HHS Public Access

Author manuscript

Cortex. Author manuscript; available in PMC 2019 June 01.

Published in final edited form as:

Cortex. 2018 June ; 103: 360–371. doi:10.1016/j.cortex.2018.03.030.

Neural Networks Supporting Audiovisual Integration for Speech: A Large-Scale Lesion Study

Gregory Hickok[1,*], **Corianne Rogalsky**[2,*], **William Matchin**[3], **Alexandra Basilakos**[4], **Julia Cai**[2], **Sara Pillay**[5], **Michelle Ferrill**[6], **Soren Mickelsen**[2], **Steven W. Anderson**[7], **Tracy Love**[6], **Jeffrey Binder**[5], and **Julius Fridriksson**[4]

[1]University of California, Irvine

[2]Arizona State University

[3]University of California, San Diego

[4]University of South Carolina

[5]Medical College of Wisconsin

[6]San Diego State University

[7]University of Iowa

Abstract

Auditory and visual speech information are often strongly integrated resulting in perceptual enhancements for audiovisual (AV) speech over audio alone and sometimes yielding compelling illusory fusion percepts when AV cues are mismatched, the McGurk-MacDonald effect. Previous research has identified three candidate regions thought to be critical for AV speech integration: the posterior superior temporal sulcus (STS), early auditory cortex, and the posterior inferior frontal gyrus. We assess the causal involvement of these regions (and others) in the first large-scale (N=100) lesion-based study of AV speech integration. Two primary findings emerged. First, behavioral performance and lesion maps for AV enhancement and illusory fusion measures indicate that classic metrics of AV speech integration are not necessarily measuring the same process. Second, lesions involving superior temporal auditory, lateral occipital visual, and multisensory zones in the STS are the most disruptive to AV speech integration. Further, when AV speech integration fails, the nature of the failure—auditory vs. visual capture—can be predicted from the location of the lesions. These findings show that AV speech processing is supported by unimodal auditory and visual cortices as well as multimodal regions such as the STS at their boundary. Motor related frontal regions do not appear to play a role in AV speech integration.

Adding a visual speech signal to an auditory speech signal can modify both the intelligibility and the nature of the perceived speech. Studies on the effect of congruent audiovisual (AV)

---

[*]co-first authors

compared to auditory speech alone (A) indicate a substantial perceptibility advantage for AV over A speech alone (Erber NP, 1969; Ross LA et al., 2007; Sumby WH and I Pollack, 1954). Research on the effect of incongruent (mismatched) audio and visual signals have revealed that visual information can even change the category of the perceived speech sound in some cases, the McGurk-MacDonald effect (Jiang J and LE Bernstein, 2011; McGurk H and J MacDonald, 1976).

The neural basis of AV speech integration has been investigated using functional brain imaging (Bernstein LE and E Liebenthal, 2014; Calvert GA and R Campbell, 2003; Matchin W et al., 2014; Nath AR and MS Beauchamp, 2011; Skipper JI et al., 2007; Venezia JH et al., 2017) and transcranial magnetic stimulation (TMS) (Beauchamp MS et al., 2010; Watkins KE et al., 2003). Three hypotheses have emerged from this work. One is that AV speech integration is supported by multisensory areas in the posterior superior temporal sulcus region (pSTS) (Beauchamp MS, KE Lee, et al., 2004; Nath AR and MS Beauchamp, 2012; Venezia JH *et al.*, 2017). Evidence for this claim comes from functional imaging studies showing that this region responds to auditory, visual, and AV speech signals often with response profiles that suggest integration rather than simple summing (e.g., supra-additive, AV>A+V, responses) (James TW and RA Stevenson, 2012), that its activation correlates with measures of fusion and with individual differences in McGurk-MacDonald susceptibility (Miller LM and M D'Esposito, 2005; Nath AR and MS Beauchamp, 2012), and that TMS to the STS can modulate AV speech integration (Beauchamp MS *et al.*, 2010).

The second hypothesis is that AV speech integration is supported by motor speech-related areas, Broca's area in particular, reflecting a mapping from visual cues to motor speech gestures to phoneme-level perception (Sams M et al., 2005; Skipper JI *et al.*, 2007; Watkins KE *et al.*, 2003). Evidence for this claim comes from a range of functional imaging studies showing motor speech areas active during V or AV speech perception and modulation of activity during repetition suppression induced by veridical and illusory phonemes (Skipper JI et al., 2005; Skipper JI *et al.*, 2007) as well as from TMS work showing that face-related motor evoked potentials are enhanced when perceiving visual speech (Watkins KE *et al.*, 2003) and from behavioral evidence for a McGurk-MacDonald-like effect induced by self-articulation of speech that is incongruent with a perceived auditory signal (Sams M *et al.*, 2005); but see (Matchin W *et al.*, 2014) for contradictory evidence. For reviews of this literature see (Bernstein LE and E Liebenthal, 2014; Venezia JH et al., 2015).

A third hypothesis is that visual speech signals directly modulate acoustic analysis in early stages of cortical auditory processing, including primary auditory cortex (Arnal LH et al., 2009; Okada K et al., 2013; Schroeder CE et al., 2008; van Wassenhove V et al., 2005). Evidence for this view includes the observation that early (onset ~100 msec) auditory speech responses are modulated by visual speech (Arnal LH *et al.*, 2009; van Wassenhove V *et al.*, 2005), that primary auditory cortex activity is greater during AV speech perception than auditory speech perception alone (Okada K *et al.*, 2013), and more generally that non-auditory sensory input interacts with acoustic processing in early cortical stages in non-human primates (Ghazanfar AA, 2012; Ghazanfar AA et al., 2008; Ghazanfar AA and CE Schroeder, 2006).

Here we present the first large-scale lesion study aimed at identifying the regions causally related to AV speech integration. As part of a larger stroke and speech-language program, 100 stroke patients were recruited across three sites and their ability to perceive speech, consonant-vowel (CV) syllables, under three conditions was evaluated: auditory speech alone, congruent AV speech, and incongruent AV speech. These conditions allowed us to assess the neural correlates of the two behavioral measures that have been used to demonstrate AV speech integration: (1) the *AV speech advantage*, the improvement in perception afforded by AV speech over A (AV_{con} - A); and (2) *McGurk-MacDonald effect,* the rate at which incongruent AV signals change the percept of A alone (AV_{incon} - A). These variables were then used as dependent measures in voxel-based lesion symptom (VLSM) mapping (Bates E et al., 2003) to identify the neural structures that support AV speech integration. To preview we find support for the view that the posterior superior temporal lobe is the primary locus of AV speech integration consistent with both the STS and direct auditory modulation models but inconsistent with the Broca's area model. We also find that the AV speech advantage and McGurk-MacDonald effect measures are largely uncorrelated.

Methods

100 participants with a single event left hemisphere stroke were administered audiovisual (AV) or audio-only (A) speech perception tasks. The audiovisual stimuli were of two types, matching or mismatching (McGurk). MRI scans for 95 cases were obtained and brain lesions were mapped to a common space for voxel-based lesion symptom mapping. See supplemental material for additional details. Informed consent was obtained from each patient prior to participation in the study, and all procedures were in compliance with the Code of Ethics of the World Medical Association and approved by the Institutional Review Boards of UC Irvine, University of South Carolina, San Diego State University, Medical College Wisconsin, University of Iowa and Arizona State University.

