NeuroImage 44 (2009) 1210–1223

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition

Ryan A. Stevenson *, Thomas W. James

Department of Psychological and Brain Sciences, Indiana University, 1101 East Tenth Street, Room 293, Bloomington, IN 47405, USA

ARTICLE INFO

Article history:
Received 18 April 2008
Revised 8 August 2008
Accepted 9 September 2008
Available online 10 October 2008

Keywords:
Multisensory integration
Vision
Superadditivity
Perception
fMRI
Neuroimaging
Auditory
Speech perception

ABSTRACT

The superior temporal sulcus (STS) is a region involved in audiovisual integration. In non-human primates, multisensory neurons in STS display inverse effectiveness. In two fMRI studies using multisensory tool and speech stimuli presented at parametrically varied levels of signal strength, we show that the pattern of neural activation in human STS is also inversely effective. Although multisensory tool-defined and speech-defined regions of interest were non-overlapping, the pattern of inverse effectiveness was the same for tools and speech across regions. The findings suggest that, even though there are sub-regions in STS that are speech-selective, the manner in which visual and auditory signals are integrated in multisensory STS is *not* specific to speech.

Introduction

Much of what we know about the neural mechanisms of human sensory integration comes from non-human models. Based on solid supporting evidence replicated in a number of studies, three rules that govern the properties of multisensory neurons have been proposed (Holmes, 2007; Stanford et al., 2005). The spatial rule states that the components of a multisensory stimulus are more effectively integrated (either increased behavioral or neural responses relative to unisensory stimuli) when they originate from congruent spatial locations (Meredith and Stein, 1986a, 1996). The temporal rule states that the components of a multisensory stimulus are more effectively integrated when they occur simultaneously, rather than consecutively (Meredith et al., 1987; Miller and D'Esposito, 2005; Senkowski et al., 2007). Finally, the rule of inverse effectiveness states that components of a multisensory stimulus are more effectively integrated when the salience of those components is relatively weak (Kayser et al., 2005; Meredith and Stein, 1983, 1986b; Perrault et al., 2005; Stanford et al., 2005).

These rules, and the models that instantiate them, have been instrumental in molding our understanding of human sensory integration, even though a majority of the evidence comes from non-human animals. While results of behavioral studies in humans suggest that these models can be generalized to humans (Bolognini et al., 2005; Frens et al., 1995; Hairston et al., 2003a,b, 2006; Mozolic et al., 2007; Serino et al., 2007), the results of neuroimaging studies in humans are more controversial (Beauchamp, 2005). While the spatial, temporal, and inverse effectiveness rules have been used to characterize the properties of neuronal convergence (where individual neurons receive information from both modality-specific streams) in animal studies of sensory integration, they have not been well explored in humans. Specifically, while experimenters have begun to investigate the spatial and temporal rules in humans (Bushara et al., 2001; Macaluso et al., 2004), neuronal inverse effectiveness has never been shown. To incorporate fully what we have learned from animal models of sensory integration with human studies of the same phenomena, these properties must be clarified. Here, we will specifically address the lack of evidence for neuronal inverse effectiveness with audiovisual stimuli in human cortex.

A second outstanding issue in the field of sensory integration is whether or not speech stimuli are integrated in the same manner as other environmental sounds. For instance, it has been hypothesized that speech, or stimuli perceived as speech, are integrated through special mechanisms (Kuhl et al., 1991; Stekelenburg and Vroomen, 2007; Tuomainen et al., 2005). In neuroimaging studies, this has been supported both anatomically and functionally within superior temporal sulcus (STS). Anatomically, responses in STS to visual facial movements, particularly of the mouth and eyes, are more anterior than other forms of biological motion (Johnson-Frey et al., 2005; Materna et al., 2007; Pelphrey et al., 2005; Puce et al., 1998; Winston et al., 2004). Perhaps not coincidentally, intelligible auditory-speech stimuli activate anterior STS whereas non-intelligible, speech-like

* Corresponding author. Fax: (812) 855 4691.
E-mail address: stevenra@indiana.edu (R.A. Stevenson).

1053-8119/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2008.09.034

stimuli do not (Scott et al., 2000). Functionally, presentations of audiovisual speech have shown superadditive responses within STS (Calvert et al., 2000), whereas objects have only been shown to elicit superadditivity when they are highly degraded (Stevenson et al., 2007), and have often been shown to *not* evoke a superadditive response otherwise (Beauchamp, 2005; Beauchamp et al., 2004a,b).

The goal of the following studies was twofold: first, to attempt to characterize inverse effectiveness in human cortex, and second, to investigate possible differences and similarities between mechanisms of integration of the auditory and visual sensory streams in human STS with regards to speech and object recognition. Two experiments are reported investigating BOLD responses in human STS, an area known to show multisensory integration for audiovisual stimuli in humans (Beauchamp, 2005; Calvert, 2001; Calvert et al., 2000; Stevenson et al., 2007), non-human primates (Barraclough et al., 2005; Benevento et al., 1977; Bruce et al., 1981; Hikosaka et al., 1988) and other mammals (Allman and Meredith, 2007). These two experiments use a range of stimulus-salience levels, which produce parametrically varying levels of behavioral performance, to test whether or not BOLD activation in human STS follows the law of inverse effectiveness with both speech and object stimuli. Changes in unisensory activation are compared to changes in multisensory activation to assess inverse effectiveness in multisensory integration. Additionally, functional localization of regions of STS involved in the integration of audiovisual speech and independent localization of STS regions involved in audiovisual object integration allow for direct anatomical and functional comparisons of these regions within STS.

Methods and materials

Experiment 1: Fast event-related design with tool stimuli

Participants

Participants included eleven right-handed subjects (6 female, mean age = 25.9). Experimental protocol was approved by the Indiana University Institutional Review Board and Human Subjects Committee.

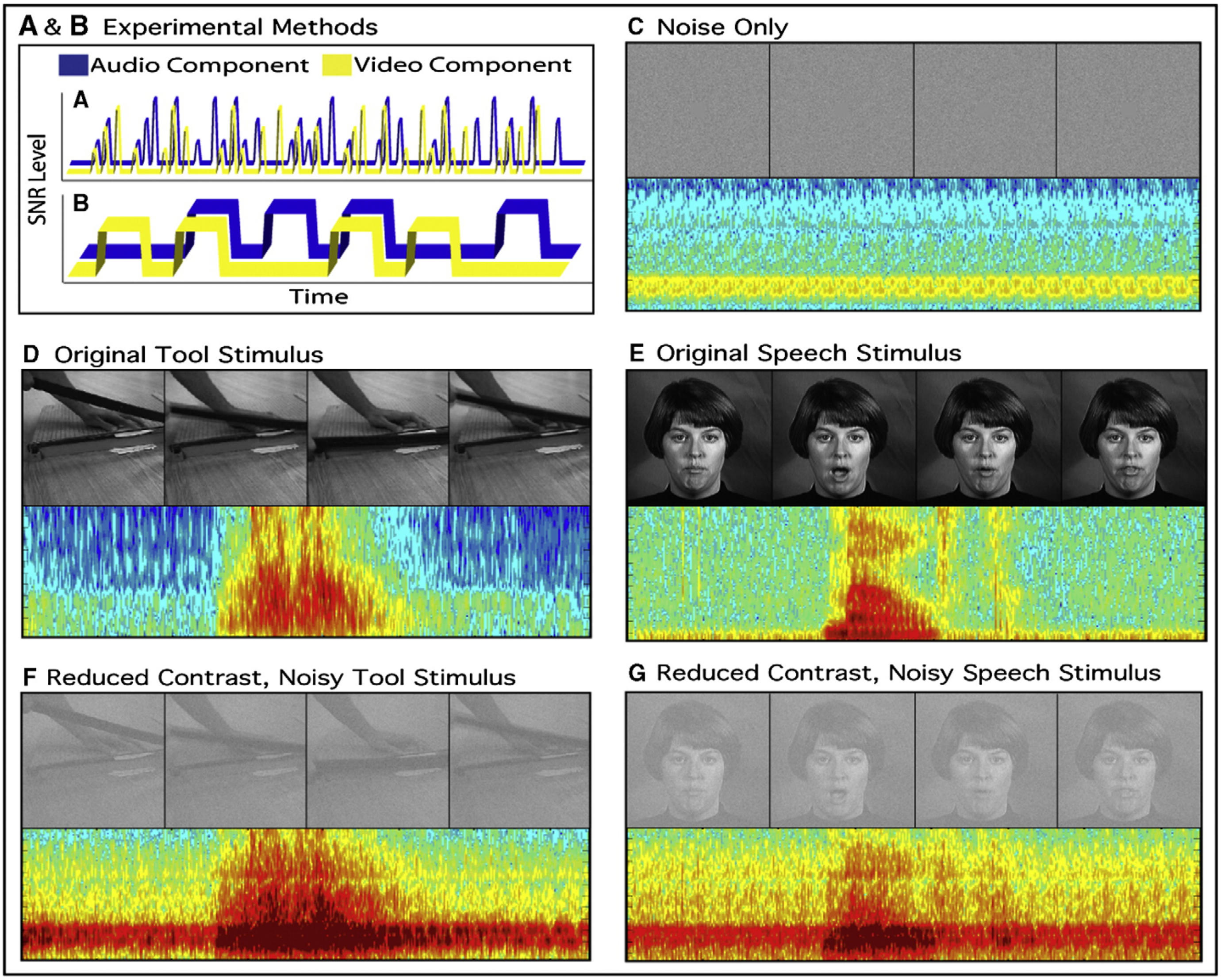

A & B Experimental Methods
Audio Component
Video Component
A
B
SNR Level
Time
C Noise Only
D Original Tool Stimulus
E Original Speech Stimulus
F Reduced Contrast, Noisy Tool Stimulus
G Reduced Contrast, Noisy Speech Stimulus

Fig. 1. Methods and stimuli. Experimental runs were conducted using an event-related design with stimulus presentations at various SNR-levels (A), while localizer runs used a blocked design with high-SNR level stimuli (B). Stimuli from Experiment 1 consisted dynamic videos and sounds of hand-held tools (D), with the video presented here as individual frames, and the audio as a spectrogram (red = high power spectral density, blue = low power spectral density). Visual white noise and audio scanner noise (C) were added to the original stimuli, and the SNR was reduced for experimental runs (see F for an example). Stimuli from Experiment 2 consisted of dynamic videos and sounds of single word utterances (D). Visual white noise and audio scanner noise (C) were added to the original stimuli, and the SNR was reduced for experimental runs (see G for an example).

Stimuli

Stimuli consisted of two-second, dynamic AV recordings of manual tools including hammers and paper cutters (16 stimuli in all were used, with varying rhythms, viewpoints, and speeds, to ensure that recognition was based upon the object itself as opposed to a specific exemplar) (see Fig. 1D). Participants' individual psychophysical thresholds were found while in an MRI simulator designed to mimic the actual fMRI. Interleaved two-down-one-up and one-down-one-up staircases were used. Participants performed a two-alternative forced-choice (2AFC) decision: whether the stimulus was a hammer or a paper cutter. During the audio task, pre-recorded scanner noise was played at an equal decibel level to the actual MRI. For both the A and V tasks, dynamic noise (standard deviation=0.1 for A and 0.0118 for V) was added to the signal (see Fig. 1C). Signal level was measured as root mean square contrast for both A and V stimuli. Signal level was varied across staircase trials while noise contrast was held constant, thus varying the signal-to-noise ratio (SNR) (see Fig. 1F). Participants completed 400 trials in each modality. The data gathered during the staircase procedures were then fitted with psychometric Weibull functions. Stimulus levels used during the fMRI testing procedure were derived from the fitted functions for each participant for A and V stimuli at five SNR levels corresponding to 55, 65, 75, 85, and 95% accuracy.

