bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Increased connectivity among sensory and motor regions during visual and audiovisual speech perception

Jonathan E. Peelle[1], Brent Spehar[1], Michael S. Jones[1], Sarah McConkey[1], Joel Myerson[2], Sandra Hale[2], Mitchell S. Sommers[2], Nancy Tye-Murray[1]

[1] Department of Otolaryngology, Washington University in St. Louis, St. Louis MO USA

[2] Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis MO USA

Running title: Audiovisual speech perception in noise

Please address correspondence to:

Dr. Jonathan Peelle
Department of Otolaryngology
Washington University in Saint Louis
Saint Louis, MO 63110
jpeelle@wustl.edu

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

In everyday conversation, we usually process the talker's face as well as the sound of their voice. Access to visual speech information is particularly useful when the auditory signal is degraded. Here we used fMRI to monitor brain activity while adults (n = 60) were presented with visual-only, auditory-only, and audiovisual words. As expected, audiovisual speech perception recruited both auditory and visual cortex, with a trend towards increased recruitment of premotor cortex in more difficult conditions (for example, in substantial background noise). We then investigated neural connectivity using psychophysiological interaction (PPI) analysis with seed regions in both primary auditory cortex and primary visual cortex. Connectivity between auditory and visual cortices was stronger in audiovisual conditions than in unimodal conditions, including a wide network of regions in posterior temporal cortex and prefrontal cortex. Taken together, our results suggest a prominent role for cross-region synchronization in understanding both visual-only and audiovisual speech.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Understanding spoken language in the presence of background noise is notoriously challenging, and when visual speech information is available, listeners make use of it—performance on audiovisual (AV) speech in noise is routinely better than that for auditory-only speech in noise (Sumby & Pollack, 1954). Although there is consensus that listeners make use of visual information during speech perception, there is little agreement either on the neural mechanisms that support visual speech processing or on the way in which visual and auditory speech information are combined during audiovisual speech perception.

One longstanding perspective on processing audiovisual speech information has been that auditory and visual information are processed through separate channels, and then integrated at a separate processing stage (Grant & Seitz, 1998; Massaro & Palmer, 1998). Audiovisual integration is thus often considered an individual ability that some people are better at and some people are worse at, regardless of their unimodal processing abilities (Magnotti & Beauchamp, 2015; Mallick et al., 2015).

However, more recent data have brought this traditional view into question. In one study, Tye-Murray and colleagues (2016) showed that unimodal auditory-only and visual-only word recognition scores accurately predicted AV performance, and factor analyses revealed two unimodal ability factors with no evidence of a separate integrative ability factor. These findings suggest that rather than a separate stage of audiovisual integration, AV speech perception may depend most strongly on the coordination of auditory and visual inputs.

Theoretical perspectives on audiovisual integration have also informed cognitive neuroscience approaches to AV speech perception. Prior functional neuroimaging studies of audiovisual speech processing have largely focused on identifying regions supporting integration. One possibility is that the posterior superior temporal sulcus (pSTS) is responsible for combining auditory and visual information during speech perception. The pSTS is anatomically positioned between auditory cortex and visual cortex, and has the functional properties of a multisensory convergence zone (Beauchamp et al., 2004a). During many audiovisual speech processing tasks, the pSTS is differentially activated by matching and mis-matching auditory-visual information, consistent with a role in integration (Stevenson & James, 2009). Moreover, functional connectivity between the pSTS and primary sensory regions varies with the reliability of the information in a modality (Nath & Beauchamp, 2011a), suggesting that the role of the pSTS may be related to combining or weighing information from different senses.

A complementary proposal is that regions of premotor cortex responsible for representing articulatory information are engaged in processing speech (Okada & Hickok, 2009). The contribution of motor regions to speech perception is hotly debated. Evidence consistent with a motor contribution includes a self-advantage in both visual-only and AV speech perception (Tye-Murray et al., 2013, 2015), and effects of visual speech training on speech production (Fridriksson et al., 2009; Venezia et al., 2016). However, premotor activity is not consistently observed in neuroimaging studies of speech perception, and in some instances may also reflect non-perceptual

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

processing (Nuttall et al., 2016; Szenkovits et al., 2012). Although it is also possible that premotor regions are only engaged in certain types of speech perception situations (for example, when there is substantial background noise, or when lipreading). Individual differences in neural strategy related to hearing sensitivity or lipreading ability also may affect the involvement of premotor cortex.

In addition to looking for discrete brain regions that may support visual-only or AV speech perception, in the current fMRI study we broaden our approach to study the role played by functional connectivity between auditory, visual, and motor regions. Our view is informed by behavioral studies suggesting that a separate integration "stage" is not necessary. We hypothesized that in the absence of an integration stage, increased functional connectivity may serve to coordinate multisensory inputs to produce a unified percept. We tested auditory-only speech perception and AV speech perception at a range of signal-to-noise ratios (SNRs) and obtained out-of-scanner measures of lipreading ability from our participants (**Figure 1**). Our goal was to identify patterns of brain activity that were associated with visual-only or audiovisual speech perception.