Participants

The total sample size of this research was 100 and participants (28 women) were enrolled by laboratories at six institutions: University of South Carolina (n=55); University of Iowa (n=16); Medical College of Wisconsin (N=11); San Diego State University (n=10); Arizona State University (n=7); and University of California, Irvine (n=1). Five participants did not have neuroimaging data. Therefore, the behavioral data analyses included all 100 participants whereas the neuroimaging data analyses included 95 participants. The participants were recruited as part of large, ongoing aphasia studies in each laboratory that included participants completing extensive neuropsychological and aphasia test batteries. As a result, the present study includes mostly patients with perisylvian lesions due to middle cerebral artery stroke. All participants had ischemic strokes, with the exception of two patients with hemorrhagic strokes. The average age at testing was 61 (SD=10.7) and the average age at stroke was 55.95 (SD=11.42). All participants were right-handed (pre-stroke), native English speakers who had incurred a single event left hemisphere stroke affecting the left hemisphere and were at least 6 months post-stroke at the time of testing. Exclusion criteria included: significant anatomical abnormalities other than the signature lesion of their vascular event, signs of multiple strokes, prior history of psychological or neurological

disease, self-reported significant visual deficits, and inability to follow task instructions. The greatest lesion overlap among the 95 participants who underwent MRI scanning was in the anterior longitudinal fasciculus (MNI coordinates: −37,−5,22) where 50 participants had damage (Figure S1; top panel).

Stimuli

We recorded videos (with simultaneous audio) of a native, male speaker of English (age 24), producing the following single syllables in isolation: /pa/ and /ka/. We collected several repetitions of each syllable, finally choosing a single (1) token for each stimulus for use in the experiment. Separately, we recorded high quality auditory productions of these same syllables from the same speaker within an anechoic chamber, designed to match the acoustic properties of the sounds produced during the video recording.

The incongruent AV stimulus was created as follows. First, we used iMovie to load the raw video of /ka/ which included simultaneously recorded audio. Then we imported separately recorded high-quality audio of the speaker producing /pa/. We then aligned the onset of the consonant burst for /pa/ with the embedded audio for natural /ka/ (Figure S2). Given that the VOT time for the two sounds were similar (/pa/: 50ms, /ka/: 55 ms), there was minimal discrepancy between the natural /ka/ and the /pa/ that replaced it. With informal testing, this produced a robust McGurk illusion. To create the congruent AV /pa/ and /ka/ videos, we performed the same procedure but with congruent auditory and visual stimuli. One token of each stimulus was used in the experiment.

The audio recordings were digitized at 44,100 Hz and RMS normalized. We added Gaussian noise of low amplitude windowed by an onset/offset duration of 20 ms to ensure the robustness of the McGurk illusion (10% RMS of speech). Movies were presented at 30 FPS. Video resolution: 768 × 576 pixels. Total stimulus duration: 3000 ms. Syllable onset occurred at approximately 1835 ms after start of video. Syllable duration: approximately 330 ms. Preparatory movement before syllable onset, /ka/ videos: approximately 633 ms, /pa/ videos: approximately 500 ms.

Behavioral testing

Participants completed two speech perception tasks as part of a larger test battery: (1) an AV speech perception task (i.e. a McGurk-MacDonald task; (McGurk and MacDonald, 1976)) and (2) an auditory-only speech perception task. Individual tests within the battery were presented in a non-fixed pseudorandom order. Given the illusory nature of the McGurk stimulus, the audiovisual task was always presented first followed by the audio-only task to reduce the likelihood that participants realized that no auditory /ta/ existed in the audiovisual task. Stimuli were delivered through a computer via PowerPoint (Microsoft Office) software or Matlab (Mathworks, Inc.). The computer screen was placed at a distance comfortable to the participant. To familiarize participants with each task, one sample trial was presented prior to each task. In both tasks, trials were presented in a fixed random order to all participants using PowerPoint, with the exception of 11 participants tested at MCW who were presented the trials in a randomized order unique to each participant using Matlab. The clinical settings and hardware of some of the testing sites were not conducive to running

Matlab; thus a PowerPoint version was implemented to maximize our ability to collect data from a large group of participants.

Audiovisual Speech Perception (AV).—This task is based on the classic McGurk-MacDonald paradigm (McGurk and MacDonald, 1976). The AV task consisted of 30 trials in which participants were asked to indicate which of three auditory stimuli, /pa/ /ka/ or /ta/, was presented. The 30 trials consisted of 20 congruent trials (in ten trials the auditory and visual stimuli both were /pa/, and in ten trials both were /ka/), and 10 incongruent trials (auditory stimulus was /pa/, visual stimulus was /ka/). These /pa/-/ka/ incongruent trials reliably generate a perception of /ta/ in control subjects.

Each trial consisted of the words “Get Ready” presented for 1000ms in the middle of the computer screen, followed by an “X” appearing for 1200ms, followed by the AV stimulus. After each stimulus, three printed response options were displayed horizontally across the computer screen, “Pa Ta Ka”, with the serial positions of the three options presented in a fixed random order across trials for each participant. Participants were asked to point to the letters corresponding to the sound that they heard. Responses were self-paced. A mouse click began the next trial. Participants were instructed to pay close attention to both the face in the video and the auditory stimulus. The total duration of each video recording was 3000ms and the auditory stimulus (and visual speech movements) began at the midpoint of the video. Each of the 20 congruent AV stimuli were generated by simultaneously presenting the same consonant’s auditory and visual stimulus (either /pa/ or /ka/), whereas the incongruent AV stimuli were generated by simultaneously presenting the auditory stimulus /pa/ and the visual stimulus /ka/.

Auditory-Only Speech Perception.—The auditory-only task was designed in the same manner as the AV task, but no visual speech stimuli were presented. The auditory-only task consisted of 30 trials. Participants were given the same instructions as in the AV task, i.e. to indicate which sound they heard. As in the AV task, three response options (“Pa, Ta, Ka”) were presented on the screen and the serial position of the response options was in a fixed random order for each participant. In 20 trials the auditory stimulus /pa/ was presented, and in 10 trials /ka/ was presented.

Scoring

A participant’s first response was recorded, and a stimulus was presented only once per trial. The following performance measures were computed:

1. *A_only /pa/*: proportion correct for auditory only trials using the /pa/ stimulus.
2. *A_only /ka/*: proportion correct for auditory only trials using the /ka/ stimulus.
3. *AV /pa/:* proportion correct for the AV /pa/ congruent trials, i.e. proportion of AV trials in which both the auditory and visual stimuli were /pa/ and the participant responded /pa/.
4. *AV /ka/:* proportion correct for the AV /ka/ congruent trials, i.e. proportion of AV trials in which both the auditory and visual stimuli were /ka/ and the participant responded /ka/.

5. *AV_{con}-A /pa/*: difference score comparing proportion correct for AV /pa/ (measure #3 above) minus A only /pa/ (measure #1 above). This is our AV advantage measure.

6. *AV_{incon}-A* /pa/: difference score comparing proportion of /pa/ responses for AV incongruent stimulus (auditory /pa/, visual /ka/) minus A_only /pa/ (measure #1 above). This is our global measure of the effect of conflicting AV cues on auditory perception, which does not distinguish type of effect, fusion versus visual capture.

7. *AV_{fusion} rate*: proportion of AV incongruent trials (auditory /pa/, visual /ka/) in which the participant responded /ta/.

8. *Visual capture rate:* proportion of AV incongruent trials (auditory /pa/, visual /ka/ trials) in which the participant responded /ka/, which was the visual stimulus.

9. *Auditory capture rate:* proportion of AV incongruent trials (auditory /pa/, visual /ka/ trials) in which the participant responded /pa/, which was the auditory stimulus.

