Published in final edited form as:
Neuroimage. 2011 April 1; 55(3): 1339–1345. doi:10.1016/j.neuroimage.2010.12.063.

Discrete Neural Substrates Underlie Complementary Audiovisual Speech Integration Processes

Ryan A. Stevenson[1,2,3], **Ross M. VanDerKlok**[1], **David B. Pisoni**[1], and **Thomas W. James**[1,2]

[1] Department of Psychological and Brain Sciences, Indiana University

[2] Program in Neuroscience, Indiana University

[3] Department of Hearing and Speech Sciences, Vanderbilt University

Abstract

The ability to combine information from multiple sensory modalities into a single, unified perceptis a key element in an organism's ability to interact with the external world. This process of perceptual fusion, the binding of multiple sensory inputs into a perceptual gestalt, is highly dependent on the temporal synchrony of the sensory inputs. Using fMRI, we identified two anatomically distinct brain regions in the superior temporal cortex, one involved with processing temporal-synchrony, and one with processing perceptual fusion of audiovisual speech. This dissociation suggests that the superior temporal cortex should be considered a "neuronal hub" comprised of multiple discrete subregions that underlie an array of complementary low-and high -level multisensory integration processes. In this role, abnormalities in the structure and function of superior temporal cortex provide a possible common etiology for temporal-processing and perceptual -fusion deficits seen in a number of clinical populations, including individuals with autism spectrum disorder, dyslexia, and schizophrenia.

Keywords

Multisensory integration; STS; fMRI; speech perception; temporal processing; autism; dyslexia; schizophrenia

Introduction

The integration of multiple sensory signals into a single, fused percept is reliant upon a number of signal properties, including temporal synchrony(Bishop and Miller, 2009; Macaluso et al., 2004; Meredith, 2002; Meredith et al., 1987; Miller and D'Esposito, 2005; Stevenson et al., 2010). Temporal synchrony, along with other signal properties, allows an organism to properly fuse sensory signals originating from a single external event(perceptual fusion, or binding), and likewise, to properly dissociate signals that originate from distinct external events. The relationship between synchrony and fusion is demonstrated by the well -known finding that greater temporally asynchrony of a pair of sensory signals leads to a lower probability of perceptual fusion and perceived synchrony (Conrey and Pisoni, 2006;

Correspondence to: Ryan A. Stevenson, Department of Hearing and Speech Sciences, Vanderbilt University, MRB3, Suite 7110, Nashville, TN, 37232, ryan.andrew.stevenson@gmail.com, Phone: (615) 936-7108.

van Atteveldt et al., 2007; van Wassenhove et al., 2007). Although the correlation between synchrony and fusion establishes a relationship between them, it also makes it difficult to isolate the effects of each factor using only behavioral measurements.

Neuroimaging measures have also been used to study the effects of temporal synchrony and perceptual fusion with multisensory stimuli. Synchrony of speech modulates brain activation in a number of multisensory regions(Macaluso et al., 2004; Miller and D'Esposito, 2005; Stevenson et al., 2010), but multiple studies have found different effects, ranging from enhancement to suppression. With one exception(Miller and D'Esposito, 2005), the neuroimaging studies of speech synchrony, like the behavioral studies, have not attempted to isolate the effects of temporal synchrony from the effects of perceptual fusion. Furthermore, perceptual fusion has been shown to modulate multisensory brain activation when manipulated in isolation from synchrony (Bushara et al., 2003). Thus, it remains unclear whether or not the effects of synchrony on brain activation in multisensory brain regions are due to the concomitant changes in perceptual fusion caused by manipulations of temporal synchrony.

In this study, we focus on BOLD activation in m ultisensory superior temporal cortex (STC), one of the most commonly studied multisensory regions. Activation in this region is modulated both by temporal synchrony (Macaluso et al., 2004; Miller and D'Esposito, 2005; Stevenson et al., 2010)and perceptual fusion (Bushara et al., 2003). Furthermore, there is now other converging evidence that mSTC may be comprised of several subregions with distinct responses based on input modality (Beauchamp et al., 2004)which are differentially driven by synchrony, with one subregion responding only when sensory inputs are precisely synchronous, and another parametrically varying with asynchrony level(Stevenson et al., 2010).

Further evidence that mSTC plays a role in audiovisual temporal synchrony processing and perceptual fusion has been found in several clinical populations, such as autism spectrum disorder (ASD), schizophrenia, and dyslexia. Impairments in the ability to detect asynchronies in audiovisual speech has been reported in ASD (Bebko et al., 2006), schizophrenia (Foucher et al., 2007), and dyslexia (Hairston et al., 2005). All three groups also show impairments in perceptual fusion of audiovisual inputs, particularly as indexed by the McGurk effect (Bastien-Toniazzo et al., 2009; Bleich-Cohen et al., 2009; Boddaert and Zilbovicius, 2002; de Gelder et al., 2003; Foss-Feig et al., 2009; Gervais et al., 2004; Mongillo et al., 2008; Pearl et al., 2009; Pelphrey and Carter, 2008a, [b]; Ross et al., 2007; Smith and Bennetto, 2007; Surguladze et al., 2001; Szycik et al., 2009) and a reduction of effects of congruency in mSTC and auditory brain regions (Blau et al., 2010; Blau et al., 2009). In addition to sharing common impairments of temporal processing and perceptual fusion of audiovisual speech, individuals with ASD(Bastien-Toniazzo et al., 2009; Boddaert et al., 2004; Boddaert and Zilbovicius, 2002; Gervais et al., 2004; Levitt et al., 2003; Pelphrey and Carter, 2008a, b), schizophrenia (Boddaert and Zilbovicius, 2002; Gervais et al., 2004; Shenton et al., 2001), and dyslexia(Pekkola et al., 2006; Richards et al., 2008) also show atypical anatomy in STC and atypical activation in mSTC during language processing. The co-occurrence of anatomical and functional differences in mSTC between typically developing and atypical individuals has led to the hypothesis that dysfunction in mSTC may lead to the impairments of both temporal processing and perceptual fusion of audiovisual speech in ASD(Brunelle et al., 2009; Zilbovicius et al., 2000), schizophrenia (Ross et al., 2007), and dyslexia (Wallace, 2009), providing further evidence that mSTC is involved in these two processes.

