Published in final edited form as:
Hum Brain Mapp. 2014 November ; 35(11): 5587–5605. doi:10.1002/hbm.22572.

An ALE Meta-Analysis on the Audiovisual Integration of Speech Signals

Laura C. Erickson[a,b], **Elizabeth Heeg**[a], **Josef P. Rauschecker**[b], and **Peter E. Turkeltaub**[a,c]

[a]Department of Neurology, Georgetown University Medical Center, Washington DC, 20057 USA

[b]Department of Neuroscience, Georgetown University Medical Center, Washington DC, 20007 USA

[c]Research Division, MedStar National Rehabilitation Hospital, Washington, DC 20010 USA

Abstract

The brain improves speech processing through the integration of audiovisual (AV) signals. Situations involving AV speech integration may be crudely dichotomized into those where auditory and visual inputs contain 1) equivalent, complementary signals (validating AV speech), or 2) inconsistent, different signals (conflicting AV speech). This simple framework may allow for the systematic examination of broad commonalities and differences between AV neural processes engaged by various experimental paradigms frequently used to study AV speech integration. We conducted an activation likelihood estimation (ALE) meta-analysis of 22 functional imaging studies comprising 33 experiments, 311 subjects, and 347 foci examining "conflicting" versus "validating" AV speech. Experimental paradigms included content congruency, timing synchrony, and perceptual measures, such as the McGurk effect or synchrony judgments, across AV speech stimulus types (sub-lexical to sentence). Co-localization of conflicting AV speech experiments revealed consistency across at least two contrast types (e.g., synchrony and congruency) in a network of dorsal-stream regions in the frontal, parietal, and temporal lobes. There was consistency across all contrast types (synchrony, congruency, and percept) in the bilateral posterior superior/middle temporal cortex. Although fewer studies were available, validating AV speech experiments were localized to other regions, such as ventral-stream visual areas in the occipital and inferior temporal cortex. These results suggest that while equivalent, complementary AV speech signals may evoke activity in regions related to the corroboration of sensory input, conflicting AV speech signals recruit widespread dorsal-stream areas likely involved in the resolution of conflicting sensory signals.

Keywords

cross-modal; language; superior temporal sulcus; activation likelihood estimation (ALE); multisensory; auditory dorsal stream; inferior frontal gyrus; asynchronous; incongruent

Corresponding Author: Peter E. Turkeltaub, 4000 Reservoir Rd, NW, Building D, Suite 165, Washington DC, 20057 USA, turkeltp@georgetown.edu, 202-784-1764.

This research was performed at Georgetown University Medical Center.

1. Introduction

During speech processing, the brain enhances comprehension through the incorporation of both auditory and visual sensory signals, i.e., audiovisual (AV) integration. In most natural settings for speech, auditory and visual sensory inputs are equivalent in content and timing, so integration of these complementary cues can provide validation of sensory information. In other instances, auditory and visual sensory inputs may contribute inconsistent speech signals; conflicting in content and/or timing, in which case, neural processes must resolve the discrepancy for understanding. Common everyday examples include trying to have a conversation with someone in a noisy setting (Nath and Beauchamp, 2011; Sumby and Pollack, 1954), or viewing a dubbed foreign language film or poorly downloaded/synchronized video. Deficits and differences in AV speech integration are associated with several disorders, such as schizophrenia (Ross et al., 2007; Szycik et al., 2009b), Alzheimer's disease (Delbeuck et al., 2007), autism spectrum disorders (Irwin et al., 2011; Smith and Bennetto, 2007; Woynaroski et al., 2013), dyslexia (Blau et al., 2010; Blau et al., 2009; Pekkola et al., 2006) and other learning disabilities (Hayes et al., 2003), and have been found in some cases of focal brain injury (Baum et al., 2012; Hamilton et al., 2006). Thus, understanding the normal processes and brain regions consistently related to AV speech processing may provide insight into the underlying biological substrates associated with these disorders.

AV speech integration can be examined in detail by manipulating the content and timing of auditory and visual signals relative to each other. These types of stimulus manipulations are commonly reported in the multisensory literature (Beauchamp, 2005; Hocking and Price, 2008). Many functional neuroimaging studies across languages have used different types of speech signals (e.g., sub-lexical, words, sentences), manipulations of the AV sensory signals, and measurements of the perceived signals (see Table I for example studies). Manipulations of stimulus sensory characteristics have often included content congruency (e.g., contributing different auditory and visual signals) and timing synchrony (e.g., shifting the onset of the auditory signal relative to the visual signal). AV speech integration can also be assessed based on the perceived signal, which may actually differ from both the auditory and visual signal presented, such as the McGurk effect (McGurk and MacDonald, 1976). The McGurk effect occurs when an entirely new, merged speech percept (e.g., "da"), called the McGurk percept, arises from the resolution of conflicting auditory (e.g., "ba") and visual cues (e.g., "ga"). In general, other percepts, called non-McGurk percepts, can be typically described as the perception of the speech sound (e.g., "ba") or the visual-only facial movements (e.g., "ga"), although other AV combinations have been reported (McGurk and MacDonald, 1976). Similarly, the judgment of fusion of AV sensory events in time is another perceptual measure, which is examined by varying the onset timing of auditory and visual stimuli (Lee and Noppeney, 2011; Miller and D'Esposito, 2005; Noesselt et al., 2012). The fusion percept is the perception of only one sensory event in time and occurs during synchronous or near-synchronous AV speech, while increasingly asynchronous AV speech can lead to perception of two distinct sensory events in time, much like the example of viewing a poorly synched video, where the lips appear to move separately from the speech sounds.

The goal of the current study was to evaluate the neuroimaging literature on AV speech integration through the examination of the brain activity patterns associated with commonly used paradigms including AV stimulus manipulations and percept measurements. While many approaches have been used to assess AV speech integration, when considering the big picture, the results of these different approaches have not been systematically and quantitatively compared. Formal comparisons could demonstrate commonalities or differences in results that would suggest either common or discrete types of AV speech computations associated with different types of AV conflict. Because of the variety of specific experimental manipulations used within the common paradigms, i.e., content, timing, percept reports, a simplified framework was needed to systematically examine co-localization of activity within broadly similar studies.

Although the different manipulations and measurements used to examine AV speech in neuroimaging experiments certainly involve different specific processes in AV integration, in broad terms, these experiments can be thought of as stressing, to varying degrees, two general and fundamental types of operations that are in direct opposition with each other: resolution of *conflict* between discrepant AV sensory signals versus *validation* of the same, complementary AV sensory signals. When the content or timing of the stimulus is equivalent (e.g., sound “ba” is presented synchronously with the visual articulation of “ba”), neural processes related to sensory validation are stressed, since there is no conflict between AV signals. By contrast, when the content or timing of the stimulus is inconsistent (e.g., sound “hotel” is paired with the visual articulation “island” (Szycik et al., 2009a), or sound “tree” is presented 240 ms before visual articulation of “tree”, see (Macaluso et al., 2004)), it is likely that neural operations related to processing conflicting auditory and visual inputs are more strongly stressed compared to when auditory and visual cues are congruent and synchronous.

In experiments examining the perceived AV signal, while there is potentially stress on both conflict resolution and validation processes, the relative stress on each may likely differ depending on the percept. Both the McGurk and non-McGurk percept occur during conflicting AV stimulation, but we suggest that in general the McGurk percept may serve as a behavioral outcome indicating more stress on neural systems responsible for processing AV conflict resolution, represented by the merging of disparate sensory inputs, and less strain on reinforcement of one sensory signal or the other. Conversely, the non-McGurk percept compared to the McGurk percept may suggest relatively less stress on resolution of AV conflict between the sensory signals, and more bias toward bolstering one sensory modality, resulting typically in the perception of either the speech sound or facial movements. Similarly, in the conflicting versus validating framework, the fusion percept may reflect relatively more validation of sensory cues than conflict, whereas the non-fusion percept of asynchronous sensory input may reflect relatively more conflict between auditory and visual input.

Whether AV speech integration is examined based on the sensory stimulus presented or the percept reported, the neural computations of commonly used contrasts across studies can be broadly considered within the conflicting versus validating framework. This meta-analytic framework does group several specific computation types present within the AV speech

literature, synchrony versus congruency versus percept. However, it may still provide an acceptable scheme to integrate findings, and allow for the critical evaluation of the degree of overlap versus the difference among distinct contrast and AV stimulus types across a variety of experimental paradigms in the field. Despite these frequently used approaches, previous studies have mainly focused on specific contrasts and have not typically asked whether there may be more general processing demands inherent to AV speech integration regardless of the specific stimulus or contrast type. Thus, using the proposed conflicting versus validating framework to categorize experiments for meta-analysis is not only useful, but also novel. The conflicting versus validating framework has the potential to inform hypotheses regarding the types of neural operations performed in these brain regions, influence existing models of speech processing (Hickok, 2012; Hickok and Poeppel, 2007; Rauschecker, 2011; Rauschecker and Scott, 2009; Skipper et al., 2007), and allow for the broad-view quantitative examination of neural systems involved in AV speech integration, which is, to the best of our knowledge, lacking in the current literature.

Many brain regions are involved in processing AV speech signals including areas within the auditory dorsal and ventral streams (Hickok and Poeppel, 2007; Rauschecker, 2011; Rauschecker and Scott, 2009; Rauschecker and Tian, 2000), such as the posterior superior temporal sulcus (STS) (Beauchamp et al., 2004a; Beauchamp et al., 2010; Calvert et al., 2000; Hein and Knight, 2008; Raij et al., 2000), the frontal motor areas (Skipper et al., 2005; Skipper et al., 2007), and the inferior frontal gyrus (Ojanen et al., 2005; Sekiyama et al., 2003). Even relatively early sensory areas have demonstrated multimodal speech processes (Bavelier and Neville, 2002; Calvert et al., 1997; Driver and Noesselt, 2008; Hackett and Schroeder, 2009; Pekkola et al., 2005; Sams et al., 1991). However, the extent and constraint of AV computation types occurring within these regions have not been completely examined. Thus, a systematic and quantitative evaluation of the common experimental paradigms within the AV speech literature is needed.