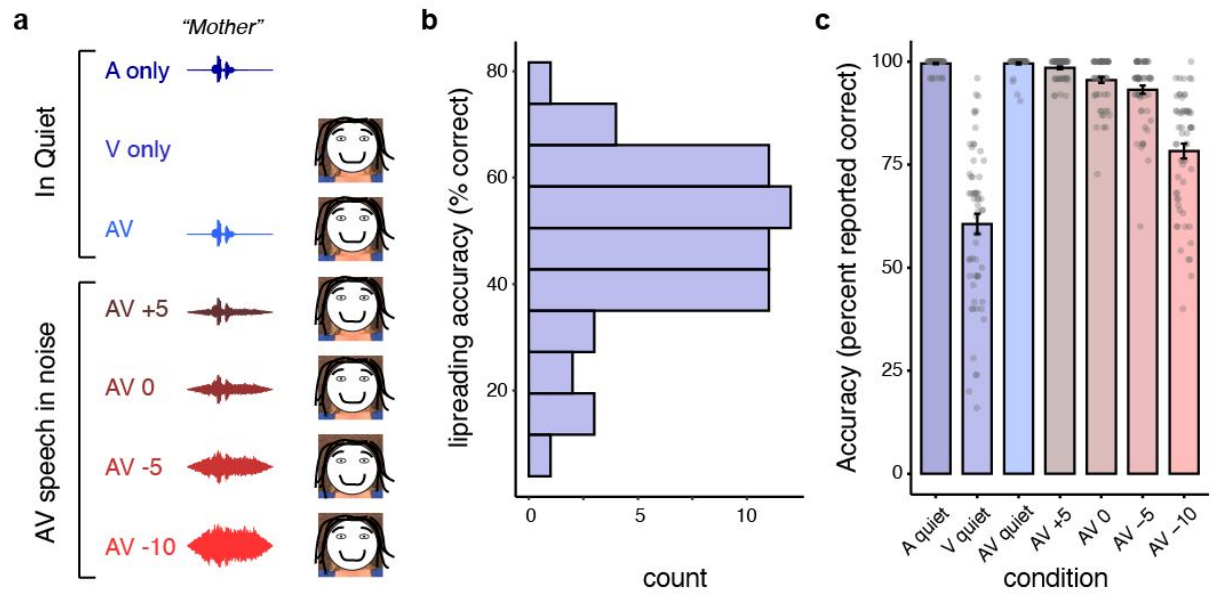

Figure 1. a. Experimental conditions with auditory-only speech, visual-only speech, and audiovisual speech. (Note: Visual stimuli anonymized according to bioRxiv requirements. Actual stimuli were videos of our talker.) **b.** Lipreading ability outside the scanner. **c.** Within-scanner behavioral performance (subjective ratings of understanding); individual participants shown in dots, mean ± SE displayed in bars..

If a dedicated brain region is necessary to combine auditory and visual speech information, we would expect to see activity in this region during audiovisual speech. If changes in task-based synchronized activity—that is, effective connectivity (Friston, 1994; Stephan & Friston, 2010)—underlie visual-only or audiovisual speech processing,

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

we would expect to see greater connectivity between speech-related regions during these conditions relative to auditory-only speech. Given the substantial variability in lipreading ability across people (Feld & Sommers, 2009; Sommers et al., 2005), one might also expect to see individuals who are better lipreaders engage different brain regions than those who are poorer. Finally, if task difficulty modulates the brain regions supporting visual-only and audiovisual speech processing, we might find that the above effects depend on the difficulty of the speech signal (with visual-only speech and audiovisual speech in noise presenting the most challenge).

Method

Stimuli, behavioral data, and analysis scripts are available from https://osf.io/qxcu8/.

Materials

We created seven lists of 50 words. The stimuli were recordings of a female actor speaking single words. The talker sat in front of a neutral background and spoke words along with the carrier phrase "Say the word _____" into the camera. The actor was instructed to allow her mouth to relax to a slightly open and neutral position before each target word was spoken. The edited versions of the recordings used in the experiment did not include a carrier phrase and were each 1.5 seconds long. Recordings were made using a Canon Elura 85 digital video camera and showed the talker's head and shoulders. Digital capture and editing were done using Adobe Premiere Elements. The original capture format for the video was uncompressed .avi; the final versions used in the study were compressed as high quality .wmv files. Audio was leveled to ensure that each word had the same root mean squared (RMS) amplitude using Adobe Audition. Conditions that included background noise used RMS leveled six-talker babble that was mixed and included in the final version of the file.

The 350 recordings used in the study were selected from a corpus of 970 recordings of high frequency words (log HAL frequency 7.01–14.99) from the English Lexicon Project (Balota et al., 2007). The words that were selected for presentation in the lipreading (visual-only) or audiovisual (AV) conditions in varying signal-to-noise ratios (SNR) were selected from the larger corpus based on visual-only behavioral performance on each word from 149 participants (22–90 years old) who were tested using the entire corpus. The words selected ranged from 10%–93% correct in the lipreading-only behavioral tests. They were distributed among the six conditions that included visual information (AV in Quiet, AV +5 SNR, AV 0 SNR, AV -5 SNR, AV -10, and visual-only) so they would, on average, be equivalent for lipreading difficulty. The words used in the auditory-only condition were selected from the remaining words.