Neuroimaging

All participants underwent MRI scanning using a 3T or 1.5T MRI system. Lesion mapping proceeded along one of two paths, depending on testing site. For participants tested at the University of South Carolina (n=55), T1-MRI and T2-MRIs with 1 mm^3 resolution were collected, and used in the following way: the chronic stroke lesion was demarcated on T2-weighted images staff experienced in doing lesion studies. The T2 image was co-registered to the T1 image, and these parameters were used to reslice the lesion into the native T1 space. The resliced lesion maps were smoothed with a 3mm full-width half maximum Gaussian kernel to remove jagged edges associated with manual drawing. Enantiomorphic normalization (Nachev et al., 2008) used SPM12 and ‘in house’ Matlab scripts: A mirrored image of the T1 image (reflected around the midline) was coregistered to the native T1 image. Then, we created a chimeric image based on the native T1 image with the lesioned tissue replaced by tissue from the mirrored image (using the smoothed lesion map to modulate this blending, feathering the lesion edge). SPM12's unified segmentation-normalization (Ashburner and Friston, 2005) was used to warp this chimeric image to standard space, with the resulting spatial transform applied to the actual T1 image as well as the lesion map. The normalized lesion map was then binarized, using a 50% probability threshold.

Participants tested at the remaining sites (n=40) underwent MRI scanning including a T1-MRI with $1mm^3$ resolution. Lesion mapping was performed using MAP-3 lesion analysis methods (Damasio and Damasio, 2003) implemented on Brainvox software (Frank et al., 1997). This method is described in detail elsewhere (Damasio, 2000). Briefly, the method entails the transfer of the brain areas of chronic lesion as seen in the patient's T1-MRI into the common space of a template brain. To do so, the template brain is resliced such that each slice maximally corresponds to each slice in the lesion's native space, based on anatomical markers (e.g. sulci and subcortical structures). The lesion is then manually demarcated on

the template brain's corresponding slice respecting the identifiable landmarks. Each lesion map was completed by an individual with extensive training in this technique, and supervised by an expert neuroanatomist (HD). The resulting lesion maps are binary (with no probability threshold). The template brain and lesion maps were then transformed into Talairach space by Brainvox and resampled to a voxel size of $1mm^3$. The MAP-3 method has been shown to have high inter- and intra-rater reliability and in some cases higher accuracy than some automated methods (Fiez et al., 2000; Pantazis et al., 2010). Prior to being combined with the lesion maps from the University of South Carolina, the remaining sites' maps were transformed into MNI space using the NiiStat toolbox for Matlab (http://www.nitrc.org/projects/niistat).[1]

Voxel Based Lesion-Symptom Mapping

Univariate and multivariate VLSM analyses were completed to identify localized brain damage associated with aspects of audio/visual speech perception. Univariate analyses relied on general linear modeling whereas multivariate analyses used Freedman-Lane testing in which the permutation threshold is computed for each behavior using each and every other behavior as a nuisance regressor. Univariate analyses were conducted for the following variables: A-only /pa/ performance, AV fusion rate, and AV advantage (i.e. AV_{con} – A_only /pa/). Two multivariate analyses were calculated: A_only /pa/ with AV fusion rate, and visual capture rate with auditory capture rate.

Permutation thresholding included 4,000 permutations to correct for multiple comparisons ($p < 0.05$ controlled for familywise error). As voxels that are infrequently damaged will have low statistical power while increasing the number of comparisons, only voxels where at least ten participants had damage were included in the analyses. A map that shows areas where at least 10 participants had damage is included in Figure S1, bottom panel. All of the VLSM routines used here are integrated into the NiiStat toolbox for Matlab (http://www.nitrc.org/projects/niistat).

Results

Behavioral results

Table S1 presents descriptive statistics for the A and AV conditions. Overall performance was quite good and skewed toward ceiling scores (see Figure S3 for distribution plots). Proportion correct for the /pa/ and /ka/ stimuli in the A condition did not differ ($t(99) = 0.32$, $p = 0.75$; BF_{10} 0.16[2]). Adding a congruent visual signal, the AV_{con} conditions, significantly improved perception of /pa/ ($t(99) = 5.94$, $p < .001$; $BF_{10} = 278992.059$) but not /ka/ ($t(99) = 0.19$, $p = 0.85$; $BF_{10} = 0.11$); this is consistent with the salience of the visual cues for these

[1]Collecting our large dataset required the collaboration of several aphasia laboratories spanning multiple years. An inherent side effect of this arrangement is that different imaging and mapping techniques have been employed. Any inter-lab lesion mapping differences would only add noise to our dataset and reduce our power. Nonetheless, the two lesion mapping approaches yield very similar lesion maps. As a check, we compared 12 lesions mapped with both techniques; visual inspection found no qualitative differences; Volumes varied by an average of 15% ± 9%. The differences in the maps are likely due to inter-personal mapping differences, as both techniques require trained personnel to determine what tissue is labelled as "lesion; previous studies have found similar inter-personal mapping differences even within the same mapping technique (Fiez et al. 2000).

[2]For behavioral analyses we report both frequentist and Bayesian statistics implemented in JASP (https://jasp-stats.org/).

syllables, /pa/ more salient than /ka/. For this reason, subsequent analyses reported below were carried out on the /pa/ stimuli.

Presenting incongruent auditory and visual cues in the AV_{incon} condition—auditory /pa/ paired with visual /ka/--strongly affected the mean perceptual response in our sample. This is evident in the sign reversal in the mean and median /pa/ responses for the AV_{incon}-A difference scores in Table S1. Figure 1 shows this effect graphically with congruent AV resulting in an approximately 10% increase in /pa/ responses (correctly reporting the auditory signal) compared to A /pa/ alone, and incongruent AV resulting in an approximately 69% decrease in /pa/ responses compared to A alone. A statistical contrast comparing the rate of /pa/ responses in AV_{con} versus AV_{incon} yielded very high significance values ($t(99) = 27.68$, $p < 0.001$; BF = 8.19e+44).

The above analysis shows that incongruent AV signals are having a major impact on perceptual responses in our sample at the group level. For the incongruent stimuli, there are two ways this might occur. One is via audiovisual fusion in which the auditory /pa/ and visual /ka/ combine perceptually to yield /ta/, the classic McGurk-MacDonald fusion effect. The other is visual capture in which the visual signal overrides the auditory signal completely. The mean fusion rate in the AV_{incon} condition was 0.73 (median = 0.80) whereas the mean visual capture rate was 0.13 (median = 0). This indicates that at the population level, incongruent AV signals are predominantly integrated perceptually (Figure S4 displays the population distribution for AV fusion).

A priori one would expect that both the AV advantage effect (AV_{con}-A) and AV fusion effect (fusion rate for AV_{incon}) should provide a measure of AV integration, which predicts that the two variables will be correlated. Recent behavioral work has questioned this assumption, however (Van Engen KJ et al., 2017). In our sample the two variables were only weakly correlated in our full sample and in an unexpected negative direction (Pearson's $r = -0.22$, $p < 0.026$; $BF_{10} = 1.44$). The negative correlation indicates that a greater AV advantage predicts a lower fusion rate. However, an examination of the scatterplot of this relation (Figure S5) revealed two outliers with a *negative* AV advantage (substantially worse performance on AV_{con} compared to A). These two cases were therefore removed and statistics recomputed. This yielded a stronger and still negative correlation ($r = -0.409$, $p < .001$; $BF_{10} = 682.0$). Closer examination of the data showed that 53% of our sample were performing at ceiling (≥95%) in the A condition and that these participants were fusing at a rate of 88% compared to 57% in the non-ceiling cases. These cases could be driving the negative correlation via an anchoring effect with approximately half of the sample showing no AV advantage (because they can't as they are already at ceiling on A alone) and high fusion rates. To eliminate this possibility, we excluded all the ceiling cases and re-calculated the statistics. This completely abolished the correlation ($N=44$, $r = -0.032$, $p = 0.84$; $BF_{10} = 0.19$; Figure 2). We conclude from these analyses that contrary to initial expectation and consistent with recent work (Van Engen KJ *et al.*, 2017), the two typical measures of AV integration—the AV advantage and McGurk-MacDonald fusion—are not measuring the same AV processes.

