showing greater multisensory gains in response times relative to the race model as salience decreased showed a respective decrease of multisensory gain in their P1 and P1-N2 deflections, and increased multisensory gain in their N1-P2 deflections relative to the summed unisensory deflections

ERPs, exploratory voltage maps were calculated for additive-factor interactions (Fig. 8a). With salience as the added factor, differences in audiovisual responses across salience levels ( $\Delta AV$ ) were compared with summed unisensory differences ( $\Delta A + \Delta V$ ). A difference voltage map was created, and three regions of the scalp were identified that showed reliable interactions, including the midline occipital pole, right temporo-parietal, and vertex. While statistical tests performed on the waveforms extracted from these electrodes (Fig. 8b–d) would be, by definition, non-independent (Kriegeskorte et al. 2009), these areas warrant further investigation for multisensory interactions, particularly related to interactions across salience levels.

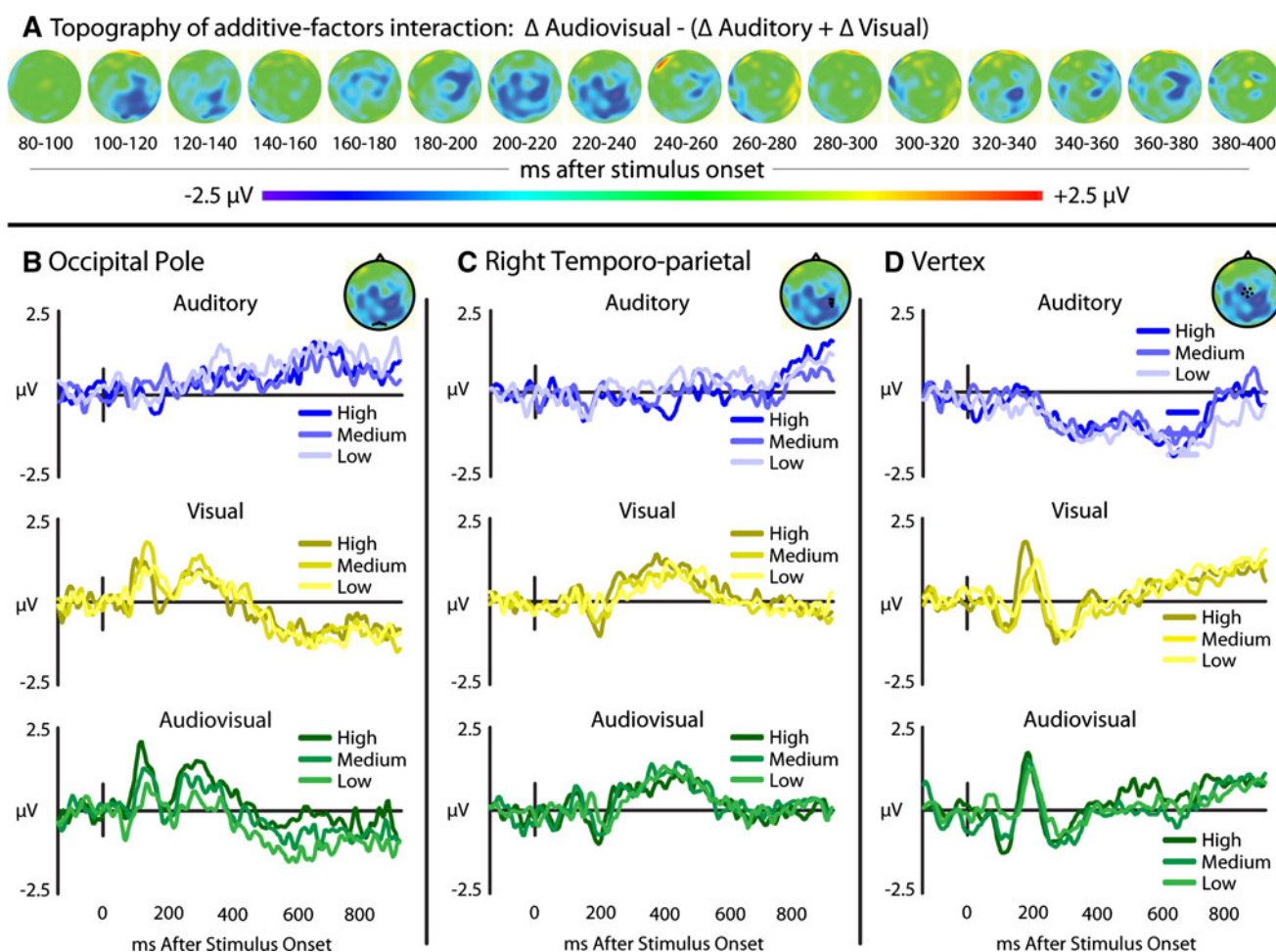
## Discussion

Recent studies have demonstrated that multisensory interactions can influence sensory processes that were previously thought to be unisensory in nature (Fort et al. 2002b; Giard and Peronnet 1999; Molholm et al. 2002; Senkowski et al. 2011). Here, we found that the visual P1-N1-P2 complex is reliably modulated by the salience of