Participants

We collected data from 60 participants ranging in age from 18–34 years (M = 22.42, SD = 3.24, 45 female). All were right-handed native speakers of American English (no other languages other than English before age 7) who self-reported normal hearing and an absence of neurological disease.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Procedure

Before being tested in the fMRI scanner all participants were consented, completed a safety screening, and completed an out-of-scanner lipreading assessment. The behavioral lipreading assessment consisted of 50 single word clips selected in the same way and taken from the same corpus of recorded material used in the scanner. The lipreading assessment was complete by presenting each video clip to the participant using a laptop. Participants were encouraged to verbally provide a guess for each clip. Only verbatim responses to the stimuli were considered correct.

Participants were positioned in the scanner with insert earphones and a viewing mirror placed above the eyes to see a two-sided projection screen located at the head-side of the scanner. Those that wore glasses were provided scanner-friendly lenses that fit their prescription. Participants were also given a response box that they held in a comfortable position on their torso during testing. Each of the sequences presented included trials with recordings of audio, visual-only, audiovisual speech stimuli, or printed text via an image projected on the screen that was visible to the participant through the viewing mirror. A camera positioned at the entrance to the scanner bore was used to monitor participant movement. A well-being check and short conversation occurred before each sequence and, if needed, participants were reminded to stay alert and asked to try to reduce their movement during the session.

Six runs were completed during the session. Each run lasted approximately 5.5 minutes. The first five runs contained 98 trials each. The stimuli were presented in blocks of five experimental trials plus two null trials for each condition. The total result was 14 blocks resulting in 70 experimental trials plus 28 null trials. All trials included 800 ms of quiet without a visual presentation before the stimuli began. During the null trials participants were presented with a fixation cross instead of the audiovisual presentation. The auditory-only condition did not include visual stimuli; instead a black screen was presented. The blocks were quasi-randomized so that two blocks from the same condition were never presented one right after the other and one null trial never occurred right after another.

To keep attention high, half of the experimental trials required a response from the participant. On response trials, a set of two dots appeared on the screen after the audiovisual/audio presentation. The right-side dot was green and the left-side dot was red. The participant was instructed to use the right-hand button on the response box to indicate “yes” if they were confident that they had been able to identify the previous word and to use the left-hand button if they felt they had not identified the previous word correctly.

After the initial five runs, a final run of 60 trials was presented in which participants saw a series of words on projected on the screen. The items were the same 50 words used for the behavioral visual-only assessment, and 10 null trials. Each word stayed on the screen for 2.3 seconds, followed by two green dots that appeared for 2.3 seconds. Participants were asked to say aloud the word that was presented during the period that the dots were on the screen. Ten null trials were randomly distributed throughout the sequence. Null trials lasted 1.5 seconds and included a fixation cross on the screen.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

MRI data acquisition and analysis

MRI images were acquired on a Siemens Prisma 3T scanner using a 32 channel head coil. Structural images were acquired using a T1-weighted MPRAGE sequence (details) with a voxel size of 1 x 1 x 1 mm. Functional images were acquired using a multiband sequence (Feinberg et al., 2010) with an acceleration factor of 8. Each volume took 0.770 s to acquire. We used a sparse imaging paradigm (Edmister et al., 1999; Hall et al., 1999) with a repetition time of 2.47 s, leaving 1.7 s of silence on each trial. We presented words during this silent period, and during the repetition task, instructed participants to speak during a silent period to minimize the influence of head motion on the data.

Analysis of the MRI data was performed using Automatic Analysis version 5.4.0 (Cusack et al., 2014) (RRID:SCR_003560) that scripted a combination of SPM12 (Wellcome Trust Centre for Neuroimaging) version 7487 (RRID:SCR_007037) and FSL (FMRIB Analysis Group; Jenkinson et al., 2012) version 6.0.1 (RRID:SCR_002823). Functional images were realigned, co-registered with the structural image, and spatially normalized to MNI space (including resampling to 2 mm voxels) using unified segmentation before smoothing with an 8 mm FWHM Gaussian kernel.

Contrast images from single subject analyses were analyzed at the second level using permutation testing (FSL *randomise;* 5000 permutations) with a cluster-forming threshold of $p < .001$ (uncorrected) and results corrected for multiple comparisons based on cluster extent ($p < .05$). Anatomical localization was performed using converging evidence from author experience (Devlin & Poldrack, 2007) viewing statistical maps overlaid in MRIcroGL (Rorden & Brett, 2000), supplemented by atlas labels (Tzourio-Mazoyer et al., 2002).

Regions of interest

We defined regions of interest (ROIs) for the left posterior temporal sulcus (pSTS), left premotor cortex, left primary auditory cortex (A1), and left primary visual cortex (V1). For the pSTS, the ROI was defined as a 10 mm radius sphere centered at MNI coordinates (x=-54, y=-42, z=4) previously reported to be activated during audiovisual speech processing (Venezia et al., 2017). For the premotor ROI, we used subject-specific activations obtained during the read-aloud task (voxelwise threshold $p < 0.001$, uncorrected) masked by the left primary motor cortex as defined in the SPM Anatomy Toolbox (Eickhoff et al., 2005). The ROI for AI and V1 were defined as the conjugation of Anatomy Toolbox Areas TE1.0, TE 1.1, and TE 1.2 in the left hemisphere (Morosan et al., 2001) and the left half of area hOC1, respectively.