We first hypothesized that, across experiments, the conflicting versus validating framework would capture two general computational characteristics of AV speech integration, which should be reflected in consistent patterns of activity within each type of contrast, and different patterns when comparing conflict versus validation. We hypothesized further that AV speech integration contrasts that stress conflict over validation would require involvement of multisensory regions, such as the posterior STS (Beauchamp et al., 2004b; Beauchamp et al., 2010; Man et al., 2012; Watson et al., 2014), and a larger network of regions proposed in speech-related feedback/error processing, such as auditory dorsal stream areas (Hickok, 2012; Hickok and Poeppel, 2007; Rauschecker, 2011; Rauschecker and Scott, 2009), or in regions proposed to process domain-general conflict resolution and response selection, such as inferior frontal cortex (Novick et al., 2005; Novick et al., 2010). In contrast, we hypothesized that experiments emphasizing validation of AV input over conflict would consistently recruit regions more proximal to sensory areas as compared to frontal and parietal regions hypothesized for processing conflicting AV speech, where sensory areas were defined in terms of relative location to A1 or V1 as compared to conflicting AV speech. This hypothesis was supported by previous studies that have shown increased activity for congruent AV speech in auditory areas (Okada et al., 2013; van Atteveldt et al., 2004; van Atteveldt et al., 2007), and increased activity for non-native,

second language processing of congruent AV speech in visual areas (Barros-Loscertales et al., 2013).

To interrogate these hypotheses, we conducted an activation likelihood estimation (ALE) (Turkeltaub et al., 2002) meta-analysis of 22 functional imaging studies comprising 33 experiments, 311 subjects, and 347 activation foci. These experiments examined conflicting versus validating AV speech including paradigms of content congruency, timing synchrony, and perceptual measures, such as the McGurk effect and other perceptual fusions related to synchrony judgments. These experiments utilized AV speech stimulus types that ranged from sub-lexical to sentence in various languages. Specifically, across experiments we distinguished the brain areas more consistently active when there were discrepancies in sensory signals (conflicting AV speech) versus brain areas more consistently active when sensory signals were in agreement (validating AV speech). We then examined the specific experiments driving the ALE activation patterns to determine the degree to which various specific manipulations of content, timing, and perception overlap in their processing and to what degree these specific experiments differ. Lastly, we assessed the specificity of each ALE cluster for conflicting and validating AV speech, which was examined based on the proximity of foci from validating experiments to conflicting AV speech ALE peaks, and vice versa.

2. Materials and Methods

2.1 Literature Search

Studies published through September 2013 were identified through online searches of PubMed, using EndNote software (endnote.com) and Google Scholar databases for functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) studies using the following key words: “speech”, “audiovisual”, “auditory”, “visual”, “integration”, “cross modal”, “crossmodal”, “McGurk”, and “multisensory” in various combinations. References from studies identified and review articles were also reviewed for additional publications.

2.2 Inclusion Criteria

Studies were included with the following criteria: 1) conducted experiments using fMRI or PET imaging modalities; 2) subjects were normal, healthy participants; 3) stimuli consisted of AV speech, that is speech sounds consisting of either sub-lexical parts of speech (e.g., phonemes, syllables, vowel-consonant-vowel (VCV) tokens, etc.), or words, or sentences, paired with visual stimuli consisting of either video of a speaker or text (e.g., letters; only studies #20, #21); 4) contrasts could be classified to identify activity for conflicting AV stimuli, validating AV stimuli, or differences between them; 5) AV stimuli could be classified as conflicting or validating based on content (incongruent versus congruent) or timing (asynchronous versus synchronous); 6) perceptual measures that could be classified included the McGurk percept, or other perceptual fusions associated with judgments of AV synchrony (e.g., perception of one sensory event or two sensory events close in time); 7) results reported foci in a stereotactic/standard 3-dimensional coordinate system (Talairach or MNI) or foci coordinates were provided by the author (only one study, #1); and 8)

experiments examined the whole brain, or used large slabs covering frontal, temporal, parietal, and occipital cortex (only studies #13, #19, #21), or included functional localizers that were not anatomically restricted to a specific brain region and allowed for the possibility of activity to be found across the whole brain. Among the included studies that reported handedness, all subjects were right-handed with the exception of study #9, where two of the 28 subjects were left-handed. Three studies (#6, #11, #22) did not report handedness. All included experiments used univariate designs. All included studies are listed in Table I with study characteristics noted.

2.3 Exclusion Criteria

Single-subject reports, experiments that assessed non-native/second language processing, and experiments that appeared to report foci within anatomically restricted brain regions were excluded from the meta-analysis. Studies that met all inclusion criteria, but did not report results in the form of 3-dimensional stereotactic coordinates (Talairach or MNI) were also excluded.

2.4 Experiment Classification

Based on the framework described in the Introduction, each individual experiment that met inclusion criteria was broadly classified as contrasting conflict over validation, or validation over conflict in AV signals. A study was defined as a distinct set of subjects. An experiment was defined as a distinct set of subjects tested on a specific AV contrast type, where a distinct set of subjects could be tested on more than one AV contrast type (e.g., study #11). AV contrast types were classified into eight categories and included stimulus contrasts (i.e., incongruent versus congruent, asynchronous versus synchronous), and percept contrasts (i.e., McGurk versus non-McGurk percept, non-fusion versus fusion percept).

Focusing on stimulus contrast types, conflicting AV speech was categorized as discordant AV speech stimuli, either in content incongruence, where auditory and visual speech signals were not the same, and/or presented asynchronously, where the timing was offset between the auditory and visual signals. Conflicting AV speech experiments were classified as contrasts that assessed neural activity related to the comparison of processing incongruent > congruent or asynchronous > synchronous AV speech stimuli. In contrast, validating AV speech was categorized as equivalent auditory and visual speech signals, either in content congruence and/or presented synchronously. In other words, validating AV speech experiments were classified as contrasts that assessed neural activity related to processing when the auditory and visual speech stimuli were the same compared to inconsistent, i.e., congruent > incongruent and synchronous > asynchronous.

For perceptual measures, the contrast of McGurk > non-McGurk percept was classified as conflicting AV speech, and non-McGurk > McGurk percept was classified as validating AV speech. These AV contrast types applied to correlations of activity with number of McGurk responses, where positive correlations were classified as McGurk > non-McGurk percept and negative correlations were classified as non-McGurk > McGurk percept. While both the McGurk percept contrasts have some level of conflict inherent in the AV stimuli, intended to elicit the McGurk percept, we suggest that there may be more conflict processing when the

McGurk percept is reported, which may lead to the merged resolution of AV signals. In contrast, we suggest that the non-McGurk percept may have more processing related to the perception of a particular AV signal, i.e., typically either the sound or the visual input, and less conflict processing related to integration of disparate AV signals. During timing synchrony paradigms, non-fusion > fusion percept was classified as conflicting AV speech, and fusion > non-fusion percept was classified as validating AV speech. Here, the fusion percept was described as the perception of one sensory event and the non-fusion percept was described as the perception of two sensory events in succession. In parallel with McGurk percept processing, we suggest that regardless of the stimulus characteristics of the AV signal, perception of one sensory event indicates relatively more validation than conflict processing related to the timing of the AV signal, whereas the perception of two sensory events in time may represent more conflict present between the AV signals. Importantly, since the perceptual contrasts, McGurk versus non-McGurk percept and non-fusion versus fusion percept, were less clearly accommodated within the conflicting versus validating framework, supplementary ALE analyses were conducted with the exclusion of the percept contrasts.

2.5 ALE Methods

ALE is a quantitative meta-analysis technique that assesses co-localization across neuroimaging (fMRI and PET) studies in the brain using coordinates of activation foci reported in the literature (Turkeltaub et al., 2002; Turkeltaub et al., 2012). To summarize, ALE operates on the assumption that there is "uncertainty" regarding the actual location of foci reported in standardized, stereotaxic brain space (Talairach, MNI). For each set of experiments organized by distinct subject groups, ALE creates a whole-brain map of localization probabilities modeled by three-dimensional Gaussian probability densities distributions. Across experiments, whole-brain voxel-wise cumulative probabilities are calculated to generate an overall ALE map. The voxel-wise ALE value is equal to the probability that at least one study should have activity/foci located there (Turkeltaub et al., 2012); the larger the ALE value, the higher the probability of activity being reported in that location. Significance is assessed using a random-effects significance test against the null hypothesis that localization of activity is independent between studies. Detailed methodological descriptions of the ALE equations and algorithms have been published elsewhere (Eickhoff et al., 2012; Eickhoff et al., 2009; Turkeltaub et al., 2002; Turkeltaub et al., 2012).

We determined the localization of conflicting and validating AV speech in the brain through the assessment of two separate ALE analyses on each set of experiments. Every experiment included in this meta-analysis contrasted two AV conditions that differed in their degree of conflict versus validation, thus, each of the ALE analyses presented represent contrasts between conflict and validation processes in AV integration. ALE analyses were performed using GingerALE 2.1 (www.brainmap.org). Coordinates of foci reported in Talairach space were transformed to MNI space in the GingerALE 2.1 platform, using *tal2icbm* (Laird et al., 2010; Lancaster et al., 2007) or Brett *tal2mni* transform if the coordinates of foci appeared to be previously transformed using this method. GingerALE provides coordinate conversions between Talairach and MNI stereotactic space in both directions. Coordinates of foci were

organized by subject group to eliminate false positives due to within-group effects, as described in Turkeltaub et al. (2012). Significant activation likelihood clusters met the following criteria to reduce false positives (type I errors): 1) a false discovery rate (FDR) $q < .01$ was applied, 2) at least two experiments contributed to each cluster, and 3) a cluster extent threshold > 100 mm^3. The applied cluster extent threshold is commonly used in the ALE literature and has previously been demonstrated to show good sensitivity while reducing false positives (Turkeltaub et al., 2012). Experiments reporting foci within three standard deviations of the calculated localization uncertainty from a peak in the ALE map were considered contributors to that peak (see (Turkeltaub et al., 2011)).

To confirm the specificity of these clusters for conflicting or validating AV speech, we examined whether experiments in the opposite category (validating or conflicting AV speech, respectively) also reported foci within three standard deviations of the calculated localization uncertainty from each ALE peak. For example, validating AV speech experiments reporting foci within three standard deviations of the calculated localization uncertainty from each conflicting AV speech ALE peak were reported as "Nearby Validating Experiments"; whereas, conflicting AV speech experiments containing foci within three standard deviations of each validating AV speech ALE peak were reported as "Nearby Conflicting Experiments".

