Aalto University

OPEN ACCESS

This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Saalasti, Satu; Alho, Jussi; Lahnakoski, Juha M.; Bacha-Trams, Mareike; Glerean, Enrico; Jääskeläinen, Iiro P.; Hasson, Uri; Sams, Mikko

Lipreading a naturalistic narrative in a female population: Neural characteristics shared with listening and reading

Published in:
Brain and Behavior

DOI:
10.1002/brb3.2869

Published: 01/02/2023

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Saalasti, S., Alho, J., Lahnakoski, J. M., Bacha-Trams, M., Glerean, E., Jääskeläinen, I. P., Hasson, U., & Sams, M. (2023). Lipreading a naturalistic narrative in a female population: Neural characteristics shared with listening and reading. *Brain and Behavior*, *13*(2), Article e2869. https://doi.org/10.1002/brb3.2869

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Check for updates

Received: 11 May 2022 | Revised: 29 November 2022 | Accepted: 6 December 2022

DOI: 10.1002/brb3.2869

Brain and Behavior Open Access WILEY

ORIGINAL ARTICLE

Lipreading a naturalistic narrative in a female population: Neural characteristics shared with listening and reading

Satu Saalasti[1,2,3] | Jussi Alho[2] | Juha M. Lahnakoski[2,4,5,6] | Mareike Bacha-Trams[2] | Enrico Glerean[2,7] | Iiro P. Jääskeläinen[2] | Uri Hasson[7] | Mikko Sams[8,9]

[1]Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland

[2]Brain and Mind Laboratory, Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland

[3]Advanced Magnetic Imaging (AMI) Centre, Aalto NeuroImaging, School of Science, Aalto University, Espoo, Finland

[4]Independent Max Planck Research Group for Social Neuroscience, Max Planck Institute of Psychiatry, Munich, Germany

[5]Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7), Research Center Jülich, Jülich, Germany

[6]Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[7]Department of Psychology and the Neuroscience Institute, Princeton University, Princeton, USA

[8]Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland

[9]Aalto Studios - MAGICS, Aalto University, Espoo, Finland

Correspondence
Satu Saalasti, Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland.
Email: satu.saalasti@helsinki.fi

Funding information
Academy of Finland, Grant/Award Number: 276643; Alfred Kordelin Foundation; The Finnish Cultural Foundation, Grant/Award Number: 150496

Abstract

Introduction: Few of us are skilled lipreaders while most struggle with the task. Neural substrates that enable comprehension of connected natural speech via lipreading are not yet well understood.

Methods: We used a data-driven approach to identify brain areas underlying the lipreading of an 8-min narrative with participants whose lipreading skills varied extensively (range 6–100%, mean = 50.7%). The participants also listened to and read the same narrative. The similarity between individual participants' brain activity during the whole narrative, within and between conditions, was estimated by a voxel-wise comparison of the Blood Oxygenation Level Dependent (BOLD) signal time courses.

Results: Inter-subject correlation (ISC) of the time courses revealed that lipreading, listening to, and reading the narrative were largely supported by the same brain areas in the temporal, parietal and frontal cortices, precuneus, and cerebellum. Additionally, listening to and reading connected naturalistic speech particularly activated higher-level linguistic processing in the parietal and frontal cortices more consistently than lipreading, probably paralleling the limited understanding obtained via lip-reading. Importantly, higher lipreading test score and subjective estimate of comprehension of the lipread narrative was associated with activity in the superior and middle temporal cortex.

Conclusions: Our new data illustrates that findings from prior studies using well-controlled repetitive speech stimuli and stimulus-driven data analyses are also valid for naturalistic connected speech. Our results might suggest an efficient use of brain areas dealing with phonological processing in skilled lipreaders.

KEYWORDS
fMRI, language, lipreading, speech, speechreading

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2022 The Authors. *Brain and Behavior* published by Wiley Periodicals LLC.

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1 | INTRODUCTION

To some extent, everyone can extract phonetic information from a speaker's lips, tongue, jaw, teeth, and cheeks, and use this ability, known as lipreading or speechreading, to support speech comprehension (Bernstein & Liebenthal, 2014; Files et al., 2015). This is particularly useful for individuals with hearing impairment (Strelnikov et al., 2009; Suess et al., 2022) even after cochlear implantation (Anderson, Wiggins, Kitterick, & Hartley, 2017), but also for normal hearing individuals when speech is difficult to comprehend, for example, due to acoustic noise (Ross et al., 2007; Sumby & Pollack, 1954). For most normal hearing individuals, comprehending speech solely via lipreading is challenging, due to poor visibility of some sounds (e.g., the velopharyngeal sounds /k/ and /h/), and due to the similarity between the lip shapes of different phonemes (e.g., the bilabial phonemes /m/, /b/, and /p/) (Altieri et al., 2011; Auer & Bernstein, 2007; Bernstein & Liebenthal, 2014; Files et al., 2015; Summerfield et al., 1992). Consequently, lipreading accuracy varies greatly (from 0% to 65%, mean 18.57%; SD = 13.18) in individuals with normal hearing (Altieri et al., 2011), and the benefits of good lipreading are not available in the same way for all. A better understanding of the neural basis of lipreading comprehension, especially the comprehension of naturalistic speech is critical for understanding how the brain accesses linguistic meaning via different means. This will help in developing more efficient intervention methods and allow more people to access the benefits of fluent lipreading.

In order to comprehend speech via lipreading, visual speech movements need to be recognized by mapping sensory or phonological representations to their meanings, similarly to the manner in which familiar speech sounds are recognized in auditory speech (DeWitt & Rauschecker, 2012; Hickok & Poeppel, 2007), and familiar orthographic forms in text comprehension (Dehaene & Cohen, 2011). Different theories postulate possible visual speech processing pathways. One theory suggests convergence with auditory speech pathways, which is supported by findings showing that viewing naturally moving speaking faces versus still images of a resting face activates auditory speech areas (BA 41/42; lateral parts of Heschl's gyrus, Calvert & Campbell, 2003). Similarly, lipreading meaningless syllables (Möttönen et al., 2002; Pekkola et al., 2005; Sams et al., 1991) and numbers (Calvert et al., 1997; MacSweeney et al., 2000) as well as words (Calvert et. al., 1997) has been found to activate the auditory cortex, including its primary areas. Alternatively, it has also been suggested that the phonetic integration of visual speech cues would be processed bottom-up within the late vision system (Bernstein et al., 2004; Bernstein et al., 2011; Paulesu, 2003). Potentially both auditory and visual cortices have a functionally specific form, and brain engages both visual and auditory pathways for comprehension (Bröhl et al., 2022). Or visual speech processing would be based on multisensory integration in the superior temporal lobe (Okada et al., 2013; Beuchamp et al., 2004; Calvert et al., 2000). Furthermore, participation of the prefrontal speech motor areas in the middle frontal gyrus (MFG), inferior frontal gyrus (IFG), supramarginal gyrus (SMG), and premotor cortex has been found when viewing orofacial movements (Nishitani & Hari, 2002), indicating the involvement of a mirror neuron-like system. The speech motor system was activated when participants attempted to lipread vowels (Callan et al., 2014), syllables (Skipper et al., 2007), words (Paulesu et al., 2003; Watkins et al., 2003), and short stories (Skipper et al., 2005). More recently, it has been suggested that a specific temporal visual speech area (TVSA), in the posterior temporal cortex, ventral and posterior to the multisensory posterior superior temporal sulcus (pSTS), supports sensory–motor integration (Bernstein & Liebenthal, 2014; Bernstein et al., 2011), because its strength of activation has been found to be related to the number of correctly recognized nonsense syllables (n = 10). But further brain areas have been found to covary with lipreading accuracy measures.

Capek et al. (2008) found lipreading a skill-dependent activity in the right lingual gyrus, the posterior cingulate (pCING), and in the right postcentral and inferior temporal gyri in normal hearing individuals (N = 13). The authors suggested that this was due to the relatively greater involvement of face processing and articulatory skill. Furthermore, an early functional magnetic resonance imaging (fMRI) study, based on data from nine participants, suggested that poor versus good lipreaders had less activation in the superior temporal gyrus (STG) and middle temporal gyrus (MTG) while lipreading sentences (Ludman et al., 2000). In a more comprehensive study, 33 normally hearing participants with lipreading skills ranging from 7% to 89% (quantified with a sentence-based lipreading test) watched silent, naturally spoken, isolated sentences during fMRI (Hall et al., 2005). Irrespective of lipreading skills, activity was found in the IFG, MFG, inferior parietal lobule (IPL), particularly in the left hemisphere, and MTG (peak activity in the posterior part) bilaterally. A correlation analysis suggested that small voxel clusters (4–11 voxels) in the mSFG, IFG, fusiform gyrus, pCING cortex, and lingual gyrus (bilaterally) were associated with lipreading skills. Additionally, the authors found that the activity strength in the left STG showed a significant linear relationship with lipreading skills when using individually thresholded activation maps and restricting the analysis to the STG (the outermost anatomical boundaries of Heschl's gyrus and the planum temporale). The authors suggested that phonological processing mechanisms were important for successful lipreading.

Previous findings suggest that lipreading sublexical material such as pseudowords and syllables elicits activation that is more restricted to areas sensitive to visual motion (Bernstein et al., 2011; Campbell et al., 2001). Instead, lipreading words (Paulesu et al., 2003) and sentences (Hall et al., 2005; Ludman et al., 2000) has been found to elicit activation that extends to the temporal cortex, including MTG in skilled lipreaders. The involvement of MTG increases with the linguistic complexity of the lipread material (Ludman et al., 2000; Paulesu et al., 2003). That said the brain areas that are specifically important for lipreading naturalistic, connected speech are not yet properly identified. Working with naturalistic narratives has become increasingly common (Simony et al., 2016; Wilson et al., 2008; Yeshurun et al., 2017a; for a review, see Jääskeläinen et al., 2021, and for related M/EEG research Crosse et al., 2016; Golumbic et al., 2013; Haider et al., 2022). Listening to naturalistic narratives synchronizes brain

