

5-16-2023

Benefit of Visual Speech Information for Word Comprehension in Post-stroke Aphasia

Anna Krason

Gabriella Vigliocco

Marja-Liisa Mailend

Harrison Stoll

Rosemary Varley

See next page for additional authors

Follow this and additional works at: <https://jdc.jefferson.edu/mossrehabfp>

 Part of the [Speech Pathology and Audiology Commons](#)

[Let us know how access to this document benefits you](#)

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Moss Rehabilitation Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: JeffersonDigitalCommons@jefferson.edu.

Authors

Anna Krason, Gabriella Vigliocco, Marja-Liisa Mailend, Harrison Stoll, Rosemary Varley, and Laurel J. Buxbaum



Research Report

Benefit of visual speech information for word comprehension in post-stroke aphasia



Anna Krason ^{a,b,*}, Gabriella Vigliocco ^{a,b}, Marja-Liisa Mailend ^{b,c},
Harrison Stoll ^{b,d}, Rosemary Varley ^e and Laurel J. Buxbaum ^{b,f}

^a Experimental Psychology, University College London, UK

^b Moss Rehabilitation Research Institute, Elkins Park, PA, USA

^c Department of Special Education, University of Tartu, Tartu Linn, Estonia

^d Applied Cognitive and Brain Science, Drexel University, Philadelphia, PA, USA

^e Language and Cognition, University College London, UK

^f Department of Rehabilitation Medicine, Thomas Jefferson University, Philadelphia, PA, USA

ARTICLE INFO

Article history:

Received 5 November 2022

Reviewed 12 February 2023

Revised 13 March 2023

Accepted 22 April 2023

Action editor Swathi Kiran

Published online 16 May 2023

Keywords:

Language

Audiovisual speech

Comprehension

Aphasia

ABSTRACT

Aphasia is a language disorder that often involves speech comprehension impairments affecting communication. In face-to-face settings, speech is accompanied by mouth and facial movements, but little is known about the extent to which they benefit aphasic comprehension. This study investigated the benefit of visual information accompanying speech for word comprehension in people with aphasia (PWA) and the neuroanatomic substrates of any benefit. Thirty-six PWA and 13 neurotypical matched control participants performed a picture-word verification task in which they indicated whether a picture of an animate/inanimate object matched a subsequent word produced by an actress in a video. Stimuli were either audiovisual (with visible mouth and facial movements) or auditory-only (still picture of a silhouette) with audio being clear (unedited) or degraded (6-band noise-vocoding). We found that visual speech information was more beneficial for neurotypical participants than PWA, and more beneficial for both groups when speech was degraded. A multivariate lesion-symptom mapping analysis for the degraded speech condition showed that lesions to superior temporal gyrus, underlying insula, primary and secondary somatosensory cortices, and inferior frontal gyrus were associated with reduced benefit of audiovisual compared to auditory-only speech, suggesting that the integrity of these fronto-temporo-parietal regions may facilitate cross-modal mapping. These findings provide initial insights into our understanding of the impact of audiovisual information on comprehension in aphasia and the brain regions mediating any benefit.

© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author. Department of Experimental Psychology, University College London, 26 Bedford Way, WC1H 0AP, London, UK.
E-mail address: anna.krason.15@ucl.ac.uk (A. Krason).

<https://doi.org/10.1016/j.cortex.2023.04.011>

0010-9452/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Post-stroke aphasia is a language disorder most frequently associated with difficulties with speech production and/or comprehension (Stroke Association UK, 2021). However, face-to-face communication goes beyond speech as it also involves processing a great deal of other communicative information, including mouth and facial movements. We know very little about whether these movements benefit the comprehension of people with aphasia (PWA) and if particular brain regions mediate any benefit. Studies with neurotypical individuals have shown that observing mouth movements facilitates auditory comprehension, particularly when speech is challenging to process due to message complexity (Arnold & Hill, 2001; Reisberg et al., 1987) or additional noise (Krason et al., 2021; Ma et al., 2009; Ross et al., 2007; Schwartz et al., 2004; Sumby & Pollack, 1954; Tye-Murray et al., 2007). This benefit is thought to occur because mouth movements support temporal and phonological encoding of the auditory speech information, as well as constrain lexical competition (for a review see Peelle & Sommers, 2015). For instance, during a conversation in a busy restaurant, mouth movements inform the listener about when to attend to others' speech and complement auditory signals by disambiguating the place of articulation of a consonant (e.g., /bæt/ versus /cæt/).

Studies with PWA have primarily investigated audiovisual speech processing using the McGurk and MacDonald paradigm (McGurk & MacDonald, 1976). In this paradigm, simultaneous mismatching information from speech acoustics (e.g., “pa”) and mouth movements (e.g., “ka”) induce an audiovisual illusion in which individuals perceive a fused percept (e.g., “ta”; McGurk & MacDonald, 1976). Despite great individual variability in susceptibility to the McGurk effect (Brown et al., 2018), most neuroanatomically healthy individuals and PWA perceive a fused percept during mismatching presentations, which has been interpreted in terms of audiovisual integration mechanisms (see Alsius, Paré, & Munhall, 2018 for a review). However, processing mismatching information from mouth and auditory speech is of unknown relevance to word comprehension and may be driven by different cognitive mechanisms (Hickok et al., 2018; Van Engen et al., 2017).

Notably, and of greater potential relevance to comprehension, PWA also benefit from mouth movements when acoustic and visual speech cues match, e.g., when “pa” is produced both auditorily and visually, relative to when “pa” is produced auditorily only (Andersen & Starrfelt, 2015; Baum et al., 2012; Campbell et al., 1990; Hessler et al., 2012; Hickok et al., 2018; Michaelis et al., 2020; but see also Youse et al., 2004). However, in a study assessing the ability of individuals with left hemisphere stroke to extract visual speech information, Schmid and Ziegler (2006) showed that PWA did not benefit from audiovisual relative to auditory-only stimuli and were impaired in matching asynchronous stimuli across auditory and visual modalities. This was particularly the case for individuals with poor verbal repetition skills and apraxia of speech (i.e., a motor speech planning disorder), suggesting that these factors may be important for successful encoding of phonological information from mouth movements and

integration with auditory speech. However, as with studies of the McGurk and McDonald illusion, the relevance of these findings for naturalistic speech comprehension may be limited: stimuli were nonsense syllables and non-speech sounds (e.g., whistling), as well as matching of asynchronous cross-modal information. Finally, associations between lesion site and behavior were not assessed.

Although lesion information is often not available in behavioral studies with clinical populations, it may strongly influence performance. Functional neuroimaging studies with neurotypical individuals generally, but not exclusively, report that three brain regions play central roles in audiovisual speech processing (for a review see Peelle, 2019). The left posterior superior temporal sulcus/gyrus (STS/STG) displays enhanced activation for audiovisual speech (with visible mouth and facial movements) relative to combined responses to auditory-only and visual-only stimuli (Callan et al., 2003; Calvert & Campbell, 2003; Calvert et al., 2000; Erickson et al., 2014; Nath & Beauchamp, 2012; Sekiyama et al., 2003; Skipper et al., 2005, 2007; Venezia et al., 2017; Wright et al., 2003), suggesting that it contributes to multisensory integration, including cross-modal integration for speech (Amedi et al., 2005; Beauchamp, 2005; Beauchamp et al., 2004; Baum et al., 2012; see also Hocking & Price, 2008; Olson, Gatenby, & Gore, 2002 for contradictory results). Some fMRI studies have also reported increased activation in the auditory cortex, including primary auditory cortex (A1), for visual speech relative to silent non-speech movements (Calvert et al., 1997; Pekkola et al., 2005). Similar findings from electrophysiological experiments show that visual cues modulate oscillations in A1 (Crosse et al., 2015; Luo et al., 2010) and that this modulation starts early, i.e., approximately 100–300 ms before speech onset, which is often related to mouth opening/closing (Chandrasekaran et al., 2009). Finally, the left inferior frontal cortex, including ventral premotor cortex (PMv) and inferior frontal gyrus (IFG), has also been associated with audiovisual processing (Calvert & Campbell, 2003; Erickson et al., 2014; Skipper et al., 2007; Watkins et al., 2003). These inferior frontal regions have been argued to play a role in mapping articulatory gestures to phoneme representations (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009), with some suggesting that observing mouth movements while listening to speech evokes activity in similar frontal brain regions as during speech production (see Skipper et al., 2017 for a review).

There is very little converging evidence from PWA that those regions are involved in audiovisual processing, and the studies that exist are also focused on perception and not comprehension. Hickok et al. (2018) conducted a large-scale voxel-based lesion-symptom mapping study assessing performance of PWA with McGurk-type stimuli. They found that left posterior superior and middle temporal regions, insula (INS), as well as parts of the occipital cortex, but not the IFG, are associated with audiovisual integration (Hickok et al., 2018). More recently, Michaelis et al. (2020) tested audiovisual integration abilities of PWA using asynchronous auditory and visual signals. Lesions to the left supramarginal gyrus (SMG) and planum temporale of the STG were associated with reduced temporal sensitivity to the asynchronous audiovisual signal, indicating that these regions are important for temporal perception that mediates audiovisual integration.

Although these findings provide important initial insights into the mechanisms driving audiovisual processing in PWA, both studies used the McGurk paradigm and are therefore subject to the criticisms raised above, i.e., they investigated syllable perception rather than comprehension.