Supplementary ALE analyses were also conducted: 1) the exclusion of percept contrasts for conflicting AV speech, and 2) the exclusion of percept contrasts for validating AV speech. The supplementary findings are reported with a false discovery rate (FDR) $q < .01$ and a cluster extent threshold > 100 mm^3.

All cluster anatomical locations were verified through a combination of the Automated Anatomical Labeling (AAL) atlas and the Colin27 brain anatomy in MRIcron (http://www.mccauslandcenter.sc.edu/mricro/index.html). Results are displayed on surface renderings and slices of the Colin27 brain using MRIcron.

3. Results

We classified 33 fMRI and PET experiments that met our inclusion criteria. These experiments derived from 22 imaging studies that comprised a total of 311 subjects and 347 foci (Table I). The fMRI/PET experimental designs included block, event-related, and adaptation. Of the 33 experiments, there were 21 sub-lexical level (e.g., phonemes, vowels, syllables, etc.), five word level, and seven sentence level AV speech stimulus types. These studies used active and passive tasks that assessed conflicting versus validating AV speech through the manipulation of sensory stimulus characteristics that differed in content congruency (incongruent versus congruent) or timing synchrony (asynchronous versus synchronous), or through perceptual measures (e.g., the McGurk percept or judgments of the AV fusion percept of sensory events in time). Detailed information regarding study characteristics is located in Table I.

3.1 Localization of Conflicting AV Speech

The ALE analysis of conflicting AV speech included 210 foci from 20 experiments. Ten experiments were incongruent > congruent contrasts, four were asynchronous > synchronous contrasts, three were McGurk > non-McGurk percept contrasts, and three were non-fusion > fusion percept contrasts. The ALE analysis resulted in nine clusters of significant activation likelihood in areas of the frontal, temporal, and parietal lobes (Table II; Figure 1). These ALE findings were consistent whether or not the percept contrasts were included (Supplementary Figure 1). All 17 conflicting AV speech ALE peaks were derived from both content and timing conflicts, 15 from incongruent > congruent and asynchronous > synchronous, and two from incongruent > congruent and non-fusion > fusion percept contrast types.

Two large clusters were identified in the left posterior superior/middle temporal cortex that spanned superior temporal gyrus (STG) through the superior temporal sulcus (STS) to the middle temporal gyrus (MTG), with peak ALE values of 0.0256 and 0.0153, and Y values of −44 and −26, respectively. These ALE clusters derived from all contrast types (stimulus and percept contrasts, Table II), and large range of AV speech stimulus types from sub-lexical to sentence. Nine different experiments in total contributed to the larger left posterior STG/STS/MTG cluster (2008 mm^3). Of the three peaks, the highest ALE peak derived from seven experiments composed of equal number of incongruent > congruent, asynchronous > synchronous, and McGurk > non-McGurk percept contrast types, and only one non-fusion > fusion percept contrast type. Note that two of the three McGurk > non-McGurk percept experiments included in the meta-analysis reported foci here. Nine experiments also contributed to the smaller left posterior STG/STS/MTG cluster (1424 mm^3), with most being either incongruent > congruent (four experiments) or asynchronous > synchronous contrast types (three experiments). Note that three of the four total asynchronous > synchronous contrasts in the conflicting AV speech ALE were localized in this cluster (#1, #6a, #11a). Similar to the left posterior STG/STS/MTG clusters, the right posterior STG/STS/MTG cluster was also derived from all contrast types, more frequently incongruent > congruent (three of seven experiments) and asynchronous > synchronous (two of seven experiments), utilizing both sub-lexical and sentence AV stimulus types.

Among the other clusters outside the temporal lobe, the SMA cluster derived mainly from incongruent > congruent experiments (four of five contributing experiments), comprised mostly of sub-lexical AV speech stimulus types with only one experiment that used a word AV stimulus type, disyllabic nouns. Bilateral dorsal IFG clusters and one smaller ventromedial left IFG cluster were also identified. The left dorsal IFG cluster (1104 mm^3) was most frequently derived from incongruent > congruent experiments (six of eight overall contributing experiments) and these experiments used sub-lexical AV speech stimulus types with the exception of one experiment using disyllabic nouns. One study (#11) contributed foci to this cluster from both asynchronous > synchronous and non-fusion > fusion percept contrast types using sentence stimuli. The right dorsal IFG cluster showed a relatively similar pattern of contributing experiments and AV stimulus types, with four of seven contributing experiments classified as incongruent > congruent. Lastly, the left IPL was

derived from four experiments in which three experiments were classified as incongruent > congruent, and one experiment was classified as asynchronous > synchronous.

3.2 Localization of Validating AV Speech

A smaller ALE analysis of validating AV speech included 137 foci from 13 experiments. Three experiments were congruent > incongruent contrasts, five were synchronous > asynchronous contrasts, three were non-McGurk > McGurk percept contrasts, and two were fusion > non-fusion percept contrasts. The ALE analysis revealed six clusters of significant activation likelihood (Table III; Figure 1) in sensory areas including bilateral fusiform gyrus (FFG), left inferior occipital lobe, and bilateral middle superior temporal gyrus (mid-STG). These findings remain largely the same whether percept contrasts were included or not, with some exceptions as described below (Supplementary Figure 1).

Validating AV speech brain areas were identified in bilateral posterior FFG. A right posterior FFG cluster derived from activity reported in five experiments using sub-lexical and word level AV speech stimuli in two contrast types: synchronous > asynchronous and non-McGurk > McGurk percept. Note that two of the three non-McGurk > McGurk percept experiments within the list of validating AV speech experiments reported activity here. The left posterior FFG cluster derived from similar contrast types, also including two of the three non-McGurk > McGurk percept contrasts in the validating AV speech analysis, as well as congruent > incongruent, using only sub-lexical AV speech stimuli.

The largest cluster (656 mm^3) was located in the right mid-STG with an ALE value of 0.0176 and derived from three experiments classified as congruent > incongruent and fusion > non-fusion percept, using word and sentence AV speech stimulus types. Comparably, a smaller cluster was found in the left mid-STG with an ALE value of 0.0127, derived from two experiments also classified as congruent > incongruent and fusion > non-fusion percept, using sentence AV speech stimuli. This cluster overlaps with the anterior edge of the left STG cluster in the conflicting AV speech ALE map (Figure 1). It was no longer significant when percept contrasts were excluded (Supplementary Figure 1). One other medial cluster was also found in the left mid-STG region, extending deep into inferior white-matter regions, derived from three experiments utilizing three different contrast types, congruent > incongruent, synchronous > asynchronous and non-McGurk > McGurk percept, and both sub-lexical and sentence AV stimulus types.

3.3 Specificity for AV Speech Conflict and Validation

The ALE analyses above revealed largely non-overlapping networks for conflicting and validating AV speech processing. The only area of overlap was in the left mid-STG, where a small cluster in the validating AV speech ALE map overlapped with the anterior edge of a cluster in the conflicting AV speech ALE map. Despite the apparent differences observed in the ALE maps, it remains possible that activity in validating AV speech experiments was reported in the areas identified as involved in AV conflict, but failed to reach significance in the ALE analysis due to lower power in the validating AV speech ALE analysis. In general, threshold effects in the ALE analyses may lead to a false impression of specificity. One approach to address this issue is an ALE subtraction analysis in which the two datasets are

directly compared. The current analysis is underpowered for a direct ALE subtraction, and this approach still may not provide a full picture of the degree of specificity in different areas of the brain. Therefore, to assess the specificity of the ALE results for conflicting and validating AV speech, we examined each ALE peak in both maps and identified "nearby experiments" from the other dataset. "Nearby" was defined by the same criterion used to determine whether experiments in each dataset contribute to their own ALE maps (see Materials and Methods). In other words, we asked "if this validating AV experiment had been included in the conflicting AV dataset, would it have contributed to this ALE result?" and we asked "if this conflicting AV experiment had been included in the validating AV dataset, would it have contributed to this ALE result?".

For the conflicting AV speech map, this specificity analysis demonstrated that all clusters outside the temporal lobes were specific to AV conflict. That is, no validating AV speech experiments reported activity near any of the conflicting AV speech ALE clusters in the parietal or frontal lobes (Table II). In the mid-posterior superior temporal lobe of both hemispheres, a few validating AV speech experiments reported foci near most of the conflicting AV speech ALE peaks (range 0–3 nearby validating experiments). One validating AV speech experiment (#11b) using a fusion > non-fusion percept contrast was responsible for much of this overlap. Notably, substantially more conflicting AV speech experiments compared to validating AV speech experiments reported foci within the temporal lobe (see Table II).

In parallel to the findings for the conflicting AV speech map, the validating AV speech map showed no specificity for validating AV speech in the mid-STG ALE clusters. Eight nearby conflicting AV speech experiments were identified for one left mid-STG ALE peak, one nearby conflicting AV speech experiment for the other medial left mid-STG ALE peak, and two nearby conflicting AV speech experiments were identified for the right mid-STG ALE peak (Table III). These findings suggest that the mid-STG may not be involved in processes exclusive to conflicting or validating AV speech. The left inferior occipital lobe ALE cluster had one nearby conflicting AV speech experiment. In contrast, no foci from conflicting AV speech experiments were found near the left or right FFG clusters identified in the validating AV speech ALE analysis, suggesting these areas may be engaged in processes specific to AV speech validation (see Table III).

3.4 Complementary Findings with the Removal of Percept Contrast Types

Although the conflicting versus validating dichotomization clearly captures key processing differences based on our results, there are gray areas around the boundary between the categories. The gray areas are particularly related to the percept contrast classifications, although only a small number of percept contrasts were included in each analysis. For example, some conflict-related activity might be expected in a non-McGurk > McGurk percept contrast, even though the experiment was classified as validating AV speech. In general, this issue should have diluted our findings, creating apparent overlap between processes related to AV conflict and validation. However, this was not the case, as there was not sufficient overlap to warrant excluding percept studies all together. Regardless, to address this potential shortcoming, we conducted an additional ALE analysis with the

exclusion of percept contrasts (Supplementary Figure 1). Excluding percept contrasts did not significantly alter the main findings, suggesting that the overall observed patterns do not critically depend on relatively subjective decisions, like the assignment of percept contrasts within the conflicting versus validating framework. As discussed above, the main difference of note was that the left mid-STG validating AV speech cluster, which overlapped with the anterior edge of the conflicting AV speech cluster, was not identified with the exclusion of percept contrasts. This result was not surprising since the cluster was derived from two experiments, including one percept contrast.