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

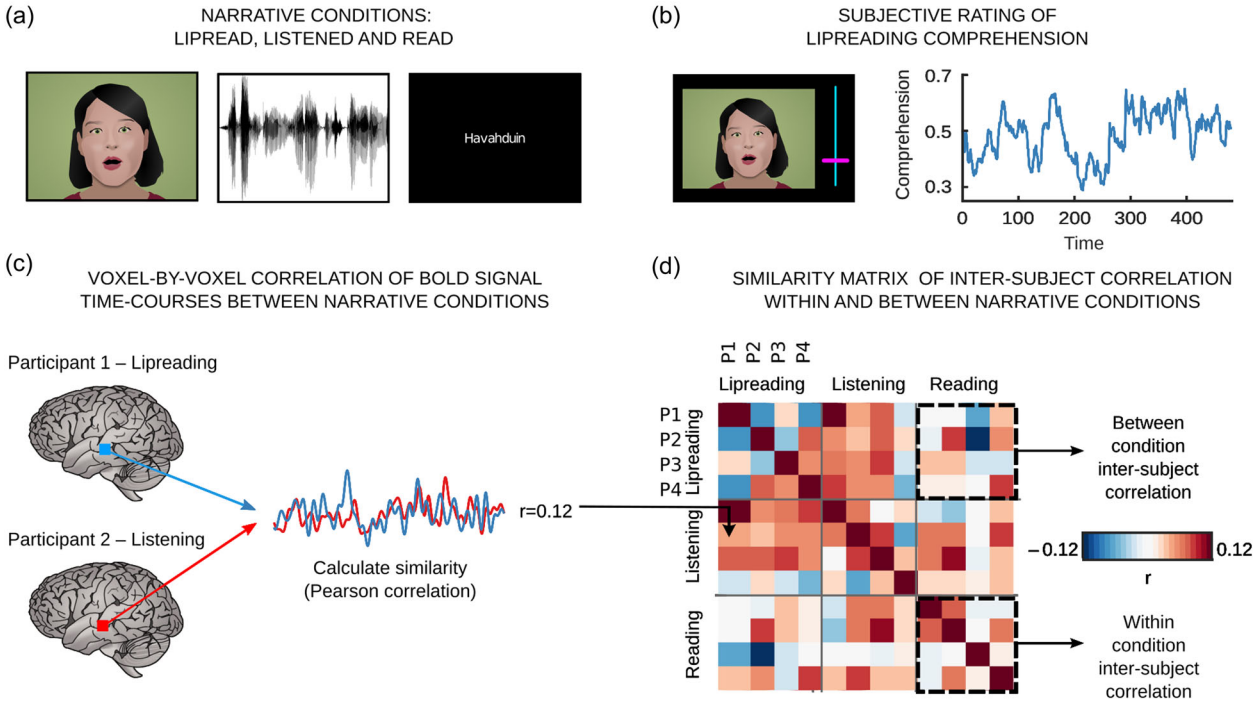

FIGURE 1 Illustration of the experimental design and between-narrative-type inter-subject correlation (ISC) (A cartoon face is used to conceal the identity of the speaker.) (a) The participants lipread, listened to, and read the same narrative. (b) After scanning, the participants lipread the narrative again and rated how well they comprehended what the speaker said. (c) We used the BOLD signal time courses of one participant during listening as a model for the activity time course of other participants in the same voxel to identify brain areas with similar time courses during lipreading: we computed the *r* statistics voxel-by-voxel between the signals and repeated the process for all participant pairs. (d) The similarity matrix of four participants depicting between- and within-condition inter-subject correlations from one voxel in the posterior temporal cortex. In contrast to previous studies, we analyzed the between-condition ISCs

activation in higher-level brain functions reflecting semantic and socio-emotional processing, for example (Lerner, Honey, Silbert, & Hasson, 2011; Nastase, Gazzola, Hasson, & Keysers, 2019; Regev et al., 2013; Wilson et al., 2008). Interestingly, Regev et al. (2013) found that the neural responses were both selective and invariant to spoken and written narratives. Here we address two research questions. First, to what extent lipreading naturalistic, connected speech engages neural mechanisms underlying lower- and higher-level psycholinguistic processing similar to listening and reading? Second, which brain areas are most relevant in lipreading connected speech?

In this study, our aim was (1) to identify cortical processing areas underlying lipreading naturalistic, connected speech, and (2) to evaluate to what extent the higher-level brain areas of listening and reading comprehension are recruited by lipreading the same narrative. Furthermore, we aimed (3) to identify brain areas where activity is predicted by better comprehension of lipread connected speech. To this end, participants with large inter-individual variation in lipreading skills had their brain activity measured with 3-T fMRI while they lipread an 8-min narrative from a silent video showing the speaker's face, listened to the same narrative without seeing the face of the speaker, and read a time-locked transcript of the narrative (Figure 1).

We hypothesized that lipreading a naturalistic narrative would involve cortical areas in the occipital, temporal, and parietal cortices. In addition to these low-level speech processing areas, we hypothesized that cortical areas responsible for lexical–semantic and higher-level linguistic processing in the parietal and frontal cortices would be involved, similarly to the way they are involved during listening to and reading the narrative. However, depending on how well the narrative is comprehended via lipreading, we expected that the activation in the latter areas might be spatially more restricted during lipreading than during listening to and reading the narrative. Finally, we hypothesized that better comprehension of the lipread narrative would be enabled by the same cortical processing areas in the temporal cortex that are used in coding heard speech.

2 | METHODS

2.1 | Participants

The volunteer participants were 31 healthy native Finnish-speaking females (mean age 30.9 years, range 20–49), all reported normal hearing. The narrative was related from the female first-person perspective to maximize the engagement of the participants with the narrative. There is also evidence suggesting that women and men may use different neural mechanisms, because females had more activity in the left auditory area while lipreading silently articulated numbers, although their number recognition accuracy was similar

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to males (Ruytjens et al., 2006; Ruytjens et al., 2007). Thus, to focus on individual differences in lipreading skills, we wanted to exclude between-sex variability in the current study. One participant's data was removed due to excessive head movement, and another's due to poor attention (eyes closed during scanning for approximately 3 min), resulting in a final sample of 29 participants. All participants were right handed (Edinburgh handedness inventory, Oldfield, 1971), and reported normal hearing and normal or corrected to normal (with contact lenses) vision, and no psychiatric or neurological disabilities. All participants signed informed consent forms, and received monetary compensation for their time. The study was approved by the research ethics committee of Aalto University and was conducted in accordance with the Helsinki Declaration for Human Studies.

We developed an online test for screening lipreading skills. We first recorded a female speaker with clear visual articulation, speaking 100 sentences that were translated into Finnish from the CID everyday sentences with varying length and sentence structure (Sims, 1975). Out of 100 sentences, 10 sentences that mimic the sentence structure of the Finnish language well were chosen for an online screening tool for participant recruitment. The online screening tool was distributed with information on participant requirements via the student mailing lists of local academies including speech therapy and sign language interpreters, and via the Finnish Federation of Hard of Hearing to reach individuals with higher lipreading skill but normal hearing. The participants were instructed to type out the words they recognized from the silent videos. The number of correctly recognized words out of a total of 57 words was used as each participant's lipreading skill score. For the current study, we then chose individuals with a large inter-individual variation in their lipreading skills, which were again tested on site with another test (Lonka, 1993).

2.2 | Stimuli and experimental design

The stimulus was a narrative (duration 7 min 54 s) related by a female speaker, portraying the events and thoughts that occurred during her day from a first-person perspective. The narrative conditions were: audio only (listening), visual only (lipreading a silent video of a speaking face), and text only (reading). The study design also contained an additional dimension: an unintelligible, gibberish version of the same narrative was presented to the participants as audio, visual, and text, so that the full experimental design consisted of six narrative conditions altogether. The gibberish was created by replacing consonants from each word of the original narrative with other consonants with a similar place of articulation, but the suffixes that indicated syntax remained unchanged. This resulted in a meaningless string of speech sounds that had very similar facial gestures and acoustic properties and structure (syntax) to the original narrative, but no content (semantics), sounding phonetically natural. The gibberish narratives always preceded the corresponding intact narratives, which were then presented in an order counterbalanced across the participants (Figure 1a). The full results related to the gibberish narrative will be reported separately. The speaker, who was chosen for her clear visual articulatory gestures, was video-recorded reading the narrative aloud with neutral prosody from a prompter in an acoustically shielded room using an additional Sennheiser EW 112P-G3-C microphone and a Canon XA10 video camera. The speaker had rehearsed reading the stories aloud at home, and the high phoneme–grapheme correspondence of the Finnish language eased the natural reading of the gibberish. Two external LED lights (Dyna-Core Elf2-DS LED) illuminated the speaker's face against a pale green background canvas to provide good visibility of articulatory movements.

The stimuli were edited with Matlab (MathWorks Inc.). The playback speed of the audio time course was dynamically adjusted to make the narrative types as similar as possible. Each paragraph's starting points were marked manually and the timing of the narrative audio tracks was matched by stretching the audio waveform to maximize the similarity of the loudness envelopes. Playback speeds were between 95% and 101% of the original speed during the paragraphs to make the changes in playback speed imperceptible to the participants (higher deviations from normal playback speeds were allowed during pauses between paragraphs to align the beginning of each paragraph between modalities). The root mean square (RMS) envelopes of the two stimuli after this transformation were highly correlated ($r = 0.57$ for RMS in 0.5-s windows; $r = 0.88$ after convolution with a canonical hemodynamic response function [HRF]). A similar action was performed on the video, and it was speeded up and slowed down similarly to the audio track using the same timestamps. The video frames were linearly interpolated according to the variable playback speed and resampled back to the original, constant frame rate. The playback speed was varied based on the manually marked timepoints at the end and beginning of each paragraph, so that the changes were implemented during natural pauses in the narrative.

The video recording was divided into visual and audio files. Silent videos of the speaker's face against a light green background were used in the lipreading condition. In the auditory condition, the narrative was presented with a blank screen in a similar shade of green to the background of the lipreading condition to avoid possible effects of still face image on speech perception (Calvert & Campbell, 2003). The participants' fixation was not controlled in order to approximate a natural listening condition.

We also created a written 614-word transcript of the narrative. The written words were presented centrally on the screen time-locked to each word of the original spoken narrative. When the duration of the words in the spoken narrative was very short, two or three words were presented simultaneously to maintain the timing and ensure the legibility of the text.

Stimuli were presented using Presentation software (Neurobehavioral Systems Inc., Albany, California, USA). The audio stimuli were played with MRI-compatible in-ear earbuds (Sensimetrics S14 insert earphones). In addition, MRI-safe protective earmuffs were placed over the earbuds for noise removal and safety. Sound intensity was adjusted for each participant to be loud enough to be heard over the scanner noise by playing example stimuli that were normalized to the same level as the auditory stories during a dummy EPI sequence

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

before the actual experiment. In the MRI scanner, the stimulus videos and texts were back-projected onto a semitransparent screen, using a Panasonic PT-DZ110XEJ projector. The viewing distance was 35 cm, the width and height of the projected face image was 380 pixels in height and 161 pixels in width (video display resolution 650 × 490), corresponding to approximately 10.9° vertical and 4.6° horizontal angle in the visual field.

Each stimulus presentation began with a fixation cross in the middle of the screen.

2.2.1 | Assessment of lipreading skills

Prior to scanning, each participant's individual lipreading skills were confirmed on-site with a sentence-based lipreading test, comprising 10 sentences of variable length (Lonka, 1993). The test stimuli were presented on a 17″ computer screen (resolution 1366 × 768) with a 40-cm viewing distance. The face image (height 9.6 cm, width 6 cm) on the screen corresponded to a 12.8° vertical and 8.5° horizontal angle in the visual field. The speaker repeated each of the 10 sentences twice: first by saying each one at a slower-than-usual speech rate, and then at a normal speech rate. Participants were instructed to write down the words they were able to recognize. The number of correctly recognized words (as a percentage out of a maximum of 50) provided the lipreading skill score. For example, if the sentence was "On Thursday we eat pancakes" and a participant wrote down "we eat," she got a score of 2 out of 5.