1.1. The current study

This study is the first to investigate, using both lesion-symptom mapping and behavioral methods, the benefit of visual speech information for spoken word comprehension in PWA. We assessed 36 PWA and 13 neurotypical controls with a computer-based picture-verification task requiring judgments about whether a spoken word from a video matched a previously seen picture. We manipulated the presence of mouth and facial movements, and speech clarity. As face-to-face interactions are typically embedded in noise (e.g., a conversation on a busy street) and such adverse listening conditions increase reliance on visual speech information in neurotypical individuals (e.g., Krasen et al., 2021; Ma et al., 2009; Ross et al., 2007; Sumby & Pollack, 1954), we compared clear speech to 6-band noise-vocoded stimuli. Finally, we assessed the neural regions associated with any benefit of visual speech information during word comprehension using Support-Vector Regression Lesion-Symptom mapping (SVR-LSM, Zhang et al., 2014).

Based on the current literature on the processing of audiovisual speech by neurotypical individuals, we predicted that performance of both groups would improve in the degraded condition when mouth movements were present thanks to the support they provide to phonological encoding of degraded auditory signals (e.g., Ross et al., 2007; Sumby & Pollack, 1954). Given limited studies on audiovisual speech processing beyond syllable level and involving individuals with post-stroke aphasia, it is unclear whether PWA would benefit from observing mouth and facial movements to a larger extent than neurotypical individuals. It is possible that PWA would use visual speech information to overcome noise (similarly to neurotypical individuals), but also to remedy any auditory speech deficits caused by aphasia. It may also be the case, however, that integrating visual and auditory channels is more challenging for PWA than for healthy adults, thus, resulting in a smaller audiovisual benefit. We hypothesized that any benefit from observing mouth and facial movements to PWA would depend on individuals' lesion location. That is, we predicted that a reduced audiovisual benefit should be observed in patients with lesions to the posterior STS/STG, a key region for multisensory integration in studies with neurotypicals (e.g., Beauchamp et al., 2004). As we considered comprehension of real words with visible mouth and facial movements, other regions including A1 and inferior frontal cortices (PMv and IFG; e.g., Calvert et al., 1997; Watkins et al., 2003) may also contribute to visual speech benefit.

2. Methods

In the methods section, we report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established

prior to data analysis, all manipulations, and all measures in the study.

2.1. Participants

Forty-nine native speakers of North American English were recruited from the Moss Rehabilitation Research Institute (MRRI) Research Registry (Schwartz et al., 2005) to participate in the study. Participants included (i) 36 individuals at least 6 months-post a single left hemispheric cerebrovascular accident who exhibited aphasia and, to ensure that they would be able to understand experimental task instructions, had a score of at least 5 (out of 10) on the auditory comprehension subtest of the WAB (Kertesz, 1982; PWA group; mean age = 62, SD = 11.55) and (ii) 13 neurotypical subjects (control group; mean age = 64, SD = 9.13) matched for age ($t(47) = .59, p = .56$) and educational level ($t(47) = -.28, p = .78$) to the PWA group. Control participants were included if they achieved a score of at least 27 on the Mini-Mental State Exam (Folstein et al., 1975). Exclusion criteria for both groups included a history of comorbid neurological disorders, psychosis, and alcohol or drug abuse. Additionally, 33 of the PWA passed a hearing screening at 50, 1000, 2000, and 4000 Hz (if they were <65 years old) or 1000 and 2000 frequency (if they were >65 years old) in both ears. To maximize the sample size, 3 PWA were included in the study despite not passing the hearing screening.¹ All participants gave informed consent before taking part in the experiment according to the guidelines of the Institutional Review Board of Einstein Healthcare Network and were compensated for their time and travel expenses. The testing sessions took place in the MRRI laboratories in Elkins Park (Pennsylvania, USA). The de-identified data from this study are publicly available on Open Science Framework (OSF) at <https://osf.io/fuscq/>.

2.1.1. Neuroimaging acquisition

Twenty-nine participants with aphasia received research-quality structural MRI (26) or CT (3) scans if the former was medically contraindicated. The MRI scans included whole-brain T1-weighted images acquired on a 3 T Siemens Trio (Erlangen, Germany) scanner with repetition time of 1620 ms, echo time of 3.87 ms, field of view of 192,256 mm, with $1 \times 1 \times 1$ mm voxels, and using a Siemens eight-channel head coil. The CT scans were obtained without contrast (60 axial slices, 3–5 mm slice thickness) on a 64-slice Siemens SOMATOM Sensation scanner.

Lesions were manually segmented on each patient's high-resolution T-1 weighted structural images. Lesioned voxels were assigned a value of 1, and preserved voxels were assigned a value of 0. Both contained grey and white matter. Binarized lesion masks were then registered to an MNI template (Montreal Neurological Institute "Colin27") using a symmetric diffeomorphic registration algorithm (Avants

¹ We tested an accuracy model including all predictors of interest (see Data Analysis section) but excluding the 3 participants who did not pass the audiometry screening. The results are consistent with the results from the accuracy model with the full sample, suggesting that the hearing factor did not influence our findings. All results are presented in the Supplementary Materials for comparison.

et al., 2008; www.picsl.upenn.edu/ANTS). First, volumes were registered to an intermediate template of healthy brain images acquired on the same scanner, and they were then mapped onto the “Colin27” template. Lesion maps were subsequently inspected by an experienced neurologist (H.B. Coslett), naive to the behavioral results of the study, to ensure mapping accuracy. The same neurologist drew the CT scans directly onto the “Colin27” template using MRICron (Rorden & Brett, 2000). To ensure maximum accuracy with high intra- and inter-rater reliability (>.85%), the pitch of the template was rotated to approximate the slice plane of each participant's scan (see e.g., Schnur et al., 2009).

3. Materials

In the experimental picture-word verification task participants indicated whether a spoken stimulus matched a previously seen picture. Experimental materials for the study consisted of 120 words, a corpus of 480 pictures with high name agreement, and 240 video-clips. The list of words and the video-clips are publicly available at <https://osf.io/fuscq/>. The picture materials could not be publicly archived due to copyright concerns.

3.1. Words

All words were concrete (Mn. 3.5 out of 5 on a concreteness scale; Brysbaert et al., 2014) and referred to common objects and living things. Words were grouped into sets of four (e.g., “cow”, “ear”, “egg”, “pie”) and items within a set were matched on number of syllables and as closely as possible on number of phonemes, lexical frequency (Brysbaert & New, 2009), age of acquisition (AoA; Kuperman et al., 2012), and phonological neighborhood density (Luce & Pisoni, 1998). Each participant saw all 120 words, but the words within a group were presented in different conditions (see below) to different participants. For example, participant 1 heard the word “cow” in the clear condition with visible mouth movements, whereas participant 2 heard the same word in the clear condition, but with no visible facial cues. The sets of four words remained constant across participants and experimental conditions.

3.2. Pictures

Pictures with high name agreement, selected through a naming experiment or inter-rater agreement analysis, were selected from one of three databases (Druks & Masterson, 2000; Hebart et al., 2019; Snodgrass & Vanderwart, 1980). The pictures could refer to: (i) a target word (e.g., “chair”); (ii) a semantically related object (e.g., “table”); (iii) an object with a phonologically related (rhyming) name (e.g., “bear”); or (iv) an unrelated object (e.g., “shoe”). Object names for (ii), (iii), and (iv) had different onset phonemes than the target words on 90% of occasions.

3.3. Video-clips

The video-clips were recorded in a professional, well-lit sound-proof booth at University College London. They depicted a

female native speaker of American English with visible head and shoulders uttering target words. The videos were further manipulated in iMovie (version 10.1.12) in the following way. First, we extracted the audio from the video files and combined it with a still image of a female silhouette. As a result, each video had two versions: with (audiovisual) and without (auditory-only) visual cues. This contrast is analogous to real-life scenarios in which interlocutors have face-to-face versus telephone conversations. The decision to use a still image of a silhouette as an auditory-only baseline, rather than a still picture of a speaker or video of a speaker with a blurred lip area, was driven by the concern that the auditory-visual mismatch would create expectancy conflicts that would actively disrupt processing rather than serving as a truly neutral condition. In addition, blurring different parts of the face to control for their role in speech processing is ecologically less valid.

Second, we moderately noise-vocoded the audio in Praat (Boersma, 2021) using a 6-band pass filter following Drijvers and Özyürek (2017) and Krasen et al. (2021). Noise-vocoded speech is a type of degraded speech in which pitch-related information is manipulated to simulate the listening experience of someone with a cochlear implant (Shannon et al., 1995). Six-band filtering makes the speech challenging, but still intelligible (to a certain degree) and has been previously shown to increase neurotypical individuals' use of visual cues in word recognition tasks (Drijvers & Özyürek, 2017; Krasen et al., 2021). The final set of videos was therefore presented in one of the following conditions: clear audiovisual (clear audio + visible mouth and facial movements), degraded audiovisual (noise-vocoded audio + visible mouth and facial movements), clear auditory-only (clear audio + no visual cues), and degraded auditory-only (noise-vocoded audio + no visual cues). Fig. 1 depicts the experimental conditions and trial types used in the study.

The stimuli were displayed on a 24-inch monitor with 1920 × 1080 resolution. The videos occupied the upper 2/3 of the screen, and the pictures occupied the lower part.

3.4. Procedure

The experiment was programmed in Gorilla (<https://gorilla.sc/>). Participants wore high-quality headphones during the experiment. Participants' task was to indicate whether a spoken stimulus matched a previously seen picture. Each trial started with a still image of an actress (or a female silhouette in the auditory-only modality) and a fixation cross beneath it. After 500 ms, a picture of an object or living thing appeared in place of the fixation cross. After another 1500 ms, a 200 ms beep tone was played indicating the beginning of a ~1500 ms video. The picture remained in view until the end of a video, after which a new screen with a question (“Does the speech match the picture?”) and two response boxes appeared. Participants used their left hand (i.e., the unaffected hand in the PWA) to indicate their responses using “z” and “x” buttons on the keyboard with corresponding colored stickers (“z” [yellow sticker] = “yes”, “x” [blue sticker] = “no”). See Fig. 2 for an example of the trial sequence.