audiovisual speech signals, and specifically that the multisensory interactions measured were also modulated by salience. In addition to identifying a novel set of multisensory interactions within the visual ERP, this study made use of a novel approach to measuring multisensory interactions in electrophysiological measures by adapting an additive-factors paradigm which has been successfully implemented in fMRI studies (James and Stevenson 2012; James et al. 2009, 2012; Kim and James 2010; Stevenson et al. 2010). Finally, the present results characterized the modulations of the P1-N1-P2 complex across parametrically varied levels of stimulus salience which were found to be correlated with behavioral responses. Here it was found that the degree of inverse effectiveness as measured in reaction times correlated positively with multisensory interactions in the N1-P2 and negatively with the P1 and P1-N2.

Our additive-factors ERP analysis revealed two distinct interactions (Fig. 6). N1-P2 peak-to-peak amplitude exhibited the multisensory interaction known as inverse effectiveness. As stimulus salience was decreased, multisensory gain increased. That is, the decrease in multisensory activity was small relative to the decrease predicted by the summed unisensory activity decreases, which is the formal indicator of inverse effectiveness. The P1 and P1-N1 peak-to-peak amplitudes also showed significant differences in the pattern of activity change between unisensory and multisensory conditions, however, those differences were not in the direction of inverse effectiveness. As stimulus salience was decreased, multisensory gain also decreased, which is the opposite of inverse effectiveness. The observation that different components of the visual ERP showed qualitatively different multisensory interactions may be surprising, given the consistency with which inverse effectiveness has been found with BOLD fMRI measures (James et al. 2009, 2012; Kim et al. 2011; Senkowski et al. 2011; Stevenson et al. 2007, 2009; Stevenson and James 2009; Werner and Noppeney 2010; but see Kim and James 2010). However, the results suggest that different salience-driven AV interactions occur at different times during the sensory analysis process, a result that could not have been predicted from or measured with BOLD fMRI. The specific interactions found here were inverse effectiveness during the N1-P2 interval and the opposite effect, previously referred to as enhanced effectiveness (Kim and James 2010), during the P1-N1 interval.

The reductions in the P1 and P1-N1 amplitudes with AV presentations relative to the summed unisensory responses across salience levels could be explained by at least three possible mechanisms. Since this pattern was driven in large part by larger changes in the AV response where there was limited change in unisensory response, the explanations will focus on the change in the AV response. First, the



**Fig. 8** Interactions in scalp responses in the occipital, right lateral temporal, and vertex electrodes. *Panel A* shows topographic maps of additive-factors interactions, where changes in multisensory gain can be seen across stimulus salience levels. Waveforms extracted from

locations showing interactions in the cortical maps including the occipital pole (*panel B*), right temporo-parietal (*panel C*), and at the vertex (*panel D*)

pattern could reflect inhibition, where the presence of an auditory stimulus reduced the P1 and P1-N1. Such an inhibitory effect has been proposed to explain reductions in auditory ERPs with the addition of visual input (Besle et al. 2004a), as well as negative BOLD changes observed in the occipital lobe with auditory input (Laurienti et al. 2002). However, if indeed the auditory stimuli were interfering with visual processing, one would predict reductions in multisensory gain found with RTs. However, the opposite effect was found, suggesting that this decrease in the P1 and P1-N1 is not a result of inhibition.

A second possible explanation for the observed pattern of ERP changes could be that the addition of a stimulus within a second sensory modality (in this case audition) increased the salience of the stimulus within the preferred sensory modality, vision (Besle et al. 2004a; Giard and Peronnet 1999). This explanation, although it may account for other multisensory interactions, is also likely not to be responsible for these results. In this paradigm, salience

effects were directly measured (Fig. 5), and whereas increases in stimulus salience (i.e., efficacy) were associated with *increases* in the P1 and P1-N1 amplitudes (Figs. 4, 5), the addition of an auditory component *decreased* the P1 and P1-N1 amplitudes relative to summed unisensory responses (Fig. 6).

A third possible explanation could be that the addition of an auditory stimulus increased the efficiency of processing the visual stimulus, allowing for a reduction of neural activity during combined stimulation (Giard and Peronnet 1999). The correlation between individuals' increases in multisensory gain relative to the reductions in their P1 and P1-N1 amplitudes (Fig. 7) provides support for this hypothesis, as a relative decrease in amplitude is associated with faster RTs. Consistent with this model, the phenomenon of repetition priming is known to speed RT both within and across modalities (Clarke and Morton 1983; Roediger and McDermott 1993), but also to reduce BOLD activation patterns (Henson 2003; Grill-Spector

et al. 2006). It has been hypothesized that the reduction in BOLD activation with priming is the result of an increase in the efficiency of the processes that underlie recognition (James and Gauthier 2006). Thus, the multisensory interaction specific to the P1 and P1-N1 amplitudes is consistent with an increased efficiency of processing, resulting in both speeded RTs and reduced neural activity.