Immediately after the fMRI experiment, the participants watched the silent visual narrative again, and rated their comprehension (on a continuous scale from very poor to very good). The instruction was to estimate how well they had comprehended the narrative when it was first presented to them in the scanner. The subjective rating was performed after fMRI acquisition to prevent any influence on the neural activity during scanning, and the participants were not aware that they would be asked to rate their comprehension afterwards. The rating was conducted using a web-based dynamic rating tool (https://version.aalto.fi/gitlab/eglerean/dynamicannotations) (see Nummenmaa et al., 2012). The data were collected at 5 Hz. Participants used a mouse to move a small cursor up (good comprehension) and down (poor comprehension) on the right-hand side of the screen. The original scale of the rating was from 0 (very poor) to 1 (very good).

2.2.2 | MRI acquisition

fMRI was performed with a 3T Magnetom Skyra whole-body scanner (Siemens Healthcare, Erlangen, Germany) and a standard 20-channel receiving head/neck coil at the Advanced Magnetic Imaging (AMI) Centre of the Aalto NeuroImaging infrastructure at Aalto University School of Science. For functional scans, images were acquired using a T2-weighted echo planer imaging (EPI) pulse sequence: repetition time (TR), 1700 ms; echo time (TE), 24 ms; flip angle, 70°, each volume comprising 33 slices of 4 mm thickness with 0 mm gap; in-plane resolution was 3 × 3 mm^2 (field of view, 192 × 192 mm^2). A total of 295 volumes were acquired from which 13 volumes were discarded from each run to exclude brain activity during the viewing of the pre-stimulus fixation cross and after ending the actual stimulus presentation. Anatomical T1-weighted structural images were acquired at a resolution of 1 × 1 × 1 mm^3 (MPRAGE pulse sequence, TR 2530 ms, TE 3.3 ms, TI 1100 ms, flip angle 7°, 256 × 256 matrix, 176 sagittal slices).

To monitor the participants' attention, their eye gaze was recorded with an EyeLink 1000 eye tracker (SR Research, Mississauga, Ontario, Canada; sampling rate 1000 Hz, spatial accuracy 0.5°). Prior to the experiment, a nine-point calibration and validation was performed.

2.3 | Data analysis

2.3.1 | Pre-processing

The fMRI data were pre-processed with FSL software (www.fmrib.ox.ac.uk/fsl) using the BRAMILA parallel pre-processing pipeline (https://version.aalto.fi/gitlab/BML/bramila). First, after correcting for slice-timing during acquisition, the EPI volumes were spatially realigned to the middle scan by rigid body transformations to correct for head movements using FSL MCFLIRT. EPI and structural images were co-registered and normalized to each individual's anatomical scan (linear transformation with nine degrees of freedom with FSL FLIRT; structural images were cleared from non-brain tissues with FSL BET) followed by a linear transformation from anatomical to standard MNI template space (12 degrees of freedom; FSL FLIRT). Finally, BOLD time series were detrended (linear detrend), motion parameters were regressed out (24 parameters expansion) as well as average signals at deep white matter, ventricles and cerebro-spinal fluid (Power et al., 2014). Finally, a temporal high-pass filter with a cut-off frequency of 0.01 Hz was applied, followed by spatial smoothing with a Gaussian kernel of 8-mm FWHM. The anonymized data that support the findings of this study are available on reasonable request from the corresponding author. The data are not publicly available due to restrictions, for example, they contain information that could compromise the privacy of research participants.

2.3.2 | Similarity of brain activity between narrative types measured with inter-subject correlation of BOLD signal time courses within and between narrative conditions

The data were analyzed with voxel-wise comparison of the BOLD signal time courses triggered by the listened to, read, and lipread narratives. We estimated the similarity of the time series using inter-subject correlation analysis (ISC; Hasson et al., 2004), examined the temporal similarity of the signals in individual voxels. ISC is a model-free, data-driven method, which has been found suitable for more ecologically valid stimulus paradigms (Hasson et al., 2004; Kauppi et al., 2010) optimal for analyzing data acquired from experiments with complex

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

stimuli by quantifying the similarity of BOLD signals of different participants (Hasson et al., 2010; Kauppi et al., 2010; Pajula et al., 2012). The data were analyzed by comparing BOLD time courses over the narrative listening, and the ISC method used one participant's brain activity as a model to predict brain activity within another participant. Correlations were used as a similarity metric between the BOLD signals of a participant pair, and statistical inference was performed by identifying the brain area where the similarities between subject pairs were consistently larger than a null distribution. A recent paper has shown that the stimulus structure can have an effect on ISC (Lu et al., 2016). Therefore, as recommended, we controlled the possible effect of silent pauses by modeling the stimulus structure based on the presence of speech as in Lahnakoski et al. (2012). We used linear regression by fitting a vector with values of one at the time points of silences. The vector was convolved with the canonical HRF to take into account the delay of the BOLD response. The residuals from the linear model were used for subsequent analyses. ISC was calculated using the ISC toolbox (Kauppi et al., 2010). First, ISC matrices were obtained for each brain voxel by calculating all pairwise Pearson's correlation coefficients (r) of the voxel time courses across the subjects, resulting in 406 unique pairwise r-values for each voxel in each condition, and corrected for multiple comparisons using the threshold-free cluster enhancement (TFCE) option (Smith & Nichols, 2009). Our main interest lay in finding similarities in brain activity during processing of the same narrative presented via different means. Thus, we additionally performed between-condition ISC: BOLD signal time courses from each condition (listening, reading, and lipreading) were used as a model to identify brain areas with similar time courses in another condition (Figure 1d). This provided a measure for those brain areas during lipreading that responded similarly to the brain responses measured during listening or reading. Specifically, given a subject pair with one participant in condition *c1* and a second participant in condition *c2*, for each voxel we compute the average ***r*** statistic as the average of all pairwise correlations between the BOLD signal time series $s(t)$ at the voxel (formula 1). Unthresholded brain maps for all can be found in Neurovault (/collections/BJRMDQXU/).

$$\mathbf{r}_{\text{between-conditionsISC}} = 1/\mathrm{N}_{\text{pairs}} \Sigma_{i=1,\ldots,N;j=1,\ldots,N;i\neq j} \text{correlations}_{i,c1}(t), s_{j,c2}(t) \quad (1)$$

Due to the fact that the similarity values are not independent (each value depends on two participants), permutation-based statistical tests were performed. To obtain the statistical significance of the between-condition ISC and to control for multiple comparisons, whole brain permutations ($N = 5000$) were performed with FSL randomize. With the TFCE option, FSL randomize performs a permutation-based one sample t-test by flipping the sign of each pairwise correlation to construct surrogate data. TFCE finds "cluster-like" structures in the data without selecting an a priori cluster-forming threshold and produces corrected voxel-level statistics (Smith & Nichols, 2009).

2.3.3 | Predicting brain activity with lipreading skills and experienced comprehension

To reveal the brain areas related to lipreading skills and experienced comprehension of the lipread narrative, we predicted the ISC between listening and lipreading against the objective lipreading test score as well as subjective experienced of comprehension. We used the mean of the subjective experienced comprehension rating to mitigate possible moment-to-moment differences in the rating of the narrative during and after scanning (Jääskeläinen et al., 2022). For each voxel, the pairwise BOLD similarity between two participants in the two different conditions was modeled with both the objective lipreading test score and the subjective comprehension ratings of the participant in the lipreading condition in each pair. To test for significance of the association between similarities, we ran a Mantel test (Mantel, 1967), which is a non-parametric test where subjects' labels are shuffled at each permutation. To obtain statistical significance of the association and to jointly control for multiple comparisons, whole brain permutations ($N = 5000$) were performed with FSL randomize using TFCE. Additionally, to verify that the associations of lipreading comprehension and ISC were not driven by prior exposure to the narrative in the reading or listening conditions, the linear effects of presentation order were regressed out both from the pairwise ISCs and lipreading ratings and the analyses were repeated. To control for the linear effects, we calculated the mean ISCs and mean lipreading scores and ratings for each of the subgroups (i.e., those for who the lipreading condition was first, second, and third, respectively) and subtracted the mean values from each participants' data before recalculating the correlations. This is equivalent to including the groups as nuisance covariates in a typical GLM analysis employed in fMRI studies. Finally, because of the dependencies between the pairwise ISCs, we repeated the statistical tests with the participants modeled as exchangeability blocks in FSL randomise. Due to the limited sample size and potential limitations of this (Beta stage) feature in accounting for the dependencies in pairwise data, we present the results with both the standard permutations (assuming independence of data) and by modeling the within-subject dependencies.

3 | RESULTS

3.1 | Lipreading skills of participants

Participants were pre-screened with a web-based lipreading test, and prior to the fMRI scanning, their lipreading skills were estimated with a previously validated test, which consisted of 50 words in 10 sentences of variable length (Lonka, 1993). Results of the lipreading test is displayed in Figure 2. Inter-individual variability was very large with the proportion of correctly recognized words ranging from 6% to 100%

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

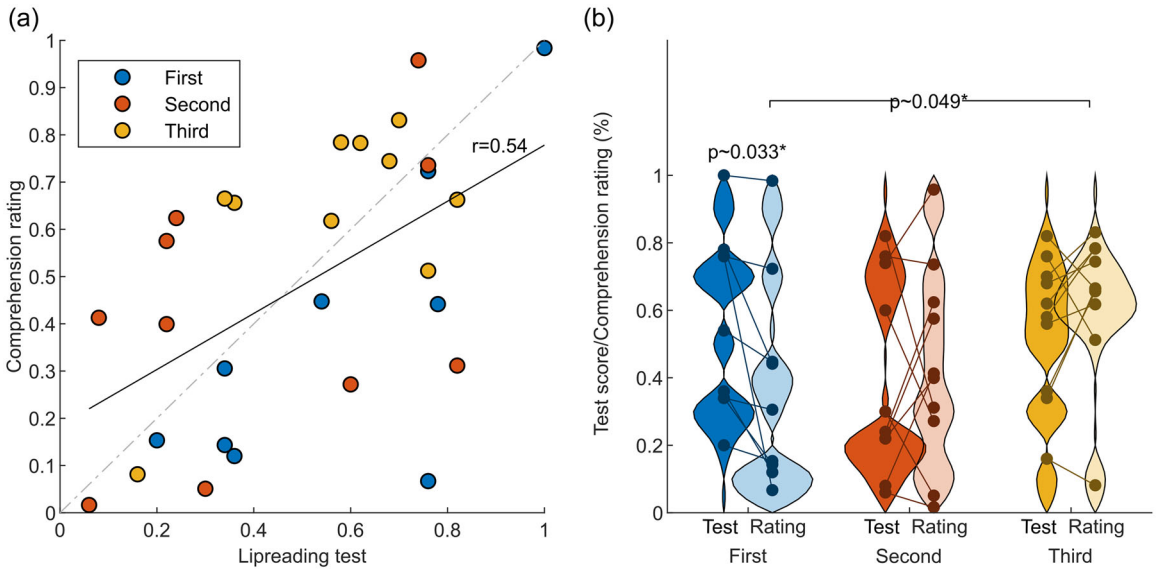

FIGURE 2 (a) Correlation between lipreading test score and subjective comprehension rating and (b) distribution of test scores and comprehension ratings as a function of presentation order. The ordinal position of the lipreading condition is depicted by the colors of dots and the distributions (blue, lipreading first, red, second, orange, third). In panel B, pale-colored distributions depict the comprehension ratings while the saturated colors show depict the test score distributions. Individual subjects' are shown as dots and test and rating data from the same participant are connected by lines

(mean = 50.7%; SD = 26). According to the Shapiro–Wilk normality test ($W = 0.94$, p-value = 0.082), the lipreading test results followed a normal distribution. After the fMRI, participants used a dynamic rating tool to provide a continuous subjective estimate of how well they could lipread the stimulus narrative. This estimate was different during different parts of the narrative (Figure 1b). The mean, calculated over the whole narrative, ranged from 0.016 to 0.98 (scale 0–1). Linear regression showed a significant ($r = 0.54$, $p < 0.003$) correlation between lipreading score and mean subjective rating of comprehension (Figure 2a), demonstrating that the lipreading score predicted comprehension of the narrative in the scanner moderately well. Because the presentation order of the different narrative modalities was randomized across participants, some participants listened to and/or read the narrative prior to lipreading. We examined the effect of the presentation order to the subjective ratings (Figure 2b) and found that the participants who had listened to and read the narrative before lipreading (yellow), had slightly better subjective comprehension than participants who lipread narrative first (blue; $p \sim 0.049$). However, examining the correlation at the level of individual participants (Figure 2b) revealed that the difference was due to significantly poorer subjective rating ($p \sim 0.033$) compared to lipreading test performance in the group who lipread the narrative first, particularly driven by two participants dropping from almost 80% lipreading test score to ~40% and ~20% comprehension rating, respectively (blue) (Figure 2b). Importantly, the correlation between test scores and comprehension ratings persisted after controlling for the presentation order ($r \sim 0.58$, $p = 0.0014$).