Prior to the main task, participants were presented with four practice trials illustrating all possible conditions (i.e., clear audiovisual, degraded audiovisual, clear auditory-only,


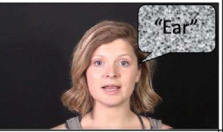






Condition	Clear Audiovisual	Degraded Audiovisual	Clear Auditory-only	Degraded Auditory-only
Video				
Picture				
Trial type	“yes” trial	“no” trial semantically related	“no” trial phonologically related	“no” trial unrelated

Fig. 1 – Schematic representation of the experimental conditions and trial types (note: grey speech bubbles represent the noise-vocoded conditions).

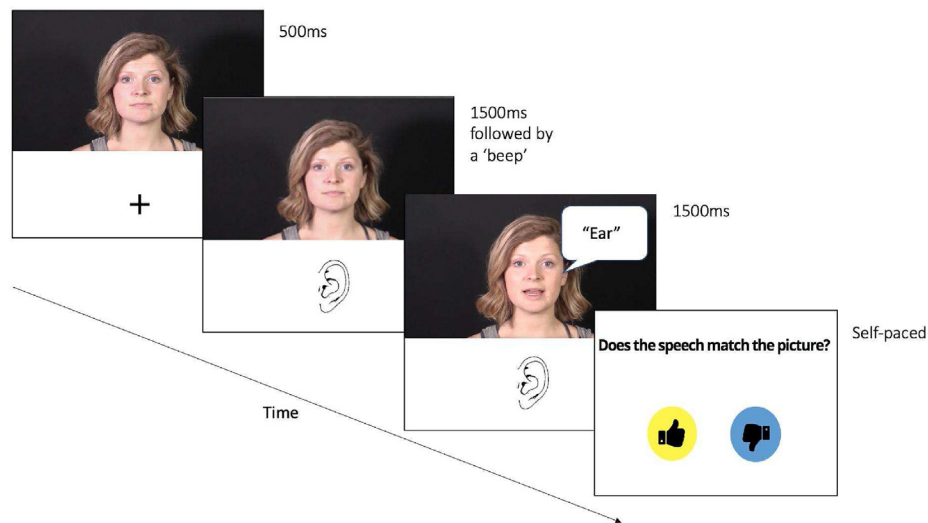


Fig. 2 – Example of a matching trial sequence in audiovisual modality with clear speech.

degraded auditory-only). The practice trials were repeated as necessary to ensure participants understood the task. Both visual and oral feedback was provided during the practice phase. In the main task, participants were exposed to all the target words twice, resulting in 240 trials, with 50% of the trials requiring a “yes” response (matching trials) and the other 50% requiring a “no” response (mismatching trials). None of the mismatching pictures appeared as targets. The second presentation of each word always appeared in a different experimental condition and in the second half of the experiment (after a 10-min break). The trials were pseudo-randomized into eight blocks of 30 trials. Each session lasted approximately 1.5 h.

3.5. Data analysis

3.5.1. Behavioral analysis

The *lme4* package (Bates et al., 2015) was used to perform a set of mixed-effect analyses in RStudio (RStudio Team, 2015). We carried out generalized logistic mixed-effect regressions (glmer) on accuracy separately for the matching and mismatching trials.² The decision to analyze matching and

mismatching trials separately was driven by the findings from neurotypical individuals showing that different cognitive processes may be involved when responding yes/no to matching and mismatching picture-word pairs, with matching trials being overall more reliable (see, e.g., Stadthagen-Gonzalez et al., 2009). Specifically, matching trials have been suggested to reflect conceptual (semantic) matching, i.e., individuals access the meaning of both spoken words and pictures (Stadthagen-Gonzalez et al., 2009). In contrast, mismatching trials have been shown to elicit much more variability in how people respond to them, which may be related to the number of additional “checks” one has to perform to decide that a word and a picture mismatch (Krueger, 1978). Potential cognitive mechanisms that may be triggered during mismatching, but less so during matching, trials are cognitive control and priming. Finally, assessing the benefit of congruent visual information is of clinical relevance.

Prior to the analyses, we removed trials with technical difficulties and trials with a phonologically related word “gauge”, because its visual speech information is identical with the visual information of its matching target word “cage” (21 trials in total). We entered the following predictors and up to three-way interactions between them in our models: Speech Clarity (clear, degraded), Modality (audiovisual,

² Reaction time data were unreliable due to a number of responses prior to the response window, i.e., while videos played.

auditory-only) and Group (PWA, neurotypicals), as well as Relation Type (semantic, phonological, unrelated) in the mismatching trial analysis. Following a design-driven approach (Barr et al., 2013), we included by-Participant and by-Item random intercepts to account for participant and item variability. We also entered random slopes for Speech Clarity and Modality both by-Participant and by-Item to better control for type I error. Random slopes of Modality were removed from the analysis of the mismatching trials due to model singularity fit. The control variables entered in the models included the Number of Syllables of the target words, Log Frequency (Brysbaert & New, 2009), AoA (Kuperman et al., 2012), and Phonological Neighborhood Density (Luce & Pisoni, 1998). We applied the “bobyqa” algorithm to optimize model convergence and speed of iterations (Powell, 2009). There was no obvious multicollinearity, with Variance Inflation Factors (VIFs) below 2.7 and 4.8 for the matching and mismatching trial analyses, respectively. Finally, the coefficients were used to interpret the size and direction of effects (Jaeger, 2008) and significance values were assessed with Laplace approximation using the *LmerTest* package (Kuznetsova et al., 2017). Plots were created using the *ggplot2* package (Wickham, 2009). The R code for the analyses is available on OSF at <https://osf.io/fuscq/>. No part of the study procedures or analyses was preregistered.

Finally, we calculated d' and c , using the *psycho* package (Makowski, 2018), to check for task sensitivity and response bias, respectively. D' was calculated by taking the difference in z -scores between hits (correct responses to “yes” trials) and false alarms (incorrect responses to “no” trials). Larger d' values indicate better sensitivity to the task, and d' values closer to 0 signify performance approximating chance level (Stanislaw & Todorov, 1999). C was calculated by looking at the number of standard deviations from the point where neither response is preferred (so-called “neutral point”), with positive values indicating a tendency towards “no” responses and negative values indicating a tendency towards “yes” responses (Stanislaw & Todorov, 1999). The d' values in our study varied between .62 and 4.35, suggesting that task sensitivity was good and all participants responded above chance level. The c values ranged from $-.72$ to $.66$ and fell well within $\pm 3SD$ from the neutral point, suggesting that participants were not biased towards “yes” or “no” responses.

3.5.2. Lesion-symptom mapping analysis

Support Vector Regression Lesion-Symptom Mapping (SVR-LSM) was performed in MATLAB using a toolbox developed by DeMarco and Turkeltaub (2018). SVR-LSM is a multivariate machine learning technique that uses a nonlinear function to determine the association between a map of lesioned voxels in the brain (rather than single voxels) and patients' behavior (Zhang et al., 2014). As compared with its predecessor, voxel based lesion-symptom mapping (VLSM), it offers better specificity and sensitivity (Mah et al., 2014) by controlling for type I and type II errors caused by correlations between neighboring voxels (Pustina et al., 2018) and multiple comparisons (Bennett et al., 2009), respectively. Importantly, SVR-LSM also outperforms VLSM if a particular behavior is associated with multiple brain regions (Herbet et al., 2015; Mah et al., 2014), as may be the case for speech comprehension.

Here, we focused on the lesions associated with reduced audiovisual benefit (as compared to auditory-only speech) in the matching trials (i.e., requiring a “yes” response) because of their clinical relevance. Based on the accuracy data distribution, we looked at the degraded speech condition only. We used residuals of the audiovisual condition with auditory-only scores regressed out as the dependent variable. We excluded any voxels that were lesioned in less than three patients ($\sim 10\%$ of the total number of patients). We regressed lesion volume from both the individual lesion masks and participants' behavioral scores to control for total lesion volume following DeMarco and Turkeltaub (2018). We generated a null distribution using 10,000 Monte Carlo permutations to determine voxelwise statistical significance. We cross-validated our model by dividing our sample into 5-folds, with four subgroups used to create a regression model and the fifth subgroup used to validate it. The resulting map was then thresholded at $p < .05$, and any clusters smaller than 500 voxels were excluded, following Garcea et al. (2019), Lacey et al. (2017), and Vigliocco et al. (2020).

Finally, we used the Johns-Hopkins DTI-based probabilistic white matter tractography atlas (Mori et al., 2008) to determine the overlap between significant voxels from the SVR-LSM analysis and major white matter tracts at a 75% probability threshold (Baldo et al., 2012; Schwartz et al., 2012).

4. Results

4.1. Behavioral results

4.1.1. Matching trials

We found significant main effects of Speech Clarity ($\beta = 1.29$, $SE = .22$, $z = 5.92$, $p < .001$) and Group ($\beta = .50$, $SE = .19$, $z = 2.59$, $p = .01$), with participants performing better on the clear speech relative to degraded speech and the control group performing more accurately than the PWA group. There was also a significant interaction between Speech Clarity and Modality ($\beta = -.30$, $SE = .13$, $z = -2.38$, $p = .02$). Pairwise comparison with Holm's corrections showed that when the speech was degraded, participants made fewer errors on audiovisual compared to auditory-only presentations ($p < .001$). There was no difference between audiovisual and auditory-only modalities when the speech was clear, likely because performance was at ceiling ($p > .05$). One control variable was also significant (Number of Syllables: $\beta = 1.10$, $SE = .29$, $z = 3.76$, $p < .001$, with participants performing better on longer words). Fig. 3 (A) shows mean accuracy scores per group for the matching trials.