In this report, the effects of temporal synchrony and perceptual fusion on BOLD activation were isolated and compared across two sub-regions of mSTC known to be modulated by

temporal offsets in audiovisual speech. Trials were conditionalized based on perceptual fusion independent of temporal synchrony. Two sub-regions of mSTC produced contrasting patterns of activation. Synchrony-defined mSTC, which was defined as responding more with synchronous than asynchronous stimulus presentations, was not sensitive to the perception of fusion. Bimodal mSTC, which was defined as responding strongly with both visual and auditory stimuli, was sensitive to the perception of fusion but was not sensitive to changes in synchrony. Data from the current study were also related to previous measures of BOLD activation with parametrically varied audiovisual speech in B-mSTC and S -mSTC. The present findings suggest that despite the strong relations between temporal synchrony and perceptual fusion behaviorally, discrete neural substrates underlie the processing of temporal synchrony and perceptual fusion. The activations reflecting these complimentary processes were found in sub-regions of mSTC, suggesting that mSTC is the site of multiple process es associated with audiovisual integration of speech signals.

Methods and Materials

Participants

Participants included 12 right-handed native English speakers (6 female, mean age = 22.3, s.d. = 2.8). The experimental protocol was approved by the Indiana University Institutional Review Board and Human Subjects Committee.

Stimulus Materials

Stimuli included dynamic, audiovisual (AV) recordings of a female speaker saying ten nouns (see Figure 1). Stimuli were selected from a previously published stimulus set, The Hoosier Audiovisual Multi-Talker Database (Sheffert et al., 1996). All stimuli were spoken by speaker F1. The stimuli selected were monosyllabic English words that had the highest 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). Words were chosen to belong to two categories. Body-part words included face, leg, mouth, neck, and teeth, and environmental words included beach, dirt, rain, rock, and sand. These same tokens were used successfully in categorization tasks in previous studies (Stevenson et al., 2010;Stevenson and James, 2009;Stevenson et al., 2009). Audio signal levels were measured as root mean square contrast and equated across all tokens.

All stimuli throughout the study 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. Visual stimuli were projected onto a frosted glass screen using a Mitsubishi XL30U projector. Visual stimuli were 200 × 200 pixels and subtended 4.8 × 4.8° of visual angle. Audio stimuli were presented using pneumatic headphones.

Behavioral pre-scan procedures

Prior to scanning, participants' individual sensitivity to asynchrony was measured in an MRI simulator designed to mimic the actual MRI. The simulator consisted of an in-house mock scanner complete with bore, sliding patient table, and rear projection, frosted glass image presentation, seen through a mirror mounted on a plastic replica headcoil, and headphones for stimulus presentation. Speakers mounted within the simulator were connected to a recording of the same EPI sequence used during functional scans. Participants were presented with the audiovisual spoken-words described above with the temporal synchrony varied parametrically in 33 ms increments from 300 ms audio preceding video to synchronous trials (see Figure 1a and b). Participants performed a two-alternative forced-choice (2AFC) task in which they reported if they perceptually fused the auditory and visual

components of the stimulus(i.e., did they perceive the auditory and visual components as a single, unified event or as two distinct events). During the task, pre-recorded scanner noise was played at a sound level equal to the actual MRI. Fifty trials were presented for each level of onset asynchrony, and responses were collected by a button press. A sigmoid function was fit to the behavioral data, and each participant's 50% perceptual-fusion threshold (where half of the trials were perceptually fused and half unfused) was identified, and later used as the level of onset asynchrony for the respective participant's ambiguous stimulus condition during experimental imaging runs in the scanner.

Scanning procedures

Each imaging session included two phases: functional localizer runs and experimental runs. Functional localizers consisted of stimuli presented in a blocked stimulus design while participants completed a 2AFC semantic-categorization task with identical single-word, audiovisual utterances used in the pre-scan behavioral session (body-part word or environmental word). Each run began with the presentation of a fixation cross for 12 s followed by six blocks of audio(audio plus visual fixation cross), visual(visual plus ambient scanner noise), or audiovisual stimuli. The auditory and visual components of the stimuli were semantically congruent as well as temporally synchronous. Each run included two 16 s blocks of each stimulus type, with blocks consisting of eight 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. This functional localizer was used to specifically target bimodal mSTC, a posterior region of the superior temporal cortex. Bimodal mSTC shows strong activation with both auditory and visual stimuli. It also shows stronger activation with audiovisual stimuli than with either visual or audio stimuli alone. Based on this activation profile, bimodal mSTC is sometimes localized by contrasting activation with an audiovisual stimulus condition to the maximum of the visual or auditory conditions (Beauchamp, 2005; Werner and Noppeney, 2009). Alternatively, bimodal mSTC can be localized with a conjunction of contrasts showing visual> rest and auditory > rest (Beauchamp et al., 2004). The latter method with the same specific parameters described here has been used successfully in several previous studies to localize bimodal mSTC (Stevenson et al., 2010; Stevenson et al., 2007; Stevenson and James, 2009). When localized using this method, bimodal mSTC has been shown to exhibit changes in BOLD response related to differences in audiovisual modality and synchrony (Stevenson et al., 2010), making it an ideal region of interest for the current study.