Results

We first examined whole brain univariate effects by condition, shown in **Figure 2**. We observed temporal lobe activity in all conditions, including visual-only, and visual cortex activity in all conditions except auditory only.

We next related the activity during visual-only speech with the out-of-scanner lipreading score (**Figure 1b**). Across participants, lipreading accuracy ranged from

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4–74% (mean = 47.75, SD = 15.49), and correlated with in-scanner ratings (Spearman rho = 0.38). We included out-of-scanner lipreading as a covariate to see whether individual differences in out-of-scanner scores related to visual-only activity; we did not find any significant relationship (positive or negative).

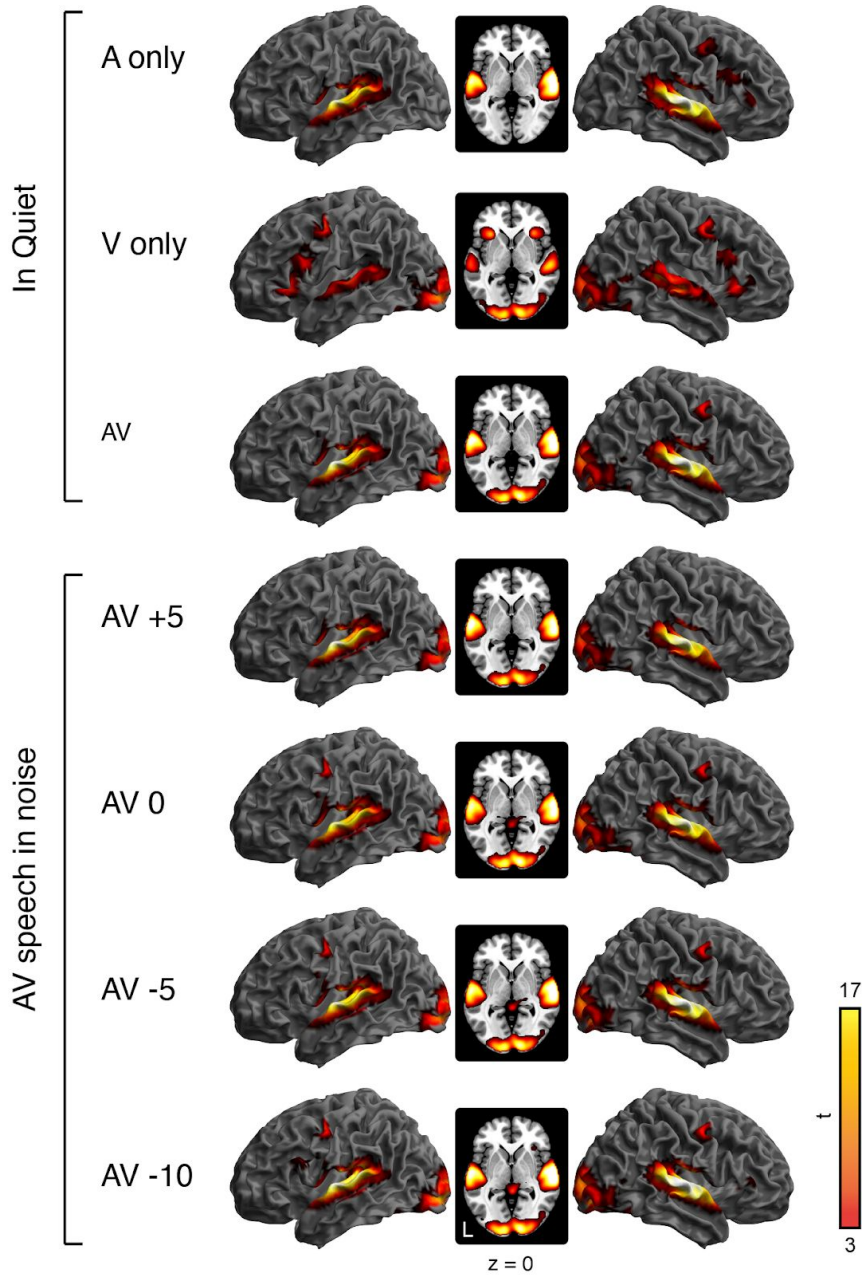

Figure 2. Univariate results for spoken word perception in all experimental conditions.

Following univariate analyses, we examined effective connectivity using psychophysiological interaction (PPI) models. We started by using a seed region in left

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

visual cortex. As seen in **Figure 3**, compared to auditory-only speech, visual-only and all audiovisual conditions showed increased connectivity with the visual cortex seed, notably including bilateral superior temporal gyrus and auditory cortex. The same was true with an auditory cortex seed, shown in **Figure 4**. Here, compared to the visual-only condition, we see increased connectivity with visual cortex in all conditions except the auditory-only condition.