4.1.2. Mismatching trials

There were significant main effects of Speech Clarity ($\beta = .56$, $SE = .13$, $z = 4.43$, $p < .001$), with fewer errors for the clear speech; Modality ($\beta = .38$, $SE = .07$, $z = 5.17$, $p < .001$), with fewer errors for audiovisual presentations; and Relation Type, with fewer errors for unrelated pictures as compared to phonologically ($\beta = -.58$, $SE = .10$, $z = -5.76$, $p < .001$) and semantically related pictures ($\beta = -.64$, $SE = .10$, $z = -6.26$, $p < .001$). There was also a significant interaction between Modality and Group ($\beta = .22$, $SE = .07$, $z = 3.18$, $p = .001$).

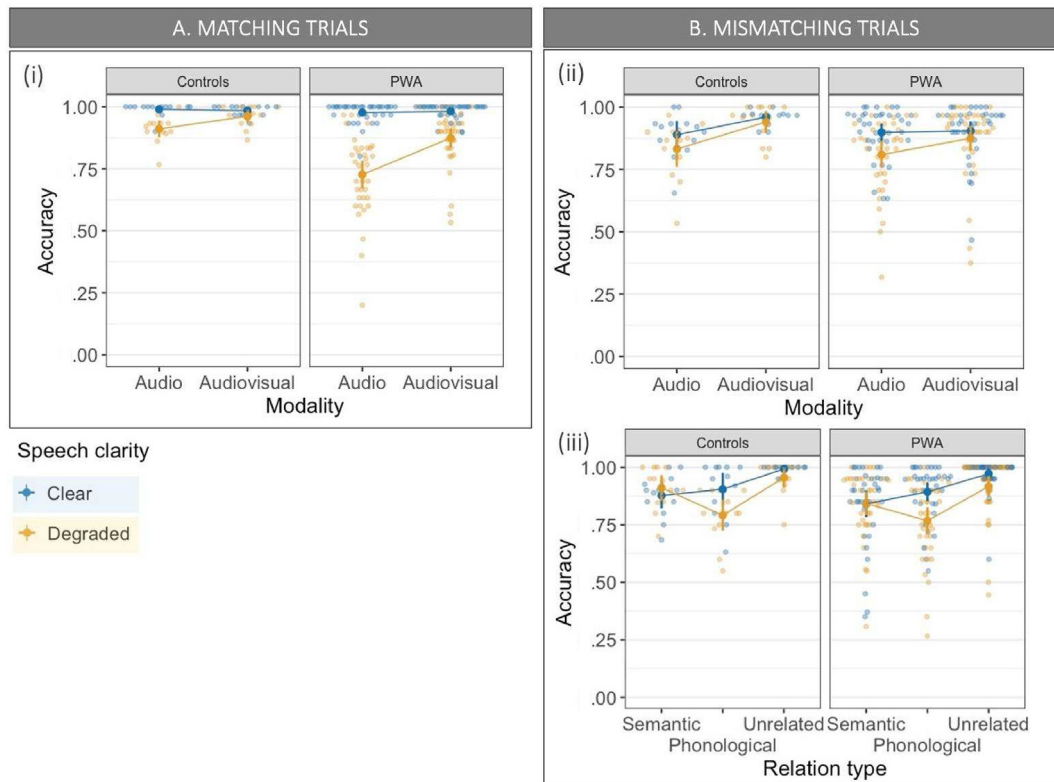


Fig. 3 – Mean accuracy scores for matching (A) and mismatching (B) trials for the control and PWA groups. Plots (i) and (ii) show Modality on the x-axis, whereas plot (iii) shows Relation Type on the x-axis. Speech Clarity is represented in colors. Error bars are standard errors of the mean.

Follow-up pairwise comparisons with Holm's corrections showed that although both groups performed better in the audiovisual modality compared to auditory-only (P 's < .04), the difference between conditions was larger for the control group ($p = .05$). This effect was further examined in a post-hoc analysis with only the PWA for whom lesion information was available (29) and including a new variable – Lesion Volume (in mm³) – to establish whether lesion size is a significant predictor of smaller benefit from the audiovisual modality. There was no effect of lesion volume on the behavioral results (see Supplementary Materials for full results).

There was also a significant interaction between Speech Clarity and Relation Type (for the phonological type with the unrelated type as a reference: $\beta = -.59$, $SE = .10$, $z = -6.00$, $p < .001$). Pairwise comparisons showed significantly better performance for the clear relative to degraded speech for phonological and unrelated pictures (P 's < .01), but not semantically related pictures ($p > .05$). When the speech was clear, participants were also more accurate on the phonological than semantic pictures ($p = .004$), but when the speech was degraded, they were more accurate on the semantic trials compared to phonological ones ($p < .001$). One control variable (Phonological Neighborhood Density) was also significant ($\beta = -.02$, $SE = .01$, $z = -2.05$, $p = .04$, with participants performing better on words with smaller phonological neighborhood density). Fig. 3 (B) shows mean accuracy scores per group for the mismatching trials.

4.1.3. Lesion-symptom mapping results

To assess which brain areas, when lesioned, are associated with reduced benefit of visual speech cues, we carried out a SVR-LSM analysis in the 29 PWA who had scans (see Table 1). The overlap map with regions lesioned in at least three participants is depicted in Fig. 4. The dependent variable was the residuals of the audiovisual condition with the auditory-only condition regressed out for degraded speech in the matching condition. The SVR-LSM analysis showed several significant clusters, including parts of the superior temporal pole (TPOsup, STG), postcentral gyrus (PoCG), SMG, INS, and IFG (pars triangularis and pars orbitalis). Table 2 and Fig. 5 summarize the results. Finally, based on the Johns-Hopkins DTI probabilistic atlas (Mori et al., 2008), we found an overlap between significant clusters and superior longitudinal fasciculus (SLF). The probabilistic location of SLF and the overlap is presented in Fig. 6. The percentage overlap between SLF and SVR-LSM results is presented in Table 3.

5. Discussion

The current study is the first to investigate the benefit of mouth and facial movements in word comprehension of people with aphasia using both behavioral and lesion-symptom mapping methods. In contrast to previous studies, we used a picture-verification task and manipulated the presence of visual speech information and the clarity of

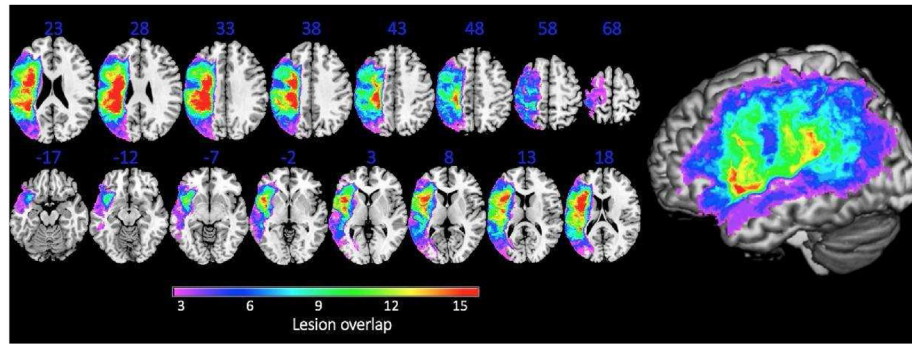


Fig. 4 – Voxelwise lesion overlap for 29 participants. Only voxels lesioned in a minimum of 3 participants are displayed.

Table 1 – Patient demographics.

ID	Aphasia Diagnosis	WAB AQ	WAB Compr.	Sex	Educat. (years)	Age At Testing	Months Since Stroke	Lesion Volume (mm ³)
P01	Broca's	43.7	7.4	F	12	55	173	111,206
P02	Conduction	71.9	8.0	M	16	60	19	189,767
P03	Transcortical motor	69.8	6.9	M	19	75	19	120,820
P04	Broca's	66.0	9.0	M	12	69	174	71,750
P05	Anomic	90.8	8.6	M	12	64	139	47,566
P06	Anomic	87.8	8.9	F	14	61	135	41,502
P07	Anomic	86.4	9.2	M	12	62	50	11,961
P08	Anomic	92.6	8.5	M	18	50	52	96,147
P09	Anomic	88.3	9.8	F	18	58	168	80,532
P10	Anomic	89.7	9.1	M	14	57	126	55,685
P11	Anomic	81.3	9.4	M	12	57	88	126,448
P12	Anomic	82.4	10.0	M	13	63	168	193,421
P13	Broca's	39.6	7.7	M	13	56	68	n/a
P14	Anomic	92.7	9.5	M	18	77	36	87,120
P15	Anomic	88.1	9.4	F	13	57	179	106,731
P16	Anomic	92.8	10.0	F	13	33	97	n/a
P17	Anomic	95.4	9.6	F	12	52	22	26,504
P18	Broca's	50.3	9.4	F	16	70	118	109,181
P19	Conduction	73.8	8.8	M	12	64	15	n/a
P20	Anomic	93.9	9.4	F	12	42	84	27,840
P21	Anomic	89.2	9.4	F	13	70	23	n/a
P22	Anomic	90.1	8.6	F	14	67	10	n/a
P23	Anomic	93.9	9.9	M	12	66	120	64,284
P24	Anomic	87.4	9.5	F	16	41	53	181,199
P25	Broca's	33.2	6.3	M	13	40	96	222,352
P26	Broca's	32.4	7.9	M	12	84	170	145,170
P27	Anomic	92.3	9.9	M	16	67	245	99,980
P28	Anomic	88.5	9.1	F	16	60	232	124,678
P29	Anomic	92.4	8.8	F	12	83	135	18,528
P30	Broca's	31.9	7.8	M	16	59	19	n/a
P31	Broca's	68.6	8.3	F	11	70	48	56,156
P32	Anomic	91.2	9.3	M	19	75	67	32,003
P33	Anomic	88.0	9.2	F	13	55	134	136,576
P34	Broca's	61.6	6.6	F	19	73	378	231,141
P35	Anomic	89.5	9.3	F	13	56	104	48,459
P36	Anomic	94.6	10	M	19	69	96	n/a
24 Anomic		M = 77.8	M = 8.9	17Fs	M = 14.3	M = 61.6	M = 106.6	M = 98,783
9 Broca's		SD = 20.0	SD = 1.0		SD = 2.6	SD = 11.5	SD = 78.4	SD = 61,483
2 Conduction 1 Transcor-tical motor		R = 31.9–95.4	R = 6.3–10		R = 11–19	R = 33–84	R = 10–378	R = 11,961–231,141