The ability of an additive-factors paradigm to identify multisensory interactions in evoked potentials provides researchers with a methodology to test other interactions between cognitive processes. The additive-factors paradigm, which was originally developed to identify selective influence and interactions in studies of RT (Sternberg 1969, 1998, 2001), has also been successfully applied to other neuroimaging methods, particularly fMRI (James et al. 2009, 2012; Kim and James 2010; Kim et al. 2011; Sartori and Umiltà 2000; Stevenson et al. 2009, 2010; Wallace and Murray 2011). This particular method, relative to the additive criterion, is less susceptible to issues of common activation (see “Appendix”). The application of additive-factors analysis illustrated in topographical maps in this study also provides a number of possible regions in which future experiments may focus their attention (see Fig. 8). It should be noted that any analysis of the ERP waveforms extracted from these regions to detect multisensory interactions would be non-independent from the selection of these electrode clusters given that the electrode clusters were identified by interactions in topographic voltage maps (Kriegeskorte et al. 2009). As such, the interactions observed via the topographic maps in these regions require direct, hypothesis-driven testing, but here, we offer possible hypotheses that may explain these interactions in the topographic voltage maps at the occipital pole, a right temporo-parietal cluster, and vertex.

Two of these regions correspond with potentials that have been historically considered unisensory in nature: a visual response measured at the occipital pole (Fig. 8b) and response at the vertex that is typically considered auditory (Fig. 8d). The interaction effect in the occipital polar scalp showed the same pattern as the (visual) temporal scalp P1-N1-P2 previously discussed. This finding provides converging evidence that auditory processing can modulate processes and brain regions that were once considered to be purely visual. The addition of an auditory stimulus to a visual stimulus has been previously suggested to modulate visual evoked potentials (Brefczynski-Lewis et al. 2009; Giard and Peronnet 1999; Joassin et al. 2004; Klucharev et al. 2003; Molholm et al. 2002), and auditory signals have been shown to modulate visual cortex excitability through enhanced perceptions of TMS-induced visual phosphenes (Romei et al. 2009). Our new data add to this body of knowledge by indicating that this effect is also modulated by stimulus salience. The mechanisms for such interactions

are not clear, however, direct connections between early auditory and visual regions have been identified (Cappe and Barone 2005; Clavagnier et al. 2004; Falchier et al. 2002; Rockland and Ojima 2003). This proposed mechanism has also been used to explain similar decreases in activity within early visual regions with concurrent auditory presentations (Haxby et al. 1994; Kawashima et al. 1995; Laurienti et al. 2002; Senkowski et al. 2011). A different, but equally interesting pattern was seen at the vertex, where the auditory-only stimulus was weak enough as to elicit very little response, yet the comparison of the responses with visual and audiovisual presentations shows a clear interaction. This interaction is in line with results showing the impact of visual inputs on auditory multisensory processing, where stimulation of the visual cortex through TMS has been shown to produce improved auditory and multisensory processing in a manner similar to combining visual stimuli with auditory stimuli (Romei et al. 2007). Coupled together, these findings provide further evidence that there are multisensory interactions in these evoked responses previously described as solely unisensory, but again, it should be noted that these hypotheses require direct testing in a statistically independent manner before more conclusive inferences can be drawn.

Multisensory interactions between auditory and visual speech across salience levels were also measured in the right temporo-parietal region. The temporo-parietal scalp (Fig. 8c) may reflect activation in superior temporal cortex (STC), a known site of audiovisual speech integration (Beauchamp et al. 2004a, b, 2008; Calvert et al. 2000, 2001; Stevenson et al. 2009, 2010, 2011; Stevenson and James 2009; Werner and Noppeney 2009, among others). However, most studies of STC report hemodynamic *increases* when multiple sensory modalities are presented simultaneously. Furthermore, our own additive-factors fMRI study (James et al. 2009, 2012; Stevenson et al. 2009) provided evidence that hemodynamic activity within STC shows an increase in multisensory gain as stimulus salience is decreased, whereas the interaction seen in lateral temporo-parietal electrodes in the current study showed a reduction of multisensory ERP amplitudes relative to summed unisensory amplitudes as stimulus salience decreased. This apparent dichotomy between measured ERP interactions and BOLD interactions might be accounted for by one of three alternatives.