4. Discussion

Using the ALE meta-analysis technique, we identified distinct brain regions that were consistently more active during the resolution of discrepancies in sensory input (conflicting AV speech) or the reinforcement of complementary sensory input (validating AV speech) in a large number of neuroimaging studies across several languages. The conflicting versus validating framework allowed for the critical evaluation of localization overlap among different contrast and AV stimulus types reflective of the AV literature. Overall, there was consistency in localization within each of these groups of experiments, more convincingly for conflicting AV speech, despite the wide variation in experimental methods (e.g., task, design) and kinds of AV stimulation (e.g., manipulations of timing versus content, sub-lexical versus sentence). These findings remained largely the same whether or not percept contrasts were included in the meta-analysis (Supplementary Figure 1). In general, these results indicate a partial dichotomy of AV processes that serve to resolve conflict between discrepant AV signals versus those that serve to validate equivalent AV signals, which is reflected by a reliance on distinct brain regions. Within these broad brain networks, patterns were observed wherein specific types of speech signals or contrast types (e.g., synchrony versus congruency, conflict versus validation) were more likely to activate specific regions than others. These differences may inform the specific roles of these regions in AV integration beyond the simple conflict versus validation dichotomy. These findings are relevant to current sensorimotor speech models (Hickok, 2012; Hickok and Poeppel, 2007; Rauschecker, 2011; Rauschecker and Scott, 2009; Skipper et al., 2007), and indicate that the auditory dorsal stream may be important during AV speech conflict processing.

4.1 Recruitment of Bilateral Posterior Temporal Areas for Processing of Conflicting AV Speech Integration

Every contrast type contained within this analysis, including experiments examining conflicts in content and timing of AV signals and perceptual measures, that ranged from sub-lexical to sentence level AV speech stimulus types, consistently activated the same regions of the bilateral posterior STG/STS/MTG, in general spanning an area more lateral and posterior to the validating AV speech clusters. While a few validating AV speech experiments reported foci near the posterior temporal clusters, most experiments reporting foci here used conflicting AV speech contrasts. These results provide preliminary evidence that the posterior STG/STS region may be involved in general AV sensory integration processes that are stressed by the presence of conflict between auditory and visual signals.

The posterior STS conflict-related activation likelihood was left lateralized both in terms of peak ALE values and the total volume of significant ALE clusters. Left and to a lesser degree right posterior STG/STS has been argued to provide storage of and access to phonological representations of speech ((Hickok and Poeppel, 2007), but see (DeWitt and Rauschecker, 2012)), and is activated in auditory speech studies without visual input (Turkeltaub and Coslett, 2010). It could be argued that the posterior STS might play no role in AV integration, but that conflicting AV signals induce competing co-activation of multiple phonemic or lexical representations, placing stress on the left posterior STG/STS storage/access system and resulting in greater brain activity in this area. However, the stimuli included here represent an array of speech signals, and recent meta-analytic evidence suggests that auditory speech representations reside farther anterior with sublexical units, words and phrases hierarchically arrayed along a gradient from the mid-to-anterior STG/STS (DeWitt and Rauschecker, 2012). Also, competing co-activation of speech representations could theoretically cause a decrease rather than an increase in activity in these storage/access areas, if the conflict results in mutual inhibition of the two competing representations.

If posterior temporal areas serve a different purpose in speech processing, it remains possible that conflict in AV signals places strain on more general processes which contribute to speech that is served by the posterior STG/STS region, such as phonological working memory (Leff et al., 2009), resulting in greater activity in this area for conflicting AV speech signals. Validating AV contrasts may engage these general processes as well, albeit to a lesser degree, resulting in inconsistent activity in the posterior STG/STS, as we observed here. However, the posterior STS has also been implicated in AV integration for non-speech signals (Beauchamp et al., 2004a; Beauchamp et al., 2004b; Man et al., 2012), making this unlikely.

Rather, the consistency of activity in the bilateral posterior STS observed here, across all contrast types, particularly those that stress AV conflict, likely suggests a direct role for this region in comparison of auditory and visual inputs not specific to speech stimuli. This may result in greater fMRI signal when there is discrepancy between the inputs, which could be related to the recruitment of more neural processes responding to the different auditory signal, visual signal, or both (see (Hocking and Price, 2008)). This is supported by previous work in demonstrating connections between the STS and auditory/visual areas (Beer et al., 2011; Falchier et al., 2002; Rockland and Pandya, 1981), and that the STS does indeed have a “patchy” organization containing both AV and unisensory areas (Beauchamp et al., 2004a; Dahl et al., 2009). Other neuroimaging studies also provide further support. A multivariate pattern analysis of posterior STS identified similar neural patterns for both the sound and video of particular objects (Man et al., 2012). One study of non-speech AV stimuli showed effective connectivity changes between posterior STS and auditory/visual areas after AV synchrony discrimination training (Powers et al., 2012), suggesting that STS may help to discriminate timing-related perceptions of AV sensory events. Sensory signal accuracy may contribute to STS connectivity patterns; one study showed increased reliability of speech sounds compared to visual speech movements correlated with the increased functional connectivity between posterior STS and auditory cortex (Nath and Beauchamp, 2011), indicating that posterior STS may evaluate which sensory input is more likely to be accurate.

In another study, the left pSTS was recruited with the addition of noise to conflicting AV speech stimuli (Sekiyama et al., 2003). Lastly, a transcranial magnetic stimulation (TMS) study (Beauchamp et al., 2010) and a case study on a patient with damage to the left STS (Baum et al., 2012) provide further evidence for bilateral STS involvement in AV conflict processing. Beauchamp et al. (2010) found that inhibitory TMS of the left pSTS greatly reduced the number of "fused" McGurk percept reports within a specific time window. Baum et al. (2012) reported on a patient that could still perceive the McGurk effect with left STS damage and with an intact right STS, where this patient had increased right STS activity compared to healthy controls during McGurk stimuli presentation. These findings (Baum et al., 2012; Beauchamp et al., 2010) suggest that the left and right STS may have complementary functions in processing AV conflicts. Overall, previous studies suggest that the neural computations performed by the STS are necessary for interpreting, and in some cases, resolving AV sensory inconsistencies.

The posterior STS was identified in the meta-analysis through the overlap of mostly different conflicting AV speech contrast and stimulus types, however, a few validating AV speech experiments reported foci nearby. Thus, it could be that the posterior STS is involved in more general sensory processes not specific to conflict between AV signals, or restricted to multisensory AV inputs. The STS, of both the left and right hemisphere, may have a role in numerous types of computations (Hein and Knight, 2008), both unimodal and multimodal (Allison et al., 2000; Beauchamp et al., 2004a; Beauchamp et al., 2004b; Beauchamp et al., 2008; Bidet-Caulet et al., 2005; Calvert et al., 2000; Giese and Poggio, 2003; Grossman and Blake, 2002; Lahnakoski et al., 2012; Man et al., 2012; Noesselt et al., 2007; Peelen et al., 2010; Pelphrey et al., 2003; Pelphrey et al., 2004; Raij et al., 2000; Redcay, 2008; Watson et al., 2014), suggesting the STS may merge different kinds of sensory information, and possibly, allow for the identification of general sensory discrepancies. Future experiments are needed to test whether these potential conflict detection/resolution processes are domain-general and extend beyond speech processes, perhaps to sensorimotor actions (Rauschecker, 2011; Rauschecker and Scott, 2009). It also remains possible that the posterior STS region may compute comparisons between conflicting stimuli, and specific neuronal populations that receive inputs from different types of signals may be intermingled or spatially segregated (synchrony versus congruency or conflict versus validation). Some studies have started to distinguish discrete processing regions in the superior temporal cortex and STS (Beauchamp et al., 2004a; Noesselt et al., 2012; Stevenson et al., 2010; Stevenson and James, 2009; Stevenson et al., 2011; van Atteveldt et al., 2010). However, because the current ALE study did not have high enough spatial resolution, we could not reliably identify small differences in localization of activity for different types of signal (e.g., synchrony versus congruency or conflict versus validation). With attention to specific localization of various unimodal and cross-modal computations within individual subjects (Beauchamp et al., 2010; Nath and Beauchamp, 2012), future studies using other more precise parcellation methods are clearly needed to continue to investigate the specific functions and organization of the STG/STS.

4.2 Dorsal Stream Structures Involved in Conflicting AV Speech Integration

In addition to the posterior STG/STS regions, conflicting AV speech consistently activated frontal and parietal regions within the dorsal "how/where" auditory stream (Hickok, 2012; Hickok and Poeppel, 2007; Rauschecker, 2011; Rauschecker and Scott, 2009; Rauschecker and Tian, 2000), including the left IPL, SMA, right precentral gyrus, and bilateral dorsal IFG. This network of dorsal stream regions identified outside of the temporal lobe may be specific to AV conflict, since no foci from validating AV speech experiments were identified near these conflicting AV speech ALE peaks.

The auditory/language dorsal stream may constitute a sensorimotor feedback system, whereas the ventral stream may process inputs related to object recognition and comprehension (Hickok and Poeppel, 2007; Rauschecker, 2011; Rauschecker and Scott, 2009). A central sensorimotor mechanism of the dorsal stream, includes an error detection process (Rauschecker, 2011; Rauschecker and Scott, 2009), which suggests that the dorsal stream may be well-suited to contribute to conflict resolution. It is likely that conflict in the AV signal stresses these sensorimotor feedback systems, because the auditory and visual signals are composed of different information, and these sensorimotor interactions in the dorsal stream may help to resolve the discrepancy (Hickok, 2012; Hickok and Poeppel, 2007; Rauschecker, 2011; Rauschecker and Scott, 2009). In general, for these reasons, dorsal stream regions may be linked to the interpretation of ambiguous or inconsistent sensory input. For example, one auditory dorsal stream model suggests that the IFG, premotor areas, IPL, and posterior superior temporal regions contribute to these sensorimotor feedback mechanisms to minimize error and help with "disambiguation" of phonological input (Rauschecker, 2011; Rauschecker and Scott, 2009), which is likely highly significant when sensory input is in disagreement.