SNR of AV presentations, which combined A and V stimulus components into a multisensory stimulus, were not determined from AV staircases. Instead, AV presentations were derived by combining A and V components of the same signal level. For instance, the 75% AV condition was the simultaneous presentation of the 75% A and the 75% V stimuli.

All stimuli were presented using MATLAB 5.2 (MATHWORKS Inc., Natick, MA) software with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997), running on a Macintosh computer. V stimuli were projected onto a frosted glass screen using a Mitsubishi XL30U projector. V stimuli were 200×200 pixels and subtended 10.3×10.3 of visual angle. Audio stimuli were presented using pneumatic headphones.

Scanning procedures

Each imaging session included two phases: functional localizers and experimental scans. Functional localizers consisted of high-SNR stimuli (i.e., not degraded by lowering contrast, see Fig. 1F) presented in a blocked stimulus design (see Fig. 1B) while participants completed an identification task. Each run began with the presentation of a fixation cross for 12 s followed by six blocks of A, V, or AV stimuli. AV stimuli were always congruent, which has been shown to be a critical factor in multisensory facilitation (Laurienti et al., 2004). Each run included two blocks of each stimulus type, with blocks consisting of eight, two-second stimulus presentations, separated by 0.1 s inter-stimuli intervals (ISI). New blocks began every 28 s separated by fixation. Runs ended with 12 s of fixation. Block orders were counterbalanced across runs and participants. Each participant completed two functional localizer runs.

During experimental runs, stimuli were presented in a fast event-related design (see Fig. 1A) in which participants performed a 2AFC task, discriminating hammers from paper cutters. Each run included A, V, and AV stimuli from three of the five SNR levels (recall that AV stimuli were a combination of A and V stimuli at their respective unisensory SNR levels), as well as blank trials, which consisted of A and V noise without an A or V signal (henceforth to be referred to as 'baseline' in experimental runs) (see Fig. 1C). Runs began with the presentation of a fixation cross for 12 s, followed by seven trials of each stimulus type, for a total of 70 trials per run. For the seven trials of each stimulus type, four trials were preceded by a two-second ISI, two preceded by a four-second ISI, and one by a six-second ISI, with ISIs consisting of a static visual fixation cross. Runs concluded with 12 s of fixation. Trial and ISI orders were counterbalanced across runs and run order was counterbalanced across participants. Each participant completed 10 experimental runs, for a total of 42 trials per condition.

Imaging parameters and analysis

Imaging was carried out using a Siemens Magnetron Trio 3-T whole body scanner, and collected on an eight-channel phased-array head coil. The field of view was 22×22×9.9 cm, with an in plane resolution of 64×64 pixels and 33 axial slices per volume (whole brain), creating a voxel size of 3.44×3.44×3 mm. Images were collected using a gradient echo EPI (TE=30 ms, TR=2000 ms, flip angle=70°) for BOLD imaging. High-resolution T1-weighted anatomical volumes were acquired using Turbo-flash 3-D (TI=1100 ms, TE=3.93 ms, TR=14.375 ms, flip angle=12°) with 160 sagittal slices with a thickness of 1 mm and field of view of 224×256 (voxel size=1×1×1 mm).

Imaging data were pre-processed using Brain Voyager ™ 3-D analysis tools. Anatomical volumes were transformed into a common stereotactic space (Talaraich and Tournoux, 1988). Functional data were aligned to the first volume of the run closest in time to the anatomical data collection. Each functional run was then aligned to the transformed anatomical volumes, transforming the functional data to a common stereotactic space across participants. Functional data underwent a linear trend removal, 3-D spatial Gaussian filtering (FWHM 6 mm), slice scan time correction, and 3-D motion correction.

Whole-brain, random-effects (RFX) statistical parametric maps (SPM) were calculated using Brain Voyager™ general linear model (GLM) procedure. Event-related averages (ERA), consisting of aligning

Table 1
Region of interests with tool and speech stimuli

	ROI	*X*-coordinate	*Y*-coordinate	*Z*-coordinate	# of voxels
Experiment 1 (tool)	R audiovisual	50.0	−40.1	11.3	642.4
	L audiovisual	−47.7	−40.1	10.6	754.3
	R audio	52.3	−15.1	8.9	3267.9
	L audio	−50.0	−18.9	7.8	3948.6
	R visual	29.0	−73.7	1.9	3054.7
	L visual	−33.1	−70.5	0.6	3348.9
Experiment 2 (tool)	R audiovisual	54.3	−45.0*	15.4*	2047.4
	L audiovisual	−47.5	−46.6*	11.9	1395.3
	R audio	53.8	−17.5	10.7	13019.1
	L audio	−51.0	−22.7*	11.3*	11755.6
	R visual	30.7	−74.5	−2.8	17796.4
	L visual	−27.1	−75.98	0.3	15737.6
Experiment 2 (speech)	R audiovisual	50.8	−35.8*	9.8*	3315.6
	L audiovisual	−48.2	−38.2*	11.1	1861.9
	R audio	56.0	−16.6	10.5	4041.6
	L audio	−43.5	−18.1*	8.9*	1895.4
	R visual	25.9	−81.9	−1.5	6288.8
	L visual	−23.5	−80.5	−3.9	5543.0

* Indicates a significant difference between tool and speech ROIs in Experiment 2.

Table 2
Group SPM regions of activation with tool stimuli (event-related runs)

Contrast	Region	BA	*X*	*Y*	*Z*	mm3
Audio-only:	[a]Left Heschel's gyrus	42				
High-SNR speech >	[b]Right Heschel's gyrus	13	−52	−20	10	9875
Low-SNR speech	[a]Left superior temporal gyrus	21/22				
(Experiment 2)	[a]Left superior temporal sulcus	42				
	[b]Right superior temporal gyrus	13	52	−20	10	10037
	[b]Right superior temporal sulcus	21/22				
	Left inferior occipital gyrus/Left lingual gyrus	19	−24	−76	−13	355
	Right inferior occipital gyrus/Right lingual gyrus	19	22	−83	−10	328
	Left middle occipital gyrus	18	−24	−91	4	90
	Right middle occipital gyrus	18	24	−91	6	456
Visual-only:	Left fusiform gyrus	37	−23	−44	−9	282
High-SNR Speech >	Right fusiform gyrus	37	36	−40	−14	539
Low-SNR Speech	Left anterior inferior frontal gyrus	45	−32	32	−4	288
(Experiment 2)	Right anterior inferior frontal gyrus	45	33	32	1	23
	Left posterior inferior frontal gyrus	9	−46	9	22	929
	Right posterior inferior frontal gyrus	9	48	13	27	27
	Left medial occipital gyrus	19	−38	−70	−3	19
	Right medial occipital gyrus	19	44	−66	−11	1577
	Left anterior superior temporal gyrus	38	−45	10	−12	1446
	Right anterior superior temporal gyrus	38	47	16	−9	598
	Left anterior superior temporal sulcus	21	−56	−18	−5	2728
	Left superior temporal sulcus	22	−42	−58	13	520
	Right superior temporal sulcus	22	−49	−54	20	215
Audiovisual:	Left anterior cingulate gyrus	24	−5	17	44	1130
High-SNR Speech >	Right anterior cingulate gyrus	24	2	6	45	288
Low-SNR Speech	Left fusiform gyrus	37	−25	−48	−11	41
(Experiment 2)	Right fusiform gyrus	37	35	−59	−13	184
	[c]Left Heschel's gyrus	42				
	[c]Left superior temporal gyrus	13	−52	−22	8	25434
	[c]Left superior temporal sulcus	21/22				
	[d]Right Heschel's gyrus	42				
	[d]Right superior temporal gyrus	13	54	−20	10	18656
	[d]Right superior temporal sulcus	21/22				
	Left medial occipital gyrus	19	−39	−70	−5	344
	Right medial occipital gyrus	19	36	−67	0	357

[a] Superscripted letters indicate contiguous activations over several adjacent brain regions.

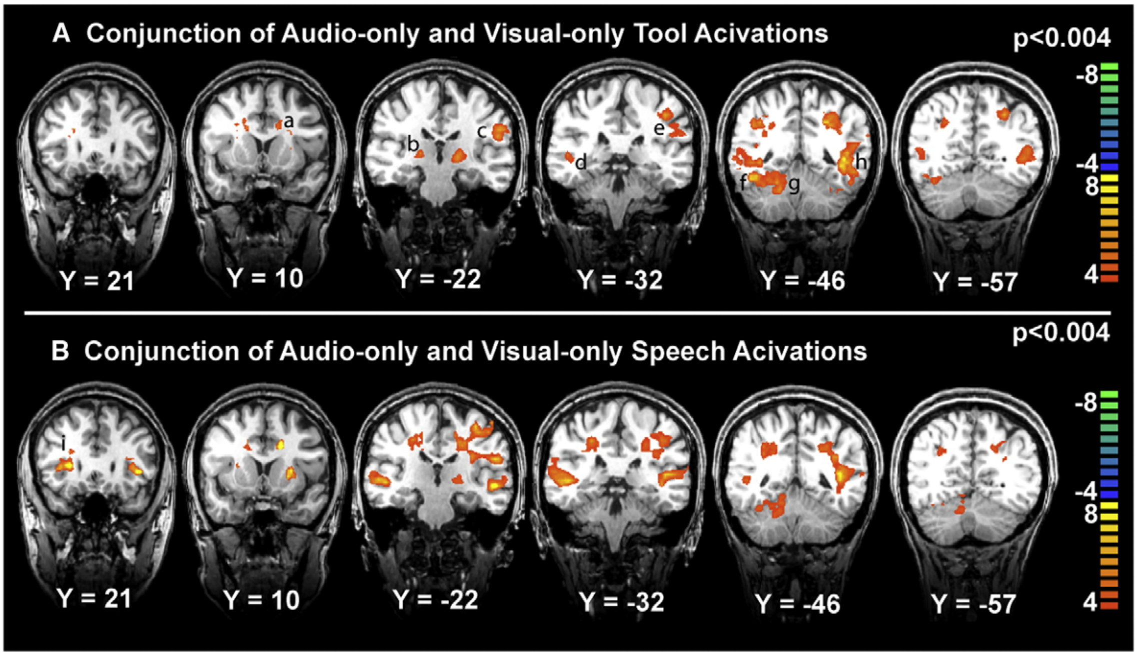

A Conjunction of Audio-only and Visual-only Tool Acivations
p<0.004
-8
-4
8
4
a
b
c
d
e
f
g
h
Y = 21
Y = 10
Y = -22
Y = -32
Y = -46
Y = -57
B Conjunction of Audio-only and Visual-only Speech Acivations
p<0.004
-8
-4
8
4
i
Y = 21
Y = 10
Y = -22
Y = -32
Y = -46
Y = -57

Fig. 2. Conjunction of activations in response to unisensory presentations of audio-only and visual-only tools and speech. Whole-brain, RFX SPMs of activations defined by the contrast (A>rest) ∩ (V>rest) from blocked localizer runs with both tools (A) and speech (B) stimuli in Experiment 2. Activations were similar, and in both cases indicated bilateral STS activations, which were subsequently defined separately for each individual as multisensory ROIs according to that individual's data. Other regions of activation are labeled by letters, and include: (a) anterior cingulated gyrus, (b) thalamus, (c) post-central gyrus, (d) superior temporal sulcus, (e) inferior parietal lobule, (f) fusiform gyrus, (g) cerebellum, (h) Heschel's gyrus, superior temporal gyrus, and superior temporal sulcus, and (i) insula. For region of activity details, see Table 2.

and averaging all trials from each condition to stimulus onset, were created based on stimulus type for both the localizer and the experimental study. Hemodynamic response amplitudes were defined as the arithmetic mean of the time course within a time window 6–16 s after block onset for the localizer runs, and a window of 4–6 s after trial onset for the fast event-related experimental runs.