First, the relationship between neural activity, particularly the voltage changes associated with ERP signals, and the BOLD hemodynamic response, is not clearly understood, and it is unlikely that there is a linear relationship between the two (Attwell and Iadecola 2002; Heeger and Ress 2002; Logothetis 2002, 2003; Raichle and Mintun 2006). As such, one cannot immediately assume that neural

activity producing an increase in ERP amplitudes would also produce increases in the BOLD signal. Second, comparing ERP interactions to BOLD interactions is problematic due to the differences in temporal resolution of each. While interactions in the P1 and P1-N1 can be isolated in time using ERP, any interaction measured in the BOLD signal will also include activity spanning temporal epochs well beyond the P1-N1. Third, the ERP interaction measured in lateral temporo-parietal electrodes may originate from another brain region. For example, auditory ERPs that are known to originate in the primary auditory cortex generated ERP activity that is maximal at the vertex (Giard and Perronnet 1999; Ponton et al. 2002; Scarff et al. 2004), whereas fMRI localizes these processes to the superior temporal region (e.g., Brefczynski-Lewis et al. 2009). Differentiating between these three possibilities is beyond the scope of this current study, which has produced testable hypotheses for future studies employing combined EEG-fMRI data collection with neural source modeling.

It is worth noting here that while inverse effectiveness has been a remarkably consistent finding in multisensory studies, there are instances in which it has not been observed. Particularly relevant to the current study is a behavioral experiment in which multisensory gain did not parametrically increase as stimulus salience decreased, but rather peaked at an intermediate salience level (Ross et al. 2007). In contrast, our data showed no evidence of such a “sweet spot” in either behavioral or ERP responses to multisensory speech stimuli. However, while these two paradigms have much in common, there are also a number of significant differences that may have resulted in these seemingly discrepant findings. The paradigm reported here parametrically varied both the auditory and visual SNR and measured multisensory gain relative to both auditory- and visual-only performance. The study of Ross and colleagues varied only the auditory SNR with the visual component of the audiovisual stimuli presented without noise, and with multisensory gain measured only relative to the auditory-only performance. Additionally, differences in SNR level, task, and set size may also produce the differences in findings. While these differences make it difficult to directly compare these two discrepant findings, these differences also highlight a number of future directions that need to be explored.

In summary, the results from this study provide a number of new insights into the neural processing of multisensory speech. We found evidence that the salience of the speech signal modulates the manner with which auditory and visual processes interact. In particular, the multisensory gain seen upon the addition of an auditory stimulus has an inverse effect on the visually evoked N1-P2 responses; gain increases at lower salience levels. The opposite is found on the P1 and P1-N1, where gain

decreases with lower salience speech. These neural interactions as indexed via EEG were also correlated with behavioral measures of RT, with reductions in the P1 and P1-N1 components being associated with greater multisensory gain in RTs, suggesting that the decreases in ERP amplitude can be facilitatory as opposed to inhibitory. It is argued that this negative correlation may represent an increase in processing efficiency when audiovisual speech is processed relative to unisensory visual processing. In addition to more traditional approaches, these results were obtained using an additive-factors experimental design. While this paradigm has been successfully applied in RT and fMRI studies, to our knowledge this is the first application in an ERP study. The additive-factors approach provides a possible means to reduce confounds of common activation and differences in sensory modality-directed attention that have often been critiqued in ERP studies of multisensory integration (Besle et al. 2004b; Giard and Besle 2010; Gondon and Röder 2006). These results, while found with speech, are not necessarily specific to speech. Further studies investigating possible differences between multisensory processes impacted by stimulus salience are warranted to answer such questions.

**Acknowledgments** This research was supported in part by a grant to T. W. James from Indiana University's Faculty Research Support Program administered by the office of the vice provost for research, an NIH NIDCD grant to R. A. Stevenson, 1F32 DC011993, an NIH NIDCD grant to M. T. Wallace, R34 DC010927, and NIH NIDCD training grant T32 DC000012 Training in Speech, Hearing, and Sensory Communication. Thanks to Laurel Stevenson, Karin Harman James, and Jeanne Wallace for their support, to Luis Hernandez for the stimuli, and to Vera Blau for help with the manuscript.

## Appendix

The most commonly used metric of multisensory interactions in ERP research is the additive metric, or model (Barth et al. 1995; Berman 1961; Besle et al. 2004b). The additive model's null hypothesis asserts that the multisensory response should equal the summed responses measured with both unisensory presentations in isolation, an assertion based on the law of superposition of electrical fields (Besle et al. 2004b, 2009; Giard and Besle 2010). While this metric is useful, there are a number of situations where it may produce spurious superadditive, or subadditive results. In particular, variations in attention across sensory modality (including variations in difficulty across sensory modalities or divided attention in bi-sensory conditions relative to unisensory conditions) and common activity create serious concerns about the use of the additive metric in multisensory paradigms (for an in-depth discussion, see Besle et al. 2004b; Giard and Besle 2010;