Abbreviations: WAB AQ = Western Aphasia Battery Aphasia Quotient; WAB Compr. = Western Aphasia Battery Auditory Comprehension Score; Educat. = Education; M = mean; SD = standard deviation; R = range; n/a = not applicable.

Table 2 – SVR-LSM results with X, Y and Z centers of mass associated with reduced benefit from the audiovisual speech relative to auditory-only in the degraded listening condition for the matching trials. Regions with clusters of >500 voxels were identified by Automated Anatomical Labeling (AAL).

Regions	Abbrev.	Number of Voxels in Damaged Region	Percentage of Voxels in Damaged Region	MNI Centers of Mass		
				X	Y	Z
Temporal Pole: Superior Temporal Gyrus	TPOsup	1630	15.94	−49	14	−3
Superior Temporal Gyrus	STG	1433	7.83	−54	−43	21
Postcentral Gyrus	PoCG	1297	4.18	−42	−17	39
Supramarginal Gyrus	SMG	1038	10.48	−49	−29	31
Insula	INS	1001	6.66	−30	19	−8
Inferior Frontal Gyrus, Pars Triangularis	IFGtriang	975	4.85	−55	21	3
Inferior Frontal Gyrus, Pars Orbitalis	IFGorb	719	5.29	−51	28	−2

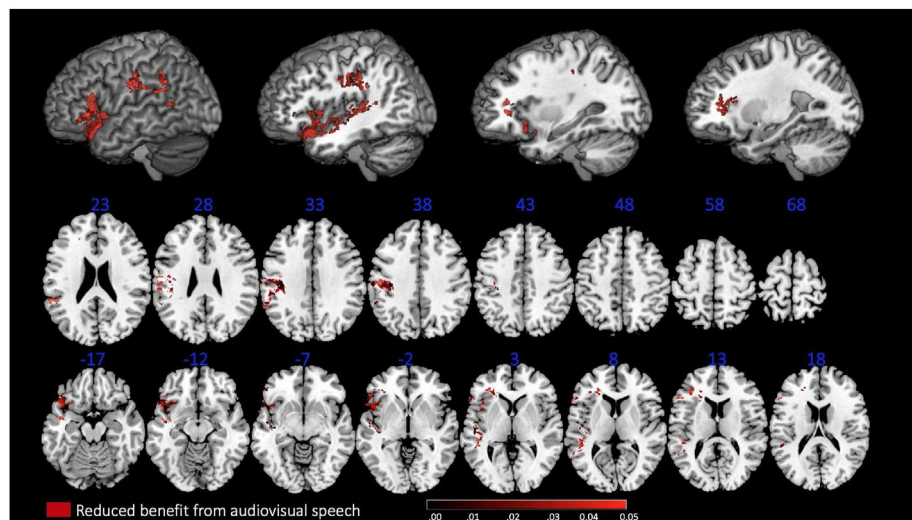


Fig. 5 – SVR-LSM results depicting significant voxels (in red), which when lesioned, are associated with reduced benefit from audiovisual presentation relative to auditory-only presentation during degraded listening condition for the matching trials. Voxelwise threshold set to $p < .05$ with 10,000 Monte Carlo permutations and 5-fold cross-validation. Clusters of <500 contiguous 1 mm^3 voxels were excluded.

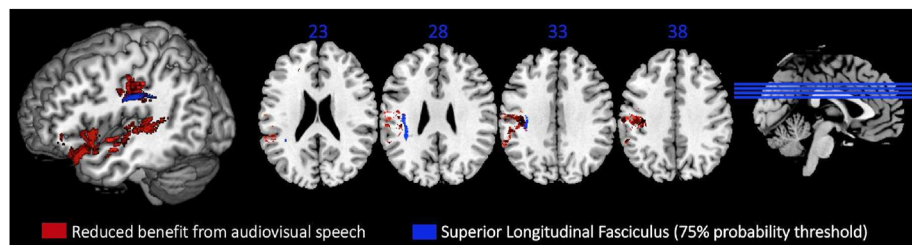


Fig. 6 – Probabilistic location of the white matter tracts based on the JHU white matter atlas overlaid onto SVR-LSM results. The dependent variable was the amount of benefit from audiovisual speech relative to auditory-only speech in the degraded condition for the matching trials. White matter tract probability threshold: 75%.

Table 3 – The overlap percentage between peak voxels in MNI space identified in the SVR-LSM analysis and superior longitudinal fasciculus (SLF), as verified with the Johns Hopkins DTI-based probabilistic white matter tractography atlas.

Regions	Abbrev.	Number of Voxels in SLF	Number of Voxels Identified in SVR-LSM that Overlap	Percentage of Overlapping Voxels	MNI Centers of Mass		
					X	Y	Z
Superior Longitudinal Fasciculus	SLF	690	104	15.07	−38	−18	31

auditory signal to assess the extent to which these factors impact speech comprehension in adults with post-stroke aphasia and a neurotypical control group. We also conducted exploratory SVR-LSM to investigate the neural regions associated with any benefit of visual speech for word comprehension.

In line with previous studies assessing audiovisual comprehension of neurotypical individuals, we found that visual information accompanying speech benefits comprehension in challenging listening conditions and that such benefit is larger for the controls relative to PWA regardless of speech clarity conditions. Our SVR-LSM and tractographic analyses indicated that TPOsup, STG, SMG, PoCG, INS, IFG, and SLF may mediate the benefit of audiovisual information in comprehension.

5.1. Benefit of visual speech for aphasic comprehension

Potential benefits of visual speech information were assessed separately for matching (i.e., speech matched a previously-seen picture) and mismatching (i.e., speech mismatched a previously-seen picture) trials. For the matching trials, we showed that comprehension of degraded speech was easier when the speech was accompanied by mouth and facial movements relative to when the visual information was absent. This result is in line with previous findings with neurotypical adults (Krasen et al., 2021; Ma et al., 2009; Ross et al., 2007; Schwartz et al., 2004; Sumbly & Pollack, 1954; Tye-Murray et al., 2007), indicating that visual speech information plays a role particularly when phonological processing is more difficult. In such challenging listening conditions, mouth movements are likely to be beneficial because they support temporal predictions of the upcoming auditory speech information and constrain lexical competition (Peelle & Sommers, 2015). Our findings are also consistent with previous research showing similar performance of PWA and neurotypical adults under adverse listening conditions (Kittredge et al., 2006; Healy, Moser, Morrow-Odom, Hall, & Fridriksson, 2007). The lack of audiovisual benefit for PWA in the clear speech condition is likely to be driven by a ceiling effect; that is, like controls, these individuals with mild-moderate aphasia performed relatively well in the clear condition. For this reason, the present findings may not generalize to individuals with more severe comprehension impairments.

Additionally, we found effects of visual speech for the mismatching trials. Both groups benefited from seeing mouth and facial movements in addition to hearing speech, but the control group showed a larger advantage than the aphasic group, which may be related to the involvement of additional cognitive processes (such as cognitive control that is often impaired in PWA; Brownsett et al., 2014) during mismatching presentations. To our knowledge, only one recent unpublished study investigated audiovisual speech benefit in a sentence repetition task in PWA and found a similar pattern of larger audiovisual advantage for neurotypical individuals in one of their experimental conditions (i.e., during very high noise levels of 0 dB SNR; Raymer, Ringleb, Sandberg, & Schwartz, 2021). Although the reported methods and data analysis are insufficiently detailed to allow strong

comparisons to our findings, both our study and the study of Raymer et al. (2021) suggest the possibility that PWA may have difficulty integrating visual and auditory streams of information into a coherent percept, as would be required for mouth movements to be useful in disambiguating speech (Massaro & Jesse, 2007; Schmid & Ziegler, 2006).

Moreover, it is interesting to note that the control group in the present study also showed audiovisual benefit for the mismatching trials when the speech was clear, as well as when it was degraded. This is a different pattern than we observed in the matching trials; however, in the mismatching trials performance was “off-ceiling” in the auditory-only condition, leaving room for a benefit of visual information. Finally, neurotypical and aphasic individuals also responded more accurately to unrelated trials compared to both phonologically and semantically related trials. Moreover, the performance on the latter two relation types depended on speech clarity: Individuals performed equally well on semantic trials whether the auditory signal was clear or degraded; In contrast, they made more errors on phonological trials when speech was degraded than clear. Altogether, this finding demonstrates that phonological discriminability is reduced by noise, whereas semantic discriminability is not.