Experiment 2: Fast event-related design with speech stimuli

Participants

Participants included eleven right-handed subjects (6 female, mean age=24.4). All subjects were native speakers of English. Experimental protocol was approved by the Indiana University Institutional Review Board and Human Subjects Committee.

Stimuli

Stimuli included dynamic, AV recordings of a female actor saying ten nouns (see Fig. 1e). Stimuli were selected from a previously published stimulus set, The Hoosier Audiovisual Multi-Talker Database (Sheffert et al., 1996). All word stimuli were spoken by speaker F1. We selected words that were monosyllabic, had the highest levels of accuracy on both visual-only and audio-only discrimination (Lachs and Hernandez, 1998), and that resided in low-density neighborhoods (Sheffert et al., 1996). From the set of words that matched these criteria, we selected 10 such that they fell into two easily distinguishable semantic categories, and had mean word length approximately equal across categories. The two categories were body parts (face, leg, mouth, neck, and teeth) and environmental words (beach, dirt, rain, rock, and sand). Mean body-part-word duration was 1.562 s, and mean environmental-word duration was 1.582 s. Participants' psychophysical thresholds for A and V stimuli were found using the same staircase procedure described for Experiment 1, but the 2AFC task was to discriminate between body-part and environmental words. Means of stimuli presentation were identical to those in Experiment 1 (see Fig. 1G).

Scanning procedures

Each imaging session included two phases: functional localizers and experimental scans. Functional localizers included blocks of both tool stimuli from Experiment 1 as well as the previously described speech stimuli. Runs consisted of high-SNR stimuli presented in a blocked stimulus design (see Fig. 1B) while participants completed an identification task. Each run began with the presentation of a fixation cross for 12 s followed by 1 block of A, V, and AV tool stimuli, and 1 block of A, V, and AV speech stimuli. AV stimuli were always congruent with blocks consisting of eight, two-second stimuli presentations, separated by 0.1 s ISI. New blocks began every 28 s

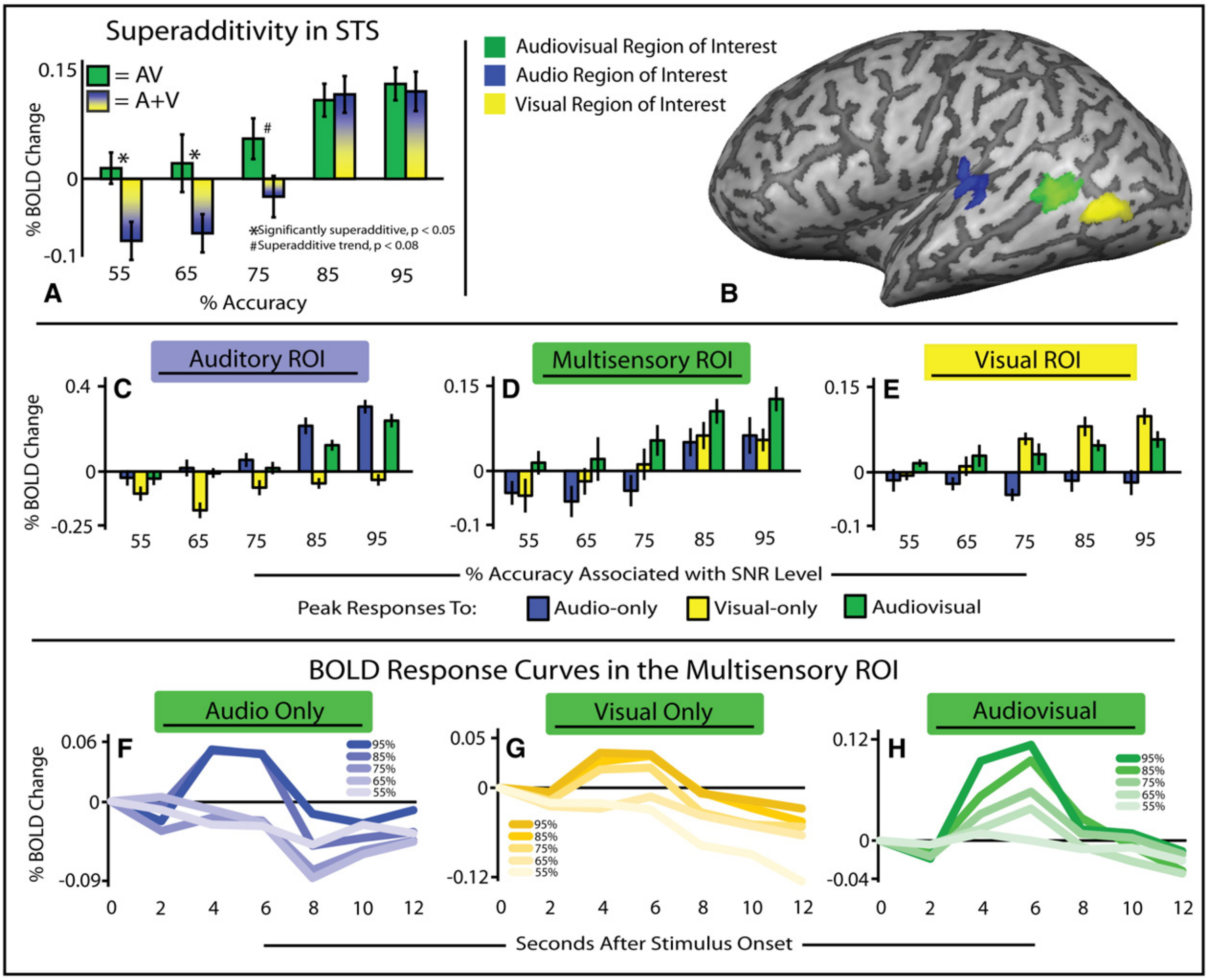


Fig. 3. Response from individual ROIs with object stimuli in Experiment 1. Unisensory audio (STG), unisensory visual (LOC), and multisensory audiovisual (STS) tool ROIs were identified in each individual (see B for an example subjects ROIs). BOLD response curves were extracted in the multisensory ROI with A-only, V-only, and AV tool stimuli at each SNR level (F, G, and H, respectively). BOLD response amplitudes were calculated in the multisensory ROI (D), as well as in the auditory (C) and visual (E) ROIs. The multisensory ROIs responses with unisensory tool stimuli were summed and compared to the response with multisensory stimuli (A). Responses with high SNR-level stimuli did not show superadditivity, but as SNR decreased, multisensory enhancement became stronger, implying inverse effectiveness.

separated by fixation. Runs ended with 12 s of fixation. Block orders were counterbalanced across runs and participants. Each participant completed 4 functional localizer runs.

Experimental runs were identical to those in Experiment 1 but used only speech stimuli, with subjects performing a discrimination task between word categories. Each participant completed 10 experimental runs, for a total of 42 trials per condition.

Imaging parameters and analysis

All imaging parameters and preprocessing analysis were identical to those in Experiment 1.

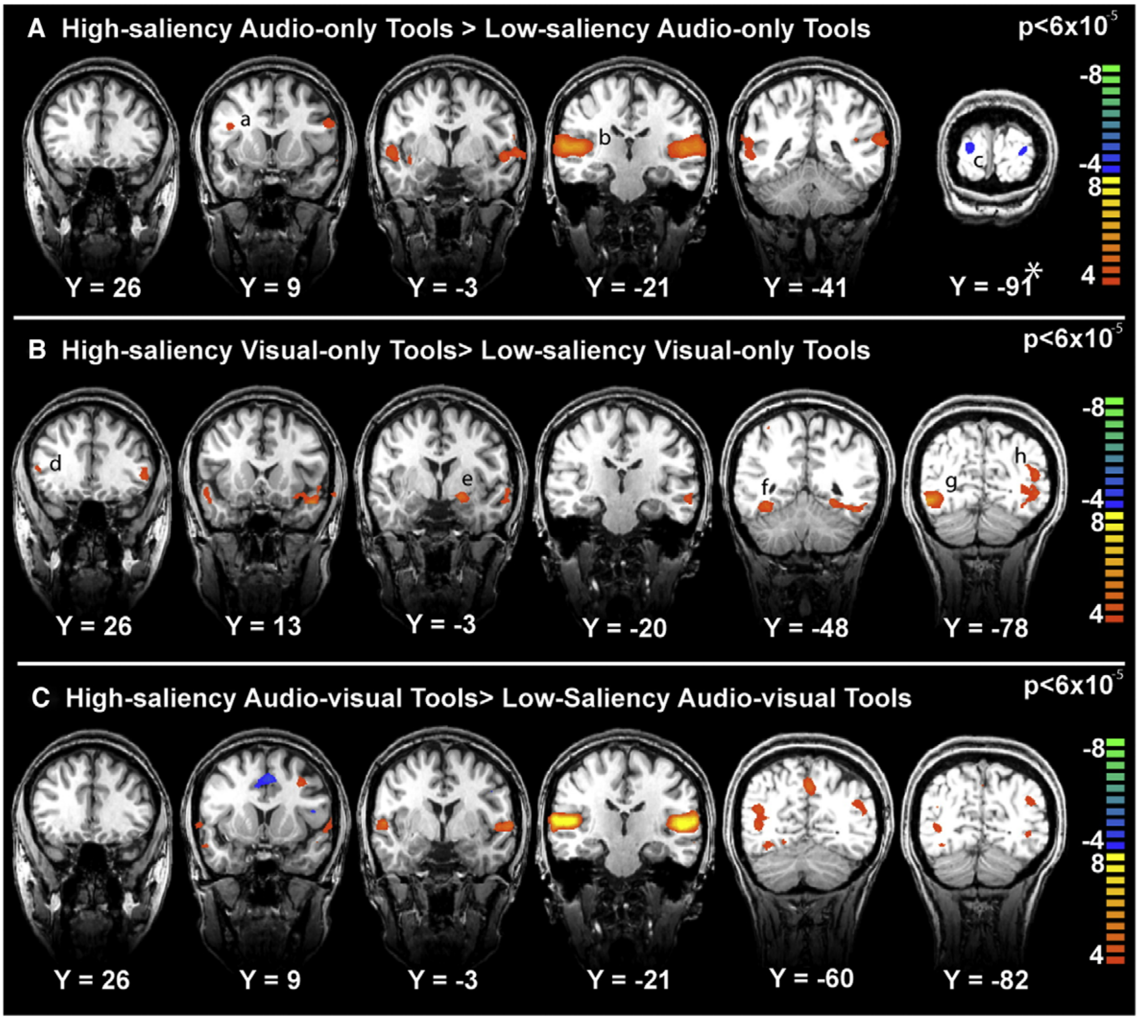
Results

Functional regions of interest

In both experiments, three functional ROIs were defined independently for each participant by combining functional information from a whole-brain SPM analysis on the localizer runs (in which blocked, high-SNR manual-tool stimuli were used) with anatomical landmarks.