Gondan and Röder 2006). While the additive metric calculates interactions as:

$$AV \neq A + V.$$

In terms of sensory specific activations, it is more accurately written as:

$$AV + CA \neq (A + CA) + (V + CA),$$

where *CA* refers to common activation, i.e., activation of processes that commonly occur regardless of sensory input. In this case, the common activation is accounted for twice on the right side of the equation, but only once on the left, producing spurious findings of multisensory interactions (Besle et al. 2004b; Giard and Besle 2010). Importantly, the additive factors equation reduces the impact of the common activation, measuring a change in each sensory modality:

$$AV_H - AV_L \neq (A_H - A_L) + (V_H - V_L).$$

In terms of specific sensory activations, again, this equation is more accurately written as:

$$(AV_H + CA) - (AV_L + CA) \neq [(A_H + CA) - (A_L + CA)] + [(V_H + CA) - (V_L + CA)].$$

Here, it should be noted that the impact of common activations, or *CA*, are diminished relative to the classic additive metric equation. Each component of the additive-factors equation includes two common activations that are subtracted from one another, leaving only the difference between common activations associated with levels of the added factor.

Finally, it should also be noted that there are a number of other approaches that have been used to circumvent the issues associated with the additive criterion. One such example is the application of electrical neuroimaging analyses to ERPs that includes assessing not only the response amplitude and timing of responses, but also utilizes response topography. This analysis, in addition to bypassing issues of associated with the additive metric also allows the experimenter to differentiate effects cause by changes in response strength from a given set of generators from effects caused by changes in the configuration of these generators Furthermore, the use of global field power can allow for the identification of the directionality of those interactions (Cappe et al. 2010; Murray et al. 2005, 2008).

## References

- Allison T, Ginter H, McCarthy G, Nobre AC, Puce A, Luby M, Spencer DD (1994a) Face recognition in human extrastriate cortex. *J Neurophysiol* 71(2):821–825
- Allison T, McCarthy G, Nobre A, Puce A, Belger A (1994b) Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cereb Cortex* 4(5):544–554
- Allison T, Puce A, Spencer DD, McCarthy G (1999) Electrophysiological studies of human face perception. I: potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex* 9(5):415–430
- Attwell D, Iadecola C (2002) The neural basis of functional brain imaging signals. *Trends Neurosci* 25(12):621–625
- Barth DS, Goldberg N, Brett B, Di S (1995) The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex. *Brain Res* 678(1–2):177–190
- Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004a) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat Neurosci* 7(11):1190–1192
- Beauchamp MS, Lee KE, Argall BD, Martin A (2004b) Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41(5):809–823
- Beauchamp MS, Yasar NE, Frye RE, Ro T (2008) Touch, sound and vision in human superior temporal sulcus. *NeuroImage* 41(3):1011–1020. doi:10.1016/j.neuroimage.2008.03.015
- Bentin S, Allison T, Puce A, Perez E, McCarthy G (1996) Electrophysiological studies of face perception in humans. *J Cogn Neurosci* 8(6):551–565
- Berman AL (1961) Interaction of cortical responses to somatic and auditory stimuli in anterior ectosylvian gyrus of cat. *J Neurophysiol* 24:608–620
- Besle J, Fort A, Delpuech C, Giard MH (2004a) Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J Neurosci* 20(8):2225–2234. doi:10.1111/j.1460-9568.2004.03670.xEJN3670[pil]
- Besle J, Fort A, Giard M (2004b) Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cogn Process* 5:189–192
- Besle J, Bertrand O, Giard MH (2009) Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. *Hear Res*. doi:10.1016/j.heares.2009.06.016
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10(4):433–436
- Brefczynski-Lewis J, Lowitzsch S, Parsons M, Lemieux S, Puce A (2009) Audiovisual non-verbal dynamic faces elicit converging fMRI and ERP responses. *Brain Topogr* 21(3–4):193–206. doi:10.1007/s10548-009-0093-6
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. *Science* 276(5312):593–596
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10(11):649–657
- Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage* 14(2):427–438
- Cappe C, Barone P (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *Eur J Neurosci* 22(11):2886–2902. doi:10.1111/j.1460-9568.2005.04462.x
- Cappe C, Rouiller EM, Barone P (2009) Multisensory anatomical pathways. *Hear Res* 258(1–2):28–36. doi:10.1016/j.heares.2009.04.017
- Cappe C, Thut G, Romei V, Murray MM (2010) Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *J Neurosci* 30(38):12572–12580. doi:10.1523/JNEUROSCI.1099-10.2010
- Clarke R, Morton J (1983) Cross modality facilitation in tachistoscope word recognition. *Q J Exp Psychol* 35A:79–96
- Clavagnier S, Falchier A, Kennedy H (2004) Long-distance feedback projections to area V1: implications for multisensory integration,