To explore what might be driving variation on these two measures we conducted some exploratory analyses. First, we used multiple regression to assess two potential predictors of McGurk-MacDonald fusion. All participants were included in this analysis except the two negative AV advantage outliers mentioned above. We evaluated A and AV_{con} using forward stepwise regression and found that only the variable A yielded significant predictive power ($r = 0.54$, $p < 0.001$). A Bayesian linear regression analysis comparing four models (Null, A, AV, & A+AV) produced the same result: the probability of model A given the data was higher than the probability of the other three models given the data ($P(A|data) = 0.837$, $BF_A = 15.42$; cf., next best model, $P(A+AV|data) = 0.163$). This suggests that McGurk-MacDonald fusion is best predicted by relatively intact auditory perception abilities.

To visualize this effect, we plotted AV_{incon} fusion rate against A (Figure 3), which also includes threshold cutoffs for chance performance on the two variables for reference. Thresholds were calculated using a binomial probability (33% chance of guessing the correct A or fusion response, 67% chance of a miss) set at approximately 1%. Specifically, the probability of scoring 60% correct or better on the A task by chance is 0.009 and the probability of scoring 70% or better fusion responses on the AV_{incon} task is 0.015 (the difference being a function of the number of trials, 20 vs. 10, respectively). We also report the number of cases that fall in each quadrant. The figure makes several points clear beyond the general linear trend. The vast majority of fusers are also good auditory perceivers (61 compared with 3, 95.3%) and the majority of poor auditory perceivers are not good fusers (14 compared to 3, 82.4%). Similarly, the majority of good auditory perceivers are fusers (61 compared to 19, 76.3%) and only 8 individuals who scored 90% or better on the A task failed to fuse reliably (identified by the box in the graph). Given the variability of McGurk-MacDonald in the healthy population (Mallick DB et al., 2015), the existence of such cases in our sample is no surprise.

Second, we explored the basis of the AV advantage effect by using AV_{incon} non-fusion response rates. The logic here is that non-fusion response rates provides a measure of the weighting of visual versus auditory cues in individuals who are not fusing (recall we already know that fusion rate does not predict AV advantage). Using forward stepwise regression we found that visual capture rate predicted AV advantage ($r = 0.34$, $p < 0.001$) better than auditory capture rate, but that adding auditory capture rate significantly improved prediction ($r = 0.42$, $p = 0.013$). Bayesian linear regression corroborated this result: visual capture alone was a strong predictor of AV advantage on its own whereas auditory capture rate alone did so only marginally; however, the model combining these factors was the preferred model ($P(V+A\_capture)|data = 0.79$, $BF_{V+A\_capture} = 11.00$). These analyses suggest that individuals who show the greatest AV advantage are those who weight the visual speech cues more strongly in their perceptual judgments, although this is a relatively weak effect accounting for only about 12% of the variance.

Lesion mapping results

Lesion maps from the univariate analysis of the A-only condition identified a large fronto-parietal-superior temporal network that when damaged produced poorer performance (Table S2, Figure 4). The A-only /pa/ trials (and not the A-only /ka/ trials) were included in this

analysis to facilitate comparison with the AV fusion rate VLSM results (described below), which also included just an auditory /pa/ stimulus. In the temporal lobe, regions included several areas known to be important for speech recognition (Hickok G and D Poeppel, 2007; Hillis AE et al., 2017; Price CJ, 2012; Rauschecker JP and SK Scott, 2009): core and belt auditory areas in the supratemporal plane with extension ventrally in the superior temporal sulcus. The strongest effect was at the temporal-parietal boundary, including area Spt (Hickok G et al., 2003; Hickok G et al., 2009). Frontal regions included the inferior frontal and precentral gyri. The frontoparietal involvement was expected due to the non-specific task demands which involve translating an acoustic signal into a format suitable for visual (written syllable) matching, working memory, and response selection, all of which have been shown to implicate frontoparietal regions (Buchsbaum BR et al., 2011; Buchsbaum BR and M D'Esposito, 2008; Novick JM et al., 2005; Venezia JH et al., 2012).

Lesion maps resulting from the univariate analysis of AV fusion rate (regions that when damaged produce lower McGurk-MacDonald fusion responses) were more focal involving the superior temporal lobe and temporal-parietal junction with noticeably less involvement of the frontoparietal network implicated in the A-only condition (Table S3, Figure 5). The strongest effect was noted in the superior temporal gyrus corroborating the behavioral analysis showing that A-only perception is the best predictor of AV fusion in the variables we assessed. The lesion map distribution also extended posteriorly into STS (posterior STG and MTG) regions previously implicated in AV integration and biological motion perception in functional imaging and TMS studies (Beauchamp MS *et al.*, 2010; Grossman ED et al., 2005). Co-varying out the contribution of A-only perception to the AV fusion effect using a multivariate VLSM identified a smaller set of voxels located in posterior temporal-occipital cortex and underlying white matter (Table S3; Figure 6, red; A-only perception with AV fusion co-varied out is shown in green).

VLSMs were also computed for the AV advantage variable (AV_{con}-A), which we had initially thought would co-vary with AV fusion rates. This analysis failed to produce significant effects.

As noted above, fusion failures can arise from different sources. We carried out a multivariate VLSM to separately map the two possible non-fusion response types, auditory capture (reporting the veridical auditory syllable) versus visual capture (reporting the veridical visual syllable). This analysis revealed that lesions involving a wide swath of auditory and related regions in the superior temporal lobe as well as some inferior frontal and parietal regions produced a higher rate of visual capture errors whereas lesions involving a smaller focus in the posterior middle temporal/occipital regions produced a higher rate of auditory capture errors (Table S4; Figure 7).

Discussion

The present experiment is the first large scale lesion study of audiovisual speech integration. We used two primary measures of integration, AV advantage and McGurk-MacDonald fusion, and report two main results. First, the two principle measures used to assess AV integration in the field, including in this study, were found to be either weakly negatively

correlated or uncorrelated depending on whether ceiling performers are included or not, respectively, suggesting that the two tasks as they were employed here are measuring different processes. Second, superior temporal, posterior middle temporal, and occipital regions comprise the network broadly supporting AV integration, presumably including auditory analysis (superior temporal), visual analysis (occipital/posterior middle temporal), and multisensory functions (regions at the boundary of these ~pSTS). The findings are consistent with the hypothesis that AV integration involves multisensory regions in the pSTG (Beauchamp MS, KE Lee, *et al.*, 2004; Bernstein LE and E Liebenthal, 2014; Venezia JH *et al.*, 2017) and is also consistent with the hypothesis that early auditory cortical areas are important for AV integration (Ghazanfar AA, 2012; Okada K *et al.*, 2013; Schroeder CE *et al.*, 2008). We found little evidence for the hypothesis that Broca's area is a critical substrate for AV integration (Skipper JI *et al.*, 2007; Watkins KE *et al.*, 2003), consistent with a recent case report (Andersen TS and R Starrfelt, 2015).