Bilateral AV tool ROIs were defined based on a conjunction of two contrasts, activations with audio stimuli greater than baseline AND with visual stimuli greater than baseline ($p<4\times10^{-9}$). These ROIs were located on the posterior aspect of the STS (see Fig. 3B and Table 1). The AV tool ROIs described in both experiments used identical stimuli, stimulus presentation, and functional and anatomical definitions. Consistency of tool ROIs across experiments was examined using a one-way ANOVA with center of mass of ROIs as the dependent measure. No significant differences were found between tool ROIs (but see results from Experiment 2 for a comparison of tool and speech ROIs). All individual ROIs were found bilaterally ($p<4\times10^{-9}$) with the exception of 1 individual who lacked a definable right AV tool ROI.

Bilateral tool A and V ROIs were defined in both experiments. Audio ROIs were defined as areas in which the BOLD response was greater with A than with V stimuli ($p<4\times10^{-9}$), and located on the superior temporal gyrus (STG), likely corresponding to primary and secondary auditory cortex (Semple and Scott, 2003). Visual ROIs were defined as areas in which the BOLD response was greater with V than with A stimuli ($p<4\times10^{-9}$), and located on the anterior inferior occipital gyrus, likely consisting of a portion of the lateral occipital


A High-saliency Audio-only Tools > Low-saliency Audio-only Tools
p<6x10-5
-8
-4
8
4
a
b
c
Y = 26
Y = 9
Y = -3
Y = -21
Y = -41
Y = -91*
B High-saliency Visual-only Tools> Low-saliency Visual-only Tools
p<6x10-5
-8
-4
8
4
d
e
f
g
h
Y = 26
Y = 13
Y = -3
Y = -20
Y = -48
Y = -78
C High-saliency Audio-visual Tools> Low-Saliency Audio-visual Tools
p<6x10-5
-8
-4
8
4
Y = 26
Y = 9
Y = -3
Y = -21
Y = -60
Y = -82

* Contrast threshold for slice Y = -91 was lowered to t = -2.4 (p < 0.016)

Fig. 4. Regions showing higher activation with high-SNR than with low-SNR tool stimuli. Whole-brain, RFX SPMs of group activations defining regions more active with high-SNR than with low-SNR tools for audio-only (A), visual-only (B), and audiovisual (D) presentations. Importantly, all three contrasts identified bilateral STS, mirroring the results seen in individually-defined, multisensory STS. Other regions of significance are labeled by letters, and include: (a) inferior frontal gyrus, (b) a large area of activation including Heschel's gyrus, superior temporal sulcus, and superior temporal gyrus, (c) middle occipital gyrus, (d) inferior frontal gyrus, (e) thalamus, (f) fusiform gyrus, (g) lateral occipital complex, and (h) middle occipital gyrus. For region of activity details, see Table 3.

complex (LOC) (James et al., 2003; James and Gauthier, 2006; James et al., 2000, 2002; Malach et al., 1995) (see Fig. 3B and Table 1).

Speech ROIs in Experiment 2 were defined in the same manner as tool ROIs in Experiment 1 (see Fig. 4b and Table 1). Timecourse data were extracted from bilateral multisensory regions within STS, and BOLD response amplitudes were calculated (Figs. 3d, f, g, and h). Bilateral AV speech ROIs were defined in all participants (See Table 1). All individual ROIs were found bilaterally ($p<4\times10^{-9}$) with the exception of 1 individual who lacked a definable right AV speech ROI (the same subject who lacked a significant right AV tool ROI).

In Experiment 2, whole-brain, RFX statistical parametric maps (SPMs) were created using group data from localizer runs with both speech and tool stimuli. Multisensory regions were defined as regions that responded to both audio-only and visual-only stimuli (see Figs. 4 and 6) at significant levels ($t>4$, $p<0.004$) (see Table 2).

Anatomical locations of functional tool ROIs were highly consistent across experiments. Because the localizer runs for Experiment 2 used tool and speech stimuli, the coordinates of the tool-defined and speech-defined ROIs were compared directly. Consistent with previous findings (Specht and Reul, 2003), the centers of multisensory speech ROIs in the left hemisphere were significantly anterior ($p<0.02$) to the centers of multisensory tool ROIs (see Figs. 7 and 8), and significantly anterior ($p<0.03$) and ventral ($p<0.005$) in the right hemisphere. The centers of auditory speech-defined ROIs in the left hemisphere were more anterior ($p<0.03$) and ventral ($p<0.005$) than the centers of auditory tool-defined ROIs, which is also consistent with previous reports (Binder et al., 1996; Scott et al., 2000; Scott and Johnsrude, 2003; Specht and Reul, 2003; Zekveld et al., 2006). There were no significant differences in the location of the visual ROIs.

Inverse effectiveness

Experiment 1 consisted of rapid event-related runs (see Fig. 1A) in a 3×5 factorial design, with three sensory conditions and five stimulus saliencies as levels of their respective factors. A, V, and AV manual-tool stimuli (see Fig. 2C) were presented at five distinct levels of SNR, with participants performing a 2AFC object categorization task. Behaviorally, accuracies with audio- and visual-only stimuli did not differ from their estimated accuracies ($p<0.05$ for both). Behavioral accuracies with AV trials at each unisensory SNR level were significantly higher than their respective unisensory counterparts including the multisensory accuracies at SNR levels associated with 55% (accuracy=63.6%, $p<0.5$), 65% (accuracy=80.6%, $p<0.5$), 75% (accuracy=91.8%, $p<0.5$), 85% (accuracy=97.3%, $p<0.5$), and 95% (accuracy=99.3%, $p<0.5$) unisensory accuracies.

BOLD response amplitudes with A, V, and AV stimuli were averaged across trials for each subject at all five SNR levels. SNR had a strong effect on BOLD activation in the A, V and AV conditions, with high-SNR trials producing the greatest BOLD activation. In addition, linear regressions of SNR on BOLD activation showed a highly significant linear trend for all three stimulus types in STS, A ($R^2=0.78$), V ($R^2=0.92$), and AV ($R^2=0.95$). In Experiment 1, in the multisensory STS ROI, the unisensory and multisensory changes were greater than zero, indicating that BOLD activation showed a general decrease as the stimuli were more degraded, across stimulus modalities.

BOLD response amplitudes from STS were analyzed according to the established metric of superadditivity. For each SNR level, a paired-sample *t*-test was used to compare the summed unisensory responses (A+V) to the AV response. Experiment 1 showed inverse effectiveness in the levels of superadditivity (see Fig. 3A); as SNR decreased, increases were seen in effect size and statistical significance. In accord with Stevenson and colleague's findings of superadditivity at low SNR (Stevenson et al., 2007), AV responses exceeded the superadditive criterion (A+V<AV) at the 55 and 65% accuracy conditions ($p<0.03$ and 0.04, respectively), and marginally at 75% accuracy condition ($p<0.08$), but there was no superadditivity at higher SNR levels. Activation at higher SNR levels was not superadditive. This pattern of activation implies inverse effectiveness, that is, multisensory

Table 3
Group SPM regions of activation with speech stimuli (event-related runs)

Contrast	Region	BA	*X*	*Y*	*Z*	mm3
Audio-only: High-SNR speech > Low-SNR speech (Experiment 2)	[a]Left Heschel's gyrus	42				
	[a]Left superior temporal gyrus	13	−57	−24	11	2265
	[a]Left superior temporal sulcus	21/22				
	[b]Right Heschel's gyrus	42				
	[b]Right superior temporal gyrus	13	55	−17	10	6418
	[b]Right superior temporal sulcus	21/22				
	Left inferior occipital gyrus/Left lingual gyrus	19	−32	−70	−3	1333
	Right inferior occipital gyrus/Right lingual gyrus	19	24	−72	−6	1391
Visual-only: High-SNR speech > Low-SNR speech (Experiment 2)	Left fusiform gyrus	37	−41	−45	−11	2161
	Right fusiform gyrus	37	37	−47	−11	2061
	Left anterior inferior frontal gyrus	45	−39	31	7	1005
	Right anterior inferior frontal gyrus	45	45	25	12	800
	Left posterior inferior frontal gyrus	9	−46	9	25	3327
	Right posterior inferior frontal gyrus	9	37	7	32	421
	Left medial frontal gyrus	6	−5	5	50	452
	Right medial frontal gyrus	6	1	6	50	120
	Left medial occipital gyrus	19	−33	−77	−3	604
	Right medial occipital gyrus	19	36	−76	0	318
	Left superior temporal sulcus	22	−59	−37	12	302
	Right superior temporal sulcus	22	45	−39	11	560
Audiovisual: High-SNR speech > Low-SNR speech (Experiment 2)	Left fusiform gyrus	37	−40	−50	−12	3612
	Right fusiform gyrus	37	39	−53	−9	4455
	[c]Left Heschel's gyrus	42				
	[c]Left superior temporal gyrus	13	−53	−22	10	15349
	[c]Superior temporal sulcus	21/22				
	[d]Right Heschel's gyrus	42				
	[d]Right superior temporal gyrus	13	54	−18	9	15039
	[d]Right superior temporal sulcus	21/22				
	Left anterior inferior frontal gyrus	45	−43	29	8	237
	Right anterior inferior frontal gyrus	45	49	29	9	253
	Left medial occipital gyrus	19	−35	−72	−5	3652
	Right medial occipital gyrus	19	35	−75	−3	5497

[a] Superscripted letters indicate contiguous activations over several adjacent brain regions.

enhancement increases as stimulus saliency decreases. In the A and V ROIs, the unisensory and multisensory activations did not show a pattern of inverse effectiveness, that is, summed unisensory and multisensory activations were equally sensitive to degradation. This pattern implies that the two sensory streams do not converge in those brain regions. A similar analysis was performed on the group whole-brain SPM, and no voxels were found to be significantly superadditive at any level.

In order to mirror the individual-subjects analysis, stimulus-saliency-based contrasts were used to identify regions that responded more (or less) to high-saliency tool stimuli than low-saliency tool stimuli on a group level. Parametric contrast coefficients were assigned to conditions based on participants pre-measured behavioral accuracies (95%=+2, 85%=+1, 75%=−1, 65%=−2). With this contrast, SPMs for audio-only, visual-only, and audiovisual trials were created (see Fig. 4A–C, respectively), and significant activations identified ($t>4$, $p<.003$) (see Table 3). Importantly, this group analysis confirmed decreased activation in STS with unisensory and multisensory stimulus presentations.

A second experiment was run to determine if these findings on inverse effectiveness and neuronal convergence were robust across stimulus types, specifically speech stimuli, which are often considered to be distinct from other forms of audiovisual stimuli. Experiment 2 used an identical rapid event-related design as Experiment 1 with five levels of SNR, but with speech stimuli (see Fig. 1G) as opposed to object stimuli. Behaviorally, accuracies with audio- and visual-only stimuli did not differ from their estimated accuracies ($p<0.05$ for both). Behavioral accuracies with AV trials at each unisensory SNR level were significantly higher than their respective unisensory counterparts including the multisensory accuracies at SNR levels associated with 55% (accuracy=64.1%, $p<0.5$), 65% (accuracy=77.9%, $p<0.5$), 75% (accuracy=86.8%, $p<0.5$), 85% (accuracy=96.6%, $p<0.5$), and 95% (accuracy=98.9%, $p<0.5$) unisensory accuracies.