- spatial awareness, and visual consciousness. *Cogn Affect Behav Neurosci* 4(2):117–126
- Diederich A, Colonius H (2004) Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Percept Psychophys* 66(8):1388–1404
- Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci* 22(13):5749–5759
- Falchier A, Schroeder CE, Hackett TA, Lakatos P, Nascimento-Silva S, Ulbert I, Karmos G, Smiley JF (2010) Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. *Cereb Cortex* 20(7):1529–1538. doi:10.1093/cercor/bhp213
- Fort A, Delpuech C, Pernier J, Giard MH (2002a) Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cereb Cortex* 12(10):1031–1039
- Fort A, Delpuech C, Pernier J, Giard MH (2002b) Early auditory–visual interactions in human cortex during nonredundant target identification. *Brain Res Cogn Brain Res* 14(1):20–30
- Giard M, Besle J (2010) Methodological considerations: electrophysiology of multisensory interactions in humans. In: Kaiser J, Naumer MJ (eds) *Multisensory object perception in the primate brain*. Springer, New York
- Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11(5):473–490
- Gondan M, Röder B (2006) A new method for detecting interactions between the senses in event-related potentials. *Brain Res* 1073–1074:389–397. doi:10.1016/j.brainres.2005.12.050
- Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10(1):14–23. doi:10.1016/j.tics.2005.11.006
- Haxby JV, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL (1994) The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *J Neurosci* 14(11 Pt 1):6336–6353
- Heeger DJ, Ress D (2002) What does fMRI tell us about neuronal activity? *Nat Rev* 3(2):142–151. doi:10.1038/nrn730nn730[pii]
- Henson RN (2003) Neuroimaging studies of priming. *Prog Neurobiol* 70(1):53–81. doi:S0301008203000868[pii]
- Hershenson M (1962) Reaction time as a measure of intersensory facilitation. *J Exp Psychol* 63:289–293
- James TW, Gauthier I (2006) Repetition-induced changes in BOLD response reflect accumulation of neural activity. *Hum Brain Mapp* 27(1):37–46
- James TW, Stevenson RA (2012) The use of fMRI to assess multisensory integration. In: Wallace MH, Murray MM (eds) *Frontiers in the neural basis of multisensory processes*. Taylor & Francis, London
- James TW, Stevenson RA, Kim S (2009) Assessing multisensory integration with additive factors and functional MRI. In: *Proceedings of the International Society for Psychophysics*, Dublin
- James TW, Stevenson RA, Kim S (2012) Inverse effectiveness in multisensory processing. In: Stein BE (ed) *The new handbook of multisensory processes*. MIT Press, Cambridge
- Joassin F, Maurage P, Bruyer R, Crommelinck M, Campanella S (2004) When audition alters vision: an event-related potential study of the cross-modal interactions between faces and voices. *Neurosci Lett* 369(2):132–137
- Kawashima R, O'Sullivan BT, Roland PE (1995) Positron-emission tomography studies of cross-modality inhibition in selective attentional tasks: closing the “mind’s eye”. *Proc Natl Acad Sci USA* 92(13):5969–5972
- Kim S, James TW (2010) Enhanced effectiveness in visuo-haptic object-selective brain regions with increasing stimulus salience. *Hum Brain Mapp* 31(5):678–693. doi:10.1002/hbm.20897
- Kim S, Stevenson RA, James TW (2011) Visuo-haptic neuronal convergence demonstrated with an inversely effective pattern of BOLD activation. *J Cogn Neurosci* 24(4):830–842. doi: 10.1162/jocn\_a\_00176
- Klucharev V, Mottonen R, Sams M (2003) Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Brain Res Cogn Brain Res* 18(1):65–75
- Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12(5):535–540. doi:10.1038/nn.2303
- Lachs L, Hernandez LR (1998) Update: the Hoosier Audiovisual Multitalker Database. Research on spoken language processing. Speech Research Laboratory, Indiana University, Bloomington
- Laming D, Laming J (1992) F. Hegelmaier: on memory for the length of a line. *Psychol Res* 54(4):233–239
- Latinus M, VanRullen R, Taylor MJ (2010) Top-down and bottom-up modulation in processing bimodal face/voice stimuli. *BMC Neurosci* 11:36. doi:10.1186/1471-2202-11-36
- Laurienti PJ, Burdette JH, Wallace MT, Yen YF, Field AS, Stein BE (2002) Deactivation of sensory-specific cortex by cross-modal stimuli. *J Cogn Neurosci* 14(3):420–429
- Logothetis NK (2002) The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos Trans R Soc London* 357(1424):1003–1037
- Logothetis NK (2003) The underpinnings of the BOLD functional magnetic resonance imaging signal. *J Neurosci* 23(10):3963–3971
- Lovelace CT, Stein BE, Wallace MT (2003) An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Brain Res Cogn Brain Res* 17(2):447–453
- Luce PA, Pisoni DB (1998) Recognizing spoken words: the neighborhood activation model. *Ear Hear* 19(1):1–36
- Macaluso E, Frith CD, Driver J (2000) Modulation of human visual cortex by crossmodal spatial attention. *Science* 289(5482):1206–1208
- 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(7):1672–1679. doi:10.1093/cercor/bhl077
- McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264(5588):746–748
- Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. *Science* 221(4608):389–391
- Meredith MA, Stein BE (1986) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol* 56(3):640–662
- Miller J (1982) Divided attention: evidence for coactivation with redundant signals. *Cogn Psychol* 14(2):247–279
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* 14(1):115–128
- Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cereb Cortex* 14(4):452–465
- Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ (2005) Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb Cortex* 15(7):963–974. doi:10.1093/cercor/bhh197
- Murray MM, Brunet D, Michel CM (2008) Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr* 20(4):249–264. doi: 10.1007/s10548-008-0054-5