These dorsal-stream areas may not be specific to processing speech but perhaps extend to "doable" actions (Rauschecker, 2011; Rauschecker and Scott, 2009), and may be involved in comparisons between other classes of sensory stimuli. The left posterior STG and IPL have been recruited not only during comparisons of speech sounds (Turkeltaub and Coslett, 2010), but also during tasks of perceptual color discrimination (Tan et al., 2008) and have been implicated in stimulus change detection, not exclusive to speech (Zevin et al., 2010). The IPL has been associated with visual-tactile integration (Pasalar et al., 2010), and with detection of conflicting sensorimotor input, including increased activation when there is conflict between motor actions and visual feedback related to "agency" perception (Farrer et al., 2003). Similarly, beyond its classical role in speech output, the IFG has been implicated in processing visual "symbolic gestures" (Xu et al., 2009), and conflict resolution for response selection from competing options (January et al., 2009; Novick et al., 2005; Novick et al., 2010). Previous meta-analytic evidence evaluated "interference resolution" in other types of conflict-related tasks, such as stroop, and showed recruitment of some similar regions, e.g., IPL and IFG (Nee et al., 2007). Within speech processing, the IFG and pre-SMA have also been implicated in categorical processing of phonemes (Lee et al., 2012), as has the premotor cortex (Chevillet et al., 2013). Notably, the experiments that activated the IFG and SMA for AV conflict processing in this meta-analysis most frequently used sub-lexical AV stimuli, which might suggest that these areas become involved in resolving

conflict in AV speech signals because of their role in discriminating between sublexical speech units. Overall, this network of mostly dorsal stream regions (Hickok, 2012; Hickok and Poeppel, 2007; Rauschecker, 2011; Rauschecker and Scott, 2009) was co-localized across studies during processing of conflicting AV speech and showed selectivity to the conflicting AV signals. We suggest that these dorsal stream regions may be involved in the detection and resolution of sensory discrepancies among multimodal functions and in the selection of a single response among multiple viable options.

Overall, different types of conflicting AV speech contrasts co-localized across the dorsal stream network, although there was a degree of selectivity in certain brain areas. The activation likelihood in frontal and parietal areas was mainly derived from experiments using comparisons between incongruent and congruent content, likely influenced by the large number of these experiments included in this analysis (10 out of 20). For example, left IPL and SMA were activated by predominantly incongruent > congruent experiments (three experiments for left IPL, and four and three experiments for SMA ALE peaks) with only one asynchronous > synchronous experiment (#8a) identified for each ALE peak. The most likely explanation for the high influence of incongruent > congruent contrasts in the conflicting AV speech ALE findings is that the greatest degree of conflict between auditory and visual signals occurs when the content of these signals conflict, and this conflict drives activity in parietal and frontal dorsal stream areas. However, as noted above, other AV contrast types (asynchrony and percept) did identify activity in the bilateral STG/STS and overall, 15 of the 17 conflicting AV speech ALE peaks were derived from both incongruent > congruent and asynchronous > synchronous contrast types. As discussed elsewhere, percept comparisons involve relatively subtle differences in AV conflict (e.g., McGurk versus non-McGurk percept), and thus may be sufficient to activate posterior STG/STS regions specifically involved in AV integration. However, the activity in these experiments may be less robust in dorsal-stream areas involved in domain-general conflict processing and response selection.

4.3 Sensory Areas in Validating AV Speech Integration

Compared to the widespread network of brain regions recruited in processing conflicting AV speech, including frontal and parietal areas, brain areas involved in the processing of validating AV speech were localized to more proximal auditory and visual areas of the temporal and occipital cortex, including bilateral FFG, left inferior occipital lobe, and to a lesser degree bilateral mid-STG. In general, while these activation likelihoods were derived from a small number of contributing experiments, they still preliminarily establish coherence among the literature and suggest that validating compared to conflicting AV sensory inputs may generate more activity in auditory and ventral stream visual areas of the temporal lobe, including the FFG. It is possible that consistent visual speech paired with auditory speech may create a more explicit, unambiguous signal in these areas. In other words, complementary, redundant speech information contributed by each sensory input may help to boost the most accurate signal and lead to reinforcement of the correct perception (see (Ghazanfar and Schroeder, 2006)). General mechanisms of AV validation could include increased bottom-up activity in neurons receiving the same speech information from separate sensory sources, or top-down tuning in the form of inhibition of

similar, yet incorrect signals (see other AV integration (van Atteveldt et al., 2009) or multisensory models (Driver and Noesselt, 2008)). Others have proposed that many more sensory areas than previously assumed may have multimodal properties (Driver and Noesselt, 2008; Ghazanfar and Schroeder, 2006; Hackett and Schroeder, 2009), and previous studies have shown plasticity of sensory areas in blind or deaf individuals (Amedi et al., 2003; Amedi et al., 2007; Bavelier and Neville, 2002; Bedny et al., 2011; Finney et al., 2001; Rauschecker, 1995; Renier et al., 2010; Striem-Amit and Amedi, 2014; Striem-Amit et al., 2012; Weeks et al., 2000). A recent study of non-native, second language processing recruited bilateral occipital cortex during congruent versus incongruent stimulation of AV sentences (Barros-Loscertales et al., 2013). Other studies have shown FFG activation in voice/speaker recognition tasks of auditory-only speech (von Kriegstein et al., 2005), and FFG recruitment during face processing (Haxby et al., 2000; Hoffman and Haxby, 2000).

While bilateral mid-STG was recruited for validating AV speech, these ALE peaks were less conclusive. Conflicting AV speech experiments were identified near these mid-STG ALE peaks. One left mid-STG ALE peak overlapped with the anterior portion of a conflicting AV speech cluster and was not identified when percept contrasts were excluded. These findings indicate that this mid-STG region may not be exclusive to processing specific types of AV signals. Dewitt and Rauschecker (2012) have proposed that the mid-STG may correspond to the auditory lateral belt in non-human primates and Ghazanfar and Schroeder (2006) have suggested that auditory core and lateral belt are multisensory, responding to auditory, visual, and somatosensory input. Some previous experiments indicate that classical auditory areas in the STG may be involved in processing congruent AV speech signals. For example, others have shown modulation of auditory cortex during lip-reading (Calvert et al., 1997; Calvert and Campbell, 2003; Kauramäki et al., 2010; Pekkola et al., 2005), increased auditory cortex activity during congruent compared to incongruent stimulation of AV phoneme sounds presented with visual letters (van Atteveldt et al., 2004; van Atteveldt et al., 2007), increased auditory cortex activity with stimulation of congruent AV syllables compared to sounds only (Okada et al., 2013), and face/voice integration in auditory cortex in non-human primates (Ghazanfar et al., 2008). While this analysis may provide preliminary evidence for the possibility of cross-modal validation of AV speech in regions more proximal to sensory areas as compared to frontal and parietal regions found for conflicting AV speech, future studies are certainly needed to further examine the interaction of different types of sensory inputs in sensory regions in humans, particularly concerning the mid-STG region.

4.4 Limitations

While we acknowledge that the conflicting versus validating framework may not capture all nuances of the processes involved in AV speech integration, this framework did allow for the broad quantitative examination of AV speech imaging experiments. Conflicting AV speech had more robust findings with the inclusion of 20 experiments and perhaps as a result, a higher degree of co-localization across experiments. This analysis included two experiments (#20, #21) using stimuli that paired speech sounds to letters, and we recognize it is likely there are differences in neural processing related to moving versus static/

orthographic visual signals, particularly concerning attention effects and temporal components. However, both experiments did contribute to activation likelihoods found for conflicting AV speech, indicating that despite computational differences these AV integration processes may still localize to similar brain regions. The validating AV speech analysis had relatively less co-localization across experiments and less overall specificity to validating AV speech. Thus, the validating AV speech findings should be interpreted with caution pending more research in this area.

Sub-analyses related to the isolation of specialized areas for different types of computations (synchrony versus congruency versus percept; moving versus static/orthographic visual signals) were not possible due to the relatively small number of studies reporting foci for each contrast type. Because of this limitation, and because we acknowledge that the computations required for comparisons of timing and content must differ, we have provided detailed information regarding which specific contrast types contributed to each ALE cluster, demonstrating where these experiments co-localized and where they did not (Table II; Table III). Notably, though, all 17 ALE peaks identified in the conflicting AV speech analysis were recruited by both content and timing contrast types (incongruent > congruent and either asynchronous > synchronous or non-fusion > fusion percept), suggesting that there may be broad co-localization of content and timing processes within those brain regions. It is also important to note that ALE operates at a relatively low spatial resolution (roughly similar to PET resolution), and that co-localization of activity from experiments testing different types of conflict or validation (e.g., content versus timing) does not necessarily indicate that these processes rely on the same neuronal populations. The general location of the activity is the same, but more specialized sub-regions within these broader areas may specifically process one type of input or another. Using these findings as a springboard, future studies using more precise methods can further examine these possibilities, likely in within-subject comparisons.

5. Conclusions

In this ALE meta-analysis of 33 experiments, 311 subjects, and 347 foci, we identified distinct brain regions involved in the integration of conflicting versus validating AV speech, confirming that different neural computations are likely responsible for the detection and resolution of inconsistent AV speech versus the validation of equivalent, complementary AV signals. Conflicting AV speech integration revealed a network of primarily dorsal-stream regions involved in the resolution of inconsistent sensory input. In contrast, validating AV speech integration was localized to ventral-stream visual areas of the occipital and inferior temporal lobe, suggesting functional properties related to the validation of complementary AV input. Future studies can assess whether these networks translate to other communication domains, such as face/voice integration, other sensorimotor functions, biological motion, or social-related processes. Additionally, localization of AV speech integration networks for a normal, healthy population provides the foundation for future studies in populations where this network may be altered.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This material is based upon work supported by the National Science Foundation (NSF) Graduate Research Fellowship Program under Grant Nos. DGE-0903443 and DGE-1444316 to L.C.E. This work was also supported by National Institutes of Health (NIH) Grant Nos. KL2TR000102 to P.E.T., R01EY018923 to J.P.R. and NSF PIRE Grant OISE-0730255 to J.P.R., and NIH Training Grant T32 NS041231 also funded L.C.E. We would also like to thank the reviewers for their valuable comments and contributions.