During experimental runs, participants were presented with single -word audiovisual utterances at one of three temporal onset asynchronies, synchronous (0 ms), asynchronous (400 ms), or ambiguous. For the ambiguous condition, onset asynchrony was determined individually for each participant as the amount of asynchrony required for perceptual fusion to be reported 50% of the time during a behavioral prescan session. Stimuli were presented in a fast event-related design. Participants carried out a fused -unfused 2AFC task identical to that used in the prescan session. Again, subjects were asked if they perceived the auditory and visual components as a single, unified event or as two distinct events, and responded via button press. Runs began with the presentation of a fixation cross for 12 s, followed by 28 pseudorandomly ordered trials, (such that on average, each trial type was preceded by an equal distribution of trial types; $X^2 = 0.07$, $p > 0.96$; (Dale and Buckner, 1997; Serences, 2004)) including seven synchronous trials, seven asynchronous trials, and 14 ambiguous trials (which were later divided into perceptually-fused and -unfused trials based on the participant's response to the 2AFC fused-unfused task). For every seven trials of each stimulus type, four trials were preceded by a two-second ISI, two trials preceded by a four-second ISI, and one trial 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 pseudorandom and counterbalanced across runs, and run order was counterbalanced across participants. Each participant completed 10 experimental runs, for a total of 70 synchronous and 70 asynchronous trials, and 140 ambiguous trials.

Imaging parameters and analysis

Imaging was carried out using a Siemens Magnetom Trio 3-T whole-body scanner, and collected on an eight-channel phased-array head coil. The field of view was 22 × 22 × 11.2 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 × 3.4 mm. Voxels were re-sampled to 3 × 3 × 3 mm during preprocessing. Images were collected using a gradient echo EPI sequence (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 sequence (TI = 1,100 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) using an eight-point affine transformation. All functional volumes were aligned to a reference volume, which was the first volume of the functional run acquired closest in time to the anatomical series. The reference volume was co-registered to the untransformed anatomical volume. Using the parameters from these three transformations, functional volumes were transformed into the same common stereotactic space as the anatomical volumes. Before transformation, functional volumes then underwent a linear trend removal, 3-D spatial Gaussian filtering (FWHM 6 mm), slice scan-time correction, and 3-D motion correction. Runs with more than 1 mm of motion were excluded from further analysis.

Whole-brain, random-effects (RFX) statistical parametric maps (SPM) were calculated using the Brain Voyager™ general linear model (GLM) procedure. The design matrix was assembled from separate predictors for each condition of audiovisual presentations (4 predictors total, 2 s events) modeled using a canonical two-gamma hemodynamic response function (Glover, 1999). Epoch-based event-related averages (ERAs), consisting of aligning and averaging all trials from each condition to stimulus onset, were created based on experimental condition for both the localizer and the experimental study. The use of variable ISIs combined with independent, pseudo-randomized trial orders, allows for an ERA analysis with similar fidelity to a deconvolution analysis without relying on the underlying assumptions associate with such an analysis (Serences, 2004). Hemodynamic BOLD response amplitudes were defined as the arithmetic mean of the time course within a time window of 4–6 s after trial onset for the fast event-related experimental runs.

Results

Behavioral

Behavioral data from the pre-scan session in which participants reported perceptual fusion were used to identify the temporal offset at which each individual participant fused 50% of the trials (group mean = 167 ms, s.d. = 46 ms). Each individual's threshold was used in their ambiguous condition during scanning.

Imaging

Behavioral data in the experimental runs of the scanning session were calculated based on individual's reports of perceptual fusion. Across synchrony conditions, trials were subsequently conditionalized based on the participants' fusion responses. Synchronous trials

that were reported as unfused and asynchronous trials that were reported as fused were not included in the fMRI analysis. Synchronous trials showed a mean fusion probability of 89% (s.d. = 8%), and asynchronous trials showed a mean fusion probability of 7% (s.d. = 8%). Ambiguous trials were conditionalized into either *fused* or *unfused* conditions, with none of the ambiguous trials discarded. Ambiguous trials had a mean perceptual fusion probability of 54% (SD = 15%). In summary, the behavioral results of the experimental scanning runs produced four conditions, synchronous, fused, unfused, and asynchronous.

Functional data were analyzed using a region of interest (ROI) analysis to explore the relations between perceptual fusion and BOLD response in two specific subregions of mSTC, bimodal mSTC and synchrony-defined mSTC (Stevenson et al., 2010). Regions were defined on a group level, and individual's time courses reflecting BOLD percent signal change across the three synchrony conditions were extracted from these two distinct subregions of mSTC.

The first ROI was a synchrony-defined subregion of mSTC (S-mSTC), which was localized bilaterally with data from the experimental runs using a contrast of synchronous > asynchronous (see Figure 2a and Table 1) with voxels of activation deemed significant at a minimum voxel-wise p-value of 0.001, with an additional statistical constraint of a cluster threshold of 10 voxels, a volume of 270 mm^3 (based off of the cluster-size threshold estimator plugin for Brainvoyager, which calculated a necessary cluster size of 217 mm^3, or approximately 8 voxels). The cluster-threshold correction technique used here controls for false positives, with a relative sparing of statistical power (Forman et al., 1995; Thirion et al., 2007), and has been previously used to define mSTC(Stevenson et al., 2010; Stevenson et al., 2007; Stevenson and James, 2009). The location of S-mSTC was similar to the location reported previously, using the same technique(Stevenson et al., 2010), and also extended into insular regions. The second ROI, B-mSTC, was localized bilaterally using data from separate functional localizer runs. Those runs consisted of blocked presentations of unisensory visual and unisensory auditory speech signals.

Consistent with previous work (Stevenson et al., 2010; Stevenson et al., 2007; Stevenson and James, 2009; Stevenson et al., 2009), B-mSTC was defined by a conjunction of two contrasts, audio presentations > baseline and visual presentations > baseline (see Figure 3a and Table 1) with voxels of activation deemed significant at a minimum voxel-wise p-value of 0.001, with an additional statistical constraint of a cluster threshold of 10 voxels, a volume of 270 mm^3. Time courses reflecting BOLD percent signal change were extracted from both regions of interest bilaterally. A 2×4 hemisphere-by-condition ANOVA was run with S-mSTC and B-mSTC, and no main effect of hemisphere was found ($p = 0.73$ and 0.97, respectively), nor a hemisphere x condition interaction ($p = 0.46$ and 0.83, respectively). As such, all further reports have been collapsed across hemisphere.