- Musacchia G, Schroeder CE (2009) Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hear Res* 258(1–2):72–79. doi:[10.1016/j.heares.2009.06.018](https://doi.org/10.1016/j.heares.2009.06.018)
- Nelson WT, Hettinger LJ, Cunningham JA, Brickman BJ, Haas MW, McKinley RL (1998) Effects of localized auditory information on visual target detection performance using a helmet-mounted display. *Hum Factors* 40(3):452–460
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vis* 10(4):437–442
- Perrault TJ Jr, Vaughan JW, Stein BE, Wallace MT (2003) Neuron-specific response characteristics predict the magnitude of multisensory integration. *J Neurophysiol* 90(6):4022–4026. doi:[10.1152/jn.00494.2003](https://doi.org/10.1152/jn.00494.2003)
- 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(3):407–420. doi:[10.1016/j.clinph.2002.04.017](https://doi.org/10.1016/j.clinph.2002.04.017)
- Puce A, Epling JA, Thompson JC, Carrick OK (2007) Neural responses elicited to face motion and vocalization pairings. *Neuropsychologia* 45(1):93–106. doi:[10.1016/j.neuropsychologia.2006.04.017](https://doi.org/10.1016/j.neuropsychologia.2006.04.017)
- Raab DH (1962) Statistical facilitation of simple reaction times. *Trans New York Acad Sci* 24:574–590
- Raichle ME, Mintun MA (2006) Brain work and brain imaging. *Ann Rev Neurosci* 29:449–476. doi:[10.1146/annurev.neuro.29.0516.05.112819](https://doi.org/10.1146/annurev.neuro.29.0516.05.112819)
- Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. *Int J Psychophysiol* 50(1–2):19–26
- Roediger HL III, McDermott KB (1993) Implicit memory in normal subjects. In: Boller F, Grafman J (eds) *Handbook of neuropsychology*. Elsevier, Amsterdam, pp 63–131
- Romei V, Murray MM, Merabet LB, Thut G (2007) Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: implications for multisensory interactions. *J Neurosci* 27(43):11465–11472. doi:[10.1523/JNEUROSCI.2827-07.2007](https://doi.org/10.1523/JNEUROSCI.2827-07.2007)
- Romei V, Murray MM, Cappe C, Thut G (2009) Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. *Curr Biol* 19(21):1799–1805. doi:[10.1016/j.cub.2009.09.027](https://doi.org/10.1016/j.cub.2009.09.027)
- Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ (2007) Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb Cortex* 17(5):1147–1153. doi:[10.1093/cercor/bhl024](https://doi.org/10.1093/cercor/bhl024)
- Sartori G, Umiltà C (2000) The additive factor method in brain imaging. *Brain Cogn* 42(1):68–71. doi:[10.1006/brcg.1999.1164](https://doi.org/10.1006/brcg.1999.1164)
- Senkowski D, Saint-Amour D, Hofle M, Foxe JJ (2011) Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *NeuroImage*. doi:[10.1016/j.neuroimage.2011.03.075](https://doi.org/10.1016/j.neuroimage.2011.03.075)
- Scarff CJ, Reynolds A, Goodyear BG, Ponton CW, Dort JC, Eggermont JJ (2004) Simultaneous 3-T fMRI and high-density recording of human auditory evoked potentials. *NeuroImage* 23(3):1129–1142. doi:[10.1016/j.neuroimage.2004.07.035](https://doi.org/10.1016/j.neuroimage.2004.07.035)
- Sheffert SM, Lachs L, Hernandez LR (1996) The Hooiser audiovisual multitalker database. Research on spoken language processing. Speech Research Laboratory, Indiana University, Bloomington
- Smiley JF, Falchier A (2009) Multisensory connections of monkey auditory cerebral cortex. *Hear Res* 258(1–2):37–46. doi:[10.1016/j.heares.2009.06.019](https://doi.org/10.1016/j.heares.2009.06.019)
- Soto-Faraco S, Navarra J, Alsius A (2004) Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition* 92(3):B13–B23
- Stanford TR, Quessy S, Stein BE (2005) Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J Neurosci* 25(28):6499–6508
- Stein B, Meredith MA (1993) *The merging of the senses*. MIT Press, Boston
- Stein BE, Wallace MT (1996) Comparisons of cross-modality integration in midbrain and cortex. *Prog Brain Res* 112:289–299
- Stein BE, Huneycutt WS, Meredith MA (1988) Neurons and behavior: the same rules of multisensory integration apply. *Brain Res* 448(2):355–358
- Stein BE, Stanford TR, Ramachandran R, Perrault TJ Jr, Rowland BA (2009) Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. *Exp Brain Res* 198(2–3):113–126. doi:[10.1007/s00221-009-1880-8](https://doi.org/10.1007/s00221-009-1880-8)
- Stekelenburg JJ, Vroomen J (2007) Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci* 19(12):1964–1973. doi:[10.1162/jocn.2007.19.12.1964](https://doi.org/10.1162/jocn.2007.19.12.1964)
- Sternberg S (1969) Memory-scanning: mental processes revealed by reaction-time experiments. *Am Sci* 57(4):421–457
- Sternberg S (1998) Discovering mental processing stages: the method of additive factors. In: Scarborough D, Sternberg S (eds) *An invitation to cognitive science: volume 4, methods, models, and conceptual issues*, vol 4. MIT Press, Cambridge, pp 739–811
- Sternberg S (2001) Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychol* 106:147–246
- Stevenson RA, James TW (2009) Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage* 44(3):1210–1223. doi:[10.1016/j.neuroimage.2008.09.034](https://doi.org/10.1016/j.neuroimage.2008.09.034)
- Stevenson RA, Geoghegan ML, James TW (2007) Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Exp Brain Res* 179(1):85–95
- Stevenson RA, Kim S, James TW (2009) An additive-factors design to disambiguate neuronal and areal convergence: measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI. *Exp Brain Res* 198(2–3):183–194. doi:[10.1007/s00221-009-1783-8](https://doi.org/10.1007/s00221-009-1783-8)
- Stevenson RA, Altieri NA, Kim S, Pisoni DB, James TW (2010) Neural processing of asynchronous audiovisual speech perception. *NeuroImage* 49(4):3308–3318. doi:[10.1016/j.neuroimage.2009.12.001](https://doi.org/10.1016/j.neuroimage.2009.12.001)
- Stevenson RA, VanDerKlok RM, Pisoni DB, James TW (2011) Discrete neural substrates underlie complementary audiovisual speech integration processes. *NeuroImage* 55(3):1339–1345. doi:[10.1016/j.neuroimage.2010.12.063](https://doi.org/10.1016/j.neuroimage.2010.12.063)
- Sumbly WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212–215
- Teder-Salejari WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Brain Res Cogn Brain Res* 14(1):106–114
- van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci USA* 102(4):1181–1186. doi:[10.1073/pnas.0408949102](https://doi.org/10.1073/pnas.0408949102)
- Vroomen J, Stekelenburg JJ (2010) Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *J Cogn Neurosci* 22(7):1583–1596. doi:[10.1162/jocn.2009.21308](https://doi.org/10.1162/jocn.2009.21308)
- Wallace MH, Murray MM (eds) (2011) *Frontiers in the neural basis of multisensory processes*. Taylor & Francis, London
- Wallace MT, Meredith MA, Stein BE (1992) Integration of multiple sensory modalities in cat cortex. *Exp Brain Res* 91(3):484–488
- Wallace MT, Meredith MA, Stein BE (1993) Converging influences from visual, auditory, and somatosensory cortices onto output