As in Experiment 1, reduction of BOLD signal with decreased stimulus SNR was linear in STS with A ($R^2=0.91$), V ($R^2=0.92$), and AV ($R^2=0.$ 92) stimuli. All unisensory and multisensory changes were again greater than zero in the multisensory STS ROI.

As with Experiment 1, BOLD response amplitudes from STS were analyzed according to superadditivity. For each SNR level, a paired-sample *t*-test was used to compare the summed unisensory responses (A+V) to the AV response. Experiment 2 showed inverse effectiveness in the levels of superadditivity (see Fig. 5A); as SNR decreased, increases were seen in effect size and statistical significance. AV responses exceeded the superadditive criterion (A+V<AV) at the 55% accuracy level with tool stimuli ($p<0.01$), with the 65% accuracy condition showing marginal significance ($p<0.08$), while higher SNR levels produced no superadditivity. This pattern of activation again

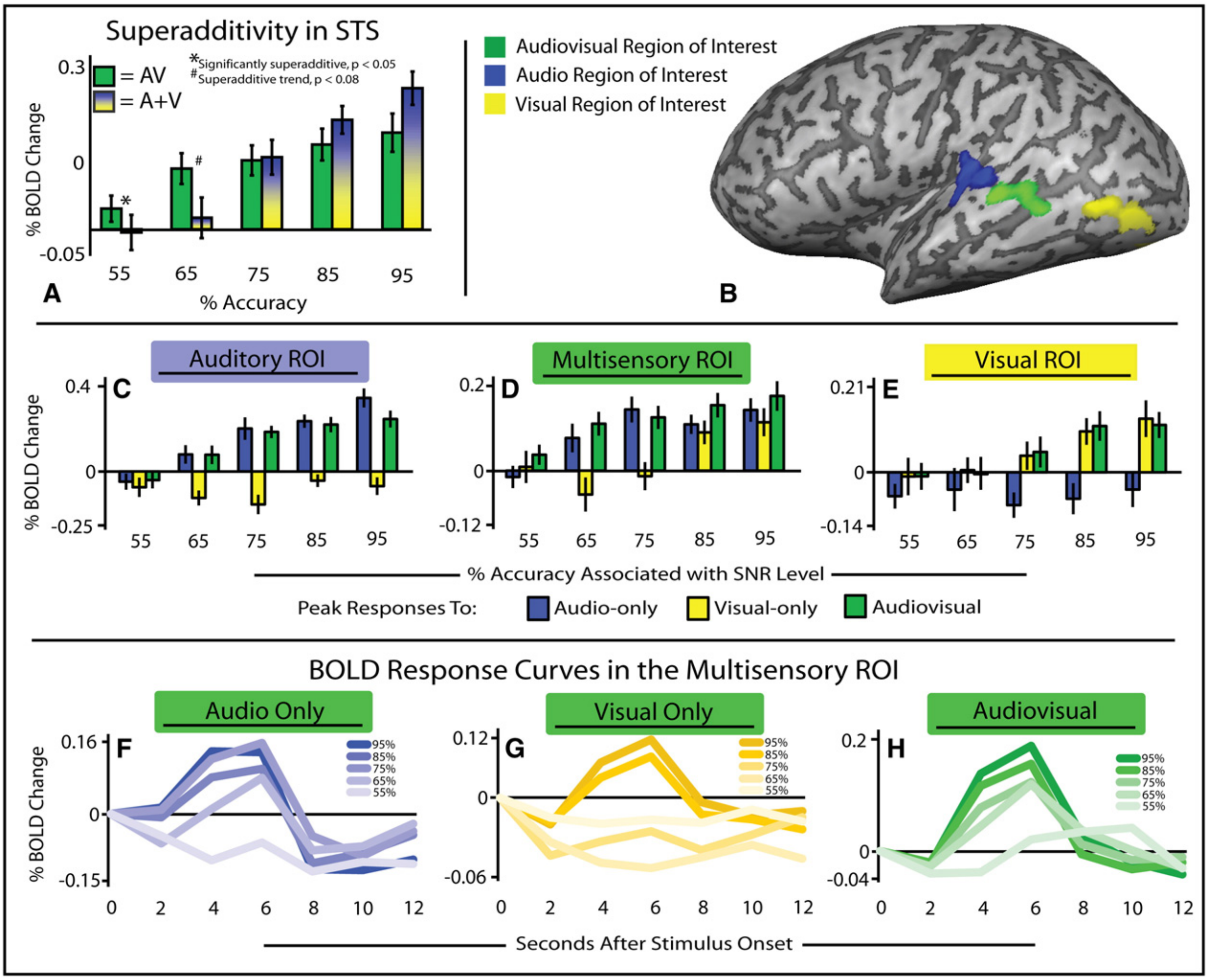


Fig. 5. Response from individual ROIs with speech stimuli in Experiment 2. Unisensory audio (STG), unisensory visual (LOC), and multisensory audiovisual (STS) speech ROIs were identified in each individual (see B for an example subjects ROIs). BOLD response curves were extracted in the multisensory ROI with A-only, V-only, and AV speech stimuli at each SNR level (F, G, and H, respectively). BOLD response amplitudes were calculated in the multisensory ROI (D), as well as in the auditory (C) and visual (E) ROIs. The multisensory ROIs responses with unisensory speech stimuli were summed and compared to the response with multisensory stimuli (A). Responses with high SNR-level stimuli did not show superadditivity, but as SNR decreased, multisensory enhancement became stronger, implying inverse effectiveness.

implies inverse effectiveness. That is, multisensory *enhancement* increases as stimulus saliency decreases. In the A and V ROIs, the unisensory and multisensory activations did not show a pattern of inverse effectiveness.

As with Experiment 1, to mirror the individual-subjects analysis, stimulus-saliency-based contrasts were used to identify regions that responded more (or less) to high-saliency speech stimuli than low-saliency speech stimuli. Parametric contrast coefficients were assigned to conditions based on participants pre-measures behavioral accuracies (95%=+2, 85%=+1, 75%=−1, 65%=−2). With this contrast, SPMs for audio-only, visual-only, and audiovisual trials were created (see Fig. 6A–C, respectively), and significant activations identified ($t>4$, $p<.003$) (see Table 4). Again, this group analysis confirmed decreased activation in STS with unisensory and multisensory stimulus presentations.

Object and speech differences

While results with both tool and speech stimuli showed inverse effectiveness, there was a marked difference between the anatomical locations of tool and speech ROIs (see Fig. 7 and 8). Due to these anatomical differences, possible differences in functional processing were assessed. Results from Experiments 1 and 2, consisting of identical event-related designs but using tool and speech stimuli, respectively, were compared in a repeated-measures ANOVA, with SNR level and modality (AV compared to A+V) as within-subjects variables, and stimulus type as a between-subjects variable. There was a significant main effect of SNR level ($p<0.001$), and a marginally significant interaction between modality and SNR level ($p=0.065$). This interaction further demonstrates inverse effectiveness. There was, however, *no* three-way interaction between modality, SNR level, and stimulus type ($p=0.89$). Thus, the level of inverse effectiveness did not vary between tools and speech.

In an attempt to verify the result of no difference in BOLD change with SNR between speech and tool stimuli, individual *t*-tests were performed on the slopes of the A, V, AV linear decreases with SNR. All of these comparisons showed no significant differences between responses to speech and tool stimuli, verifying the lack of differences between speech and tool stimuli (A, $p=0.88$; V, $p=0.78$; AV, $p=0.81$).

A final analysis attempted to address a possible discrepancy in the findings of two previous studies. The first found superadditivity for speech (Calvert et al., 2000), and the other failed to find

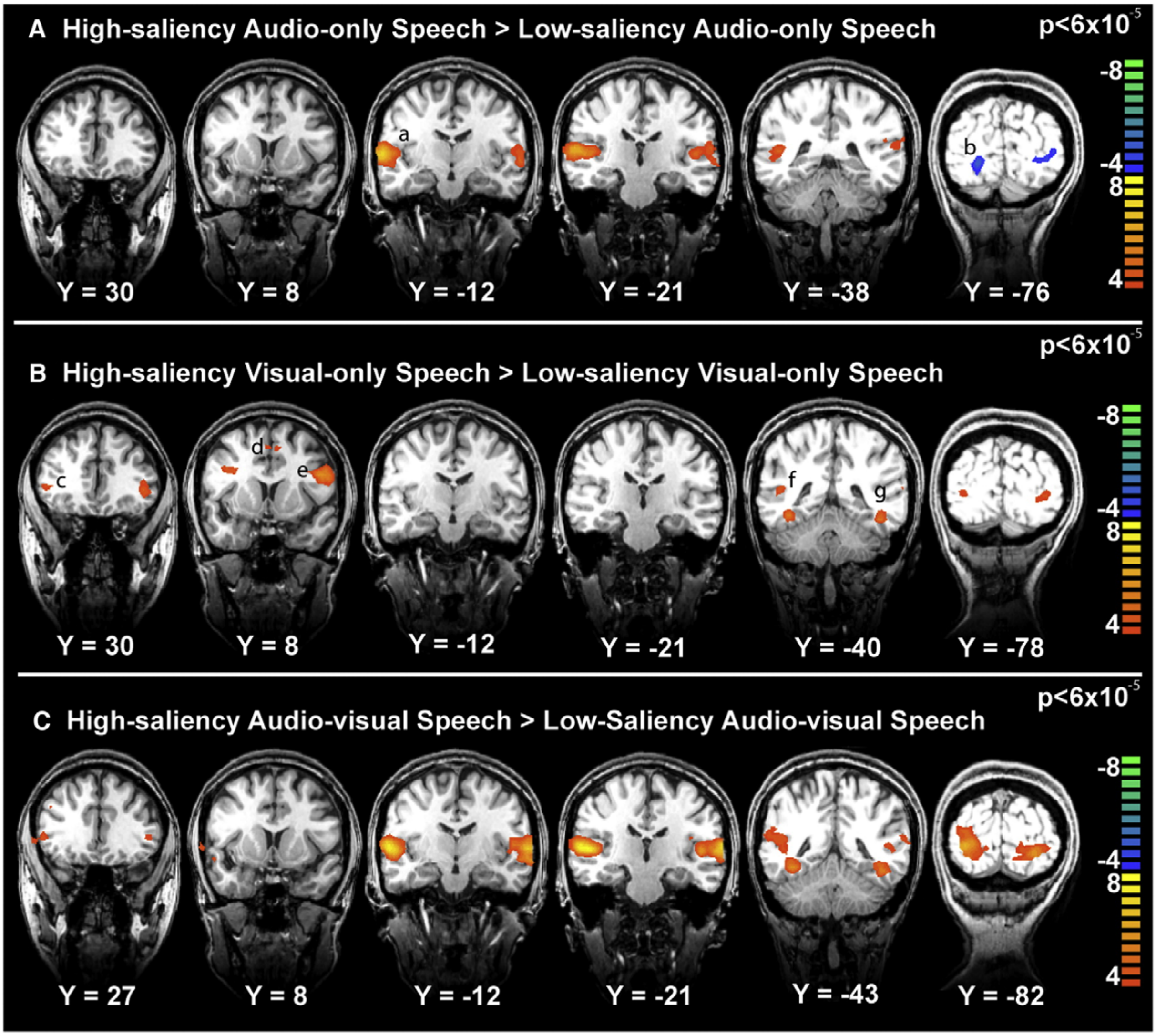

A High-saliency Audio-only Speech > Low-saliency Audio-only Speech
$p<6\times10^{-5}$
a
b
Y = 30
Y = 8
Y = -12
Y = -21
Y = -38
Y = -76
-8
-4
8
4
B High-saliency Visual-only Speech > Low-saliency Visual-only Speech
$p<6\times10^{-5}$
c
d
e
f
g
Y = 30
Y = 8
Y = -12
Y = -21
Y = -40
Y = -78
-8
-4
8
4
C High-saliency Audio-visual Speech > Low-Saliency Audio-visual Speech
$p<6\times10^{-5}$
Y = 27
Y = 8
Y = -12
Y = -21
Y = -43
Y = -82
-8
-4
8
4