To measure a difference in BOLD percent change related to perceptual fusion in S-mSTC and B-mSTC, a 2×2 ANOVA was run, revealing a 2-way interaction($F_{(40,1)} = 8.44$, $p = 0.006$) of brain region and perceptual fusion (fused and unfused). Given this interaction, a within-brain-region analysis of BOLD change across levels of perceptual fusion was conducted, showing a significant effect of fusion in B-mSTC, with unfused trials producing larger BOLD response amplitudes than fused trials ($t = 3.37$, $p = 0.0031$; Figure 3b), but no significant effect within S-mSTC ($t = 0.09$; Figure 2b). This pattern of results suggest s that one subregion of mSTC, B-mSTC, is driven by perceptual fusion, but the other subregion, S-mSTC is not.

While our main analysis showed that B-mSTC showed a significant difference between fused and unfused trials in which identical stimuli were presented, B-mSTC also showed

significantly higher levels of activation with the asynchronous condition than with the synchronous condition ($t = 3.67$, $p = 0.0015$; Figure 3b). Interestingly, there was no significant difference found ($t = 0.87$) between the unfused trials and the asynchronous trials (which were also not perceptually fused). Likewise, there was no significant difference found($t = 0.56$) between fused trials and synchronous trials (which were also perceptually fused). Taken together, the results clearly show that the differences observed between perceptually-fused and -unfused trials were not dependent on the level of synchrony between the auditory and visual presentations, suggesting that area B-mSTC is specifically driven by processes involved in producing perceptual fusion, which operate across a range temporal asynchronies.

In S-mSTC, higher levels of activation were seen with the synchronous condition than the asynchronous condition ($t = 5.41$, $p = 0.0001$), in both the fused condition ($t = 3.80$, $p = 0.0011$) and the unfused condition ($t = 3.52$ $p = 0.0021$). To some degree, these results were expected, based on the contrast (synchronous > asynchronous) used to select the voxels in the S-mSTC ROI. More precisely, posthoc tests that included the synchronous or asynchronous conditions would be considered non-independent ROI analyses(Poldrack and Mumford, 2009). However, with that noted, the lack of a difference between the fused and unfused conditions (that is, conditions that were *independent* of the ROI definition), suggests that activation in S-mSTC is driven by the level of temporal synchrony, not by the presence of perceptual fusion.

The different patterns of BOLD percent change in these two subregions of mSTC across conditions suggest that activation in each region is driven by different aspects of the perceptual experience. S-mSTC is driven by the actual temporal synchrony of the two signals and B -mSTC driven by the percept of fusion. The results suggest that activations in these two subregions represent discrete processes, both of which are involved in multisensory integration, but each using distinct, complimentary operations to compare the sensory components.

Time courses from the localizer runs were also extracted from B-mSTC and S-mSTC, however, no statistically significant multisensory interaction was observed in the BOLD responses based on either the maximum or the sum criterion, a null result that has been previously found with highly-salient, blocked AV presentations (Stevenson et al., 2007; Stevenson and James, 2009).

Discussion

We identified two subregions of multisensory STC, bimodal mSTC and synchrony-defined mSTC (Stevenson et al., 2010), which exhibit ed qualitatively different BOLD response patterns with asynchronous audiovisual speech. We showed that S-mSTC activation was driven by temporal synchrony of the two input signals regardless of perceptual fusion, while B-mSTC activation was driven by the perceptual fusion, regardless of temporal asynchrony. The distinct BOLD activation patterns in these two subregions of mSTC provide evidence that integration of speech signals involves at least two processing mechanisms, one that reflects the physical temporal alignment of auditory and visual sensory input s, and another that reflects the psychological phenomenon of perceptual fusion of separate channels into a coherent perceptual gestalt.

The evidence for multiple processes in distinct subregions of mSTC supports and extends the hypothesis that mSTC is a core region (Miller and D'Esposito, 2005) or neural hub(Hagmann et al., 2008; McIntosh and Korostil, 2008; Sporns, 2010)of a sensory integration network. Multisensory STC is known to be involved in both bottom-up

integration of sensory stimuli, receiving afferent projections from early visual and auditory cortex(as well as somatosensory) (Jones and Powell, 1970; Pandya and Yeterian, 1985; Seltzer and Pandya, 1978) and providing feedback to early sensory cortices (Pandya and Yeterian, 1985). Multisensory STC is also known to exhibit low-level multisensory interactions including inverse effectiveness (James and Stevenson, 2011; James et al., 2009, In Press; Stevenson et al., 2007; Stevenson and James, 2009; Stevenson et al., 2009), spatial congruence(Fairhall and Macaluso, 2009), and temporal synchrony(Macaluso et al., 2004; Miller and D'Esposito, 2005; Stevenson et al., 2010) effects. In addition to low -level sensory connections and interactions, mSTC also interacts with other cortical regions associated with higher-level neurocognitive multisensory interactions such as inferior frontal gyrus (including Broca's area)and inferior parietal sulcus (Romanski et al., 1999). Given these anatomical and functional connections, viewing mSTC as a neural hub (Hagmann et al., 2008; McIntosh and Korostil, 2008; Sporns, 2010)of the multisensory processing network is warranted. Our results suggest that mSTC is involved with both low-level interactions (temporal synchrony) and higher-level interactions (perceptual fusion), both of which are predicted by this hypothesized role for mSTC(Miller and D'Esposito, 2005). T he present data also suggest that mSTC is the site of multiple subregions that integrate auditory and visual information streams using complementary processes.

The results reported here provide insights into the original characterization of these two subregions of mSTC (Stevenson et al., 2010). S-mSTC exhibited a significant BOLD response only when the auditory and visual components of a stimulus were synchronous: audiovisual presentations with offset s as short as 100 ms showed no BOLD response. Our results also provide additional converging support for the previous findings suggesting that S -mSTC responds to the temporal coincidence of the sensory inputs, and extends these findings suggesting that S -mSTC is invariant to changes in perceptual fusion.