Relation between measures of AV integration

The perceptibility benefit one gets from adding visual to auditory speech, the AV advantage, and the ability of incongruent audio and visual cues to lead to a fused percept that differs from both, McGurk-MacDonald fusion, are nearly always discussed together as different examples of the same AV integration process. Our findings argue otherwise or at least that the two measures are, in practice, detecting different mixes of abilities. In our entire sample, we found the two variables to be weakly negatively correlated such that participants who showed a greater AV advantage tended to exhibit less fusion. This negative relation was found to be driven by an anchoring/ceiling effect: people with ceiling-level auditory-only perception do not show an AV advantage, but can exhibit fusion. When ceiling cases are removed, the negative correlation evaporates leaving no relation between the two measures (Figure 2). Our lesion data provided converging evidence as different lesion-symptom maps were found for the two measures. The fusion rate variable generated robust lesion maps involving the posterior superior temporal lobe whereas the AV advantage variable failed to yield any significantly related voxels. This finding is consistent with a recent study (Van Engen KJ *et al.*, 2017) that examined individual covariation between susceptibility to McGurk-MacDonald fusion and ability to use visual cues to perceive auditory sentences under a range of signal-to-noise ratios. The authors reported no relation between the measures.

Although paradoxical on first blush, the non-relation between the two measures can be understood in the following way. On one hand, the magnitude of the AV advantage measure is yoked to the magnitude of acoustic degradation (via environmental or neural noise); the worse one is at hearing speech, the more room there is for visual cues to rescue audition, the principle of inverse effectiveness (Stein B and MA Meredith, 1993; Stevenson RA et al., 2012). On the other hand, the susceptibility to fusion is yoked (at least in part) to the perceptibility of both the acoustic and visual signals; if either is substantially degraded, fusion is precluded because there is little to fuse. Conceptualized in this way, AV advantage is potentially reflecting either some level of fusion between a mildly degraded acoustic stimulus and an informative visual stimulus or it is just measuring lip reading ability (using visual speech cues to compensate for degraded audio) when the acoustic stimulus is

significantly degraded. Fusion, in contrast, is more likely a reflection of an AV integration process in this experimental setting. Our finding that visual capture rate in the AV_{incon} conditions was a dominant predictor of AV advantage is consistent with this characterization: participants who weighted visual cues heavily or could best take advantage of them benefitted most from congruent AV speech over auditory-alone speech, suggesting that in our study at least, the AV advantage measure primarily tapped lip-reading ability.

Temporal lobe networks supporting AV integration

Lesion maps associated with deficits in AV integration (decreased fusion rates) implicated superior posterior temporal regions with extension into posterior STS/middle temporal regions and occipital regions; a focus in the anterior insula/frontal operculum was also identified (Figure 5). The superior temporal gyrus was the most robustly implicated region, strongly implicating auditory perception in susceptibility to McGurk-MacDonald fusion and providing neural convergence with our behavioral finding that the best predictor of fusion in our measures is auditory-only speech perception. Visual-related cortices were also implicated, as expected. The auditory dominance in our lesion maps for fusion is likely a consequence of our sampling, which involved predominantly people with perisylvian lesions due to MCA infarcts. If we had sampled more cases with posterior temporal-occipital lesions, we may have found even stronger evidence for the role of visual-related regions in fusion susceptibility. Future studies that include patients with posterior cerebral artery strokes are needed to explore how damage to early visual cortex may affect AV speech integration, and the inclusion of anterior cerebral artery strokes would provide more insight into how attention resources, including those supported by the anterior cingulate cortex, may be critical to AV speech integration.

The posterior superior temporal sulcus is arguably the primary candidate region for AV integration based on a range of evidence including its multisensory nature in both human and nonhuman primates (Beauchamp MS, BD Argall, et al., 2004; Dahl CD et al., 2009), its activation pattern in AV speech (Bernstein LE and E Liebenthal, 2014; Venezia JH *et al.*, 2017), and TMS studies implicating pSTS in AV integration (Beauchamp MS *et al.*, 2010). The present result adds to this literature and provides strong additional support for the hypothesized role of pSTS (pSTG/pMTG) in AV speech integration. However, our results are also consistent with a role for auditory regions in the supratemporal plane as playing a role in AV speech integration, perhaps via early modulation of acoustic responses by visual speech signals (Okada K *et al.*, 2013; Schroeder CE *et al.*, 2008; van Wassenhove V *et al.*, 2005).

Also implicated in fusion “deficits” is a slightly more posterior ventral region involving the pMTG. Our multivariate analysis of fusion, which factored out the contribution of A_only performance, also identified a focus in this vicinity (Figure 6) as did our multivariate analysis of auditory capture, which factored out visual capture (Figure 7). This region has been implicated in visual speech processing in functional imaging studies (Bernstein LE et al., 2011; Venezia JH *et al.*, 2017) and has been dubbed the temporal visual speech area (TVSA) (Bernstein LE and E Liebenthal, 2014). Thus, a viable interpretation of our lesion results is that the pMTG focus interfered with the visual component necessary for fusion.

Figure 6: Multivariate analysis of A only and fusion performance. Regions predictive of A_only performance with fusion rate co-varied out are shown in green. Regions predictive of fusion rate with A_only performance co-varied out are shown in red.

Taken together, the present data, constrained by previously published findings, are consistent with the view proposed on the basis of a functional imaging results (Bernstein LE *et al.*, 2011; Venezia JH *et al.*, 2017) that dorsal STG regions in and around Heschl's gyrus are predominantly involved in unimodal auditory processing, that posterior MTG regions are predominantly involved in unimodal visual processing, and posterior STS regions are involved in processing audiovisual speech. However, our findings are also consistent with the hypothesis that AV integration can happen, in part, in early auditory cortex (Arnal LH *et al.*, 2009; Okada K *et al.*, 2013; Schroeder CE *et al.*, 2008; van Wassenhove V *et al.*, 2005). We suggest, following previous authors, that both early auditory and pSTS regions are involved, perhaps serving different functions (Arnal LH *et al.*, 2009). The anterior insula was also implicated. As this area was not predicted to be a part of the audiovisual speech network based on a priori predictions, we report it here and resist the temptation to provide a post hoc interpretation.

Bilateral speech perception and visual speech perception

Overall performance on AV speech perception and McGurk-MacDonald fusion was quite high: AV /pa/ stimuli were correctly perceived on average 90% of the time and incongruent AV stimuli led to fusion responses also at a 90% rate (Table S1). The distributions of responses were highly skewed toward intact performance (Figure S3 & S4). What this suggests is that unilateral damage tends not to produce profound chronic deficits in AV speech processing at the syllable level. This is similar to what is found with speech recognition generally, at least as measured by auditory comprehension, which has led to the hypothesis that speech recognition is neurally supported by a bilateral network in the superior temporal lobe (Hickok G and D Poeppel, 2004, 2007). Following similar logic, the present result suggests a bilateral organization for AV speech perception. This conclusion is consistent with the bilateral organization of speech recognition processes, as noted, as well as the bilateral functional imaging activation patterns found for biological motion perception (Grossman E et al., 2000) and AV speech perception (Bernstein LE *et al.*, 2011; Venezia JH *et al.*, 2017).

Role of left fronto-parieto-temporal cortex in speech perception

Lesion maps associated with A-only speech perception implicated a fronto-parieto-temporal network (Figure 4). Temporal involvement is consistent with classical and current (Hickok G and D Poeppel, 2004, 2007) models which hold that the left superior temporal lobe is the critical substrate for speech perception. But the additional involvement of fronto-parietal regions is consistent with claims that the dorsal stream, sensorimotor system plays a role as well (D'Ausilio A et al., 2009; Pulvermuller F and L Fadiga, 2010). However, it has been argued extensively that dorsal stream involvement in speech perception is driven by task demands associated with discrimination paradigms (working memory, conscious attention to phonemic level information) rather than computational systems needed to recognize speech sounds (Bishop DV et al., 1990; Hickok G, 2009, 2014; Hickok G and D Poeppel, 2007;

Rogalsky C et al., 2011; Venezia JH *et al.*, 2012). The present study used a discrimination paradigm and so involvement of the dorsal speech stream is expected in analyses that do not covary out the effects of task.