Fig. 6. Regions showing higher activation with high-SNR than with low-SNR speech stimuli. Whole-brain, RFX SPMs of group activations defining regions more active with high-SNR than with low-SNR speech for audio-only (A), visual-only (B), and audiovisual (D) presentations. Importantly, all three contrasts identified bilateral STS, mirroring the results seen in individually-defined, multisensory STS. Other regions of significance are labeled by letters, and include: (a) a large area of activation including Heschel's gyrus, superior temporal sulcus, and superior temporal gyrus, (b) inferior occipital gyrus, (c) anterior inferior frontal gyrus, (d) anterior cingulate gyrus, (e) posterior inferior frontal gyrus, (f) superior temporal sulcus, and (g) fusiform gyrus. For region of activity details, see Table 4.

superadditivity with manual tools (Beauchamp et al., 2004b). These two previous experiments used blocked designs with high-SNR stimuli, thus we analyzed our blocked-design localizer runs from Experiment 2, in which high-SNR tool and speech stimuli were presented in blocks of A, V, and AV modalities. Timecourses were extracted from multisensory ROIs (see Fig. 9A–D). Neither the left nor right multisensory STS ROIs, defined with tool or speech stimuli, elicited a superadditive response in the AV condition. In fact, all comparisons showed significant *sub-additivity*, consistent with predictions by models of multisensory BOLD response (Laurienti et al., 2005), and consistent with previous multisensory studies with tools (Beauchamp et al., 2004a,b; Stevenson et al., 2007), but inconsistent with a previous multisensory study with speech (Calvert et al., 2000).

Discussion

Our results provide the first evidence that neural activation in human cortex follows the law of inverse effectiveness. As SNR was decreased incrementally, area STS showed a relative *gain* in BOLD activation with AV stimuli over and above the activation predicted by combined responses with unisensory A and V stimuli. We assessed this interaction using superadditive changes in BOLD with unisensory and multisensory stimuli. Our results clearly show that AV tool and speech stimuli produce identical patterns of neuronal convergence across a dynamic range of stimulus saliencies, despite the fact that AV tool- and speech-defined ROIs were spatially separable.

Previous studies with non-human primates have demonstrated that STS is a site of multisensory integration (Barraclough et al., 2005; Benevento et al., 1977; Bruce et al., 1981; Hikosaka et al., 1988). Many neurons within STS show multisensory enhancement, that is, they respond more strongly to a multisensory input than they do to either unisensory input in isolation. The level of enhancement in those cells, however, can be influenced by different factors. First, multiple sensory inputs are integrated better when they originate from the same spatial location (spatial congruence); second, inputs are integrated better when they occur simultaneously or are synchronous (temporal congruence); and third, inputs are integrated better when they are degraded (inverse effectiveness).

There is evidence that integration in human STS is influenced by spatial and temporal congruence (Bushara et al., 2001; Macaluso et al., 2004), but ours is the first evidence for inverse effectiveness. The experiments reported here strongly suggest that human STS does in fact respond in an inversely-effective manner: as the SNR of stimuli decrease, BOLD activations with AV stimuli become relatively larger compared to the AV response that the unisensory responses predict. Studies of human sensory systems have demonstrated main effects of BOLD response with multisensory stimuli exceeding that of unisensory stimuli (Beauchamp, 2005; Beauchamp et al., 2004a,b; Calvert, 2001; Calvert et al., 2000, 2001; Stevenson et al., 2007), as well as increased BOLD response with increased SNR level (Stevenson et al., 2007), but have not (until now) shown the interaction between the two that defines inverse effectiveness. It should also be noted that the effect was extremely reliable, and generalized across conditions. The pattern of inverse effectiveness was the same for tool and speech stimuli and was the same in tool-defined and speech-defined ROIs. Furthermore, this interaction was seen across the entire dynamic range of behavioral responses, from 55–95% accuracy, and was shown to be linear in nature. Also, variances did not increase with higher mean BOLD responses, ensuring that this inverse effectiveness was not merely a statistical artifact (Holmes, 2007).

Inverse effectiveness can also be seen clearly in the increased number of individual's ROIs showing superadditivity at lower SNR-levels (see Fig. 7). As stimulus SNR decreased, the BOLD responses within more of these ROIs became superadditive. A group SPM analysis did not reveal any voxels that exhibited significant superadditivity, at any of the SNR levels, even at liberal statistical thresholds. This was likely due to several reasons. First, the design of this experiment divided trials into five SNR levels, whereas future studies could isolate a single low-SNR condition, resulting in five times the trials and increased power over this study. Second, as can be seen in Fig. 7, the individual-defined ROIs show a high level of variability in anatomical location. When these individuals are combined in a group analysis, the lack of anatomical overlap causes a diffusion of the superadditive effect, which in turn results in lower-than-expected statistical power for the group analysis. Unlike the loss of power

Table 4
Group SPM regions of activation (blocked runs)

Contrast	Region	BA	*X*	*Y*	*Z*	mm3
Conjunction of activation with A-only and V-only blocked tool stimuli (Experiment 2)	Left anterior cingulate gyrus	24	−14	8	11	400
	Right anterior cingulate gyrus	24	20	11	29	1443
	Right fusiform gyrus	37	33	−53	−21	412
	Left inferior parietal lobule	40	−27	−53	36	2514
	Right inferior parietal lobule	40	28	−60	32	826
	Left lateral occipital complex	37	−45	−54	1	1090
	Right lateral occipital complex	37	49	−58	4	558
	Left post central gyrus	2	−52	−25	26	2619
	Left superior temporal sulcus	21/22	−43	−50	2	2276
	Right superior temporal sulcus	22	47	−30	2	623
	Left thalamus	N.A.	−15	−17	4	1570
	Right thalamus	N.A.	19	−20	7	405
Conjunction of activation with A-only and V-only blocked speech stimuli (Experiment 2)	Left anterior cingulate gyrus	24	−14	5	34	1127
	Right anterior cingulate gyrus	24	16	12	32	378
	Right cerebellum	N.A.	17	−51	−21	3919
	Left fusiform gyrus	19/37	−33	−49	−25	447
	Right fusiform gyrus	19/37	33	−49	−20	471
	Left inferior parietal lobule	40	−32	−41	27	383
	Right inferior parietal lobule	40	26	−39	32	752
	Left insula	13	−35	23	5	608
	Right insula	13	28	21	10	871
	Left post central gyrus	2	−35	−25	44	2053
	Left superior temporal sulcus	22	−41	−40	7	1348
	Right superior temporal sulcus	22	51	−29	3	1082
	Left thalamus	N.A.	−24	4	11	1931
	Right thalamus	N.A.	25	3	17	343

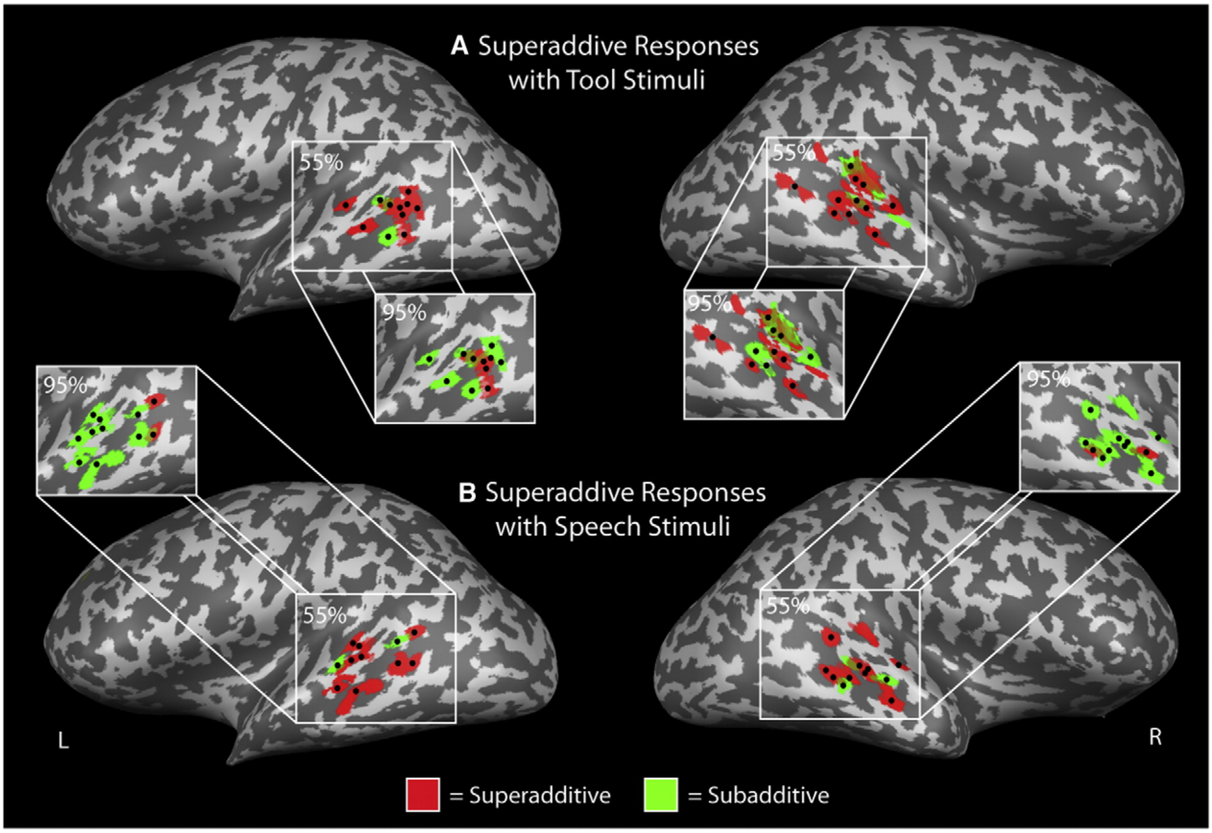


Fig. 7. Anatomically diffuse individual-defined functional ROIs show increased superadditivity at lower SNR-levels. At high SNR-levels (see insets), individuals' functionally-defined multisensory STS ROIs were more often subadditive (seen as green activations) than superadditive (seen as red activations) with both tool (A) and speech (B) stimuli. At low SNR-levels, more individuals' ROIs became superadditive than subadditive, following the law of inverse effectiveness. Individuals' multisensory ROIs (centroids represented by black dots) where highly diffuse with little overlap, resulting in a lack of statistical power in the group superadditivity analysis.

described above, which was due to the division of trials among multiple conditions, the loss of power caused by anatomical dispersion cannot be ameliorated by changing the experimental design. The choice to use individual-defined versus group-defined ROIs is a controversial one (Friston et al., 2006; Saxe et al., 2006), however, our data appear to illustrate a case where the individual-defined ROIs are able to describe the patterns in the data, whereas the group analysis cannot. This case highlights the need to consider both individual-defined and group-defined ROIs for a more complete understanding of functional neuroimaging data.