The results that B-mSTC is specifically driven by changes in perceptual fusion, rather than changes in synchrony, provide an explanation for previous results. Previously, we found that BOLD activation in B -mSTC varied parametrically with level of synchrony: the more asynchronous the presentation, the greater the BOLD response(Stevenson et al., 2010). Variations in level of synchrony, however, produce changes in the probability of perceptual fusion. Trials with greater asynchrony have a lower probability of perceptual fusion (Conrey and Pisoni, 2006; van Atteveldt et al., 2007; van Wassenhove et al., 2007). This relationship of asynchrony level and perceptual fusion, combined with the current findings that perceptual fusion is associated with *decreases* in BOLD response in B-mSTC, provides an alternate explanation for the previously found relations between B-mSTC activation and level of synchrony(Stevenson et al., 2010). Decreased BOLD activation with increasing synchrony can be explained by an increase in the probability of fusion, which in turn is related to a decreased in BOLD response.

The existence of the subregions described above that respond preferentially to changes in either temporal synchrony or perceptual fusion, also have possible clinical relevance. A number of clinical populations including ASD, schizophrenia, and dyslexia, show impairments in the temporal processing and perceptual fusion of audiovisual speech. As such, mSTC provides a locus for a possible common etiology of audiovisual speech impairments seen with ASD, SZ, and dyslexia. Individuals with ASD(Smith and Bennetto, 2007), schizophrenia (Ross et al., 2007), and dyslexia (Bastien-Toniazzo et al., 2009) show impairments in perceptual fusion as seen by decreased susceptibility to the McGurk effect(Bastien-Toniazzo et al., 2009; Mongillo et al., 2008; Pearl et al., 2009), and a decreased gain when visual speech signals are paired with auditory speech relative to healthy individuals (de Gelder et al., 2003; Ramirez and Mann, 2005; Smith and Bennetto, 2007). Furthermore, functional differences in mSTC responses have been shown in

individuals with ASD (Pelphrey and Carter, 2008a, b), schizophrenia (Szycik et al., 2009), and dyslexia (Pekkola et al., 2006) relative to typically developing individuals when perceptually fusing audiovisual speech or when integrating simple written and spoken language inputs. In addition to these functional differences, all three impaired populations are known to have anatomic abnormalities in mSTC(Boddaert et al., 2004; Levitt et al., 2003; Richards et al., 2008; Shenton et al., 2001). Functional differences found in schizophrenic patients are not limited to mSTC, but also include other regions that project directly to mSTC, including parietal regions in the dorsal visual stream (Doniger et al., 2002) such as the inferior parietal sulcus, which has itself been shown to be involved with perceptual fusion of audiovisual speech(Bishop and Miller, 2009; Miller and D'Esposito, 2005). Disruption of these parietal dorsal visual stream inputs into mSTC have also been implicated in impairments in both perceptual fusion and temporal processing of audiovisual speech in patient AWF (Hamilton et al., 2006). The possibility that a common etiology may exist for this wide range of multisensory impairments warrants further investigation in these clinical populations, as well as in healthy listeners.

In this report, we have provided evidence that mSTC is involved in lower-level temporal processing as well as higher-level perceptual fusion of audiovisual speech. These properties make mSTC a good candidate as a neural hub for multisensory processing of auditory and visual speech signals. The low-level and high-level integrative processes were localized to anatomically and functionally distinct subregions of mSTC, suggesting that mSTC should be considered a complex of regions, rather than a single region. Establishing that mSTC serves as a neural hub of multisensory processing provides a possible common etiology for a number of disorders that are associated with specific deficits in temporal processing and perceptual fusion of multimodal audiovisual speech signals.

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., 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, NIH NIDCD Training grant T32 DC000012 Training in Speech, Hearing, and Sensory Communication, NIH NIDCD Research Grant R01 DC-00111, and the Indiana University GPSO Research Grant. Thanks to Laurel Stevenson, Beth Greene, and Karin Harman James for their support, to Luis Hernandez for the stimuli, and the Indiana University Neuroimaging Group for their insights on this work. Thanks also to Vera Blau and Mark Wallace for their help with the manuscript.

References

Bastien-Toniazzo M, Stroumza A, Cavé C. Audio-visual perception and integration in developmental dyslexia: An exploratory study using the McGurk effect. Current Psychology Letters 2009:25.

Beauchamp MS. Statistical criteria in FMRI studies of multisensory integration. Neuroinformatics 2005;3:93–113. [PubMed: 15988040]

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci 2004;7:1190–1192. [PubMed: 15475952]

Bebko JM, Weiss JA, Demark JL, Gomez P. Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. J Child Psychol Psychiatry 2006;47:88–98. [PubMed: 16405645]

Bishop CW, Miller LM. A multisensory cortical network for understanding speech in noise. J Cogn Neurosci 2009;21:1790–1805. [PubMed: 18823249]

Blau V, Reithler J, van Atteveldt N, Seitz J, Gerretsen P, Goebel R, Blomert L. Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. Brain 2010;133:868–879. [PubMed: 20061325]

Blau V, van Atteveldt N, Ekkebus M, Goebel R, Blomert L. Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. Curr Biol 2009;19:503–508. [PubMed: 19285401]

Bleich-Cohen M, Hendler T, Kotler M, Strous RD. Reduced language lateralization in first-episode schizophrenia: an fMRI index of functional asymmetry. Psychiatry Res 2009;171:82–93. [PubMed: 19185468]

Boddaert N, Chabane N, Gervais H, Good CD, Bourgeois M, Plumet MH, Barthelemy C, Mouren MC, Artiges E, Samson Y, Brunelle F, Frackowiak RS, Zilbovicius M. Superior temporal sulcus anatomical abnormalities in childhood autism: a voxel-based morphometry MRI study. Neuroimage 2004;23:364–369. [PubMed: 15325384]

Boddaert N, Zilbovicius M. Functional neuroimaging and childhood autism. Pediatr Radiol 2002;32:1–7. [PubMed: 11819054]

Brainard DH. The Psychophysics Toolbox. Spat Vis 1997;10:433–436. [PubMed: 9176952]

Brunelle F, Boddaert N, Zilbovicius M. Autism and brain imaging. Bull Acad Natl Med 2009;193:287–297. discussion 297–288. [PubMed: 19718886]