3.2 | Brain imaging results

3.2.1 | Inter-subject correlation during different narrative types

We first computed the ISC of voxel-wise BOLD time courses to identify shared hemodynamic activity during each of the narrative types. During lipreading (Figure 3, left), significant ISC was mainly found in cortical areas processing different aspects of visual information (the occipital gyri, supracalcarine cortex, cuneus, lingual gyri). Small clusters with significant ISC were also found in the precuneus (PRECUN), precentral gyrus (PCG), and cerebellum (peak values in Table S1).

When the participants listened to the narrative, significant ISC was found bilaterally in the auditory and peri-auditory cortices, in the superior temporal gyrus and sulcus (STG/S), in the middle temporal gyrus and sulcus (MTG/S), the temporal pole (TP), the IPL as well as the cerebellum (Figure 3 middle, peak values Table S2). In the left hemisphere, significant ISC was found in the superior frontal gyrus (SFG), PCG, and IFG. In the cortical midline, significant ISC was found bilaterally in the PRECUN, pCING, medial prefrontal cortex (mPFC), and visual cortex.

When the participants read the narrative, significant ISC was found bilaterally in visual areas in the cuneus (CUN) and around the calcarine sulcus (CAL), as well as in MTG, IPL, IFG, and CRUS I of the cerebellum. Therefore, the pattern of significant ISCs was rather similar to that during listening, excluding the auditory and periauditory areas, but including more widespread lateral and ventral visual areas and extensive bilateral activity of the occipital midline areas around the CAL instead.

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

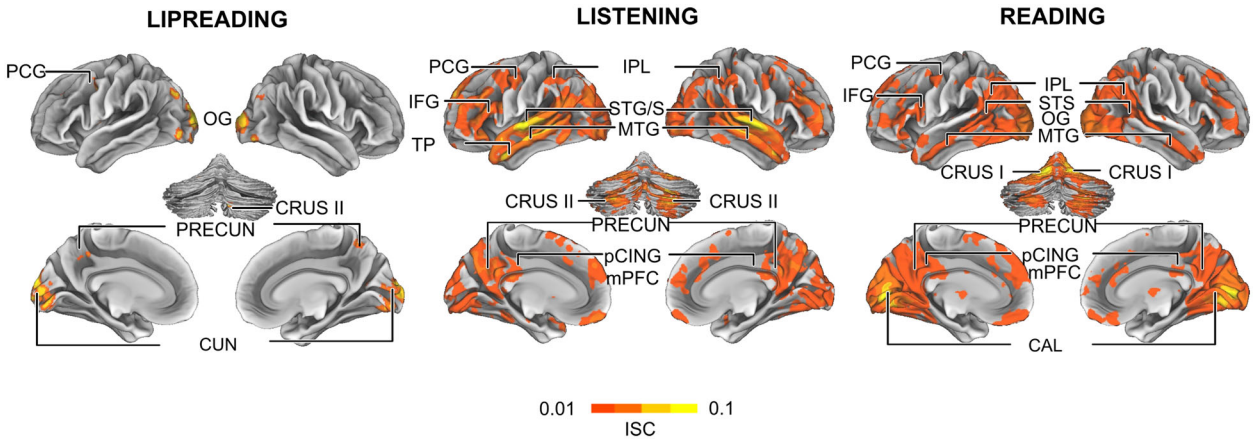

FIGURE 3 Significant inter-subject correlation (ISC) during lipreading, listening to, and reading the same narrative (cluster corrected $p < .05$). During lipreading, significant ISC was restricted to visual areas, and a few small clusters were found in the cuneus, precentral gyrus, precuneus and lingual gyri. During listening to and reading the narrative, similarity of activation extended from the temporo-parietal and frontal language areas to midline. CAL, calcarine sulcus; CRUS I, cerebellar lobule I; CRUS II, cerebellar lobule II; CUN, cuneus; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; M1, primary motor cortex; mPFC, medial prefrontal cortex; MTG, middle temporal gyrus; OG, occipital gyri; pCING, posterior cingulate; PRECUN, precuneus; STG/S, superior temporal gyrus/sulcus; TP, temporal pole

3.2.2 | Similarity of brain activity during lipreading, listening to, and reading the narrative measured with between-condition ISC

Next, we directly examined the similarity of brain activity between lipreading and listening conditions, and between lipreading and reading conditions, meaning that we calculated between-condition ISCs.

ISC between lipreading and listening (Figure 4a, peak values in Table S4, see also Figure S2) was significant in the middle and posterior STG/S, as well as along the whole bilateral MTG/S extending to the temporal pole (TP), left SMG, and right IFG. Furthermore, ISC was significant in two areas of the left primary motor cortex (M1) corresponding approximately to the hand (Yousry et al., 1997) and mouth representation areas (Fox et al., 2001). ISC was also significant in the bilateral PCUN, occipital midline areas around the calcarine sulcus (CAL) (extending to the lingual gyrus [LG] and the right fusiform gyrus [FG]), right-lateralized areas of the cerebellum, and left somatosensory cortex. By comparison (Figure S3), between-condition ISC for lipreading and listening to gibberish narratives was significant only in the left middle and posterior STG, and PCG as well as bilateral middle MTG, as well as in the cerebellum (CRUS I and II).

ISC between lipreading and reading was significant (Figure 4b, see also Figure S2 and peak values in Table S5) along the whole MTG/S extending to TP, left SMG, and bilateral lateral occipital gyri (LOG), starting from the occipital pole (OP), and extending around CS, comprising the primary visual cortex as well as other early visual areas. Furthermore, ISC was significant in the left and right cerebellum, PCUN, and cuneus, as well as in the anterior cingulate (ACC). By comparison (Figure S3), between-condition ISC for lipreading and reading gibberish narratives was significant around the LOG, bilateral IFG, and right posterior MTG.

ISC between reading and listening was significant in an extensive set of brain areas excluding the primary somatosensory cortex, ventrolateral visual cortex, subcortical areas (not shown in the figure), frontal orbital cortex, and right supra-temporal auditory cortex (Figure 4c; see also Figure S2 and peak values in Table S6). By comparison (Figure S3), between-condition ISC for reading and listening to gibberish narratives was significant, around the CAL and CUN as well as CRUS I and II in the cerebellum.

Figure 5 depicts the overlap (conjunction with minimal statistics; Nichols et al., 2005) of the three between-condition ISCs of Figure 4. These areas comprise the left middle MTG, and SFG, and bilaterally anterior MTG, posterior STS, TPJ, IFG and CS, PCUN, orbitofrontal cortex (OFC), and right cerebellum (VI and VIIA/Crus II).

For intact narratives, the observed neural similarity between lipreading and listening as well as lipreading and reading reflected higher-order information, although to a lesser extent than between listening and reading. The significant between-condition ISC of lipreading and listening to gibberish (Figure S3) was restricted to STG and STS, middle MTG and around CAL, CRUS I and II in the cerebellum. Similarly, significant ISC of lipreading and reading gibberish (Figure S3), was restricted to LOG and CUN with only small clusters in the MTG. The significant between-condition ISC during gibberish narratives did not include PRECUN, anterior MTG, and areas around the posterior STG/S, which were identified to have shared activation during all intact narrative types (Figure 5).

3.2.3 | Brain activity predicted by lipreading skills and subjective experienced comprehension

We then identified brain areas where ISC during lipreading and listening covaried with participants' lipreading skills by estimating how the