Conclusions and Loose Ends

The present report supports neural models of AV integration that hypothesize posterior superior temporal regions as the critical substrate as opposed to those that propose the involvement of Broca's area. We can extend the pSTG model by proposing that the network is to some degree bilaterally organized. We also conclude that not all AV integration tasks are the same. AV advantage tasks may not necessarily tap into AV integration processes.

Does the task dissociation mean we don't do AV integration in normal speech processing? No, it just means that the magnitude of the AV advantage as it is typically studied is not a valid measure of AV integration. Perhaps AV integration operates most robustly under mildly degraded conditions where the signal is largely intelligible except for a few local acoustic ambiguities and thus the potential for improvement in intelligibility is relatively small. More severe auditory signal degradation, where the potential for intelligibility improvement is large, may push the system toward more complete visual reliance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant support: NIH DC009659

References

Andersen TS, Starrfelt R. 2015 Audiovisual integration of speech in a patient with Broca's Aphasia. Front Psychol 6:435. [PubMed: 25972819]

Arnal LH, Morillon B, Kell CA, Giraud AL. 2009 Dual neural routing of visual facilitation in speech processing. The Journal of neuroscience : the official journal of the Society for Neuroscience. 29:13445–13453. [PubMed: 19864557]

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF. 2003 Voxel-based lesion-symptom mapping. Nature neuroscience. 6:448–450. [PubMed: 12704393]

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

Beauchamp MS, Lee KE, Argall BD, Martin A. 2004 Integration of auditory and visual information about objects in superior temporal sulcus. Neuron. 41:809–823. [PubMed: 15003179]

Beauchamp MS, Nath AR, Pasalar S. 2010 fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. The Journal of neuroscience : the official journal of the Society for Neuroscience. 30:2414–2417. [PubMed: 20164324]

Bernstein LE, Jiang J, Pantazis D, Lu ZL, Joshi A. 2011 Visual phonetic processing localized using speech and nonspeech face gestures in video and point-light displays. Hum Brain Mapp 32:1660–1676. [PubMed: 20853377]

Bernstein LE, Liebenthal E. 2014 Neural pathways for visual speech perception. Frontiers in neuroscience. 8:386. [PubMed: 25520611]

Bishop DV, Brown BB, Robson J. 1990 The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals. Journal of speech and hearing research. 33:210–219. [PubMed: 2359262]

Buchsbaum BR, Baldo J, Okada K, Berman KF, Dronkers N, D'Esposito M, Hickok G. 2011 Conduction aphasia, sensory-motor integration, and phonological short-term memory - an aggregate analysis of lesion and fMRI data. Brain and language. 119:119–128. [PubMed: 21256582]

Buchsbaum BR, D'Esposito M. 2008 The search for the phonological store: from loop to convolution. J Cogn Neurosci 20:762–778. [PubMed: 18201133]

Calvert GA, Campbell R. 2003 Reading speech from still and moving faces: The neural substrates of visible speech. Journal of cognitive neuroscience. 15:57–70. [PubMed: 12590843]

D'Ausilio A, Pulvermuller F, Salmas P, Bufalari I, Begliomini C, Fadiga L. 2009 The motor somatotopy of speech perception. Current Biology. 19:381–385. [PubMed: 19217297]

Dahl CD, Logothetis NK, Kayser C. 2009 Spatial organization of multisensory responses in temporal association cortex. J Neurosci 29:11924–11932. [PubMed: 19776278]

Erber NP. 1969 Interaction of audition and vision in the recognition of oral speech stimuli. Journal of speech and hearing research. 12:423–425. [PubMed: 5808871]

Ghazanfar AA. 2012 Unity of the Senses for Primate Vocal Communication In: The Neural Bases of Multisensory Processes (Murray MM, Wallace MT, eds.). Boca Raton (FL).

Ghazanfar AA, Chandrasekaran C, Logothetis NK. 2008 Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. J Neurosci 28:4457–4469. [PubMed: 18434524]

Ghazanfar AA, Schroeder CE. 2006 Is neocortex essentially multisensory? Trends in cognitive sciences. 10:278–285. [PubMed: 16713325]

Grossman E, Donnelly M, Price R, Pickens D, Morgan V, Neighbor G, Blake R. 2000 Brain areas involved in perception of biological motion. Journal of cognitive neuroscience. 12:711–720. [PubMed: 11054914]

Grossman ED, Battelli L, Pascual-Leone A. 2005 Repetitive TMS over posterior STS disrupts perception of biological motion. Vision research. 45:2847–2853. [PubMed: 16039692]

Hickok G 2009 Speech Perception Does Not Rely On Motor Cortex. In.

Hickok G 2014 The myth of mirror neurons: the real neuroscience of communication and cognition. New York, NY: W.W. Norton & Company.

Hickok G, Buchsbaum B, Humphries C, Muftuler T. 2003 Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. Journal of cognitive neuroscience. 15:673–682. [PubMed: 12965041]

Hickok G, Okada K, Serences JT. 2009 Area Spt in the human planum temporale supports sensory-motor integration for speech processing. Journal of neurophysiology. 101:2725–2732. [PubMed: 19225172]

Hickok G, Poeppel D. 2004 Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. Cognition. 92:67–99. [PubMed: 15037127]

Hickok G, Poeppel D. 2007 The cortical organization of speech processing. Nature Reviews Neuroscience. 8:393–402. [PubMed: 17431404]

Hillis AE, Rorden C, Fridriksson J. 2017 Brain Regions Essential for Word Comprehension: Drawing Inferences from Patients. Ann Neurol

James TW, Stevenson RA. 2012 The Use of fMRI to Assess Multisensory Integration In: The Neural Bases of Multisensory Processes (Murray MM, Wallace MT, eds.). Boca Raton (FL).

Jiang J, Bernstein LE. 2011 Psychophysics of the McGurk and other audiovisual speech integration effects. Journal of experimental psychology Human perception and performance. 37:1193–1209. [PubMed: 21574741]

Mallick DB, Magnotti JF, Beauchamp MS. 2015 Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type. Psychonomic bulletin & review. 22:1299–1307. [PubMed: 25802068]

Matchin W, Groulx K, Hickok G. 2014 Audiovisual speech integration does not rely on the motor system: evidence from articulatory suppression, the McGurk effect, and fMRI. J Cogn Neurosci 26:606–620. [PubMed: 24236768]

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

Miller LM, D'Esposito M. 2005 Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. The Journal of neuroscience : the official journal of the Society for Neuroscience. 25:5884–5893. [PubMed: 15976077]

Nath AR, Beauchamp MS. 2011 Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. The Journal of neuroscience : the official journal of the Society for Neuroscience. 31:1704–1714. [PubMed: 21289179]

Nath AR, Beauchamp MS. 2012 A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. Neuroimage. 59:781–787. [PubMed: 21787869]

Novick JM, Trueswell JC, Thompson-Schill SL. 2005 Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. Cognitive, affective & behavioral neuroscience. 5:263–281.