- neurons of the superior colliculus. *J Neurophysiol* 69(6):1797–1809
- Wallace MT, Wilkinson LK, Stein BE (1996) Representation and integration of multiple sensory inputs in primate superior colliculus. *J Neurophysiol* 76(2):1246–1266
- Wallace MT, Meredith MA, Stein BE (1998) Multisensory integration in the superior colliculus of the alert cat. *J Neurophysiol* 80(2): 1006–1010
- Watkins S, Shams L, Tanaka S, Haynes JD, Rees G (2006) Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage* 31(3):1247–1256. doi:[10.1016/j.neuroimage.2006.01.016](https://doi.org/10.1016/j.neuroimage.2006.01.016)
- Watkins S, Shams L, Josephs O, Rees G (2007) Activity in human V1 follows multisensory perception. *NeuroImage* 37(2):572–578. doi:[10.1016/j.neuroimage.2007.05.027](https://doi.org/10.1016/j.neuroimage.2007.05.027)
- Werner S, Noppeney U (2009) Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cereb Cortex*. doi:[10.1093/cercor/bhp248](https://doi.org/10.1093/cercor/bhp248)
- Werner S, Noppeney U (2010) Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *J Neurosci* 30(7):2662–2675. doi:[10.1523/JNEUROSCI.5091-09.2010](https://doi.org/10.1523/JNEUROSCI.5091-09.2010)
- Wilkinson LK, Meredith MA, Stein BE (1996) The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. *Exp Brain Res* 112(1):1–10