It is worth noting that there was no evidence for inverse effectiveness in the two ROIs we investigated that do not show inverse effectiveness in non-human primates, namely the putative unisensory brain regions. Although many areas once thought to be unisensory have now been shown to be modulated by other sensory inputs (Calvert et al., 1997; Kayser et al., 2005; Laurienti et al., 2002), these modulations tend to be small. Thus, the experimental conditions did not produce reliable changes in activation in either unisensory area when stimuli were presented bimodally as opposed to unimodally in either of the experiments, as would be expected in areas in which audio and visual streams do not significantly interact. That being stated, audio-only presentations did diminish the BOLD signal in the posterior occipital cortex with both tool and speech stimuli (see Figs. 4A and 6A, respectively). That is, the stronger the audio signal, the less activity there was in early visual areas. This finding was only seen in unisensory presentations of audio stimuli, and the reciprocal effect of visual-only stimuli modulating auditory cortex was not seen.

Although several different metrics have been previously explored for use in assessing multisensory integration, the most popular has been superadditivity. In Experiments 1 and 2, multisensory enhancement (as defined by superadditivity) increased as SNR level decreased (see Fig. 3A and 5A, respectively). With 85% and 95% accuracy thresholds, neither tools nor speech AV stimuli evoked a superadditive response. As SNR decreased from the 75% to 55% level, AV stimuli trended towards and exceeded the superadditivity criterion with both stimulus types. This change in superadditivity across SNR levels is indicative of inverse effectiveness. These findings coincide with our previous study that found superadditive responses with low-SNR AV stimuli, but not with high-SNR AV stimuli (Stevenson et al., 2007).

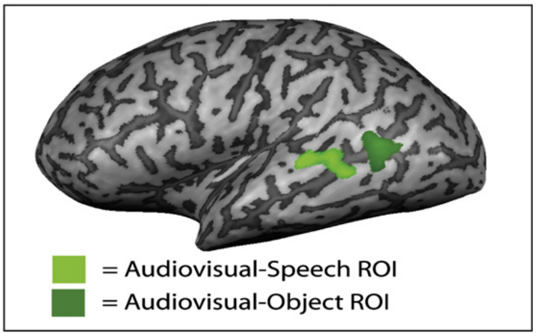


Fig. 8. Differences between anatomical location of tool-defined and speech-defined ROIs. Tool and speech stimuli identified anatomically different regions of activations in each individual (see a for an example). However, responses to both unisensory and multisensory stimuli did not vary between tool and speech regions, suggesting that while there is a difference in anatomical location, the underlying function is similar.

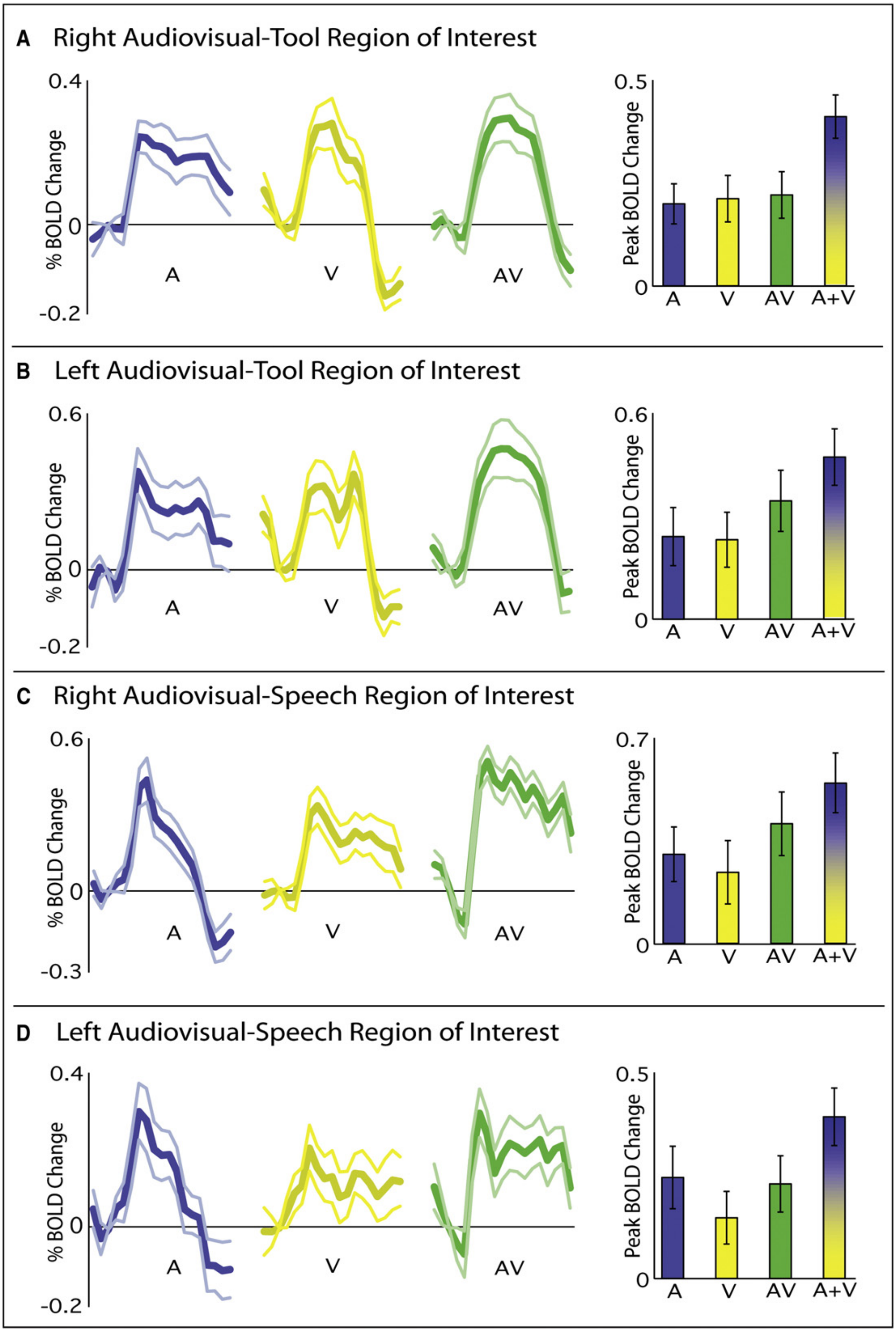


Fig. 9. Bilateral BOLD responses to blocked presentations of high-SNR tools and. BOLD response curves in response to blocked presentations of tool stimuli were extracted from multisensory ROIs in right (A) and left (B) STS (lighter curves represent standard error). Reponses with audio-only, visual-only, and audiovisual presentations were measured, and BOLD response amplitudes defined. Multisensory BOLD response amplitudes were then compared to the sum of the unisensory activation. No evidence of superadditivity was found. Similar curves were extracted, and BOLD response amplitudes calculated, for activations with unisensory and multisensory presentations of speech stimuli (C and D). Again, no superadditivity was found.

Studies of multisensory integration have been performed using a variety of different stimulus types, but one of the most studied categories has been speech. The vast majority of studies of audiovisual speech integration agree that STS is involved in this integration process (Callan et al., 2004; Calvert et al., 2000; Skipper et al., 2005). While it is generally agreed that human STS is involved in both object and speech integration (Driver and Noesselt, 2008; Hocking and Price, 2008; Stein and Stanford, 2008; Stevenson et al., 2007), it has been debated whether or not the processes underlying speech perception are unique to speech (Fowler, 1996; Liberman and Mattingly, 1985, 1989; Massaro, 1998; Moore, 2003; Remez et al., 1981, 1994; Repp, 1982) and integration (Kuhl et al., 1991; Stekelenburg and Vroomen, 2007; Tuomainen et al., 2005). Behavioral studies of speech integration suggest that the mere perception of a stimulus as speech can modify the process of audiovisual integration (Tuomainen et al., 2005), while electrophysiological studies have shown that speech and non-speech AV events indifferently modulate neural activity (Stekelenburg and Vroomen, 2007). The locations of our tool-defined and speech-defined ROIs were different and essentially non-overlapping, and coincide with previous findings of anatomical specificity for speech (Specht and Reul, 2003). Both the audio and AV ROIs in the left hemisphere were more anterior when defined with speech than when defined with tool stimuli, and AV ROIs in the right hemisphere were more anterior and ventral with speech stimuli. STS has also been shown to respond differentially to faces and facial movements, as well as hands and hand movements (Johnson-Frey et al., 2005; Materna et al., 2007; Pelphrey et al., 2005; Puce et al., 1998; Winston et al., 2004). As such, the possibility that these anatomical differences may be accounted for by the inclusion of faces in the speech condition and hands in the hand-held tools condition cannot be ruled out.

Our analysis of superadditivity further supports the lack of functional difference between speech and tool ROIs. In the Experiment 2 localizer runs, which used high-SNR stimuli, AV responses did not exceed the superadditivity criterion with either speech or tool stimuli (see Fig. 9a–d). In fact, the AV responses were actually significantly *sub-additive*. While sub-additivity has been shown reliably for tool stimuli in STS (Beauchamp, 2005; Beauchamp et al., 2004a,b; Stevenson et al., 2007), one study has reported finding *superadditivity* in STS in response to congruent audiovisual presentations of speech (Calvert et al., 2000). These discrepant findings have led to the hypothesis that speech may be integrated in a unique fashion within STS. Our data, however, fail to replicate this finding, and show that speech integration is no different from integration of stimuli from other object classes. It should be noted, however, that the speech stimuli used here were single words, while the study finding superadditivity used audiovisual reading from literary passages (Calvert et al., 2000).

In summary, human multisensory STS showed evidence of inverse effectiveness for both speech and non-speech audiovisual stimuli, similar to the inverse effectiveness seen in non-human primates at the level of single neurons. Even though the tool-defined and speech-defined ROIs were anatomically segregated, the patterns of activation across SNR levels produced in these ROIs were strikingly similar. Thus, despite anatomical specificity for stimulus type, there was no functional specificity with regard to multisensory integration. In other words, if there are substrates within STS that are specialized for processing AV speech, the manner in which they integrate the visual and auditory signals is not specific to speech stimuli.

Acknowledgments

This research was supported in part by the Indiana METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc., the IUB Faculty Research Support Program, the Indiana University GPSO Research Grant, and by the Howard Hughes Medical Institute Capstone Independent Research Opportunity Program. Thanks to Laurel Stevenson and Karin Harman James for their support with this research. Thanks to David Pisoni, Luiz Hernandez, and Nicholas Altieri for the speech stimuli; Andrew Butler, Hope Cantrell, and Luiz Pessoa for assistance with Experiment 2; and Ann Peiffer, Luiz Pessoa, Sunah Kim, Aaron Scott, James Townsend, Jennifer Willingham, Sharada Subramamian, Lindsay Dachille, and the Indiana University Neuroimaging Group for their insights on this work and manuscript.

References

Allman, B.L., Meredith, M.A., 2007. Multisensory processing in “unimodal” neurons: cross-modal subthreshold auditory effects in cat extrastriate visual cortex. J. Neurophysiol. 98, 545–549.