References

Allison T, Puce A, McCarthy G. Social perception from visual cues: role of the STS region. Trends Cogn Sci. 2000; 4(7):267–278. [PubMed: 10859571]

Amedi A, Raz N, Pianka P, Malach R, Zohary E. Early 'visual' cortex activation correlates with superior verbal memory performance in the blind. Nat Neurosci. 2003; 6(7):758–66. [PubMed: 12808458]

Amedi A, Stern WM, Camprodon JA, Bermpohl F, Merabet L, Rotman S, Hemond C, Meijer P, Pascual-Leone A. Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. Nat Neurosci. 2007; 10(6):687–9. [PubMed: 17515898]

Barros-Loscertales A, Ventura-Campos N, Visser M, Alsius A, Pallier C, Avila Rivera C, Soto-Faraco S. Neural correlates of audiovisual speech processing in a second language. Brain Lang. 2013; 126(3):253–62. [PubMed: 23872285]

Baum SH, Martin RC, Hamilton AC, Beauchamp MS. Multisensory speech perception without the left superior temporal sulcus. Neuroimage. 2012; 62(3):1825–32. [PubMed: 22634292]

Bavelier D, Neville HJ. Cross-modal plasticity: where and how? Nat Rev Neurosci. 2002; 3(6):443–52. [PubMed: 12042879]

Beauchamp MS. See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. Curr Opin Neurobiol. 2005; 15(2):145–53. [PubMed: 15831395]

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci. 2004a; 7(11):1190–1192. [PubMed: 15475952]

Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron. 2004b; 41(5):809–23. [PubMed: 15003179]

Beauchamp MS, Nath AR, Pasalar S. fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. J Neurosci. 2010; 30(7):2414–7. [PubMed: 20164324]

Beauchamp MS, Yasar NE, Frye RE, Ro T. Touch, sound and vision in human superior temporal sulcus. Neuroimage. 2008; 41(3):1011–1020. [PubMed: 18440831]

Bedny M, Pascual-Leone A, Dodell-Feder D, Fedorenko E, Saxe R. Language processing in the occipital cortex of congenitally blind adults. Proc Natl Acad Sci U S A. 2011; 108(11):4429–34. [PubMed: 21368161]

Beer AL, Plank T, Greenlee MW. Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. Exp Brain Res. 2011; 213(2–3):299–308. [PubMed: 21573953]

Bidet-Caulet A, Voisin J, Bertrand O, Fonlupt P. Listening to a walking human activates the temporal biological motion area. Neuroimage. 2005; 28(1):132–9. [PubMed: 16027008]

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(Pt 3):868–79. [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(6):503–8. [PubMed: 19285401]

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS. Activation of auditory cortex during silent lipreading. Science. 1997; 276(5312):593–6. [PubMed: 9110978]

Calvert GA, Campbell R. Reading speech from still and moving faces: the neural substrates of visible speech. J Cogn Neurosci. 2003; 15(1):57–70. [PubMed: 12590843]

Calvert GA, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol. 2000; 10(11):649–657. [PubMed: 10837246]

Chevillet MA, Jiang X, Rauschecker JP, Riesenhuber M. Automatic phoneme category selectivity in the dorsal auditory stream. J Neurosci. 2013; 33(12):5208–15. [PubMed: 23516286]

Dahl CD, Logothetis NK, Kayser C. Spatial organization of multisensory responses in temporal association cortex. J Neurosci. 2009; 29(38):11924–32. [PubMed: 19776278]

Delbeuck X, Collette F, Van der Linden M. Is Alzheimer's disease a disconnection syndrome? Evidence from a crossmodal audio-visual illusory experiment. Neuropsychologia. 2007; 45(14): 3315–23. [PubMed: 17765932]

DeWitt I, Rauschecker JP. Phoneme and word recognition in the auditory ventral stream. Proc Natl Acad Sci U S A. 2012; 109(8):E505–14. [PubMed: 22308358]

Driver J, Noesselt T. Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron. 2008; 57(1):11–23. [PubMed: 18184561]

Eickhoff SB, Bzdok D, Laird AR, Kurth F, Fox PT. Activation likelihood estimation meta-analysis revisited. Neuroimage. 2012; 59(3):2349–61. [PubMed: 21963913]

Eickhoff SB, Laird AR, Grefkes C, Wang LE, Zilles K, Fox PT. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. Hum Brain Mapp. 2009; 30(9):2907–26. [PubMed: 19172646]

Falchier A, Clavagnier S, Barone P, Kennedy H. Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci. 2002; 22(13):5749–59. [PubMed: 12097528]

Farrer C, Franck N, Georgieff N, Frith CD, Decety J, Jeannerod M. Modulating the experience of agency: a positron emission tomography study. Neuroimage. 2003; 18(2):324–33. [PubMed: 12595186]

Finney EM, Fine I, Dobkins KR. Visual stimuli activate auditory cortex in the deaf. Nat Neurosci. 2001; 4(12):1171–3. [PubMed: 11704763]

Ghazanfar AA, Chandrasekaran C, Logothetis NK. Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. J Neurosci. 2008; 28(17):4457–69. [PubMed: 18434524]

Ghazanfar AA, Schroeder CE. Is neocortex essentially multisensory? Trends Cogn Sci. 2006; 10(6): 278–85. [PubMed: 16713325]

Giese MA, Poggio T. Neural mechanisms for the recognition of biological movements. Nat Rev Neurosci. 2003; 4(3):179–92. [PubMed: 12612631]

Grossman ED, Blake R. Brain areas active during visual perception of biological motion. Neuron. 2002; 35(6):1167–75. [PubMed: 12354405]

Hackett TA, Schroeder CE. Multisensory integration in auditory and auditory-related areas of cortex. Hear Res. 2009; 258(1–2):1–3. [PubMed: 19932881]

Hamilton RH, Shenton JT, Coslett HB. An acquired deficit of audiovisual speech processing. Brain Lang. 2006; 98(1):66–73. [PubMed: 16600357]

Haxby JV, Hoffman EA, Gobbini MI. The distributed human neural system for face perception. Trends Cogn Sci. 2000; 4(6):223–233. [PubMed: 10827445]

Hayes EA, Tiippana K, Nicol TG, Sams M, Kraus N. Integration of heard and seen speech: a factor in learning disabilities in children. Neurosci Lett. 2003; 351(1):46–50. [PubMed: 14550910]

Hein G, Knight RT. Superior temporal sulcus--It's my area: or is it? J Cogn Neurosci. 2008; 20(12): 2125–36. [PubMed: 18457502]

Hickok G. Computational neuroanatomy of speech production. Nat Rev Neurosci. 2012; 13(2):135–45. [PubMed: 22218206]

Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007; 8(5): 393–402. [PubMed: 17431404]

Hocking J, Price CJ. The role of the posterior superior temporal sulcus in audiovisual processing. Cereb Cortex. 2008; 18(10):2439–49. [PubMed: 18281303]

Hoffman EA, Haxby JV. Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat Neurosci. 2000; 3(1):80–4. [PubMed: 10607399]

Irwin JR, Tornatore LA, Brancazio L, Whalen DH. Can children with autism spectrum disorders "hear" a speaking face? Child Dev. 2011; 82(5):1397–403. [PubMed: 21790542]

January D, Trueswell JC, Thompson-Schill SL. Co-localization of stroop and syntactic ambiguity resolution in Broca's area: implications for the neural basis of sentence processing. J Cogn Neurosci. 2009; 21(12):2434–44. [PubMed: 19199402]

Kauramäki J, Jääskeläinen IP, Hari R, Möttönen R, Rauschecker JP, Sams M. Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. J Neurosci. 2010; 30(4):1314–21. [PubMed: 20107058]

Lahnakoski JM, Glerean E, Salmi J, Jääskeläinen IP, Sams M, Hari R, Nummenmaa L. Naturalistic FMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. Front Hum Neurosci. 2012; 6:233. [PubMed: 22905026]

Laird AR, Robinson JL, McMillan KM, Tordesillas-Gutierrez D, Moran ST, Gonzales SM, Ray KL, Franklin C, Glahn DC, Fox PT, et al. Comparison of the disparity between Talairach and MNI coordinates in functional neuroimaging data: validation of the Lancaster transform. Neuroimage. 2010; 51(2):677–83. [PubMed: 20197097]

Lancaster JL, Tordesillas-Gutierrez D, Martinez M, Salinas F, Evans A, Zilles K, Mazziotta JC, Fox PT. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. Hum Brain Mapp. 2007; 28(11):1194–205. [PubMed: 17266101]

Lee H, Noppeney U. Long-term music training tunes how the brain temporally binds signals from multiple senses. Proc Natl Acad Sci U S A. 2011; 108(51):E1441–50. [PubMed: 22114191]

Lee YS, Turkeltaub P, Granger R, Raizada RD. Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. J Neurosci. 2012; 32(11):3942–8. [PubMed: 22423114]

Leff AP, Schofield TM, Crinion JT, Seghier ML, Grogan A, Green DW, Price CJ. The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: evidence from 210 patients with stroke. Brain. 2009; 132(Pt 12):3401–10. [PubMed: 19892765]

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(2):725–32. [PubMed: 14980575]

Man K, Kaplan JT, Damasio A, Meyer K. Sight and sound converge to form modality-invariant representations in temporoparietal cortex. J Neurosci. 2012; 32(47):16629–36. [PubMed: 23175818]

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264(5588):746–748. [PubMed: 1012311]

Miller LM, D'Esposito M. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J Neurosci. 2005; 25(25):5884–93. [PubMed: 15976077]

Nath AR, Beauchamp MS. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. J Neurosci. 2011; 31(5):1704–14. [PubMed: 21289179]

Nath AR, Beauchamp MS. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. Neuroimage. 2012; 59(1):781–7. [PubMed: 21787869]

Nee DE, Wager TD, Jonides J. Interference resolution: insights from a meta-analysis of neuroimaging tasks. Cogn Affect Behav Neurosci. 2007; 7(1):1–17. [PubMed: 17598730]

Noesselt T, Bergmann D, Heinze HJ, Münte T, Spence C. Coding of multisensory temporal patterns in human superior temporal sulcus. Front Integr Neurosci. 2012; 6:64. [PubMed: 22973202]

Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J. Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. J Neurosci. 2007; 27(42):11431–41. [PubMed: 17942738]

Novick JM, Trueswell JC, Thompson-Schill SL. Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. Cogn Affect Behav Neurosci. 2005; 5(3):263–81. [PubMed: 16396089]

Novick JM, Trueswell JC, Thompson-Schill SL. Broca's area and language processing: evidence for the cognitive control connection. Lang Linguist Compass. 2010; 4(10):906–924.