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

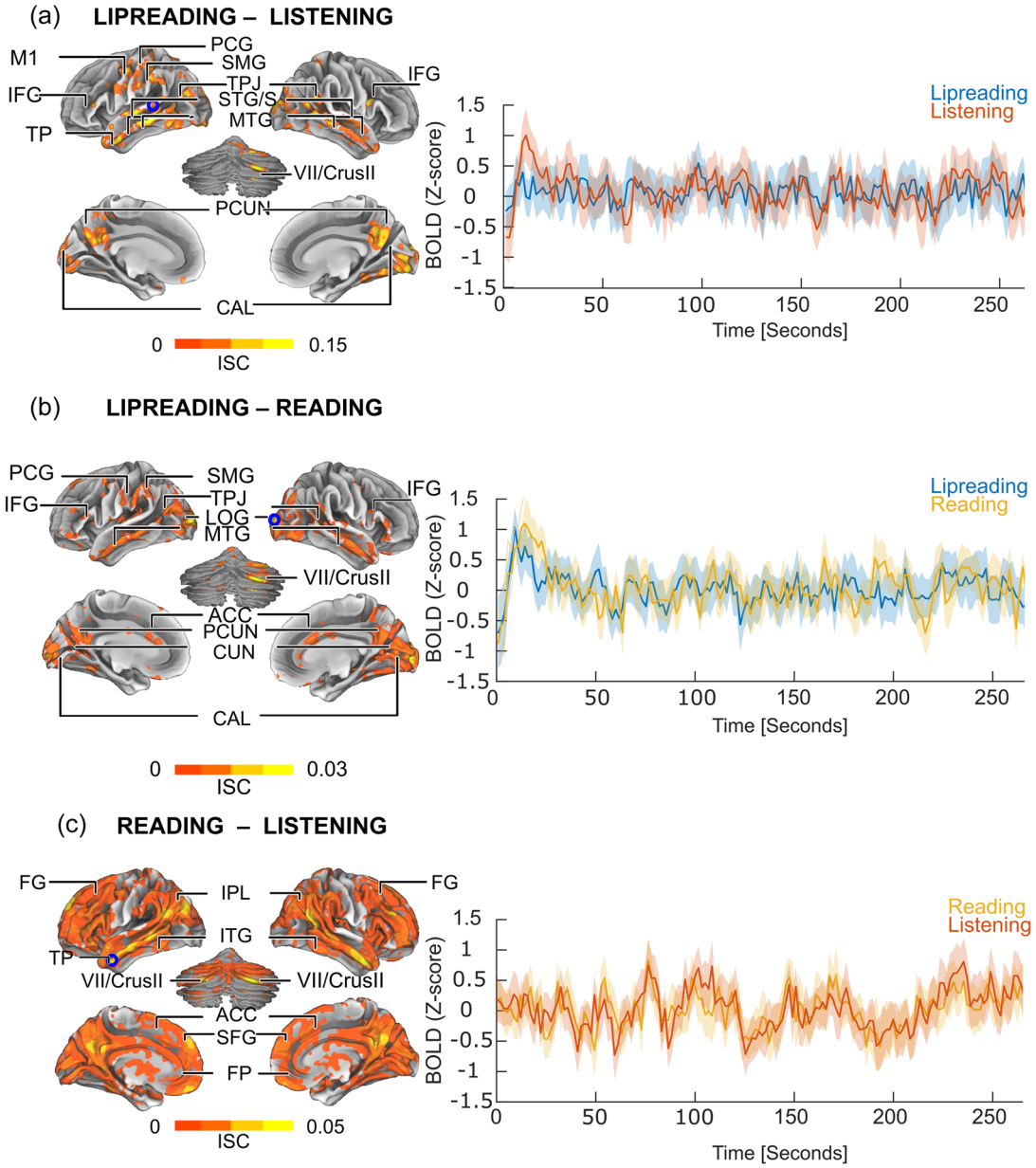

FIGURE 4 ISC between narrative types. Left: Brain areas showing significant ISC during processing of two different narrative types (permutation-based cluster correction, $p_{corrected} < .05$). Right: Time series and confidence intervals of peak values of BOLD signal strengths from the largest cluster (denoted with a blue circle) where the signals were significantly similar. (a) During lipreading and listening, maximum similarity (denoted with a circle) was centered at MNI coordinates $x = -62$, $y = -22$, $z = -4$. (b) During lipreading and reading, maximum similarity was centered at MNI coordinates $x = 22$, $y = -98$, $z = 12$. (c) During listening and reading, maximum similarity was centered at MNI coordinates $x = -54$, $y = 10$, $z = -30$. ACC, anterior cingulate cortex; CAL, calcarine sulcus; Crus II, cerebellar lobule II; CUN, cuneus; FG, inferior frontal gyrus; FP, frontal pole; IPL, inferior parietal lobule; ITG, inferior temporal gyrus; ITP, temporal pole; FP , frontal pole; M1, primary motor cortex; MTG, middle temporal gyrus; LOG, lateral occipital gyrus; PCG, postcentral gyrus; PCUN, precuneus; SFG, superior frontal gyrus; SMG, supramarginal gyrus; STG/S, superior temporal gyrus/sulcus; TPJ, temporo-parietal junction

objective lipreading test score and subjective experienced comprehension predicted the ISC between lipreading and listening (see Figure 4a). We did this by linearly regressing ISC first with the lipreading test score of the participant in each pair, and then the subjective lipreading comprehension rating of the participant in the lipreading condition in each pair. In both cases the data were thresholded with permutation-based cluster correction (TFCE corrected $p < .05$). Additionally, the mean effects of the presentation order were regressed out from the ratings and ISCs.

In a mantel test, the ISC was significantly associated with objective lipreading test score in the middle STG and Sylvian fissure close to the auditory cortex (Figure 6 white circles), and the results were significant in the left hemisphere. Instead, the ISC was significantly associated with experienced comprehension (Figure 6 orange) bilaterally in the

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

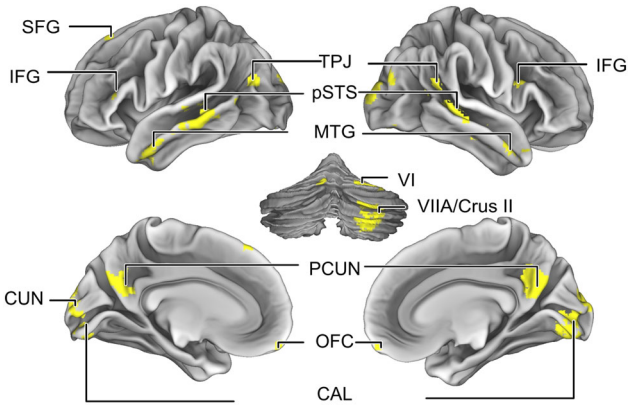

FIGURE 5 A binary mask indicating brain areas showing overlap for all between-condition ISCs: areas showing similarity were parts of the fronto-temporal (SFG, IFG, STG/S, MTG/S), parietal (TPJ), bilateral midline areas (PRECUN, OFC), and right cerebellum (VI and VIIA/Crus II. Crus II, cerebellar lobule II; IFG, inferior frontal gyrus; MTG, middle temporal gyrus; OFC, orbito-frontal cortex; PCUN, precuneus; SFG, superior frontal gyrus; STS, superior temporal sulcus; TPJ, temporo-parietal junction

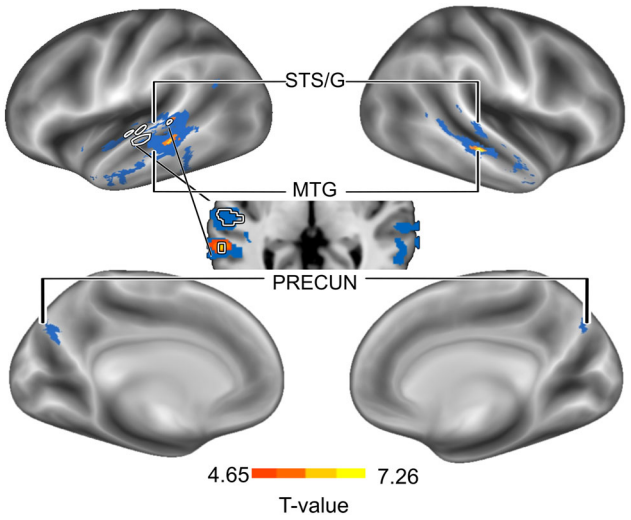

FIGURE 6 Brain areas showing higher ISC between lipreading and listening in participants with higher subjective rating of comprehension and lipreading test score (white circles). Higher pairwise subjective rating of lipreading comprehension predicted similar brain activity bilaterally in the middle STG/S and MTG when modeling within-subject dependencies (orange-yellow). Larger clusters were identified with the standard permutation test extending to bilateral superior and middle temporal areas as well as precuneus (blue), when no dependencies were modeled. Higher pairwise lipreading test score (white circles) predicted similar brain activity in clusters in the left middle and posterior STG/S and in the Sylvian fissure, when modeling within-subject dependencies. All analyses with permutation-based TFCE $p < .05$. MTG, middle temporal gyrus; PRECUN, precuneus; STS/G, superior temporal sulcus/gyrus

middle (peaks at $x = -60$, $y = -40$, $z = 3$; $x = 62$, $y = -24$, $z = 2$) and MTG (peaks at $x = -53$, $y = -37$, $z = -5$; $x = 54$, $y = -23$, $z = -7$), when modeling the within-subject dependencies in the permutation test. Using the standard permutation test that assumes independent samples (Figure 6 blue), these clusters extended to bilateral middle and posterior STG/S and middle to anterior MTG (not including the primary auditory cortex; see also Figure S2) with an additional cluster in the PRECUN. The superior temporal-cortical areas showed significant ISC when participants listened to speech, and mid-temporal cortical areas as well as PRECUN showed significant ISC in all conditions (see Figure 3).

4 | DISCUSSION

This study has illuminated the neural processes involved in lipreading naturalistic, connected speech in adults with normal hearing and varying lipreading skills by estimating similarity of brain activity during lipreading, listening to, and reading the same narrative. Significant between-condition ISC demonstrated similar processing of a lipread, listened to, and read narrative in the bilateral anterior MTG, IFG, posterior STS, TPJ, PCUN, as well as areas around Crus II of the right cerebellum (Figure 4). In addition, lipreading and listening to the narrative shared activity in the superior temporal cortex, predominantly in the left STG. Additionally, lipreading and reading activated similar areas in the visual cortex in the occipital lobe. Finally, better comprehension of connected speech via lipreading was associated with increased recruitment of the same temporo-cortical brain areas during lipreading as those during listening to speech, bilaterally (Figure 6).

4.1 | Brain areas underlying lipreading, listening to, and reading a narrative

When the participants listened to the narrative (Figure 3), within-condition ISC was significant bilaterally in the temporo-parietal (STG/S, MTG/S, IPL), frontal (SFG, PCG, IFG), as well as midline areas (PRECUN, mPFC). Previous research has reliably identified these areas as the extended linguistic network (Fedorenko et al., 2011bib29; Friederici, 2011; Hickok & Poeppel, 2007; Wilson et al., 2008). When the participants read the narrative (Figure 3), the pattern of significant within-condition ISC was largely similar to this network, but excluding the superior temporal cortex and including extensive bilateral activity of the visual cortices instead (see also Regev et al., 2013). In contrast to reading and listening, significant within-condition ISC during lipreading was restricted to visual cortices (Figure 3), with small clusters in the PCG, PRECUN, and lingual gyri most likely due to limited and idiosyncratic comprehension of the narrative. The message was unambiguous in both listening and reading conditions, explaining the strong similarity in brain activity in both within- and between-conditions ISCs. This

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

enabled us to use BOLD time courses from these conditions to model brain activity during lipreading.

Between-condition ISC (Figure 4) identified lipreading-related activation in the auditory speech processing pathways in the temporal cortex (Calvert & Campbell, 2003; Pekkola et al., 2005; Sams et al., 1991) and in the multimodal posterior STS/STG (Calvert & Campbell, 2003; Möttönen et al., 2002), also extending to more anterior and lateral areas (Bello et al., 2019). As a control, we performed between-condition ISC for lipreading and listening to gibberish narratives and it was also significant in these areas (Figure S3), suggesting similar use of the low-level speech processing areas. However, we did not find significant ISC in the primary auditory cortex in the left hemisphere (Figure 3 and Figure S2), in contrast to previous research using simple stimuli such as syllables and words (Calvert & Campbell, 2003; Pekkola et al., 2005; Sams et al., 1991). We regressed out the responses to auditory and visual transients, such as naturally occurring pauses in the narrative from the BOLD signals, as they are known to trigger activity in the primary sensory and other cortical areas (Lu et al., 2016). It is quite possible that auditory or visual transients, rather than linguistic stimulus features, caused primary auditory-cortical activation in the above-mentioned previous studies (Bernstein & Liebenthal, 2014; Bernstein et al., 2011; Paulesu et al., 2003), which may have been removed due to our efforts to remove these effects from the current data. We also found lipreading-related activation in the anterior parts of the STG and MTG (Figure 4), which have been implicated in mapping acoustic phonetic cues into lexical representations (Binder et al., 2000; Scott et al., 2000). This is in line with the recent suggestion that the STS contains pathways that process facial and vocal signals (Deen et al., 2020). Thus, the same brain areas that are involved in the sound-based coding of phonemes in the mid-STG and MTG may underlie processing of visual speech gestures. Lipreading connected speech also appears to be enabled by the convergence of auditory speech pathways (Calvert et al., 1999; Pekkola et al., 2005; Sams et al., 1991).