Barraclough, N.E., Xiao, D., Baker, C.I., Oram, M.W., Perrett, D.I., 2005. Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J. Cogn. Neurosci. 17, 377–391.

Beauchamp, M.S., 2005. Statistical criteria in FMRI studies of multisensory integration. Neuroinformatics 3, 93–113.

Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., Martin, A., 2004a. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat. Neurosci. 7, 1190–1192.

Beauchamp, M.S., Lee, K.E., Argall, B.D., Martin, A., 2004b. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41, 809–823.

Benevento, L.A., Fallon, J., Davis, B.J., Rezak, M., 1977. Auditory–visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. Exp. Neurol. 57, 849–872.

Binder, J.R., Frost, J.A., Hammeke, T.A., Rao, S.M., Cox, R.W., 1996. Function of the left planum temporale in auditory and linguistic processing. Brain 119 (Pt 4), 1239–1247.

Bolognini, N., Frassinetti, F., Serino, A., Ladavas, E., 2005. “Acoustical vision” of below threshold stimuli: interaction among spatially converging audiovisual inputs. Exp. Brain Res. 160, 273–282.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 433–436.

Bruce, C., Desimone, R., Gross, C.G., 1981. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J. Neurophysiol. 46, 369–384.

Bushara, K.O., Grafman, J., Hallett, M., 2001. Neural correlates of auditory–visual stimulus onset asynchrony detection. J. Neurosci. 21, 300–304.

Callan, D.E., Jones, J.A., Munhall, K., Kroos, C., Callan, A.M., Vatikiotis-Bateson, E., 2004. Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. J. Cogn. Neurosci. 16, 805–816.

Calvert, G.A., 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb. Cortex 11, 1110–1123.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, P.W., Iversen, S.D., David, A.S., 1997. Activation of auditory cortex during silent lipreading. Science 276, 593–596.

Calvert, G.A., Campbell, R., Brammer, M.J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr. Biol. 10, 649–657.

Calvert, G.A., Hansen, P.C., Iversen, S.D., Brammer, M.J., 2001. Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. Neuroimage 14, 427–438.

Driver, J., Noesselt, T., 2008. Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. Neuron 57, 11–23.

Fowler, C.A., 1996. Listeners do hear sounds, not tongues. J. Acoust. Soc. Am. 99, 1730–1741.

Frens, M.A., Van Opstal, A.J., Van der Willigen, R.F., 1995. Spatial and temporal factors determine auditory–visual interactions in human saccadic eye movements. Percept. Psychophys 57, 802–816.

Friston, K.J., Rothshtein, P., Geng, J.J., Sterzer, P., Henson, R.N., 2006. A critique of functional localizers. Neuroimage 30, 1077–1087.

Hairston, W.D., Laurienti, P.J., Mishra, G., Burdette, J.H., Wallace, M.T., 2003a. Multisensory enhancement of localization under conditions of induced myopia. Exp. Brain Res. 152, 404–408.

Hairston, W.D., Wallace, M.T., Vaughan, J.W., Stein, B.E., Norris, J.L., Schirillo, J.A., 2003b. Visual localization ability influences cross-modal bias. J. Cogn. Neurosci. 15, 20–29.

Hairston, W.D., Hodges, D.A., Burdette, J.H., Wallace, M.T., 2006. Auditory enhancement of visual temporal order judgment. Neuroreport 17, 791–795.

Hikosaka, K., Iwai, E., Saito, H., Tanaka, K., 1988. Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. J. Neurophysiol. 60, 1615–1637.

Hocking, J., Price, C.J., 2008. The role of the posterior superior temporal sulcus in audiovisual processing. Cereb. Cortex 18 (10), 1685–1694.

Holmes, N.P., 2007. The law of inverse effectiveness in neurons and behaviour: multisensory integration versus normal variability. Neuropsychologia 45, 3340–3345.

James, T.W., Gauthier, I., 2006. Repetition-induced changes in BOLD response reflect accumulation of neural activity. Hum. Brain Mapp. 27, 37–46.

James, T.W., Humphrey, G.K., Gati, J.S., Menon, R.S., Goodale, M.A., 2000. The effects of visual object priming on brain activation before and after recognition. Curr. Biol. 10, 1017–1024.

James, T.W., Humphrey, G.K., Gati, J.S., Menon, R.S., Goodale, M.A., 2002. Differential effects of viewpoint on object-driven activation in dorsal and ventral streams. Neuron 35, 793–801.

James, T.W., Culham, J., Humphrey, G.K., Milner, A.D., Goodale, M.A., 2003. Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. Brain 126, 2463–2475.

Johnson-Frey, S.H., Newman-Norlund, R., Grafton, S.T., 2005. A distributed left hemisphere network active during planning of everyday tool use skills. Cereb. Cortex 15, 681–695.

Kayser, C., Petkov, C.I., Augath, M., Logothetis, N.K., 2005. Integration of touch and sound in auditory cortex. Neuron 48, 373–384.

Kuhl, P.K., Williams, K.A., Meltzoff, A.N., 1991. Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. J. Exp. Psychol. Hum. Percept. Perform. 17, 829–840.

Lachs, L., Hernandez, L.R., 1998. Update: the Hoosier audiovisual multitalker database. In: Pisoni, D.B. (Ed.), Research on Spoken Language Processing. Speech Research Laboratory, Indiana University, Bloomington, IN,, pp. 377–388.

Laurienti, P.J., Burdette, J.H., Wallace, M.T., Yen, Y.F., Field, A.S., Stein, B.E., 2002. Deactivation of sensory-specific cortex by cross-modal stimuli. J. Cogn. Neurosci. 14, 420–429.

Laurienti, P.J., Kraft, R.A., Maldjian, J.A., Burdette, J.H., Wallace, M.T., 2004. Semantic congruence is a critical factor in multisensory behavioral performance. Experimental Brain Research 158 (4), 405–414.

Laurienti, P.J., Perrault, T.J., Stanford, T.R., Wallace, M.T., Stein, B.E., 2005. On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. Exp. Brain Res. 166, 289–297.

Liberman, A.M., Mattingly, I.G., 1985. The motor theory of speech perception revised. Cognition 21, 1–36.

Liberman, A.M., Mattingly, I.G., 1989. A specialization for speech perception. Science 243, 489–494.

Macaluso, E., George, N., Dolan, R., Spence, C., Driver, J., 2004. Spatial and temporal factors during processing of audiovisual speech: a PET study. Neuroimage 21, 725–732.

Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc. Natl. Acad. Sci. U. S. A. 92, 8135–8139.

Massaro, D.W., 1998. Perceiving Talking Faces: From Speech Perception to a Behavioral Principle. MIT Press, Cambridge, Mass.

Materna, S., Dicke, P.W., Their, P., 2007. Dissociable roles of the superior temporal sulcus and the intraparietal sulcus in joint attention: a functional magnetic resonance imaging study. J Cogn. Neurosci. 20 (1), 108–119.

Meredith, M.A., Stein, B.E., 1983. Interactions among converging sensory inputs in the superior colliculus. Science 221, 389–391.

Meredith, M.A., Stein, B.E., 1986a. Spatial factors determine the activity of multisensory neurons in cat superior colliculus. Brain Res. 365, 350–354.

Meredith, M.A., Stein, B.E., 1986b. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J. Neurophysiol. 56, 640–662.

Meredith, M.A., Stein, B.E., 1996. Spatial determinants of multisensory integration in cat superior colliculus neurons. J. Neurophysiol. 75, 1843–1857.

Meredith, M.A., Nemitz, J.W., Stein, B.E., 1987. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J. Neurosci. 7, 3215–3229.

Miller, L.M., D'Esposito, M., 2005. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J. Neurosci. 25, 5884–5893.

Moore, B.C.J., 2003. An Introduction to the Psychology of Hearing, 5th ed. Academic Press, Amsterdam; Boston.

Mozolic, J.L., Hugenschmidt, C.E., Peiffer, A.M., Laurienti, P.J., 2007. Modality-specific selective attention attenuates multisensory integration. Exp. Brain Res. 184 (1), 39–52.

Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. 10, 437–442.

Pelphrey, K.A., Morris, J.P., Michelich, C.R., Allison, T., McCarthy, G., 2005. Functional anatomy of biological motion perception in posterior temporal cortex: an FMRI study of eye, mouth and hand movements. Cereb. Cortex 15, 1866–1876.

Perrault Jr., T.J., Vaughan, J.W., Stein, B.E., Wallace, M.T., 2005. Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. J. Neurophysiol. 93, 2575–2586.

Puce, A., Allison, T., Bentin, S., Gore, J.C., McCarthy, G., 1998. Temporal cortex activation in humans viewing eye and mouth movements. J. Neurosci. 18, 2188–2199.

Remez, R.E., Rubin, P.E., Pisoni, D.B., Carrell, T.D., 1981. Speech perception without traditional speech cues. Science 212, 947–949.

Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S., Lang, J.M., 1994. On the perceptual organization of speech. Psychol. Rev. 101, 129–156.

Repp, B.H., 1982. Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. Psychol. Bull. 92, 81–110.

Saxe, R., Brett, M., Kanwisher, N., 2006. Divide and conquer: a defense of functional localizers. Neuroimage, 30, 1088–1099.

Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26, 100–107.

Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123 (Pt 12), 2400–2406.

Semple, M.N., Scott, B.H., 2003. Cortical mechanisms in hearing. Curr. Opin. Neurobiol. 13, 167–173.

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C.S., Woldorff, M.G., 2007. Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. Neuropsychologia 45, 561–571.

Serino, A., Bassolino, M., Farne, A., Ladavas, E., 2007. Extended multisensory space in blind cane users. Psychol. Sci. 18, 642–648.

Sheffert, S.M., Lachs, L., Hernandez, L.R., 1996. The Hoosier audiovisual multitalker database. In: Pisoni, D.B. (Ed.), Research on Spoken Language Processing. Speach Research Laboratory, Indiana University, Bloomington, IN, pp. 578–583.

Skipper, J.I., Nusbaum, H.C., Small, S.L., 2005. Listening to talking faces: motor cortical activation during speech perception. Neuroimage 25, 76–89.

Specht, K., Reul, J., 2003. Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: an auditory rapid event-related fMRI-task. Neuroimage 20, 1944–1954.

Stanford, T.R., Quessy, S., Stein, B.E., 2005. Evaluating the operations underlying multisensory integration in the cat superior colliculus. J. Neurosci. 25, 6499–6508.

Stein, B.E., Stanford, T.R., 2008. Multisensory integration: current issues from the perspective of the single neuron. Nat. Rev. Neurosci. 9, 255–266.

Stekelenburg, J.J., Vroomen, J., 2007. Neural correlates of multisensory integration of ecologically valid audiovisual events. J. Cogn. Neurosci. 19 (12), 1964–1973.

Stevenson, R.A., Geoghegan, M.L., James, T.W., 2007. Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. Exp. Brain Res. 179, 85–95.

Talaraich, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. Thieme Medical Publishers, New York, New York.

Tuomainen, J., Andersen, T.S., Tiippana, K., Sams, M., 2005. Audio-visual speech perception is special. Cognition 96, B13–22.

Winston, J.S., Henson, R.N., Fine-Goulden, M.R., Dolan, R.J., 2004. fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. J. Neurophysiol. 92, 1830–1839.

Zekveld, A.A., Heslenfeld, D.J., Festen, J.M., Schoonhoven, R., 2006. Top-down and bottom-up processes in speech comprehension. Neuroimage 32, 1826–1836.