Okada K, Venezia JH, Matchin W, Saberi K, Hickok G. 2013 An fMRI Study of Audiovisual Speech Perception Reveals Multisensory Interactions in Auditory Cortex. PLoS One. 8:e68959. [PubMed: 23805332]

Price CJ. 2012 A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage. 62:816–847. [PubMed: 22584224]

Pulvermuller F, Fadiga L. 2010 Active perception: sensorimotor circuits as a cortical basis for language. Nature reviews Neuroscience. 11:351–360. [PubMed: 20383203]

Rauschecker JP, Scott SK. 2009 Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nature neuroscience. 12:718–724. [PubMed: 19471271]

Rogalsky C, Love T, Driscoll D, Anderson SW, Hickok G. 2011 Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system. Neurocase. 17:178–187. [PubMed: 21207313]

Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ. 2007 Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cereb Cortex. 17:1147–1153. [PubMed: 16785256]

Sams M, Mottonen R, Sihvonen T. 2005 Seeing and hearing others and oneself talk. Brain Research: Cognitive Brain Research. 23:429–435. [PubMed: 15820649]

Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A. 2008 Neuronal oscillations and visual amplification of speech. Trends in cognitive sciences. 12:106–113. [PubMed: 18280772]

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

Skipper JI, van Wassenhove V, Nusbaum HC, Small SL. 2007 Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. Cereb Cortex. 17:2387–2399. [PubMed: 17218482]

Stein B, Meredith MA. 1993 The merging of the senses. Cambridge, MA: MIT Press.

Stevenson RA, Bushmakin M, Kim S, Wallace MT, Puce A, James TW. 2012 Inverse effectiveness and multisensory interactions in visual event-related potentials with audiovisual speech. Brain Topogr 25:308–326. [PubMed: 22367585]

Sumby WH, Pollack I. 1954 Visual contributions to speech intelligibility in noise. Journal of the Acoustical Society of America. 26:212–215.

Van Engen KJ, Xie Z, Chandrasekaran B. 2017 Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect. Atten Percept Psychophys 79:396–403. [PubMed: 27921268]

van Wassenhove V, Grant KW, Poeppel D. 2005 Visual speech speeds up the neural processing of auditory speech. Procedings of the National Academy of Sciences. 102:1181–1186.

Venezia JH, Matchin W, Hickok G. 2015 Multisensory integration and audiovisual speech perception In: Brain mapping: An encyclopedic reference (Toga AW, ed.), pp 565–572: Elsevier.

Venezia JH, Saberi K, Chubb C, Hickok G. 2012 Response Bias Modulates the Speech Motor System during Syllable Discrimination. Front Psychol 3:157. [PubMed: 22723787]

Venezia JH, Vaden KIJ, Rong F, Maddox D, Saberi K, Hickok G. 2017 Auditory, visual and audiovisual speech processing streams in superior temporal sulcus. Front Hum Neurosci

Watkins KE, Strafella AP, Paus T. 2003 Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia. 41:989–994. [PubMed: 12667534]

Significance Statement: The physical environment is perceived through different sensory channels that must be integrated into a single coherent percept. An everyday example of such integration is speech, where auditory and visual ("lip reading") cues are integrated to yield a substantially enhance percept. Using both matching and mismatching audiovisual speech cues (the later resulting in an illusory percept) we studied the neural basis of multisensory integration in 100 stroke patients--the first such study--thus revealing the brain networks responsible. We find surprisingly that not all multisensory tasks tap into the same brain circuits but identify a core network responsible for stitching together our sensory experience.

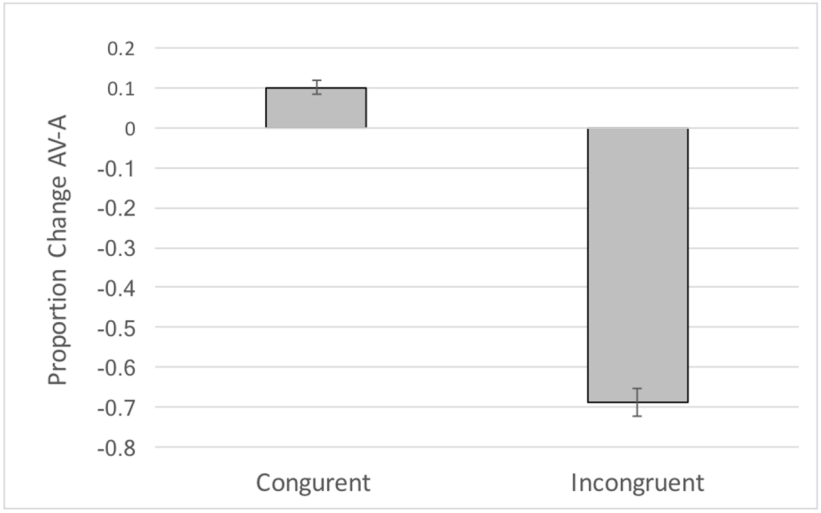

Figure 1:
Proportion change in correctly reporting the auditory /pa/ stimulus in the congruent versus incongruent AV condition compared with the A-only condition.

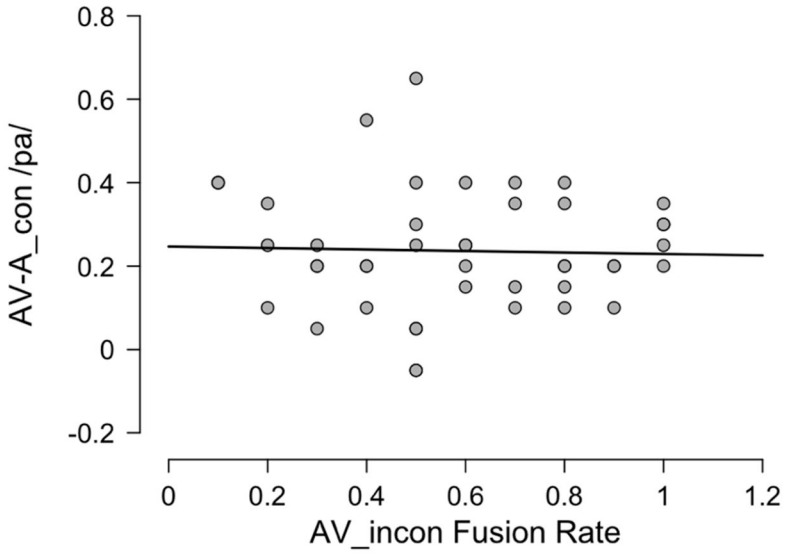

Figure 2:
AV advantage (congruent AV – A) proportion of correct /pa/ responses as a function of fusion rate on incongruent AV stimuli.

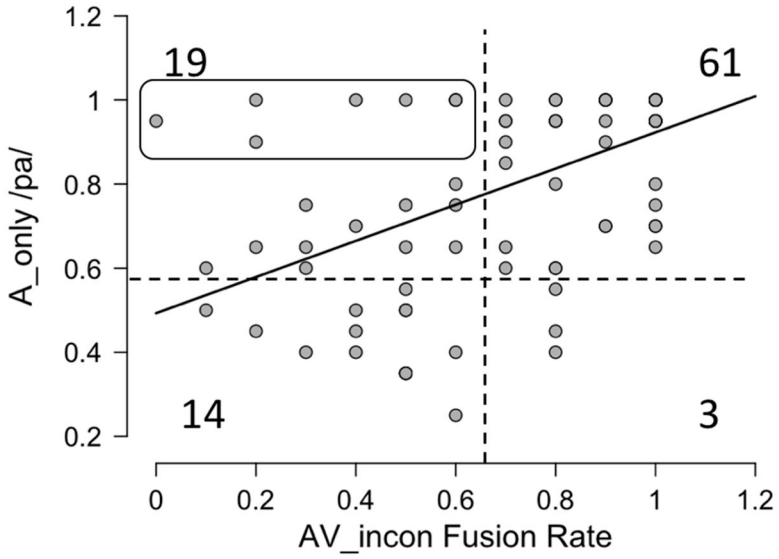

Figure 3:
Relation between A_only performance and fusion rate on the incongruent AV condition. Dotted lines represent chance performance on the two tasks. See text for details.