Bushara KO, Hanakawa T, Immisch I, Toma K, Kansaku K, Hallett M. Neural correlates of cross-modal binding. Nat Neurosci 2003;6:190–195. [PubMed: 12496761]

Conrey B, Pisoni DB. Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. J Acoust Soc Am 2006;119:4065–4073. [PubMed: 16838548]

Dale AM, Buckner RL. Selective averaging of rapidly presented individual trials using fMRI. Hum Brain Mapp 1997;5:329–340. [PubMed: 20408237]

de Gelder B, Vroomen J, Annen L, Masthof E, Hodiamont P. Audio-visual integration in schizophrenia. Schizophr Res 2003;59:211–218. [PubMed: 12414077]

Doniger GM, Foxe JJ, Murray MM, Higgins BA, Javitt DC. Impaired visual object recognition and dorsal/ventral stream interaction in schizophrenia. Arch Gen Psychiatry 2002;59:1011–1020. [PubMed: 12418934]

Fairhall SL, Macaluso E. Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. Eur J Neurosci 2009;29:1247–1257. [PubMed: 19302160]

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn Reson Med 1995;33:636–647. [PubMed: 7596267]

Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT. An extended multisensory temporal binding window in autism spectrum disorders. Exp Brain Res 2009;203:381–389. [PubMed: 20390256]

Foucher JR, Lacambre M, Pham BT, Giersch A, Elliott MA. Low time resolution in schizophrenia Lengthened windows of simultaneity for visual, auditory and bimodal stimuli. Schizophr Res 2007;97:118–127. [PubMed: 17884350]

Gervais H, Belin P, Boddaert N, Leboyer M, Coez A, Sfaello I, Barthelemy C, Brunelle F, Samson Y, Zilbovicius M. Abnormal cortical voice processing in autism. Nat Neurosci 2004;7:801–802. [PubMed: 15258587]

Glover GH. Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 1999;9:416–429. [PubMed: 10191170]

Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, Sporns O. Mapping the structural core of human cerebral cortex. PLoS Biol 2008;6:e159. [PubMed: 18597554]

Hairston WD, Burdette JH, Flowers DL, Wood FB, Wallace MT. Altered temporal profile of visual-auditory multisensory interactions in dyslexia. Exp Brain Res 2005;166:474–480. [PubMed: 16028030]

Hamilton RH, Shenton JT, Coslett HB. An acquired deficit of audiovisual speech processing. Brain Lang 2006;98:66–73. [PubMed: 16600357]

James, TW.; Stevenson, RA. The use of fMRI to assess multisensory integration. In: Wallace, MH.; Murray, MM., editors. Frontiers in the Neural Basis of Multisensory Processes. Taylor & Francis; London: 2011.

James, TW.; Stevenson, RA.; Kim, S. Assessing multisensory integration with additive factors and functional MRI. The International Society for Psychophysics; Dublin, Ireland: 2009.

James, TW.; Stevenson, RA.; Kim, S. Inverse effectiveness in multisensory processing. In: Stein, BE., editor. The New Handbook of Multisensory Processes. MIT Press; Cambridge, MA: 2011.

Jones EG, Powell TP. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain 1970;93:793–820. [PubMed: 4992433]

Lachs, L.; Hernandez, LR. Update: The Hoosier Audiovisual Multitalker Database. In: Pisoni, DB., editor. Research on spoken language processing. Speech Research Laboratory, Indiana University; Bloomington, IN: 1998. p. 377-388.

Levitt JG, Blanton RE, Smalley S, Thompson PM, Guthrie D, McCracken JT, Sadoun T, Heinichen L, Toga AW. Cortical sulcal maps in autism. Cereb Cortex 2003;13:728–735. [PubMed: 12816888]

Luce PA, Pisoni DB. Recognizing spoken words: the neighborhood activation model. Ear Hear 1998;19:1–36. [PubMed: 9504270]

Macaluso E, George N, Dolan R, Spence C, Driver J. Spatial and temporal factors during processing of audiovisual speech: a PET study. Neuroimage 2004;21:725–732. [PubMed: 14980575]

McIntosh AR, Korostil M. Interpretation of Neuroimaging data based on network concepts. Brain Imaging and Behavior 2008;2:264–269.

Meredith MA. On the neuronal basis for multisensory convergence: a brief overview. Brain Res Cogn Brain Res 2002;14:31–40. [PubMed: 12063128]

Meredith MA, Nemitz JW, Stein BE. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J Neurosci 1987;7:3215–3229. [PubMed: 3668625]

Miller LM, D'Esposito M. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J Neurosci 2005;25:5884–5893. [PubMed: 15976077]

Mongillo EA, Irwin JR, Whalen DH, Klaiman C, Carter AS, Schultz RT. Audiovisual processing in children with and without autism spectrum disorders. J Autism Dev Disord 2008;38:1349–1358. [PubMed: 18307027]

Pandya, DN.; Yeterian, EH. Architecture and connections of cortical association areas. In: Peters, A.; Jones, EG., editors. Cerebral Cortex. Plenum; New York: 1985. p. 3-61.

Pearl D, Yodashkin-Porat D, Katz N, Valevski A, Aizenberg D, Sigler M, Weizman A, Kikinzon L. Differences in audiovisual integration, as measured by McGurk phenomenon, among adult and adolescent patients with schizophrenia and age-matched healthy control groups. Compr Psychiatry 2009;50:186–192. [PubMed: 19216897]

Pekkola J, Laasonen M, Ojanen V, Autti T, Jaaskelainen IP, Kujala T, Sams M. Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: an fMRI study at 3 T. Neuroimage 2006;29:797–807. [PubMed: 16359873]

Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 1997;10:437–442. [PubMed: 9176953]

Pelphrey KA, Carter EJ. Brain mechanisms for social perception: lessons from autism and typical development. Ann N Y Acad Sci 2008a;1145:283–299. [PubMed: 19076404]

Pelphrey KA, Carter EJ. Charting the typical and atypical development of the social brain. Dev Psychopathol 2008b;20:1081–1102. [PubMed: 18838032]