Ojanen V, Möttönen R, Pekkola J, Jääskeläinen IP, Joensuu R, Autti T, Sams M. Processing of audiovisual speech in Broca's area. Neuroimage. 2005; 25(2):333–8. [PubMed: 15784412]

Okada K, Venezia JH, Matchin W, Saberi K, Hickok G. An fMRI study of audiovisual speech perception reveals multisensory interactions in auditory cortex. PLoS One. 2013; 8(6):e68959. [PubMed: 23805332]

Pasalar S, Ro T, Beauchamp MS. TMS of posterior parietal cortex disrupts visual tactile multisensory integration. Eur J Neurosci. 2010; 31(10):1783–90. [PubMed: 20584182]

Peelen MV, Atkinson AP, Vuilleumier P. Supramodal representations of perceived emotions in the human brain. J Neurosci. 2010; 30(30):10127–34. [PubMed: 20668196]

Pekkola J, Laasonen M, Ojanen V, Autti T, Jääskeläinen 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(3):797–807. [PubMed: 16359873]

Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, Tarkiainen A, Sams M. Primary auditory cortex activation by visual speech: an fMRI study at 3 T. Neuroreport. 2005; 16(2):125–128. [PubMed: 15671860]

Pelphrey KA, Mitchell TV, McKeown MJ, Goldstein J, Allison T, McCarthy G. Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. J Neurosci. 2003; 23(17):6819–25. [PubMed: 12890776]

Pelphrey KA, Morris JP, McCarthy G. Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. J Cogn Neurosci. 2004; 16(10):1706–16. [PubMed: 15701223]

Powers AR 3rd, Hevey MA, Wallace MT. Neural correlates of multisensory perceptual learning. J Neurosci. 2012; 32(18):6263–74. [PubMed: 22553032]

Raij T, Uutela K, Hari R. Audiovisual integration of letters in the human brain. Neuron. 2000; 28(2): 617–25. [PubMed: 11144369]

Rauschecker JP. Compensatory plasticity and sensory substitution in the cerebral cortex. Trends Neurosci. 1995; 18(1):36–43. [PubMed: 7535489]

Rauschecker JP. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hear Res. 2011; 271(1–2):16–25. [PubMed: 20850511]

Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci. 2009; 12(6):718–24. [PubMed: 19471271]

Rauschecker JP, Tian B. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci U S A. 2000; 97(22):11800–6. [PubMed: 11050212]

Redcay E. The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. Neurosci Biobehav Rev. 2008; 32(1):123–42. [PubMed: 17706781]

Renier LA, Anurova I, De Volder AG, Carlson S, VanMeter J, Rauschecker JP. Preserved functional specialization for spatial processing in the middle occipital gyrus of the early blind. Neuron. 2010; 68(1):138–48. [PubMed: 20920797]

Rockland KS, Pandya DN. Cortical connections of the occipital lobe in the rhesus monkey: interconnections between areas 17, 18, 19 and the superior temporal sulcus. Brain Res. 1981; 212(2):249–70. [PubMed: 7225868]

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(1–3):173–83. [PubMed: 17928202]

Sams M, Aulanko R, Hämäläinen M, Hari R, Lounasmaa OV, Lu S-T, Simola J. Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. Neurosci Lett. 1991; 127(1):141–145. [PubMed: 1881611]

Sekiyama K, Kanno I, Miura S, Sugita Y. Auditory-visual speech perception examined by fMRI and PET. Neurosci Res. 2003; 47(3):277–87. [PubMed: 14568109]

Skipper JI, Nusbaum HC, Small SL. Listening to talking faces: motor cortical activation during speech perception. Neuroimage. 2005; 25(1):76–89. [PubMed: 15734345]

Skipper JI, van Wassenhove V, Nusbaum HC, Small SL. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. Cereb Cortex. 2007; 17(10):2387–99. [PubMed: 17218482]

Smith EG, Bennetto L. Audiovisual speech integration and lipreading in autism. J Child Psychol Psychiatry. 2007; 48(8):813–21. [PubMed: 17683453]

Stevenson RA, Altieri NA, Kim S, Pisoni DB, James TW. Neural processing of asynchronous audiovisual speech perception. Neuroimage. 2010; 49(4):3308–18. [PubMed: 20004723]

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(3):1210–23. [PubMed: 18973818]

Stevenson RA, VanDerKlok RM, Pisoni DB, James TW. Discrete neural substrates underlie complementary audiovisual speech integration processes. Neuroimage. 2011; 55(3):1339–45. [PubMed: 21195198]

Striem-Amit E, Amedi A. Visual cortex extrastriate body-selective area activation in congenitally blind people "seeing" by using sounds. Curr Biol. 2014; 24(6):687–92. [PubMed: 24613309]

Striem-Amit E, Cohen L, Dehaene S, Amedi A. Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. Neuron. 2012; 76(3):640–52. [PubMed: 23141074]

Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. J Acoust Soc Am. 1954; 26(2):212.

Szycik GR, Jansma H, Münte TF. Audiovisual integration during speech comprehension: an fMRI study comparing ROI-based and whole brain analyses. Hum Brain Mapp. 2009a; 30(7):1990–9. [PubMed: 18711707]

Szycik GR, Münte TF, Dillo W, Mohammadi B, Samii A, Emrich HM, Dietrich DE. Audiovisual integration of speech is disturbed in schizophrenia: an fMRI study. Schizophr Res. 2009b; 110(1–3):111–8. [PubMed: 19303257]

Tan LH, Chan AH, Kay P, Khong PL, Yip LK, Luke KK. Language affects patterns of brain activation associated with perceptual decision. Proc Natl Acad Sci U S A. 2008; 105(10):4004–9. [PubMed: 18316728]

Turkeltaub PE, Coslett HB. Localization of sublexical speech perception components. Brain Lang. 2010; 114(1):1–15. [PubMed: 20413149]

Turkeltaub PE, Eden GF, Jones KM, Zeffiro TA. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. Neuroimage. 2002; 16(3 Pt 1):765–80. [PubMed: 12169260]

Turkeltaub PE, Eickhoff SB, Laird AR, Fox M, Wiener M, Fox P. Minimizing within-experiment and within-group effects in Activation Likelihood Estimation meta-analyses. Hum Brain Mapp. 2012; 33(1):1–13. [PubMed: 21305667]

Turkeltaub PE, Messing S, Norise C, Hamilton RH. Are networks for residual language function and recovery consistent across aphasic patients? Neurology. 2011; 76(20):1726–34. [PubMed: 21576689]

van Atteveldt N, Formisano E, Goebel R, Blomert L. Integration of letters and speech sounds in the human brain. Neuron. 2004; 43(2):271–82. [PubMed: 15260962]

van Atteveldt N, Roebroeck A, Goebel R. Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. Hear Res. 2009; 258(1–2):152–64. [PubMed: 19500658]

van Atteveldt NM, Blau VC, Blomert L, Goebel R. fMR-adaptation indicates selectivity to audiovisual content congruency in distributed clusters in human superior temporal cortex. BMC Neurosci. 2010; 11:11. [PubMed: 20122260]

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(4):962–74. [PubMed: 16751298]

von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL. Interaction of face and voice areas during speaker recognition. J Cogn Neurosci. 2005; 17(3):367–76. [PubMed: 15813998]

Watson R, Latinus M, Charest I, Crabbe F, Belin P. People-selectivity, audiovisual integration and heteromodality in the superior temporal sulcus. Cortex. 2014; 50:125–36. [PubMed: 23988132]

Weeks R, Horwitz B, Aziz-Sultan A, Tian B, Wessinger CM, Cohen LG, Hallett M, Rauschecker JP. A positron emission tomographic study of auditory localization in the congenitally blind. J Neurosci. 2000; 20(7):2664–72. [PubMed: 10729347]

Woynaroski TG, Kwakye LD, Foss-Feig JH, Stevenson RA, Stone WL, Wallace MT. Multisensory speech perception in children with autism spectrum disorders. J Autism Dev Disord. 2013; 43(12):2891–902. [PubMed: 23624833]

Xu J, Gannon PJ, Emmorey K, Smith JF, Braun AR. Symbolic gestures and spoken language are processed by a common neural system. Proc Natl Acad Sci U S A. 2009; 106(49):20664–9. [PubMed: 19923436]

Zevin JD, Yang J, Skipper JI, McCandliss BD. Domain general change detection accounts for "dishabituation" effects in temporal-parietal regions in functional magnetic resonance imaging studies of speech perception. J Neurosci. 2010; 30(3):1110–7. [PubMed: 20089919]

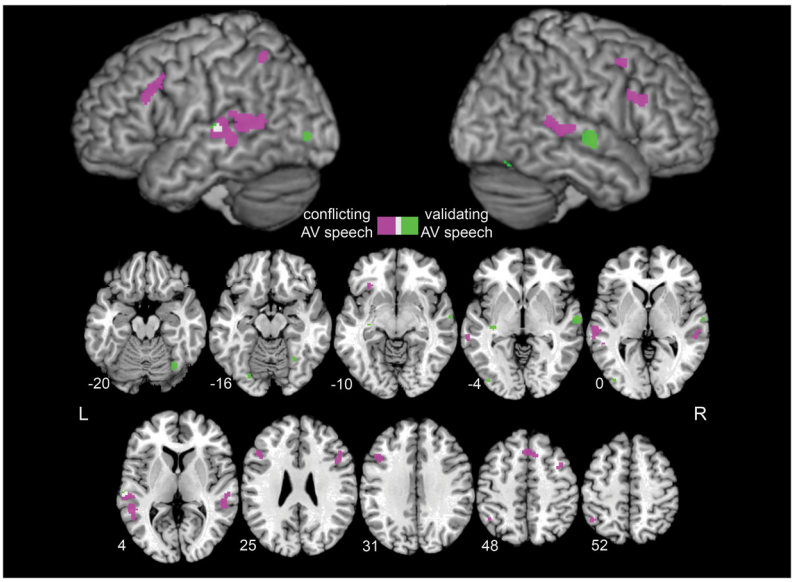

Figure 1. Significant ALE clusters for conflicting and validating AV speech

Conflicting AV speech recruited primarily dorsal stream regions, such as bilateral posterior STG/STS/MTG, bilateral dorsal IFG, left IPL, and SMA (see Table II), shown in purple. In contrast, validating AV speech generally localized to ventral-stream visual areas in the occipital and inferior temporal cortex, such as bilateral FFG and left inferior occipital lobe, as well as other regions, such as bilateral mid-STG, shown in green (see Table III). Overlap between conflicting and validating AV speech is shown in white. One left STG validating cluster was not present if percept contrasts were excluded (Supplementary Figure 1). Results are displayed on Colin27 brain with surface rendering of the left and right hemispheres, significant at FDR $q < .01$ with a cluster size > 100 mm^3. Axial slices are presented in neurological convention with the corresponding MNI Z coordinate.