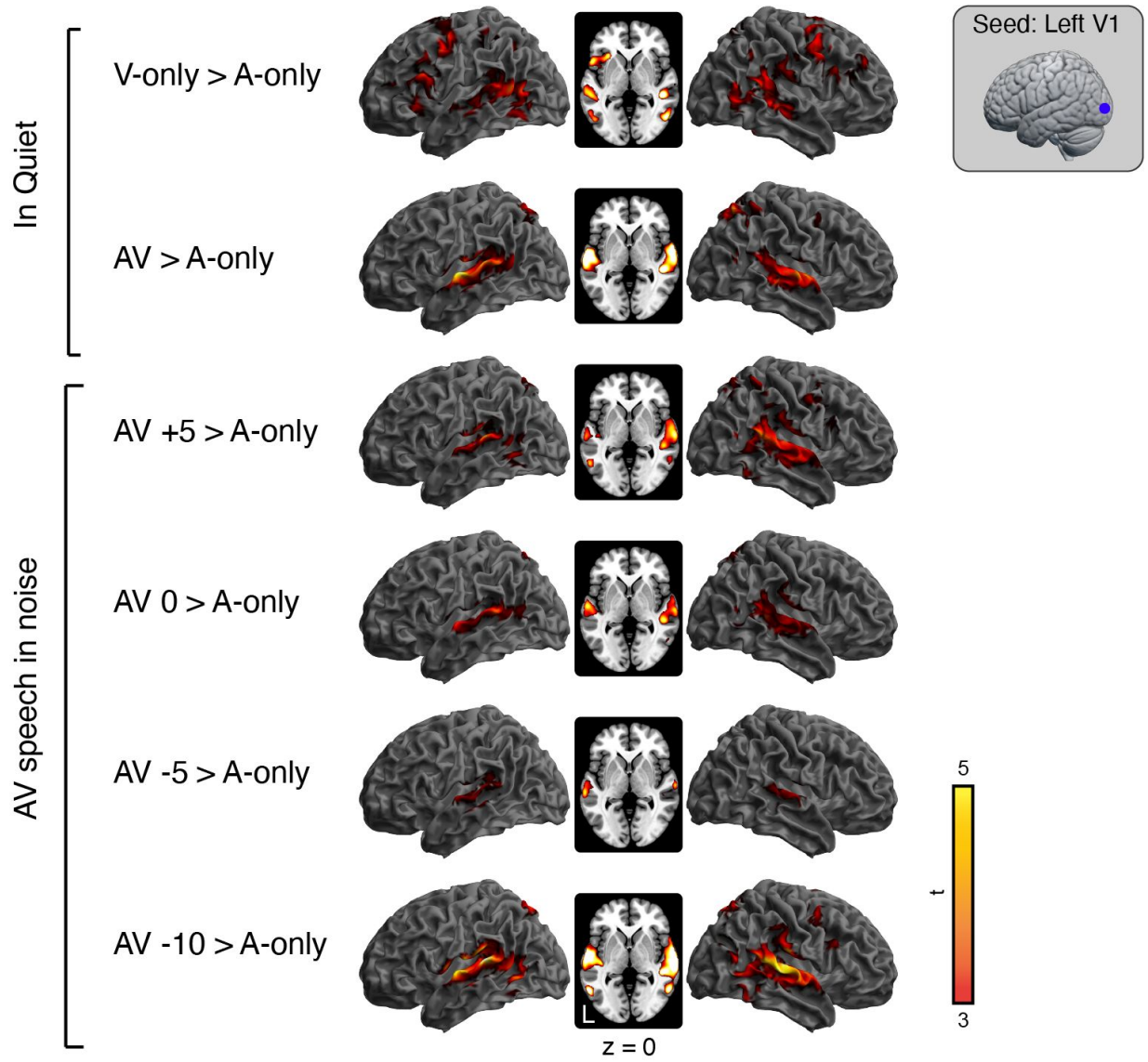

Figure 3. Psychophysiological interaction analysis for experimental conditions, using a seed from left visual cortex. Warm-colored voxels showed significantly more connectivity with visual cortex in an experimental condition than in the auditory-only condition.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