Poldrack RA, Mumford JA. Independence in ROI analysis: where is the voodoo? Soc Cogn Affect Neurosci 2009;4:208–213. [PubMed: 19470529]

Ramirez J, Mann V. Using auditory-visual speech to probe the basis of noise-impaired consonant-vowel perception in dyslexia and auditory neuropathy. J Acoust Soc Am 2005;118:1122–1133. [PubMed: 16158666]

Richards T, Stevenson J, Crouch J, Johnson LC, Maravilla K, Stock P, Abbott R, Berninger V. Tract-based spatial statistics of diffusion tensor imaging in adults with dyslexia. AJNR Am J Neuroradiol 2008;29:1134–1139. [PubMed: 18467520]

Romanski LM, Bates JF, Goldman-Rakic PS. Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. J Comp Neurol 1999;403:141–157. [PubMed: 9886040]

Ross LA, Saint-Amour D, Leavitt VM, Molholm S, Javitt DC, Foxe JJ. Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. Schizophr Res 2007;97:173–183. [PubMed: 17928202]

Seltzer B, Pandya DN. Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. Brain Res 1978;149:1–24. [PubMed: 418850]

Serences JT. A comparison of methods for characterizing the event-related BOLD timeseries in rapid fMRI. Neuroimage 2004;21:1690–1700. [PubMed: 15050591]

Sheffert, SM.; Lachs, L.; Hernandez, LR. The Hooiser Audiovisual Multitalker Database. In: Pisoni, DB., editor. Research on spoken language processing. Speech Research Laboratory, Indiana University; Bloomington, IN: 1996. p. 578-583.

Shenton ME, Dickey CC, Frumin M, McCarley RW. A review of MRI findings in schizophrenia. Schizophr Res 2001;49:1–52. [PubMed: 11343862]

Smith EG, Bennetto L. Audiovisual speech integration and lipreading in autism. J Child Psychol Psychiatry 2007;48:813–821. [PubMed: 17683453]

Sporns, O. Networks of the Brain. MIT Press; Cambridge, MA: 2010.

Stevenson RA, Altieri NA, Kim S, Pisoni DB, James TW. Neural processing of asynchronous audiovisual speech perception. Neuroimage 2010;49:3308–3318. [PubMed: 20004723]

Stevenson RA, Geoghegan ML, James TW. Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. Experimental Brain Research 2007;179:85–95.

Stevenson RA, James TW. Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. Neuroimage 2009;44:1210–1223. [PubMed: 18973818]

Stevenson RA, Kim S, James TW. 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 2009;198:183–194. [PubMed: 19352638]

Surguladze SA, Calvert GA, Brammer MJ, Campbell R, Bullmore ET, Giampietro V, David AS. Audio-visual speech perception in schizophrenia: an fMRI study. Psychiatry Res 2001;106:1–14. [PubMed: 11231095]

Szycik GR, Munte TF, Dillo W, Mohammadi B, Samii A, Emrich HM, Dietrich DE. Audiovisual integration of speech is disturbed in schizophrenia: an fMRI study. Schizophr Res 2009;110:111–118. [PubMed: 19303257]

Talaraich, J.; Tournoux, P. Co-planar stereotaxic atlas of the human brain. Theime Medical Publishers; New York, New York: 1988.

Thirion B, Pinel P, Meriaux S, Roche A, Dehaene S, Poline JB. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. Neuroimage 2007;35:105–120. [PubMed: 17239619]

van Atteveldt NM, Formisano E, Blomert L, Goebel R. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. Cereb Cortex 2007;17:962–974. [PubMed: 16751298]

van Wassenhove V, Grant KW, Poeppel D. Temporal window of integration in auditory-visual speech perception. Neuropsychologia 2007;45:598–607. [PubMed: 16530232]

Wallace MT. Dyslexia: bridging the gap between hearing and reading. Curr Biol 2009;19:R260–262. [PubMed: 19321145]

Werner S, Noppeney U. Superadditive Responses in Superior Temporal Sulcus Predict Audiovisual Benefits in Object Categorization. Cereb Cortex 2009;20(8):1829–1842. [PubMed: 19923200]

Zilbovicius M, Boddaert N, Belin P, Poline JB, Remy P, Mangin JF, Thivard L, Barthelemy C, Samson Y. Temporal lobe dysfunction in childhood autism: a PET study. Positron emission tomography. Am J Psychiatry 2000;157:1988–1993. [PubMed: 11097965]