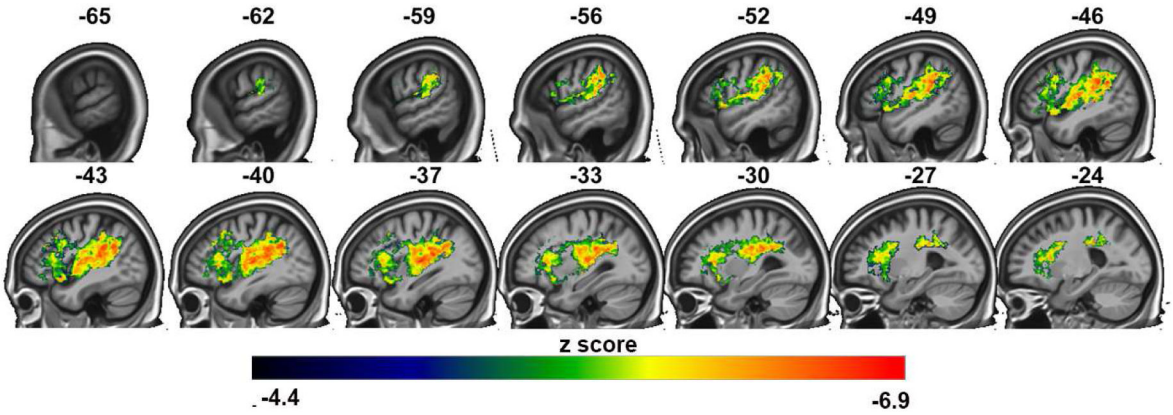

Figure 4:

Representative sagittal slices of the VLSM univariate analysis of performance on the A-only /pa/ condition, threshold of p<.05, family-wise error corrected.

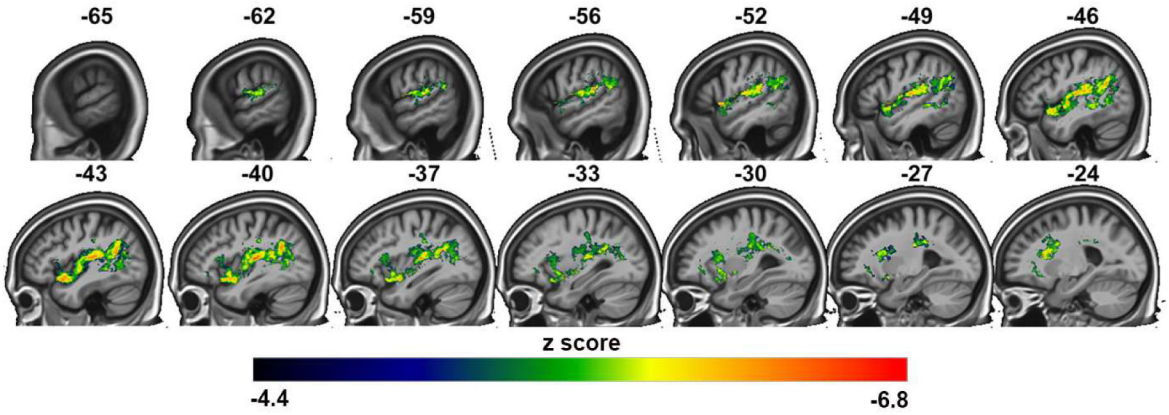

Figure 5:
Representative sagittal slices of the VLSM univariate analysis of fusion rate, threshold of p<.05, family-wise error corrected.

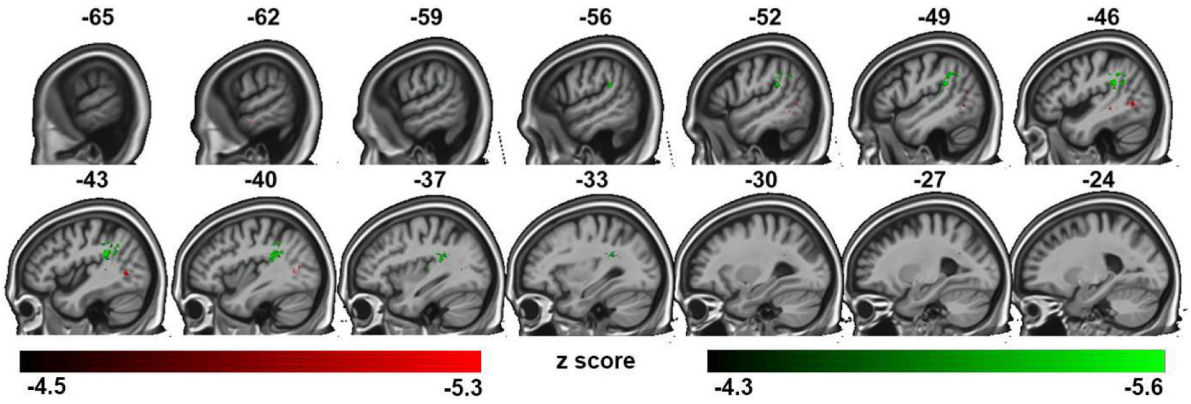

Figure 6:

Representative sagittal slices of the VLSM multivariate analysis of the A_only /pa/ condition and AV fusion performance, threshold of $p < 0.05$, family-wise error corrected. Regions predictive of A_only performance with fusion rate co-varied out are shown in green. Regions predictive of fusion rate with A_only performance co-varied out are shown in red.

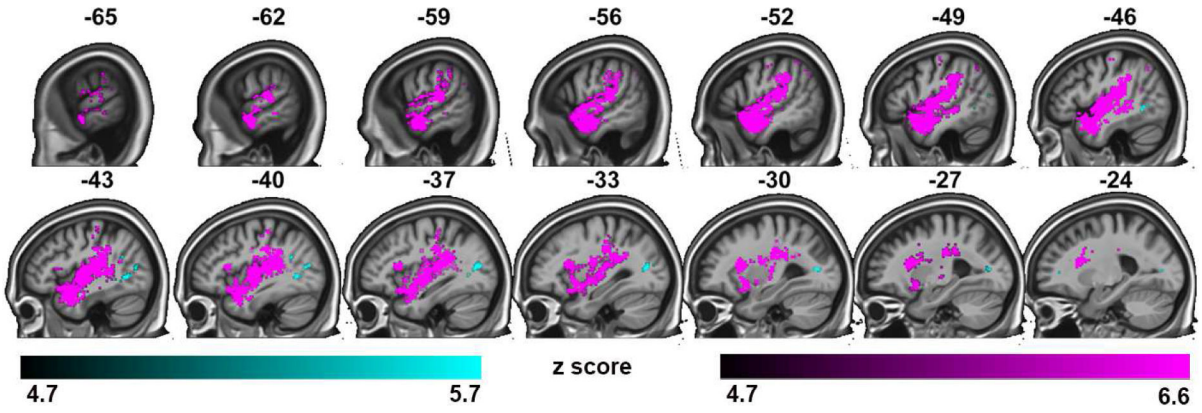

Figure 7:

Representative sagittal slices of the multivariate analysis of visual and auditory capture, threshold of $p < 0.05$, family-wise error corrected. Regions predictive of visual capture rate with auditory capture rate co-varied out are shown in magenta. Regions predictive of auditory capture rate with visual capture rate co-varied out are shown in cyan.