Table I

Studies, contrasts, foci, and categorized comparisons included in the ALE analysis

"a" and "b" designate separate contrast types from the same study and distinct group of subjects. Note that the references for the studies included in the ALE analysis are provided as a supplementary section.

Study #	Reference	N	Subjects' Language	AV Stimulus	Task	Conflicting AV Speech			Validating AV Speech		
						Contrast	# of foci	Source	Contrast	# of foci	Source
1	(Balk et al., 2010)	14	Finnish	Vowels	Target Detection	Async> Sync	**1**	Author Email	Sync >Async	**6**	Author Email
2	(Benoit et al., 2010)	15	English	McGurk Syllables	Congruency Discrimination	Incong> Cong	**40**	Table II	None	**0**	None
3	(Bishop and Miller, 2009)	25	English	VCV + babble	Speech Identification	None	**0**	None	Sync >Async	**22**	Table 1
4	(Fairhall and Macaluso, 2009)	12	Italian	Story	Selective Attention Target Detection	None	**0**	None	Cong >Incong	**6**	Table 1
5	(Jones and Callan, 2003)	12	English	McGurk VCV	Consonant Discrimination	Incong> Cong	**3**	Results text	Non-McG>McG	**1**	Results text
6	(Lee and Noppeney, 2011)	37	German	Short sentences	**a.** Passive viewing/listening (fMRI); **b.** Synchrony Judgments	**a.** Async> Sync **b.** Non-Fus>Fus	**22**	**a.** Table S1 **b.** Table 2	None	**0**	None
7*	(Macaluso et al., 2004)	8	English	Nouns	Target Detection	None	**0**	None	Sync >Async	**8**	Table 1
8	(Miller and D'Esposito, 2005)	11	English	VCV	Synchrony Judgments	**a.** Async> Sync **b.** Non-Fus>Fus	**15**	Table 1	Fus> Non-Fus	**2**	Table 1
9	(Murase et al., 2008)	28	Japanese	Vowels	Vowel Discrimination	Incong> Cong	**3**	Figure 4 (caption)	None	**0**	None
10**	(Nath et al., 2011)	17	English	McGurk Syllables	Target Detection	McG> Non-McG	**3**	Table 2	Non-McG>McG	**7**	Table 2
11	(Noesselt et al., 2012)	11	German	Sentences	Synchrony Judgments	**a.** Async> Sync **b.** Non-Fus>Fus	**42**	**a.** Table 2 **b.** Table 1, 3	**a.** Sync >Async **b.** Fus> Non-Fus	**12**	**a.** Table 2 **b.** Table 1, 3
12	(Ojanen et al., 2005)	10	Finnish	Vowels	Stimulus Change Detection	Incong> Cong	**4**	Table 1	None	**0**	None
13	(Olson et al., 2002)	10	English	McGurk Words	Passive viewing/listening Button press end of block	McG> Non-McG	**2**	Table 1	None	**0**	None
14***	(Pekkola et al., 2006)	10	Finnish	Vowels	Stimulus Change Detection	Incong> Cong	**2**	Table 3	None	**0**	None
15	(Skipper et al., 2007)	13	English	McGurk Syllables	Passive viewing/listening	Incong> Cong	**30**	Table 3	Cong >Incong	**57**	Table 4
16	(Stevenson et al., 2010)	8	English	Monosyllabic Nouns	Semantic Categorization	None	**0**	None	Sync >Async	**8**	Table 2
17	(Szycik et al., 2008)	12	German	Disyllabic Words	Target Detection	None	**0**	None	Cong >Incong	**1**	Table 2
18	(Szycik et al., 2009)	8	German	Disyllabic Nouns	Target Detection	Incong> Cong	**9**	Table 1	None	**0**	None

Study #	Reference	N	Subjects' Language	AV Stimulus	Task	Conflicting AV Speech			Validating AV Speech		
						Contrast	# of foci	Source	Contrast	# of foci	Source
19****	(Szycik et al., 2012)	7	German	McGurk Syllables	Syllable Discrimination	**a.** Incong> Cong **b.** McG> Non-McG	**23**	**a.** Table 2 **b.** Table 2, 3	None	**0**	None
20	(van Atteveldt et al., 2007) Study 2	13	Dutch	Phonemes#	Congruency Discrimination	Incong> Cong	**7**	Table 4	None	**0**	None
21	(van Atteveldt et al., 2010)	16	Dutch	Phonemes#	Target Detection	Incong> Cong	**4**	Table 1	None	**0**	None
22	(Wiersinga-Post et al., 2010)	14	Dutch	McGurk VCV	Syllable Discrimination	None	**0**	None	Non-McG>McG	**7**	Table 1
	Total:	311				20	210		13	137	

Symbols and abbreviations include:

* PET study;

** subjects were children;

*** only foci from controls were included;

**** while two of the included foci were from contrasts with n=12, n=7 was used for all foci for simplicity;

phoneme speech sounds were paired with visual text of letters (**only two studies #20, #21);** Async = asynchronous; Cong = congruent; Fus = fusion percept; Incong = incongruent; McG = McGurk percept; Non-McG = non-McGurk percept; Non-Fus = non-fusion percept; Sync = synchronous; VCV= vowel-consonant-vowel token.

Table II

Conflicting AV speech ALE analysis results

MNI coordinates are reported with an FDR q < .01 and a cluster size > 100 mm^3. Contributing experiments are reported within three standard deviations of the calculated localization uncertainty from a peak (see Materials and Methods section for details). **We determined whether any foci from validating AV speech experiments were within three standard deviations from the calculated localization uncertainty of each conflicting AV speech ALE peak ("Nearby Validating Experiments").**

Brain Region	Volume (mm^3)	ALE Value	MNI			Conflicting Contributing Experiments	Contrast Type				AV Stimulus Type			Nearby Validating Experiments
							Stimulus Contrast		Percept Contrast					
			x	y	z		Async> Sync	Incong> Cong	Non-Fus> Fus	McG> Non-McG	sub-lexical	word	sentence	
1) Left STG/STS/MTG	2008	0.0256	–54	–44	8	6a, 9, 10, 11a, 11b, 18, 19b	2	2	1	2	3	1	3	11b
		0.0176	–52	–54	8	6a, 9, 10, 11a, 11b, 18	2	2	1	1	2	1	3	11b, 15
		0.0113	–56	–34	14	11a, 11b, 19a, 19b, 21	1	2	1	1	3		2	None
2) Left STG/STS/MTG	1424	0.0153	–62	–26	2	1, 2, 6a, 9, 11a, 18, 19a, 19b	3	4		1	5	1	2	4, 11b
		0.0144	–64	–34	–4	2, 6a, 11b, 18, 19a, 19b	1	3	1	1	3	1	2	4, 11b
3) Left IFG/MFG	1104	0.0169	–44	16	28	2, 11a, 11b, 12, 14, 20	1	4	1		4		2	None
		0.0141	–42	10	36	2, 11b, 14, 15, 20		4	1		4		1	None
		0.0140	–46	22	24	11a, 11b, 12, 14, 18, 20	1	4	1		3	1	2	None
4) Right STG/STS/MTG	960	0.0160	54	–40	6	2, 6a, 8b, 9, 11a, 19a	2	3	1		4		2	11a, 11b, 16
		0.0142	58	–28	2	2, 6a, 9, 19a, 19b	1	3		1	4		1	11b
5) Right IFG	656	0.0143	48	20	22	11a, 11b, 18, 19a, 20	1	3	1		2	1	2	None
		0.0130	46	12	24	2, 8a, 11a, 18, 19a, 20	2	4			4	1	1	None
6) SMA	632	0.0155	0	24	48	2, 8a, 14, 18, 20	1	4			4	1		None
		0.0139	10	18	46	2, 8a, 14, 20	1	3			4			None
7) Right Precentral Gyrus	240	0.0145	40	8	46	2, 11a, 11b, 20	1	2	1		2		2	None
8) Left IPL	192	0.0135	–44	–56	52	5, 8a, 15, 18	1	3			3	1		None
9) Left IFG	160	0.0130	–30	26	–10	2, 11b, 18		2	1		1	1	1	None

Note that study numbers and abbreviations match those designated in Table I. Additional abbreviations include: IFG = inferior frontal gyrus; IPL = inferior parietal lobule; MFG = middle frontal gyrus; MTG = middle temporal gyrus; SMA = supplementary motor area; STG = superior temporal gyrus; and STS = superior temporal sulcus.

Table III

Validating AV speech ALE analysis results. We determined whether any foci from conflicting AV speech experiments were within three standard deviations of the calculated localization uncertainty for each validating AV speech ALE peak ("Nearby Conflicting Experiments"). All other details match those listed in the Table II legend.

Brain Region	Volume (mm^3)	ALE Value	MNI			Validating Contributing Experiments	Contrast Type				AV Stimulus Type			Nearby Conflicting Experiments
							Stimulus Contrast		Percept Contrast					
			x	y	z		Sync > Async	Cong > Incong	Fus> Non-Fus	Non-McG> McG	sub-lexical	word	sentence	
1) Right STG	656	0.0176	64	–14	–4	4, 11b, 17		2	1			1	2	11a, 21
2) Right FFG	344	0.0117	34	–68	–20	7, 10, 16, 22	2			2	2	2		None
		0.0107	28	–60	–18	1, 7, 10, 16	3			1	2	2		None
3) Left STG/Inferior White Matter	192	0.0124	–34	–22	–6	3, 4, 10	1	1		1	2		1	13
4) Left STG	168	0.0127	–62	–26	4	4, 11b		1	1				2	1, 2, 6a, 9, 11a, 18, 19a, 19b
5) Left FFG	152	0.0111	–24	–80	–16	1, 10, 15, 22	1	1		2	4			None
6) Left Inferior Occipital Lobe	128	0.0113	–40	–84	–2	3, 7, 15	2	1			2	1		2

Additional abbreviations include: FFG = fusiform gyrus.