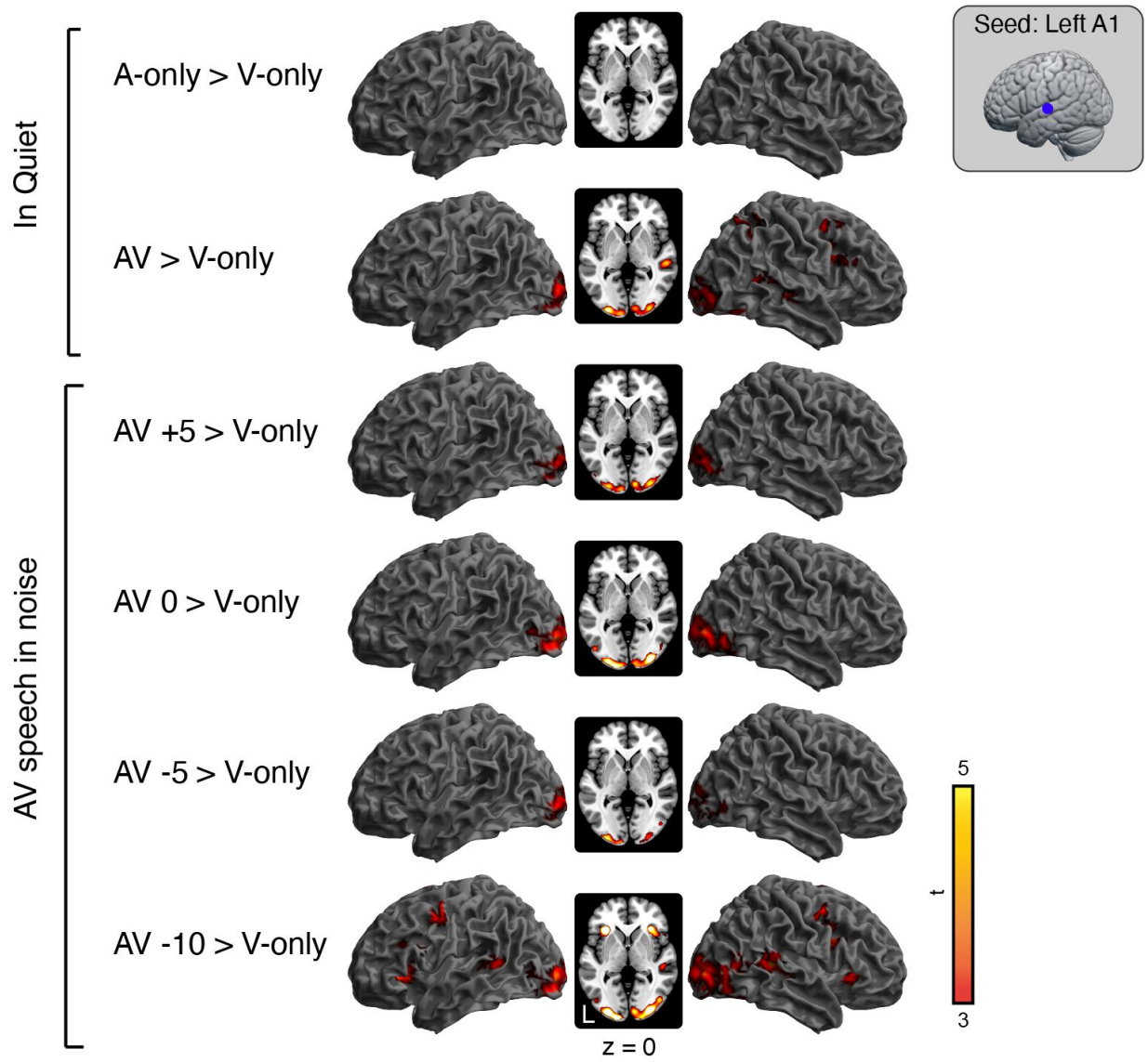

Figure 4. Psychophysiological interaction analysis for experimental conditions, using a seed from left auditory cortex. Warm-colored showed significantly more connectivity with auditory cortex in an experimental condition than in the visual-only condition.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

We studied brain activity and neural connectivity associated with visual-only and audiovisual speech perception. We found that effective connectivity between auditory, visual, and premotor cortex was enhanced during audiovisual speech processing relative to unimodal processing, and during visual-only speech processing (i.e., lipreading) relative to auditory-only speech processing. These findings are broadly consistent with a role for synchronized interregional neural activity supporting visual and audiovisual speech perception. Below we review these specific results in turn.

Dedicated regions for multisensory speech processing

Although audiovisual speech necessarily requires combining information from multiple modalities, the manner in which this happens is unclear. One possibility is that heteromodal brain regions such as the pSTS act to integrate inputs from unisensory cortices (Beauchamp et al., 2004b; Nath & Beauchamp, 2012). In addition to combining inputs to form a unitary percept, regions such as pSTS may also give more weight to more informative modalities (for example, to the visual signal when the auditory signal is noisy) (Nath & Beauchamp, 2011b).

Interestingly, we did not observe increased activity in pSTS for visual-only or AV speech. One possible reason for this may be the nature of the speech materials. Several previous studies identifying pSTS involvement in multisensory speech perception have used incongruent stimuli (i.e., a McGurk task), which differs substantially from most of our everyday speech perception experience (Van Engen et al., 2019). In at least one paper, functional connectivity between pSTS and primary sensory cortices depended on the relative clarity of the modality (Nath & Beauchamp, 2011b), and visual speech information was always clear in our stimuli. Thus, elucidating the specific conditions under which pSTS is recruited to support visual or AV speech perception will require future research.

In our univariate results, we observed activity in premotor cortex for both visual-only speech in quiet, and AV speech at more challenging signal-to-noise ratios. These findings are consistent with a flexible role for premotor cortex in speech perception, at least under some circumstances, as reported in other studies of visual and audiovisual speech perception (Venezia et al., 2017). Although our current data do not support specific conclusions, the dependence of premotor activity on task demands may explain some of the inconsistencies underlying the debates about the role of premotor cortex that permeate the speech perception literature.