We also found lipreading-related activation in the primary motor cortex in the left hemisphere similarly to listening (Figure 3a), but not reading (Figure 3b). In line with previous studies with isolated, simple linguistic stimuli, brain areas related to speech production were activated during lipreading (Callan et al., 2014; Skipper et al., 2005, 2007; Watkins et al., 2003). Apparently, motor knowledge of articulatory gestures modulates auditory-cortical processing through reciprocal sensory–motor connections (Chu et al., 2013; Kauramäki et al., 2010; Skipper et al., 2005, 2007). Furthermore, motor knowledge of our own speech production is used in lipreading others (Calvert & Campbell, 2003; Nishitani & Hari, 2002; Paulesu et al., 2003; Watkins et al., 2003). This could also be related to the involvement of the somatosensory areas (Möttönen, Järveläinen, Sams, & Hari, 2005). Furthermore, activity in the inferior parietal regions around the temporo-parietal junction (TPJ; Figure 3a–c) could be related to accessing the stored motor representations of speech during visual speech gesture decoding (Calvert & Campbell, 2003), visuomotor transformations and planning intentions (Fogassi & Luppino, 2005; Shum, Shiller, Baum, & Gracco, 2011; Desmurget, & Sirigu, 2010). There is also evidence of a specific visuomotor pathway involving the middle part of the left MTG and frontoparietal motor areas, mediating speech motor control (O'Sullivan et al., 2017; Venezia et al., 2017). Interestingly, our results also demonstrate that the right cerebellum (cerebellar lobule VI and VII/Cruss II) shows a similar activity during all narrative types (Figure 3). The cerebellum has been found to contribute to motor (production), cognitive (also speech and language), and perceptual functions, and to temporal processing (Kotz & Schwartze, 2010; Kotz et al., 2014). Here, its activity during all narrative types suggests that it is involved in complex linguistic processing and dynamic processing of the narrative stimuli (Buckner, 2013; Kotz et al., 2014; Stoodley et al., 2012; Wang et al., 2013).

We expected the brain areas underlying high-level linguistic processing during lipreading to be shared between listening and reading because previous research has suggested that they are invariant to modality (Nguyen et al., 2019; Regev et al., 2013). We found that this was true for listening and reading (Figure 4) as they similarly activated brain areas related to high-level language processing (Booth, Wood, Lu, Houk, & Bitan, 2007; Fairhall & Caramazza, 2013; Fedorenko et al., 2011; Mahowald & Fedorenko, 2016; Regev et al., 2013). Based on the extensive similarity, even cognitive operations such as experiencing emotions, imagery related to the content of the story, social reasoning, and memory (Bar, 2007; Boldt et al., 2013; Fairhall & Caramazza, 2013; Nummenmaa et al., 2014) were most likely involved. However, lipreading connected speech elicited less activation in the higher-order brain areas (Figure 3 and Figure 4). Lipreading is much more difficult than listening to speech (Altieri et al., 2011; Files et al., 2015), and comprehension of the narrative was lower and less consistent (see also Skipper et al., 2005), leading to less synchronized activity than in the simulation results (Figure S1). Nevertheless, it can be speculated that activity in the bilateral posterior STG and MTG, PCUN, as well as in the left IFG likely reflects processing of higher levels of language, and context (Fairhall & Caramazza, 2013; Wilson et al., 2008), and possibly also mentalizing (Raichle et al., 2001; Simony et al., 2016; Utevsky et al., 2014; Yeshurun et al., 2017b), but more experimental evidence is needed to verify this interpretation. Moreover, some of the brain areas activating similarly to all narrative types (Figure 5) (IPL, lateral temporal cortex, and PRECUN) belong to the default mode network (DMN) (Raichle et al., 2001), which takes part in prospective and episodic memory and self-referential decision-making (Buckner & Carroll, 2006; Rugg & Vilberg, 2013). DMN connectivity has been shown to change during narrative listening, perhaps related to accumulating and integrating information over long time scales (Simony et al., 2016).

4.2 | Brain areas enabling successful comprehension via lipreading

We found temporal-cortex activity that covaried with objective lipreading score and subjective estimate of lipreading comprehension, suggesting that these cortical areas are most important for lipreading connected speech (Figure 6). The results in respect to different measures of lipreading were in large part aligned, but the effect of comprehension ratings were stronger resulting in larger clusters

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

bilaterally. These results are in line with previous findings on the significant relationship between left STG (Capek et al., 2008; Hall et al., 2005; Ludman et al., 2000) and STS (Blank & Kriegstein, 2013) activation during lipreading isolated sentences and performance in the lipreading test. Our data with naturalistic narrative and subjective experienced comprehension extend these findings, as we also found that lipreading comprehension predicted activity in the right STG, bilateral middle STS, and middle and anterior MTG as well as PRECUN.

Previous studies have identified STG and mid-posterior STS as the core cortical region for processing acoustic features of speech and in assessing sublexical structures such as phonetic characteristics of words and syllables (DeWitt & Rauschecker, 2012; Hickok & Poeppel, 2007; Liebenthal et al., 2005; Mesgarani et al., 2015). Therefore, these data suggest that lipreading shares bilateral phonological processing pathways with auditory speech (Hickock & Poeppel, 2016). However, there are findings suggesting that the STG is multisensory (Okada et al., 2013), and the set of activated neurons may be different during listening to speech and lipreading even within the same cluster of voxels (Beauchamp et al., 2004). According to Hauswald et al. (2018), auditory regions from the dorsal stream (Hickok, 2012; Hickok & Poeppel, 2007) map the visual speech signal into an absent acoustic signal. Extending this finding, our results suggest lipreading comprehension of connected speech covaries with the activity in the anterior and middle MTG (Figure 6), which reflects subsequent selection and integration of semantic information (Binder et al., 2009; Friederici, 2011; Price, 2010) and lexical-syntactic information retrieval (Rodd et al., 2015; Snijders et al., 2009). Furthermore, we found that the temporal-cortex activity related to subjective lipreading comprehension was bilateral (Figure 6). Naturalistic auditory speech stimuli such as narratives consistently elicit more bilateral brain activity than less complex stimuli (Honey et al., 2012; Huth et al., 2016; Jung-Beeman, 2005; Regev et al., 2013). The activity of the PRECUN, the functional core of the DMN, may be related active processing of the narrative stimulus (Simony et al., 2016).

In summary, we found activity in the superior temporal cortex showing convergence with sublexical auditory speech pathways, and in the middle temporal cortex indicating the involvement of lexical and semantic processing. Furthermore, this activity was associated with both performance in the objective lipreading test as well as subjective experienced comprehension and was, to some extent, activated similarly not only by lipreading, but also by listening to and, to some extent, by reading the narrative. Therefore, these data suggest that modality-specific visual speech processing does share pathways with auditory speech stimulus processing, and this enables successful lipreading comprehension.

4.3 | Methodological considerations

It is important to note that ISC is a tool that builds on stimulus processing-evoked activity so that brain areas where such extrinsic activity is prominent and similar in participants (e.g., visual-cortex activity during viewing a feature film) show strong ISC. But brain activity is always the sum of extrinsic and intrinsic activity, where intrinsic activity refers to all brain activity in the same brain area which is not related to the experimental manipulation. The fact that we found just a few areas with significant within-condition ISC during lipreading but extensive areas showing significant between-condition ISC indicates that lipreading-triggered brain activity was not strong or consistent enough to serve as a model for other participants' brain activity (see Figure S1 for simulation). However, as also indicated in Figure 4, listening to and reading the narrative activated the brain substantially more strongly, providing such a model for studying similar processing across narrative types, increasing the sensitivity of our analyses. Potentially, pattern analyses leveraging multivariate spatial patterns, such as inter-subject functional connectivity (Simony et al., 2016) across conditions, could reveal further information on the differences and similarities of functional networks participating in language comprehension in different modalities.

Further, the presentation sequence of three narrative types were randomized across participants and some participants listened to and/or read the narrative prior to lipreading. Therefore, for some participants, previous listening to the narrative might have influenced neural processing during reading or lipreading the narrative. This might have, due to adaptation and familiarity, also influenced the ISC between listening and reading, which, however, was extensive. In addition, reading or listening to the narrative before lipreading might have made lipreading easier for some participants. Although controlling for the presentation order does not significantly alter the results in the current manuscript, it is something to consider in future studies. In addition, using only one speaker might have modified the corresponding neural processing due to familiarization (Sanchez et al., 2013).

It has also been previously shown that voxel-by-voxel cross-subject comparisons are complicated because of considerable anatomical variability, which is especially true for areas that are sensitive to low-level visual speech features (Bernstein et al., 2011). These issues should be resolved in further research, which could study a set of good lipreaders several times each to calculate ISCs. This could also illuminate the possibly idiosyncratic processing strategies of skilled lipreaders, and would enable reaching higher than moderate correlation between the lipreading test and subjective estimation of comprehension. Furthermore, a study including male participants is needed in order to verify the generalizability of results across sexes (Ruytjens et al., 2006, 2007), as more attention would be drawn to the possible differences in lipreading processing between females and males (Bernstein, 2018). In the current study, we did not coerce eye gaze, but rather expected and allowed free focus shifts between the mouth and the central area of the face and eyes during lipreading (Paré et al., 2003), because previous studies with hearing adults have suggested that individuals' ability to lipread silent spoken sentences (Lansing & McConkie, 2003) and consonant–vowel–consonant clusters (Wilson, Alsius, Paré, & Munhall, 2016) did not correlate with mouth focus. However, the effect of direct mouth gaze on lipreading should be considered more carefully in future studies.

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5 | Conclusions

Our results show that in addition to activity in visual, auditory, and speech motor areas, lipreading a naturalistic narrative activated higher-level modality-independent brain areas similarly to listening to or reading the same narrative. However, activation during lipreading was much more limited than with listening and reading, probably mirroring the more limited comprehension during lipreading. Importantly, better comprehension of lipread connected speech was enabled by bilateral activity in the superior and middle temporal gyri and sulci, suggesting an overlap in neural processes that are involved in auditory coding of phonetic speech features.

AUTHOR CONTRIBUTIONS

SS: Planned the experiment, prepared stimuli, collected and analyzed data, prepared figures, wrote the manuscript as first author. JA: Planned the experiment, collected and analyzed data, wrote the manuscript. JML: Planned the experiment, prepared and edited stimuli, analyzed data, wrote the manuscript. MB-T: Planned the experiment, wrote the manuscript. EG: Developed methods, analyzed data, wrote the manuscript. IPJ: Planned the experiment, supervised collecting data, wrote the manuscript. UH: Planned the experiment, wrote the manuscript. MS: Principal investigator of the study, planned and supervised the experiment, wrote the manuscript.

ACKNOWLEDGMENTS

We would like to express our thanks to all participants. We are also grateful to the imaging facilities personnel in the AMI Centre at Aalto University School of Science, Espoo, Finland, and particularly to Marita Kattelus for her help. We would also like to thank Eila Lonka for her suggestions while planning the study, and for providing the lipreading test.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest, and that they have no competing financial interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author SS. The data are not publicly available due to restrictions, for example, their containing information that could compromise the privacy of research participants.