5.2. Neural substrates of visual speech benefit

Our exploratory lesion-symptom mapping analysis identified several clusters in the left hemisphere that appear to be involved in audiovisual speech comprehension. These include perisylvian regions in temporal (TPOsup, STG), insular (INS) and inferior frontal (IFG) cortices, as well as parts of parietal (SMG) and somatosensory cortices (PoCG). Although the SVR-LSM was conducted on a relatively small sample size (see Ivanova et al., 2021) and replication is needed, our results are consistent with previous findings in neurotypical populations suggesting involvement of a large fronto-temporo-parietal network, including STG, STS, INS, superior and inferior frontal cortex, as well as SMG and IPL, in sensorimotor speech interactions (Calvert et al., 2001; Campbell, 2008; Dick et al., 2010; Peelle, 2019; Bernstein and Liebenthal, 2014). Our findings also indicate that both ventral and dorsal streams may contribute to the benefit of visual speech for word comprehension. Portions of the ventral stream, and in particular, posterior superior and middle temporal cortex, have been associated with sound-to-meaning mapping. In the present study we found a cluster of regions distributed along the lateral and medial surfaces of the STG to be associated with audiovisual speech comprehension. The STG is known for its multifunctionality and heteromodality (Hein & Knight, 2021; Venezia et al., 2017), and previous studies have found that posterior STG/STS play a crucial role in audiovisual and visual speech processing (Callan et al., 2003; Calvert & Campbell, 2003; Calvert et al., 2000; Erickson et al., 2014; Nath & Beauchamp, 2012; Okada & Hickok, 2009; Sekiyama et al., 2003; Skipper et al., 2005, 2007; Venezia et al., 2016; Wright et al., 2003), likely because of its multisensory integration properties (Amedi et al., 2005; Beauchamp, 2005; Beauchamp et al., 2004). Less is known about the involvement of the temporal pole in the processing of visual speech cues. The temporal pole has primarily been linked with higher-order

cognitive processes, such as naming (e.g., Rice, Hoffman, & Lambon Ralph, 2015), word retrieval (e.g., Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004), and semantic processing (Lambon Ralph, 2013; Patterson, Nestor, & Rogers, 2007). A few studies have suggested a role for the anterior STG in audiovisual speech processing (Hertrich et al., 2011; Lee & Noppeney, 2011; Ozker et al., 2017). For example, Hertrich et al. (2011) showed that relatively anterior parts of STG are linked with the processing of visual speech information (e.g., syllables "pa" and "ta") and more posterior STG is associated with cross-modal integration with non-speech stimuli (e.g., moving shapes and tones). Although the stimuli in these studies were not directly relevant to comprehension, it is of interest to note the convergence of our results with these findings.

The dorsal stream, by contrast, including portions of the posterior-frontal and parietal-temporal cortices, has been previously associated with sound-to-articulatory mapping in speech production. Here, we showed that insular regions medial to the superior temporal surface and fronto-parietal regions of the dorsal stream may play a role in visual speech comprehension, in line with previous literature (Callan et al., 2003; Calvert et al., 2001; Hickok et al., 2018; Skipper et al., 2007). For instance, Hickok et al. (2018) found associations between INS and susceptibility to perceiving fused perceptions with McGurk stimuli, while Callan et al. (2003) reported INS to be involved in mouth movement processing when the auditory signal is degraded or absent. Although the precise role of the insula in audiovisual speech comprehension is still debated, these findings indicate that it may act as a mediator during cross-modal interactions and/or executive demand processing under challenging conditions (Callan et al., 2003; Calvert et al., 2001; Hickok et al., 2018; Skipper et al., 2007). Given that our stimuli consisted of videos of a speaker's full face rather than solely mouth movements, the involvement of the insular cortex found in the current study may also be related to processing socio-emotional facial cues (Rae et al., 2018).

We also found that primary somatosensory cortex (PoCG) and parietal association areas (SMG) appear to mediate the benefit of visual speech information for comprehension. These regions may be engaged in encoding phonological information from mouth movements (Möttönen et al., 2005; Skipper et al., 2005, 2007) and binding it with the auditory signal (Bernstein, Auer, Wagner, & Ponton, 2008; Bernstein & Liebenthal, 2014; Jones & Callan, 2003; Michaelis et al., 2020). Additionally, we showed that IFG may be associated with the benefit of mouth movements, which is in line with Skipper et al. (2005, 2007) and Watkins et al. (2003), but not other recent studies with PWA (Andersen & Starrfelt, 2015; Hickok et al., 2018). These findings may be discrepant because the involvement of IFG in audiovisual processing is task specific (for a review see Peelle, 2019). For example, when speech encoding is more challenging, IFG may play a compensatory role in supporting the extraction of visual information from the mouth. It is important to note that although our findings are consistent with the prior literature in our identification of multiple fronto-temporal brain regions involved in audiovisual processing, our sample was small for a robust SVR-LSM analysis and future studies may identify additional regions.

Another limitation of the present study was that our sample of chronic patients largely consisted of anomic aphasics and lacked individuals with Wernicke's or transcortical sensory aphasia. Although these aphasia types are less common in the chronic than acute phases of recovery, future research may benefit from a more diverse sample of PWA.

Finally, our results are also in line with a recent study of Zhang and Du (2022), showing involvement of the dorsal stream, including PMv, IFG, SMG and the underlying white matter tracts of the arcuate fasciculus, in phonological encoding from mouth movements during audiovisual speech perception. Their findings are also consistent with our white matter tractographic analysis demonstrating that the SLF is associated with the benefit of visual speech information. In particular, the parts of the SLF connecting superior temporal with inferior frontal regions have been found to be critical for phonological processing (Dick et al., 2014; Glasser & Rilling, 2008). Thus, a disruption to phonological processing caused by lesions to SLF may lead to cross-modal integration failure, which could explain the reduced benefit from audiovisual speech relative to the auditory signal alone. Future studies should investigate how the connectivity between these fronto-temporo-parietal regions, as well as between these regions and the right hemisphere, impacts audiovisual speech comprehension in aphasia.

6. Conclusions

The current study brings together behavioral and lesion-symptom mapping profiles of people with aphasia to establish the benefit of visual speech information for word comprehension. We have demonstrated that mouth and facial movements are more beneficial for the comprehension of neurotypical individuals than adults with aphasia, and are more beneficial for both groups when listening conditions are challenging. We have also provided preliminary evidence that the integrity of a number of specific inferior frontal, temporal, parietal regions as well as fronto-temporal connection via the superior longitudinal fasciculus may be associated with this benefit, consistent with the previously-demonstrated role of these regions in cross-modal mapping. Although studies of spoken word comprehension have typically focused on the auditory modality, our findings suggest that future investigations should consider whether and how visual speech information impacts comprehension in aphasia.

Open Practices Section

The study in this article earned Open Data badge for transparent practices. The data from this study are publicly available on Open Science Framework (OSF) at <https://osf.io/fuscq/>

Author information

Authors declare no conflict of interest.

CRediT author statement

AK: Conceptualization, methodology, validation, formal analysis, investigation, writing original draft, visualization, funding acquisition; GV: Conceptualization, methodology, resources, writing–review & editing, supervision, project administration, funding acquisition; MM: Methodology, writing–review & editing; HS: Software, validation, investigation; RV: Methodology, writing–review & editing, supervision; LB: Conceptualization, methodology, resources, writing–review & editing, supervision, project administration, funding acquisition.

Acknowledgments

We would like to thank Erica Middleton for sharing picture stimuli with us; Rachel Metzgar for helping with recruitment and testing; Frank Garcea for his expertise in lesion-symptom mapping analysis; H. Branch Coslett for help with lesion segmentation; Linda Drijvers for sharing the Praat script; and all participants who took part in our study.

This research was supported by Peer Review Committee funding (PRC FY19-2) awarded to LB and GV. The work was further supported by a European Research Council Advanced Grant (ECOLANG, 743035). While working on this project, GV was supported by a Royal Society Wolfson Research Merit Award (WRM\R3\170016). This research was also supported by the UCL Bogue Research Fellowship awarded to AK.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cortex.2023.04.011>.