Effective connectivity and multisensory speech processing

A different perspective comes from a focus on multisensory effects in auditory and visual cortex (Peelle & Sommers, 2015). Much of the support for this “early integration” view comes from electrophysiology studies showing multimodal effects in primary sensory regions (e.g., Schroeder & Foxe, 2005). For example, Lakatos and colleagues (2007) found that somatosensory input reset the phase of ongoing neural oscillations in auditory cortex, which was hypothesized to increase sensitivity to auditory stimuli. In at

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

least one human MEG study, audiovisual effects appear sooner in auditory cortex than in pSTS (Möttönen et al., 2004), and visual speech appeared to speed processing in auditory cortex (van Wassenhove et al., 2005). These findings suggest that multisensory effects are present in primary sensory regions, and that auditory and visual information do not need a separate region in which to "integrate".

In the current data, we observed stronger connectivity between auditory and visual cortex for visual-only and audiovisual speech conditions than for unimodal auditory-only speech; and stronger connectivity in audiovisual speech conditions than in unimodal visual-only speech. That is, using a visual cortex seed we found differential increases in effective connectivity with auditory cortex, and when using an auditory cortex seed we found differential increases in effective connectivity with visual cortex. These complementary findings indicate that functionally coordinated activity between primary sensory regions that is increased during audiovisual speech perception.

Beyond primary sensory cortices, we also observed effective connectivity changes to premotor cortex for both visual-only speech (**Figure 3**) and the most difficult audiovisual condition (**Figure 4**, -10 dB SNR). The functional synchronization between visual cortex, auditory cortex, and premotor cortex is consistent with a distributed network that orchestrates activity in response to visual-only and audiovisual speech.

What might be the function of such distributed, coordinated activity? Visual and audiovisual speech appear to rely on multisensory representations (and perhaps on motor representations under certain conditions). For audiovisual speech, it may seem obvious that successful perception requires combining auditory and visual information. However, visual-only speech has been consistently associated with activity in auditory cortex (Calvert et al., 1997; Okada et al., 2013). These activations may correspond to visual-auditory associations, and auditory-motor associations, learned from audiovisual speech that are automatically reactivated, even when the auditory input is absent.

Interestingly, our out-of-scanner lipreading scores did not correlate with any of the whole brain results. It should be noted, however, that our sample size, while large for fMRI studies of audiovisual speech processing, may still be too small to reliably detect individual differences in brain activity patterns (Yarkoni & Braver, 2010). Moreover, there may be multiple ways that brains can support better lipreading, and such heterogeneity in brain patterns would not be evident in our current analyses. Future studies with larger sample sizes may be needed to quantitatively assess the degree to which users' activity might fall into neural strategies, and the degree to which these are related to lipreading performance.

Different perspectives on multisensory integration during speech perception

An enduring challenge for understanding multisensory processing during speech perception can be found in differing uses of the word "integration". During audiovisual speech perception, listeners use both auditory and visual information, and so from one perspective both kinds of information are necessarily "integrated" into a listener's (unified) perceptual experience. However, such use of both auditory and visual information does not necessitate a separable cognitive stage for integration (Tye-Murray et al., 2016), nor does it necessitate a region of the brain devoted to integration. The interregional coordination we observed here may accomplish the task of integration in

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

that both auditory and visual modalities are shaping perception. In this framework, there is no need to first translate visual and auditory speech information into some kind of common code (see also Altieri et al., 2011). The connectivity between sensory and motor regions itself appears to be sufficient for combining auditory and visual speech information (and perhaps speech motor information as well).

With any study it is important to consider how the specific stimuli used influenced the results. Here, we examined processing for single words. Visual speech can inform perception in multiple dimensions (Peelle & Sommers, 2015), including by providing clues to the speech envelope (Chandrasekaran et al., 2009). These clues may be more influential in connected speech (e.g., sentences) than in single words, as other neural processes may come into play with connected speech.

Conclusion

Our findings demonstrate the scaffolding of connectivity between auditory, visual, and premotor cortices that supports visual-only and audiovisual speech perception. These findings suggest that the binding of multisensory information need not be restricted to heteromodal brain regions (e.g., pSTS), but may also emerge from coordinated unimodal activity throughout the brain.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgments

This work was supported by grants R56 AG018029 and R01 DC016594 from the US National Institutes of Health. The multiband echo planar imaging sequence was provided by the University of Minnesota Center for Magnetic Resonance Research.

Competing interests

The authors have no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Altieri, N., Pisoni, D. B., & Townsend, J. T. (2011). Some behavioral and neurobiological constraints on theories of audiovisual speech integration: a review and suggestions for new directions. *Seeing and Perceiving*, *24*(6), 513–539.

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004a). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. In *Nature Neuroscience* (Vol. 7, Issue 11, pp. 1190–1192). https://doi.org/10.1038/nn1333

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004b). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience*, *7*(11), 1190–1192.

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

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*, e1000436.