ORCID

Satu Saalasti https://orcid.org/0000-0003-1544-1215
Jussi Alho https://orcid.org/0000-0001-7853-4115
Juha M. Lahnakoski https://orcid.org/0000-0002-5223-7822
Enrico Glerean https://orcid.org/0000-0003-0624-675X
Iiro P. Jääskeläinen https://orcid.org/0000-0001-6001-6950
Uri Hasson https://orcid.org/0000-0002-3599-7168
Mikko Sams https://orcid.org/0000-0002-1992-484X

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/brb3.2869.

REFERENCES

Altieri, N. A., Pisoni, D. B., & Townsend, J. T. (2011). Some normative data on lip-reading skills (L). *The Journal of the Acoustical Society of America*, *130*, 1–4. https://doi.org/10.1121/1.3593376

Anderson, C. A., Wiggins, I. M., Kitterick, P. T., & Hartley, D. E. (2017). Adaptive benefit of cross-modal plasticity following cochlear implantation in deaf adults. *Proceedings of the National Academy of Sciences*, *114*(38), 10256–10261. https://doi.org/10.1073/pnas.1704785114

Auer, E. T., & Bernstein, L. E. (2007). Enhanced visual speech perception in individuals with early-onset hearing impairment. *Journal of Speech, Language, and Hearing Research*, *50*(5), 1157–1165. https://doi.org/10.1044/1092-4388(2007/080)

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289. https://doi.org/10.1016/j.tics.2007.05.005

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

Bello, J. P., García-Peña, C., Modroño, C., Hernández-Martín, E., Pérez-Martín, Y., Marcano, F., & González–Mora, J. L. (2019). Visual inputs decrease brain activity in frontal areas during silent lipreading. *PLoS ONE*, *14*(10), e0223782.

Bernstein, L. E. (2018). Response errors in females' and males' sentence lipreading necessitate structurally different models for predicting lipreading accuracy. *Language learning*, *68*, 127–158. https://doi.org/10.1111/lang.12281

Bernstein, L. E., Auer, E. T., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication*, *44*(1–4), 5–18. https://doi.org/10.1016/j.specom.2004.10.011

Bernstein, L. E., Jiang, J., Pantazis, D., Lu, Z.-L., & Joshi, A. (2011). Visual phonetic processing localized using speech and nonspeech face gestures in video and point-light displays. *Human Brain Mapping*, *32*(10), 1660–1676. https://doi.org/10.1002/hbm.21139

Bernstein, L. E., & Liebenthal, E. (2014). Neural pathways for visual speech perception. *Frontiers in Neuroscience*, *8*, 1–18. https://doi.org/10.3389/fnins.2014.00386

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*, 2767–2796. https://doi.org/10.1093/cercor/bhp055

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgovan, P. S. F., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, *10*(5), 512–528. https://doi.org/10.1093/cercor/10.5.512

Blank, H., & von Kriegstein, K. (2013). Mechanisms of enhancing visual–speech recognition by prior auditory information. *Neuroimage*, *65*, 109–118. https://doi.org/10.1016/j.neuroimage.2012.09.047

Boldt, R., Malinen, S., Seppä, M., Tikka, P., Savolainen, P., Hari, R., & Carlson, S. (2013). Listening to an audio drama activates two processing networks, one for all sounds, another exclusively for speech. *PLoS ONE*, *8*(5), 1–10. https://doi.org/10.1371/journal.pone.0064489

Booth, J. R., Wood, L., Lu, D., Houk, J. C., & Bitan, T. (2007). The role of the basal ganglia and cerebellum in language processing. *Brain Research*, *1133*, 136–144. https://doi.org/10.1016/j.brainres.2006.11.074

Bröhl, F., Keitel, A., & Kayser, C. (2022). MEG activity in visual and auditory cortices represents acoustic speech-related information during silent lip reading. *Eneuro*, *9*(3), 00–00. https://doi.org/10.1523/ENEURO.0209-22.2022

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Buckner, R. L. (2013). The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging. *Neuron*, *80*(3), 807–815. https://doi.org/10.1016/j.neuron.2013.10.044

Buckner, R. L., & Carroll, D. C. (2006). Self-projection and the brain. *Trends in Cognitive Sciences*, *11*(2), 49–57. https://doi.org/10.1016/j.tics.2006.11.004

Callan, D. E., Jones, J. A., & Callan, A. (2014). Multisensory and modality specific processing of visual speech in different regions of the premotor cortex. *Frontiers in Psychology*, *5*, 389. https://doi.org/10.3389/fpsyg.2014.00389

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*(5312), 593–596.

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*, *10*(11), 649–657. https://doi.org/10.1016/S0960-9822(00)00513-3

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

Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, J., … David, A. S. (2001). Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Cognitive Brain Research*, *12*(2), 233–243. https://doi.org/10.1016/S0926-6410(01)00054-4

Capek, C. M., MacSweeney, M., Woll, B., Waters, D., McGuire, P. K., David, A. S., Brammer, M. J., & Campbell, R. (2008). Cortical circuits for silent speechreading in deaf and hearing people. *Neuropsychologia*, *46*(5), 1233–1241. https://doi.org/10.1016/j.neuropsychologia.2007.11.026

Chu, Y. H., Lin, F. H., Chou, Y. J., Tsai, K. W. K., Kuo, W. J., & Jääskeläinen, I. P. (2013). Effective cerebral connectivity during silent speech reading revealed by functional magnetic resonance imaging. *PLoS ONE*, *8*(11), e80265. https://doi.org/10.1371/journal.pone.0080265

Crosse, M. J., Di Liberto, G. M., & Lalor, E. C. (2016). Eye can hear clearly now: Inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. *Journal of Neuroscience*, *36*(38), 9888–9895.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in cognitive sciences*, *15*(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

Deen, B., Saxe, R., & Kanwisher, N. (2020). Processing communicative facial and vocal cues in the superior temporal sulcus: Faces and voices in the STS. *Neuroimage*, *221*, 117191.

Desmurget, M., & Sirigu, A. (2012). Conscious motor intention emerges in the inferior parietal lobule. *Current opinion in neurobiology*, *22*(6), 1004–1011. https://doi.org/10.1016/j.conb.2012.06.006

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*, *109*(8), E505–E514. https://doi.org/10.1073/pnas.1113427109

Fairhall, S. L., & Caramazza, A. (2013). Brain regions that represent amodal conceptual knowledge. *Journal of Neuroscience*, *33*(25), 10552–10558. https://doi.org/10.1523/JNEUROSCI.0051-13.2013

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011b). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(39), 16428–16433. https://doi.org/10.1073/pnas.1

Files, B., Bosco, S. T., Jiang, J., & Bernstein, L. E. (2015). Visual speech discrimination and identification of natural and synthetic stimuli. *Frontiers in Psychology*, *6*, 878. https://doi.org/10.3389/fpsyg.2015.00878

Files, B. T., Tjan, B., & Bernstein, L. E. (2015). Visual speech discrimination and identification of natural and synthetic stimuli. *Frontiers in Psychology*, *6*(July), 1–18. https://doi.org/10.3389/fpsyg.2015.00878

Fogassi, L., & Luppino, G. (2005). Motor functions of the parietal lobe. *Current opinion in neurobiology*, *15*(6), 626–631.

Fox, P. T., Huang, a., Parsons, L. M., Xiong, J. H., Zamarippa, F., Rainey, L., & Lancaster, J. L. (2001). Location-probability profiles for the mouth region of human primary motor-sensory cortex: Model and validation. *Neuroimage*, *13*, 196–209. https://doi.org/10.1006/nimg.2000.0659

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, *91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "Cocktail party". *Neuron*, *77*(5), 980–991. https://doi.org/10.1016/j.neuron.2012.12.037

Golumbic, E. Z., Cogan, G. B., Schroeder, C. E., & Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party". *Journal of Neuroscience*, *33*(4), 1417–1426. https://doi.org/10.1523/JNEUROSCI.3675-12.2013

Haider, C. L., Suess, N., Hauswald, A., Park, H., & Weisz, N. (2022). Masking of the mouth area impairs reconstruction of acoustic speech features and higher-level segmentational features in the presence of a distractor speaker. *Neuroimage*, *252*, 119044. https://doi.org/10.1016/j.neuroimage.2022.119044

Hall, D. A., Fussell, C., & Summerfield, Q. (2005). Reading fluent speech from talking faces: Typical brain networks and individual differences. *Journal of Cognitive Neuroscience*, *17*(6), 939–953. https://doi.org/10.1162/0898929054021175

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40–48. https://doi.org/10.1016/j.tics.2009.10.011

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*, 1634–1640.

Hauswald, A., Lithari, C., Collignon, O., Leonardelli, E., & Weisz, N. (2018). A visual cortical network for deriving phonological information from intelligible lip movements. *Current Biology*, *28*(9), 1453–1459.e3. https://doi.org/10.1016/j.cub.2018.03.044

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. (2012). Computational neuroanatomy of speech production. *Nature reviews neuroscience*, *13*(2), 135–145. https://doi.org/10.1038/nrn3158

Hickok, G., & Poeppel, D. (2016). Neural basis of speech perception. *Neurobiology of language*, 299–310. https://doi.org/10.1016/B978-0-12-407794-2.00025-0

Honey, C. J., Thompson, C. R., Lerner, Y., & Hasson, U. (2012). Not lost in translation: Neural responses shared across languages. *Journal of Neuroscience*, *32*(44), 15277–15283. https://doi.org/10.1523/JNEUROSCI.1800-12.2012

Huth, A. G., Heer, W. A. D., Griffiths, T. L., Theunissen, F. E., & Jack, L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637

Jääskeläinen, I. P., Ahveninen, J., Klucharev, V., Shestakova, A. N., & Levy, J. (2022). Behavioral experience-sampling methods in neuroimaging studies with movie and narrative stimuli. *Frontiers in Human Neuroscience*, *16*, 813684.

Jääskeläinen, I. P., Sams, M., Glerean, E., & Ahveninen, J. (2021). Movies and narratives as naturalistic stimuli in neuroimaging. *Neuroimage*, *224*, 117445https://doi.org/10.1016/j.neuroimage.2020.117445

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences*, *9*(11), 512–518. https://doi.org/10.1016/j.tics.2005.09.009

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Kauppi, J.-P., Jääskeläinen, I. P., Sams, M., & Tohka, J. (2010). Inter-subject correlation of brain hemodynamic responses during watching a movie: Localization in space and frequency. *Frontiers in Neuroinformatics*, *4*, 5. https://doi.org/10.3389/fninf.2010.00005

Kauramäki, J., Jääskeläinen, I. P., Hari, R., Möttönen, R., Rauschecker, J. P., & Sams, M. (2010). Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. *Journal of Neuroscience*, *30*(4), 1314–1321. https://doi.org/10.1523/JNEUROSCI.1950-09.2010

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in cognitive sciences*, *14*(9), 392–399. https://doi.org/10.1016/j.tics.2010.06.005

Kotz, S. A., Stockert, A., Schwartze, M., & Kotz, S. A. (2014). Cerebellum, temporal predictability and the updating of a mental model. *Philosophical Transactions of the Royal Society B. Biological Scieences*, *369*(1658), https://doi.org/10.1098/rstb.2013.0403

Lahnakoski, J. M., Glerean, E., Salmi, J., Jääskeläinen, I. P., Sams, M., Hari, R., & Nummenmaa, L. (2012). Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. *Frontiers in Human Neuroscience*, *6*, 1–14. https://doi.org/10.3389/fnhum.2012.00233

Lansing, C. R., & McConkie, G. W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Perception & Psychophysics*, *65*(4), 536–552. https://doi.org/10.3758/bf03194581

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. *Journal of Neuroscience*, *31*(8), 2906–2915. https://doi.org/10.1523/jneurosci.3684-10.2011

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, *15*(10), 1621–1631.