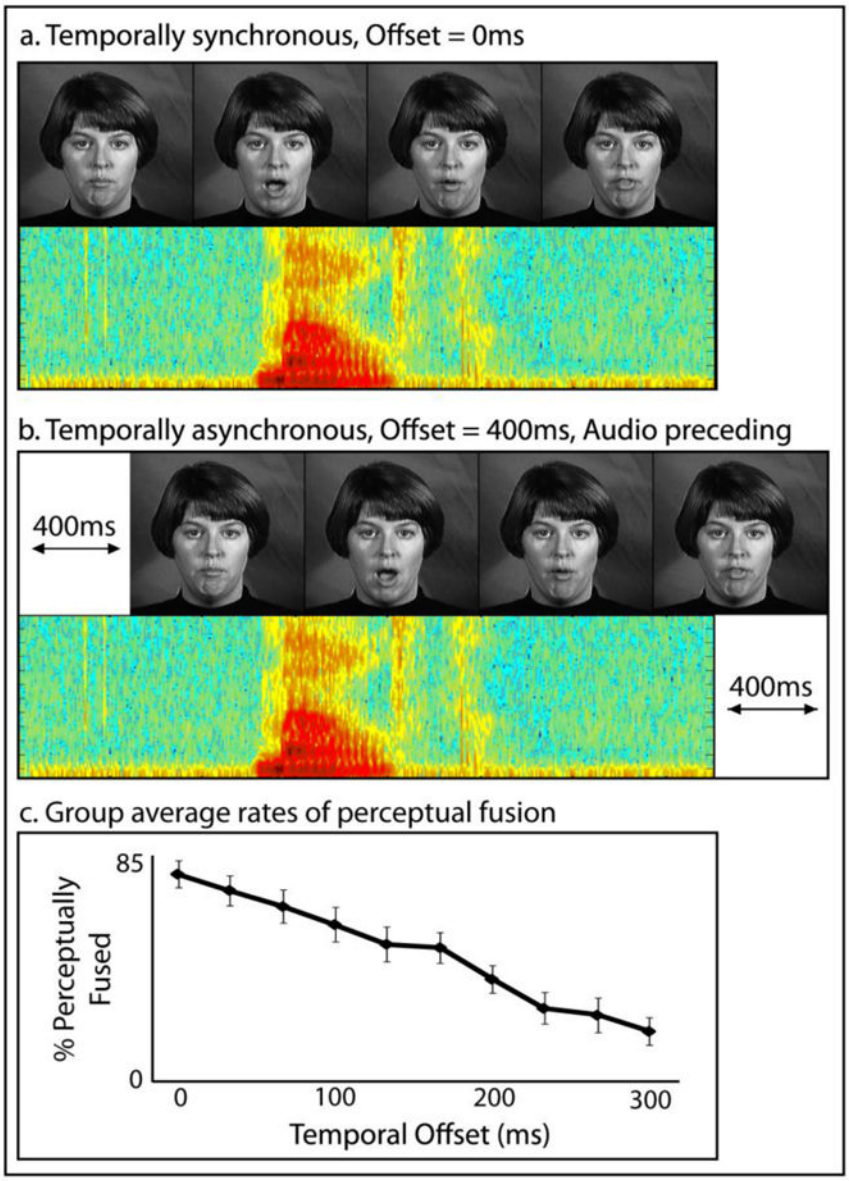

Figure 1. Stimuli

Stimuli included audiovisual presentations of single spoken words presented synchronously (a) or to varying degrees of asynchrony (b), including each individual's 50% perceptual-fusion threshold. Averaged perceptual fusion rates for each level of temporal offset are presented (c).

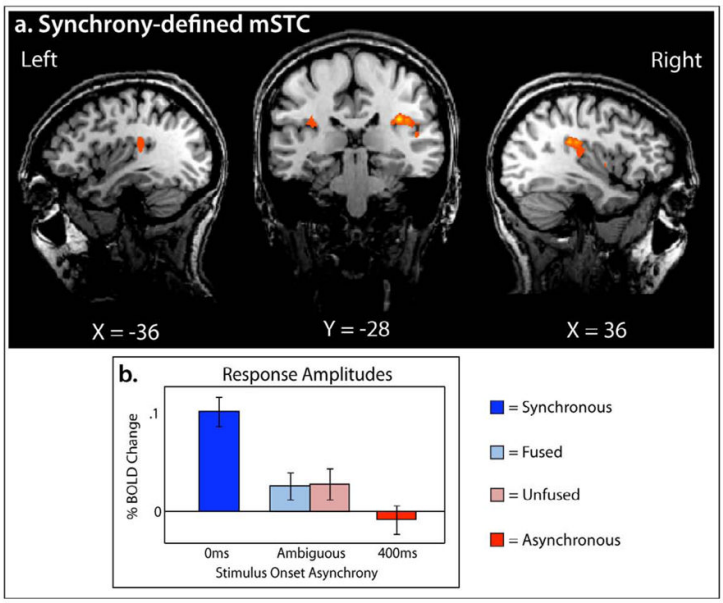

Figure 2. Synchrony-defined mSTC

Synchrony -defined mSTC was defined using a group-average synchronous > asynchronous contrast from the experimental runs. S -mSTC was defined bilaterally (a), and BOLD response amplitudes extracted varied according to temporal synchrony(b).

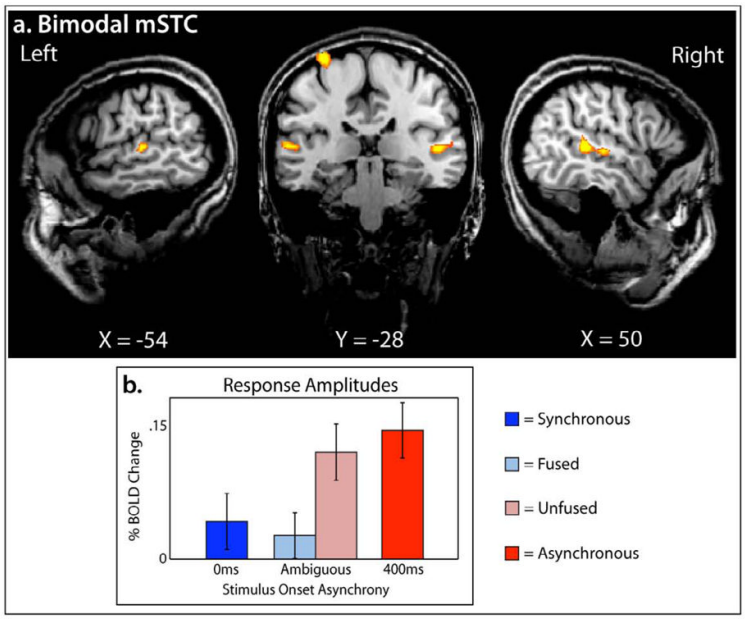

Figure 3. Bimodal mSTC

Bimodal mSTC was defined using a conjunction of two group-average contrasts from the functional localizer runs, audio > baseline and visual > baseline. B-mSTC was defined bilaterally (a), and BOLD response amplitudes varied according to perceptual fusion(b).

Table 1

Multisensory STC subregions

ROI definition	Hemisphere	X	Y	Z	Voxels	t	p
B-mSTC (A ∩ V)	Right	52	−21	11	2656	13.7	$1.73\ e^{-15}$
	Left	−52	−20	10	3807	15.3	$2.09\ e^{-16}$
S-mSTC (Synchrony)	Right	41	−25	20	2502	12.8	$7.94\ e^{-8}$
	Left	−33	−27	21	1000	6.2	$5.07\ e^{-5}$