REFERENCES

- Alsius, A., Paré, M., & Munhall, K. G. (2018). Forty Years After Hearing Lips and Seeing Voices: The McGurk Effect Revisited. *Multisensory Research*, 31(1–2), 111–144. <https://doi.org/10.1163/22134808-00002565>.
- Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research*, 166(3–4), 559–571. <https://doi.org/10.1007/s00221-005-2396-5>.
- Andersen, T. S., & Starrfelt, R. (2015). Audiovisual integration of speech in a patient with Broca's Aphasia. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00435>.
- Arnold, P., & Hill, F. (2001). Bisenory augmentation: A speechreading advantage when speech is clearly audible and intact. *British journal of psychology (London, England: 1953)*, 92(Part 2), 339–355.
- Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1), 26–41. <https://doi.org/10.1016/j.media.2007.06.004>.
- Baldo, J. V., Katseff, S., & Dronkers, N. F. (2012). Brain regions underlying repetition and auditory-verbal short-term memory deficits in aphasia: Evidence from voxel-based lesion symptom mapping. *Aphasiology*, 26(3–4), 338–354. <https://doi.org/10.1080/02687038.2011.602391>.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Medicine and Life*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Baum, S. H., Martin, R. C., Hamilton, A. C., & Beauchamp, M. S. (2012). Multisensory speech perception without the left superior temporal sulcus. *Neuroimage*, 62(3), 1825–1832. <https://doi.org/10.1016/j.neuroimage.2012.05.034>.
- Beauchamp, M. S. (2005). See me, hear me, touch me: Multisensory integration in lateral occipital-temporal cortex. *Current Opinion in Neurobiology*, 15(2), 145–153. <https://doi.org/10.1016/j.conb.2005.03.011>.
- Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, 41(5), 809–823. [https://doi.org/10.1016/S0896-6273\(04\)00070-4](https://doi.org/10.1016/S0896-6273(04)00070-4).
- Bennett, C. M., Wolford, G. L., & Miller, M. B. (2009). The principled control of false positives in neuroimaging. [Social Cognitive and Affective Neuroscience Electronic Resource], 4(4), 417–422. <https://doi.org/10.1093/scan/nsp053>.
- Bernstein, L. E., Auer, E. T., Wagner, M., & Ponton, C. W. (2008). Spatio-temporal Dynamics of Audiovisual Speech Processing. *NeuroImage*, 39(1), 423–435. <https://doi.org/10.1016/j.neuroimage.2007.08.035>.
- Bernstein, L. E., & Liebenthal, E. (2014). Neural pathways for visual speech perception. *The Florida Nurse*, 8, 386. <https://doi.org/10.3389/fnins.2014.00386>.
- Boersma, Paul & Weenink, David (2021). Praat: doing phonetics by computer [Computer program]. Version 6.3.10, retrieved 3 May 2023 from <http://www.praat.org/>
- Brown, V. A., Hedayati, M., Zanger, A., Mayn, S., Ray, L., Dillman-Hasso, N., & Strand, J. F. (2018). What accounts for individual differences in susceptibility to the McGurk effect? *Plos One*, 13(11), Article e0207160. <https://doi.org/10.1371/journal.pone.0207160>.
- Brownsett, S. L. E., Warren, J. E., Geranmayeh, F., Woodhead, Z., Leech, R., & Wise, R. J. S. (2014). Cognitive control and its impact on recovery from aphasic stroke. *Brain: a Journal of Neurology*, 137(1), 242–254. <https://doi.org/10.1093/brain/awt289>.
- Brysbaert, M., & New, B. (2009). Moving beyond kučera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990. <https://doi.org/10.3758/BRM.41.4.977>.
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904–911. <https://doi.org/10.3758/s13428-013-0403-5>.
- Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport*, 14(17), 2213–2218. <https://doi.org/10.1097/00001756-200312020-00016>.
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D., & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science (New York, N.Y.)*, 276(5312), 593–596.
- Calvert, G. A., & Campbell, R. (2003). Reading speech from still and moving faces: The neural substrates of visible speech. *Journal*

- of Cognitive Neuroscience, 15(1), 57–70. <https://doi.org/10.1162/089892903321107828>
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology: CB*, 10(11), 649–657.
- Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage*, 14(2), 427–438. <https://doi.org/10.1006/nimg.2001.0812>
- Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1001–1010. <https://doi.org/10.1098/rstb.2007.2155>
- Campbell, R., Garwood, J., Franklin, S., Howard, D., Landis, T., & Regard, M. (1990). Neuropsychological studies of auditory-visual fusion illusions. Four case studies and their implications. *Neuropsychologia*, 28(8), 787–802.
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *Plos Computational Biology*, 5(7), Article e1000436. <https://doi.org/10.1371/journal.pcbi.1000436>
- Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. *Journal of Neuroscience*, 35(42), 14195–14204. <https://doi.org/10.1523/JNEUROSCI.1829-15.2015>
- Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, 92(1–2), 179–229. <https://doi.org/10.1016/j.cognition.2002.07.001>
- DeMarco, A. T., & Turkeltaub, P. E. (2018). A multivariate lesion symptom mapping toolbox and examination of lesion-volume biases and correction methods in lesion-symptom mapping. *Human Brain Mapping*, 39(11), 4169–4182. <https://doi.org/10.1002/hbm.24289>
- Dick, A. S., Bernal, B., & Tremblay, P. (2014). The Language connectome: New pathways, new concepts. *The Neuroscientist: a Review Journal Bringing Neurobiology, Neurology and Psychiatry*, 20(5), 453–467. <https://doi.org/10.1177/1073858413513502>
- Dick, A. S., Solodkin, A., & Small, S. L. (2010). Neural development of networks for audiovisual speech comprehension. *Brain and Language*, 114(2), 101–114. <https://doi.org/10.1016/j.bandl.2009.08.005>
- Drijvers, L., & Özyürek, A. (2017). Visual context enhanced: The joint contribution of iconic gestures and visible speech to degraded speech comprehension. *Journal of Speech, Language, and Hearing Research: JSLHR*, 60(1), 212–222. https://doi.org/10.1044/2016_JSLHR-H-16-0101
- Druks, J., & Masterson, J. (2000). *An object and action naming Battery*. Psychology Press.
- Erickson, L. C., Heeg, E., Rauschecker, J. P., & Turkeltaub, P. E. (2014). An ALE meta-analysis on the audiovisual integration of speech signals: ALE meta-analysis on AV speech integration. *Human Brain Mapping*, 35(11), 5587–5605. <https://doi.org/10.1002/hbm.22572>
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189–198.
- Garcea, F. E., Stoll, H., & Buxbaum, L. J. (2019). Reduced competition between tool action neighbors in left hemisphere stroke. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior*, 120, 269–283. <https://doi.org/10.1016/j.cortex.2019.05.021>
- Glasser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cerebral Cortex*, 18(11), 2471–2482. <https://doi.org/10.1093/cercor/bhn011>
- Healy, E. W., Moser, D. C., Morrow-Odom, K. L., Hall, D. A., & Fridriksson, J. (2007). Speech Perception in MRI Scanner Noise by Persons With Aphasia. *Journal of Speech, Language, and Hearing Research*, 50(2), 323–334. [https://doi.org/10.1044/1092-4388\(2007/023\)](https://doi.org/10.1044/1092-4388(2007/023))
- Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Wicklin, C. V., & Baker, C. I. (2019). Things: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *Plos One*, 14(10), Article e0223792. <https://doi.org/10.1371/journal.pone.0223792>
- Hein, G., & Knight, R. T. (2021). *Superior Temporal Sulcus—It's My Area: Or Is It?*, 20(12), 12.
- Herbet, G., Lafargue, G., & Duffau, H. (2015). Rethinking voxel-wise lesion-deficit analysis: A new challenge for computational neuropsychology. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior*, 64, 413–416. <https://doi.org/10.1016/j.cortex.2014.10.021>
- Hertrich, I., Dietrich, S., & Ackermann, H. (2011). Cross-modal interactions during perception of audiovisual speech and nonspeech signals: An fMRI study. *Journal of Cognitive Neuroscience*, 23(1), 221–237. <https://doi.org/10.1162/jocn.2010.21421>
- Hessler, D., Jonkers, R., & Bastiaanse, R. (2012). Processing of audiovisual stimuli in aphasic and non-brain-damaged listeners. *Aphasiology*, 26(1), 83–102. <https://doi.org/10.1080/02687038.2011.608840>
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402. <https://doi.org/10.1038/nrn2113>
- Hickok, G., Rogalsky, C., Matchin, W., Basilakos, A., Cai, J., Pillay, S., et al. (2018). Neural networks supporting audiovisual integration for speech: A large-scale lesion study. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior*, 103, 360–371. <https://doi.org/10.1016/j.cortex.2018.03.030>
- Hocking, J., & Price, C. J. (2008). The Role of the Posterior Superior Temporal Sulcus in Audiovisual Processing. *Cerebral Cortex*, 18(10), 2439–2449. <https://doi.org/10.1093/cercor/bhn007>
- Ivanova, M. V., Herron, T. J., Dronkers, N. F., & Baldo, J. V. (2021). An empirical comparison of univariate versus multivariate methods for the analysis of brain-behavior mapping. *Human Brain Mapping*, 42(4), 1070–1101. <https://doi.org/10.1002/hbm.25278>
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Medicine and Life*, 59(4), 434–446. <https://doi.org/10.1016/j.jml.2007.11.007>
- Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *NeuroReport: For Rapid Communication of Neuroscience Research*, 14(8), 1129–1133. <https://doi.org/10.1097/00001756-200306110-00006>
- Kertesz, A. (1982). *The Western Aphasia Battery*. New York :Grune & Stratton.
- Kittredge, A., Davis, L., & Blumstein, S. E. (2006). Effects of nonlinguistic auditory variations on lexical processing in Broca's aphasics. *Brain and Language*, 97(1), 25–40. <https://doi.org/10.1016/j.bandl.2005.07.012>
- Krason, A., Fenton, R., Varley, R., & Vigliocco, G. (2021). The role of iconic gestures and mouth movements in face-to-face communication. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-021-02009-5>
- Krueger, L. E. (1978). A theory of perceptual matching. *Psychological Review*, 85, 278–304. <https://doi.org/10.1037/0033-295X.85.4.278>
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4), 978–990. <https://doi.org/10.3758/s13428-012-0210-4>

- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13). <https://doi.org/10.18637/jss.v082.i13>
- Lacey, E. H., Skipper-Kallal, L., Xing, S., Fama, M., & Turkeltaub, P. (2017). Mapping common aphasia assessments to underlying cognitive processes and their neural substrates. *Neurorehabilitation and Neural Repair*, 31(5), 442–450. <https://doi.org/10.1177/1545968316688797>
- Lambon Ralph, M. A. (2013). Neurocognitive insights on conceptual knowledge and its breakdown. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 20120392. <https://doi.org/10.1098/rstb.2012.0392>
- Lee, H., & Noppeney, U. (2011). Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. *Journal of Neuroscience*, 31(31), 11338–11350. <https://doi.org/10.1523/JNEUROSCI.6510-10.2011>
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19(1), 1–36.
- Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLOS Biology*, 8(8), Article e1000445. <https://doi.org/10.1371/journal.pbio.1000445>
- Mah, Y.-H., Husain, M., Rees, G., & Nachev, P. (2014). Human brain lesion-deficit inference remapped. *Brain: a Journal of Neurology*, 137(9), 2522–2531. <https://doi.org/10.1093/brain/awu164>
- Makowski, D. (2018). The psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science. *Journal of Open Source Software*, 3(22), 470. <https://doi.org/10.21105/joss.00470>
- Massaro, D. W., & Jesse, A. (2007). *Audiovisual speech perception and word recognition*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198568971.013.0002>
- Ma, W. J., Zhou, X., Ross, L. A., Foxe, J. J., & Parra, L. C. (2009). Lip-Reading aids word recognition most in moderate noise: A bayesian explanation using high-dimensional feature space. *Plos One*, 4(3), e4638. <https://doi.org/10.1371/journal.pone.0004638>
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748. <https://doi.org/10.1038/264746a0>
- Michaelis, K., Erickson, L. C., Fama, M. E., Skipper-Kallal, L. M., Xing, S., Lacey, E. H., ... Turkeltaub, P. E. (2020). Effects of age and left hemisphere lesions on audiovisual integration of speech. *Brain and Language*, 12.
- Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., et al. (2008). Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *Neuroimage*, 40(2), 570–582. <https://doi.org/10.1016/j.neuroimage.2007.12.035>
- Möttönen, R., Järveläinen, J., Sams, M., & Hari, R. (2005). Viewing speech modulates activity in the left SI mouth cortex. *Neuroimage*, 24(3), 731–737. <https://doi.org/10.1016/j.neuroimage.2004.10.011>
- 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. <https://doi.org/10.1016/j.neuroimage.2011.07.024>
- Okada, K., & Hickok, G. (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neuroscience Letters*, 452(3), 219–223. <https://doi.org/10.1016/j.neulet.2009.01.060>
- Olson, I. R., Gatenby, J. C., & Gore, J. C. (2002). A comparison of bound and unbound audio–visual information processing in the human cerebral cortex. *Cognitive Brain Research*, 14(1), 129–138. [https://doi.org/10.1016/S0926-6410\(02\)00067-8](https://doi.org/10.1016/S0926-6410(02)00067-8)
- Ozker, M., Schepers, I. M., Magnotti, J. F., Yoshor, D., & Beauchamp, M. S. (2017). A double dissociation between anterior and posterior superior temporal gyrus for processing audiovisual speech demonstrated by electrocorticography. *Journal of Cognitive Neuroscience*, 29(6), 1044–1060. https://doi.org/10.1162/jocn_a_01110
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews. Neuroscience*, 8(12), 976–987. <https://doi.org/10.1038/nrn2277>
- Peelle, J. E. (2019). *The neural basis for auditory and audiovisual speech perception*, 25.
- 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. <https://doi.org/10.1016/j.cortex.2015.03.006>
- Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., Tarkiainen, A., & Sams, M. (2005). Primary auditory cortex activation by visual speech: An fMRI study at 3 T. *Neuroreport*, 16(2), 125–128.
- Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives*, 39.
- Pustina, D., Avants, B., Faseyitan, O. K., Medaglia, J. D., & Coslett, H. B. (2018). Improved accuracy of lesion to symptom mapping with multivariate sparse canonical correlations. *Neuropsychologia*, 115, 154–166. <https://doi.org/10.1016/j.neuropsychologia.2017.08.027>
- Rae, C. L., Polyanska, L., Gould van Praag, C. D., Parkinson, J., Bouyagoub, S., Nagai, Y., Seth, A. K., Harrison, N. A., Garfinkel, S. N., & Critchley, H. D. (2018). Face perception enhances insula and motor network reactivity in Tourette syndrome. *Brain: a Journal of Neurology*, 141(11), 3249–3261. <https://doi.org/10.1093/brain/awy254>
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724. <https://doi.org/10.1038/nn.2331>
- Raymer, A., Ringleb, S., Sandberg, H., & Schwartz, K. (2021). *Visual Influences on Auditory Processing in Noise in Aphasia* (No. 6403). Article 6403. <https://easychair.org/publications/preprint/B69h>
- Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In *Hearing by eye: The psychology of lip-reading* (pp. 97–113). Lawrence Erlbaum Associates, Inc.
- Rice, G. E., Hoffman, P., & Lambon Ralph, M. A. (2015). Graded specialization within and between the anterior temporal lobes: Graded specialization within and between the ATLs. *Annals of the New York Academy of Sciences*, 1359(1), 84–97. <https://doi.org/10.1111/nyas.12951>
- Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology*, 12(4), 191–200. <https://doi.org/10.1155/2000/421719>
- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral cortex (New York, N.Y.: 1991)*, 17(5). <https://doi.org/10.1093/cercor/bhl024>
- RStudio Team, 2015. RStudio: Integrated Development Environment for R, Boston, MA. Available at: <http://www.rstudio.com/>.
- Schmid, G., & Ziegler, W. (2006). Audio-visual matching of speech and non-speech oral gestures in patients with aphasia and apraxia of speech. *Neuropsychologia*, 44(4), 546–555. <https://doi.org/10.1016/j.neuropsychologia.2005.07.002>
- Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area.

- Proceedings of the National Academy of Sciences, 106(1), 322–327. <https://doi.org/10.1073/pnas.0805874106>
- Schwartz, J.-L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition*, 93(2), B69–B78. <https://doi.org/10.1016/j.cognition.2004.01.006>
- Schwartz, M. F., Brecher, A. R., Whyte, J., & Klein, M. G. (2005). A patient Registry for cognitive rehabilitation research: A strategy for balancing patients' privacy rights with researchers' need for access. *Archives of Physical Medicine and Rehabilitation*, 86(9), 1807–1814. <https://doi.org/10.1016/j.apmr.2005.03.009>
- Schwartz, M. F., Faseyitan, O., Kim, J., & Coslett, H. B. (2012). The dorsal stream contribution to phonological retrieval in object naming. *Brain: a Journal of Neurology*, 135(12), 3799–3814. <https://doi.org/10.1093/brain/aws300>
- Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, 47(3), 277–287. [https://doi.org/10.1016/S0168-0102\(03\)00214-1](https://doi.org/10.1016/S0168-0102(03)00214-1)
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science (New York, N.Y.)*, 270(5234), 303–304.
- Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language*, 164, 77–105. <https://doi.org/10.1016/j.bandl.2016.10.004>
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: Motor cortical activation during speech perception. *Neuroimage*, 25(1), 76–89. <https://doi.org/10.1016/j.neuroimage.2004.11.006>
- Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17(10), 2387–2399. <https://doi.org/10.1093/cercor/bhl147>
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 174–215. <https://doi.org/10.1037/0278-7393.6.2.174>
- Stadthagen-Gonzalez, H., Damian, M. F., Pérez, M. A., Bowers, J. S., & Marín, J. (2009). Name–picture verification as a control measure for object naming: A task analysis and norms for a large set of pictures. *The Quarterly Journal of Experimental Psychology: QJEP*, 62(8), 1581–1597. <https://doi.org/10.1080/17470210802511139>
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137–149. <https://doi.org/10.3758/BF03207704>
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26(2), 212–215. <https://doi.org/10.1121/1.1907309>
- Tye-Murray, N., Sommers, M., & Spehar, B. (2007). Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification*, 11(4), 233–241. <https://doi.org/10.1177/1084713807307409>
- Van Engen, K. J., Xie, Z., & Chandrasekaran, B. (2017). Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect. *Attention, Perception, & Psychophysics*, 79(2), 396–403. <https://doi.org/10.3758/s13414-016-1238-9>
- Venezia, J. H., Fillmore, P., Matchin, W., Lisette Isenberg, A., Hickok, G., & Fridriksson, J. (2016). Perception drives production across sensory modalities: A network for sensorimotor integration of visual speech. *Neuroimage*, 126, 196–207. <https://doi.org/10.1016/j.neuroimage.2015.11.038>
- Venezia, J. H., Vaden, K. I., 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. <https://doi.org/10.3389/fnhum.2017.00174>
- Vigliocco, G., Krason, A., Stoll, H., Monti, A., & Buxbaum, L. (2020). Multimodal comprehension in left hemisphere stroke patients [Preprint]. *PsyArXiv*. <https://doi.org/10.31234/osf.io/umgk3>
- Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41(8), 989–994.
- Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Dordrecht: Hadley Wickham.
- Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex*, 13(10), 1034–1043. <https://doi.org/10.1093/cercor/13.10.1034>
- Youse, K. M., Cienkowski, K. M., & Coelho, C. A. (2004). Auditory-visual speech perception in an adult with aphasia. *Brain Injury*, 18(8), 825–834. <https://doi.org/10.1080/02699000410001671784>
- Zhang, L., & Du, Y. (2022). Lip movements enhance speech representations and effective connectivity in auditory dorsal stream. *NeuroImage*, 257, 119311. <https://doi.org/10.1016/j.neuroimage.2022.119311>
- Zhang, Y., Kimberg, D. Y., Coslett, H. B., Schwartz, M. F., & Wang, Z. (2014). Multivariate lesion-symptom mapping using support vector regression. *Human Brain Mapping*, 35(12), 5861–5876. <https://doi.org/10.1002/hbm.22590>