Lonka, E. (1993). Hard-of-hearing adult and learning to speechread. *Thesis for the Degree of Licentiate*, 00–00.

Lu, K.-H., Hung, S.-C., Wen, H., Marussich, L., & Liu, Z. (2016). Influences of high-level features, gaze, and scene transitions on the reliability of BOLD responses to natural movie stimuli. *PLoS ONE*, *11*(8), e0161797. https://doi.org/10.1371/journal.pone.0161797

Ludman, C. N., Summerfield, a Q., Hall, D., Elliott, M., Foster, J., Hykin, J. L., Bowtell, R., Morris, & P, G. (2000). Lip-reading ability and patterns of cortical activation studied using fMRI. *British Journal of Audiology*, *34*(4), 225–230. https://doi.org/10.3109/03005364000000132

MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., Williams, S. C., Woll, B., & Brammer, M. J. (2000). Silent speechreading in the absence of scanner noise: An event-related fMRI study. *Neuroreport*, *11*(8), 1729–1733. https://doi.org/10.1097/00001756-200006050-00026

Mahowald, K., & Fedorenko, E. (2016). Reliable individual-level neural markers of high-level language processing: A necessary precursor for relating neural variability to behavioral and genetic variability. *Neuroimage*, *139*, 74–93. https://doi.org/10.1016/j.neuroimage.2016.05.073

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*(2), 209–220.

Mesgarani, N., Cheung, C., Johnson, K., Chang, E. F., & Francisco, S. (2015). *HHS Public Access*, *343*(6174), 1006–1010. https://doi.org/10.1126/science.1245994

Möttönen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research*, *13*(3), 417–425. https://doi.org/10.1016/S0926-6410(02)00053-8

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

Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience*, *14*(6), 667–685. https://doi.org/10.1093/scan/nsz037

Nguyen, M., Vanderwal, T., & Hasson, U. (2019). NeuroImage shared understanding of narratives is correlated with shared neural responses. *Neuroimage*, *184*, 161–170. https://doi.org/10.1016/j.neuroimage.2018.09.010

Nichols, T., Brett, M., Andersson, J., Wager, T., & & Poline, J. B (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, *25*(3), 653–660. https://doi.org/10.1016/j.neuroimage.2004.12.005. PMID: 15808966.

Nishitani, N., & Hari, R. R. (2002). Viewing lip forms: Cortical dynamics. *Neuron*, *36*, 1211–1220.

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., & Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences*, *109*(24), 9599–9604. https://doi.org/10.1073/pnas.1206095109

Nummenmaa, L., Saarimäki, H., Glerean, E., Gotsopoulos, A., Jääskeläinen, I. P., Hari, R., & Sams, M. (2014). Emotional speech synchronizes brains across listeners and engages large-scale dynamic brain networks. *Neuroimage*, *102*(P2), 498–509. https://doi.org/10.1016/j.neuroimage.2014.07.063

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*(6), e68959. https://doi.org/10.1371/journal.pone.0068959

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

O'Sullivan, A. E., Crosse, M. J., Di Liberto, G. M., & Lalor, E. C. (2017). Visual cortical entrainment to motion and categorical speech features during silent lipreading. *Frontiers in Human Neuroscience*, *10*, 679. https://doi.org/10.3389/fnhum.2016.00679

Pajula, J., Kauppi, J. P., & Tohka, J. (2012). Inter-subject correlation in fMRI: Method validation against stimulus-model based analysis. *PLoS ONE*, *7*(8), e41196. https://doi.org/10.1371/journal.pone.0041196

Paré, M., Richler, R. C., ten Hove, M., & Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the mcgurk effect. *Perception & psychophysics*, *65*(4), 553–567.

Park, H., Kayser, C., Thut, G., & Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *ELife*, *5*, 1–17. https://doi.org/10.7554/eLife.14521

Paulesu, E., Perani, D., Blasi, V., Silani, G., Borghese, N. A., De Giovanni, U., … Fazio, F. (2003). A functional-anatomical model for lipreading. *Journal of Neurophysiology*, *90*(3), 2005–2013. https://doi.org/10.1152/jn.00926.2002

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 3T. *Neuroreport*, *16*(2), 125–128.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, *1191*, 62–88. https://doi.org/10.1111/j.1749-6632.2010.05444.x

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 676–682. https://doi.org/10.1073/pnas.98.2.676

Regev, M., Honey, C. J., Simony, E., & Hasson, U. (2013). Selective and invariant neural responses to spoken and written narratives. *The Journal of*

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Neuroscience: The Official Journal of the Society for Neuroscience, *33*(40), 15978–15988. https://doi.org/10.1523/JNEUROSCI.1580-13.2013

Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An activation likelihood estimation meta-analysis. *Brain and Language*, *141*, 89–102. https://doi.org/10.1016/j.bandl.2014.11.012

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*, *17*, 1147–1153. https://doi.org/10.1093/cercor/bhl024

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current opinion in neurobiology*, *23*(2), 255–260. https://doi.org/10.1016/j.conb.2012.11.005

Ruytjens, L., Albers, F., Van Dijk, P., Wit, H., & Willemsen, A. (2006). Neural responses to silent lipreading in normal hearing male and female subjects. *European Journal of Neuroscience*, *24*(6), 1835–1844. https://doi.org/10.1111/j.1460-9568.2006.05072.x

Ruytjens, L., Albers, F., Van Dijk, P., Wit, H., & Willemsen, A. (2007). Activation in primary auditory cortex during silent lipreading is determined by sex. *Audiology and Neurotology*, *12*(6), 371–377. https://doi.org/10.1159/000106480

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., & Sing-Teh Lu, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, *10*(1), 141–145.

Sanchez, K., Dias, J. W., & Rosenblum, L. D. (2013). Experience with a talker can transfer across modalities to facilitate lipreading. *Attention, Perception, and Psychophysics*, *75*(7), 1359–1365. https://doi.org/10.3758/s13414-013-0534-x

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Clinical Trials*, *123*, 2400–2406.

Shum, M., Shiller, D. M., Baum, S. R., & Gracco, V. L. (2011). Sensorimotor integration for speech motor learning involves the inferior parietal cortex. *European Journal of Neuroscience*, *34*(11), 1817–1822. https://doi.org/10.1111/j.1460-9568.2011.07889.x

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications*, *7*, 12141. https://doi.org/10.1038/ncomms12141

Sims, D. (1975). The validation of the CID everyday sentence test for use with the severely hearing impaired. *Journal of the Academy of Rehabilitative Audiology*, *8*, 70–79.

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

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference FMRIB technical report TR08SS1. *Neuroimage*, *44*(1), 83–98.

Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J. A., Petersson, K. M., & Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: An fMRI study using word-category ambiguity. *Cerebral Cortex*, *19*(7), 1493–1503. https://doi.org/10.1093/cercor/bhn187

Stoodley, C. J., Valera, E. M., & Schmahmann, J. D. (2012). NeuroImage functional topography of the cerebellum for motor and cognitive tasks: An fMRI study. *Neuroimage*, *59*(2), 1560–1570. https://doi.org/10.1016/j.neuroimage.2011.08.065

Strelnikov, K., Rouger, J., Barone, P., & Deguine, O. (2009). Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants. *Scandinavian journal of psychology*, *50*(5), 437–444. https://doi.org/10.1111/j.1467-9450.2009.00741.x

Strelnikov, K., Rouger, J., Lagleyre, S., Fraysse, B., Deguine, O., & Barone, P. (2013). Deafness after adult cochlear implantation. *Brain*, *136*, 3682–3695. https://doi.org/10.1093/brain/awt274

Suess, N., Hauswald, A., Zehentner, V., Depireux, J., Herzog, G., Rösch, S., & Weisz, N. (2022). Influence of linguistic properties and hearing impairment on visual speech perception skills in the German language. *PLoS ONE*, *17*(9), e0275585. https://doi.org/10.1371/journal.pone.0275585

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

Summerfield, Q., Bruce, V., Cowey, A., Ellis, A. W., & Perrett, D. I. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 71–78. https://doi.org/10.1098/rstb.1992.0009

Utevsky, A. V., Smith, D. V., & Huettel, S. A. (2014). Precuneus is a functional core of the default-mode network. *Journal of Neuroscience*, *34*(3), 932–940. https://doi.org/10.1523/JNEUROSCI.4227-13.2014

Wang, D., Buckner, R. L., & Liu, H. (2013). Cerebellar asymmetry and its relation to cerebral asymmetry estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *109*(1), 46–57. https://doi.org/10.1152/jn.00598.2012

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*, 989–994. https://doi.org/10.1016/S0028-3932(02)00316-0

Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2008). Beyond superior temporal cortex: Intersubject correlations in narrative speech comprehension. *Cerebral Cortex*, *18*(1), 230–242. https://doi.org/10.1093/cercor/bhm049

Woodhouse, L., Hickson, L., & Dodd, B. (2009). Review of visual speech perception by hearing and hearing-impaired people: Clinical implications. *International Journal of Language & Communication Disorders*, *44*, 253–270. https://doi.org/10.1080/13682820802090281

Wilson, A. H., Alsius, A., Paré, M., & Munhall, K. G. (2016). Spatial Frequency Requirements and Gaze Strategy in Visual-Only and Audiovisual Speech Perception. *Journal of Speech, Language, and Hearing Research*, *59*(4), 601–615. https://doi.org/10.1044/2016_jslhr-s-15-0092

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017a). Same story, different story: The neural representation of interpretive frameworks. *Psychological Science*, *28*(3), 307–319. https://doi.org/10.1177/0956797616682029

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017b). Same story, different story. *Psychological Science*, *28*(3), 307–319. 095679761668202. https://doi.org/10.1177/0956797616682029

Yousry, T. A., Schmid, U. D., Alkadhi, H., Schmidt, D., Peraud, A., Buettner, A., & Winkler, P. (1997). Localization of the motor hand area to a knob on the precentral gyrus. A new landmark. *Brain*, *120*(Pt1), 141–157.

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Venezia, J. H., Fillmore, P., Matchin, W., Isenberg, A. L., Hickok, G., & Fridriksson, J. (2017). Perception drives production across sensory modalities: A Network for sensorimotor integration of visual speech. *Neuroimage*, *126*(949), 196–207. https://doi.org/10.1016/j.neuroimage.2015.11.038

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Saalasti, S., Alho, J., Lahnakoski, J. M., Bacha-Trams, M., Glerean, E., Jääskeläinen, I. P., Hasson, U., & Sams, M. (2023). Lipreading a naturalistic narrative in a female population: Neural characteristics shared with listening and reading. *Brain and Behavior*, *13*, e2869. https://doi.org/10.1002/brb3.2869

21579032, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/brb3.2869 by Aalto University, Wiley Online Library on [23/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License