Cusack, R., Vicente-Grabovetsky, A., Mitchell, D. J., Wild, C. J., Auer, T., Linke, A. C., & Peelle, J. E. (2014). Automatic analysis (aa): efficient neuroimaging workflows and parallel processing using Matlab and XML. *Frontiers in Neuroinformatics*, *8*, 90.

Devlin, J. T., & Poldrack, R. A. (2007). In praise of tedious anatomy. *NeuroImage*, *37*(4), 1033–1041.

Edmister, W. B., Talavage, T. M., Ledden, P. J., & Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Human Brain Mapping*, *7*, 89–97.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335.

Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., Miller, K. L., Ugurbil, K., & Yacoub, E. (2010). Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PloS One*, *5*, e15710.

Feld, J. E., & Sommers, M. S. (2009). Lipreading, processing speed, and working memory in younger and older adults. *Journal of Speech, Language, and Hearing Research: JSLHR*, *52*(6), 1555–1565.

Fridriksson, J., Baker, J. M., Whiteside, J., Eoute, D., Jr, Moser, D., Vesselinov, R., & Rorden, C. (2009). Treating visual speech perception to improve speech production in nonfluent aphasia. *Stroke; a Journal of Cerebral Circulation*, *40*(3), 853–858.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, *2*, 56–78.

Grant, K. W., & Seitz, P. F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *The Journal of the Acoustical Society of America*, *104*, 2438–2450.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E. M., & Bowtell, R. W. (1999). “Sparse” temporal sampling in auditory fMRI. *Human Brain Mapping*, *7*(3), 213–223.

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790.

Lakatos, P., Chen, C.-M., O’Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, *53*, 279–292.

Magnotti, J. F., & Beauchamp, M. S. (2015). The noisy encoding of disparity model of the McGurk effect. *Psychonomic Bulletin & Review*, *22*(3), 701–709.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Mallick, D. B., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review*, *22*(5), 1299–1307.

Massaro, D. W., & Palmer, S. E., Jr. (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. MIT Press.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage*, *13*, 684–701.

Möttönen, R., Schürmann, M., & Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. *Neuroscience Letters*, *363*, 112–115.

Nath, A. R., & Beauchamp, M. S. (2011a). 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*(5), 1704–1714.

Nath, A. R., & Beauchamp, M. S. (2011b). 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*(5), 1704–1714.

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage*, *59*(1), 781–787.

Nuttall, H. E., Kennedy-Higgins, D., Hogan, J., Devlin, J. T., & Adank, P. (2016). The effect of speech distortion on the excitability of articulatory motor cortex. *NeuroImage*, *128*, 218–226.

Okada, K., & Hickok, G. (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neuroscience Letters*, *452*, 219–223.

Okada, K., Venezia, J. H., Matchin, W., Saberi, K., & Hickok, G. (2013). An fMRI study of audiovisual speech perception reveals multisensory interactions in auditory cortex. *PloS One*, *8*, e68959.

Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *68*, 169–181.

Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology*, *12*, 191–2000.

Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, “unisensory” processing. *Current Opinion in Neurobiology*, *15*, 454–458.

Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hearing*, *26*, 263–275.

Stephan, K. E., & Friston, K. J. (2010). Analyzing effective connectivity with functional magnetic resonance imaging. *Wiley Interdisciplinary Reviews. Cognitive Science*, *1*(3), 446–459.

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, *44*(3), 1210–1223.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*, 212–215.

Szenkovits, G., Peelle, J. E., Norris, D., & Davis, M. H. (2012). Individual differences in premotor and motor recruitment during speech perception. *Neuropsychologia*, *50*(7), 1380–1392.

Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (2016). Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.15.422726; this version posted December 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

integration. *Psychology and Aging*, *31*(4), 380–389.

Tye-Murray, N., Spehar, B. P., Myerson, J., Hale, S., & Sommers, M. S. (2013). Reading your own lips: Common-coding theory and visual speech perception. *Psychonomic Bulletin & Review*, *20*, 115–119.

Tye-Murray, N., Spehar, B. P., Myerson, J., Hale, S., & Sommers, M. S. (2015). The self-advantage in visual speech processing enhances audiovisual speech recognition in noise. *Psychonomic Bulletin & Review*, *22*, 1048–1053.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*, 273–289.

Van Engen, K. J., Dey, A., Sommers, M., & Peelle, J. E. (2019). *Audiovisual speech perception: Moving beyond McGurk*. https://doi.org/10.31234/osf.io/6y8qw

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181–1186.

Venezia, J. H., Fillmore, P., Matchin, W., Isenberg, A. L., Hickok, G., & Fridriksson, J. (2016). Perception drives production across sensory modalities: A network for sensorimotor integration of visual speech. *NeuroImage*, *126*, 196–207.

Venezia, J. H., Vaden, K. I., Jr., Rong, F., Maddox, D., Saberi, K., & Hickok, G. (2017). Auditory, Visual and Audiovisual Speech Processing Streams in Superior Temporal Sulcus. *Frontiers in Human Neuroscience*, *11*, 174.

Yarkoni, T., & Braver, T. S. (2010). Cognitive neurosciences approaches to individual differences in working memory and executive control: Conceptual and methodological issues. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of Individual Differences in Cognition* (pp. 87–107). Springer.