Visual speech cues enhance neural speech tracking in right auditory cluster leading to improvement in speech in noise comprehension in older adults with hearing impairment

Vanessa Frei[1], Raffael Schmitt[2], Martin Meyer[1], and Nathalie Giroud[2]

[1]University of Zurich
[2]Universitat Zurich

March 1, 2023

Abstract

Comprehending speech in noisy situations is difficult and suffers with aging, hearing loss, and cognitive decline. Using visual speech cues improves perception and comprehension of speech in noise (SiN) in older adults with hearing impairment. In the current study we investigated neurofunctional correlates (i.e., neural tracking of speech) of the visual enhancement effect in a naturalistic setting in which older adults with hearing impairment (N=67) listened to sentences in babble noise in audio-only and audiovisual conditions and performed an intelligibility and a comprehension task, while EEG was recorded. Behaviorally, we found that visual speech cues improved intelligibility and comprehension of speech presented in babble noise and that particularly individuals with high working memory capacity benefitted. Furthermore, we found that the visual enhancement effect was accompanied by increased cortical tracking of the speech envelope particularly in individuals who were low performers in the audio-only condition (i.e., < 70% correct) leading to improved speech in noise comprehension in an audiovisual setting. Further, these low auditory-only performers showed deficient neural speech tracking in a right auditory topographical cluster, which improved significantly when visual speech cues were presented leading to more bilateral processing of speech. Overall, our results show that older adults with hearing impairment who have difficulties with speech in noise perception use visual speech cues to improve performance which is reflected in more bilateral processing of speech.

Visual speech cues enhance neural speech tracking in right auditory cluster leading to improvement in speech in noise comprehension in older adults with hearing impairment

Vanessa Frei[*1 6], Raffael Schmitt[1 6 7], Martin Meyer[2 3 4 5 7], Nathalie Giroud[1 3 6 7]

[1] Computational Neuroscience of Speech and Hearing, Dept. of Computational Linguistics, University of Zurich

[2] University of Zurich, University Research Priority Program (URPP) Dynamics of Healthy Aging, Zurich, Switzerland

[3] Center for Neuroscience Zurich, University and ETH of Zurich, Zurich, Switzerland

[4] Evolutionary Neuroscience of Language, Department of Comparative Language Science, University of Zurich

[5] Alpen-Adria University, Cognitive Psychology Unit, Klagenfurt, Austria

[6] International Max Planck Research School for the Life Course: Evolutionary and Ontogenetic Dynamics (LIFE)

[7] Language & Medicine Center Zurich, Competence Center of Medical Faculty and Faculty of Arts and Sciences, University of Zurich

[*] Corresponding author

Short title: Audio-visual speech cues enhance neural speech representation and improve speech in noise comprehension Keywords: Neural speech tracking; Speech in noise; Audio-visual speech; Age-related hearing loss; EEG

Email addresses:

VF: vanessa.frei@uzh.ch RF: raffael.schmitt@uzh.ch MM: martin.meyer@uzh.ch NG: nathalie.giroud@uzh.ch

Correspondence to: Vanessa Frei, Department of Computational Linguistics, Computational Neuroscience of Speech & Hearing, University of Zurich, Switzerland

Data availability statement: The auditory and audio-visual stimuli as well as the behavioral and neural data used in this study are publicly available in the study's Open Science Framework repository (https://osf.io/sb4mu/).

Funding: This work was supported by grants from the Swiss National Science Foundation (grant no. 105319_169964 to MM and grant no. PR00P1_185715 to NG).

Declaration of competing interest: The authors have no conflict of interest to declare.

Ethics approval statement: The study was conducted ethically, in compliance with the Declaration of Helsinki and approved by the local ethics committee (Cantonal Ethics Committee Zurich, application no. 2017-00284).

Date: 01. March 2023

Abstract

Comprehending speech in noisy situations is difficult and suffers with aging, hearing loss, and cognitive decline. Using visual speech cues improves perception and comprehension of speech in noise (SiN) in older adults with hearing impairment. In the current study we investigated neurofunctional correlates (i.e., neural tracking of speech) of the visual enhancement effect in a naturalistic setting in which older adults with hearing impairment (N=67) listened to sentences in babble noise in audio-only and audiovisual conditions and performed an intelligibility and a comprehension task, while EEG was recorded.

Behaviorally, we found that visual speech cues improved intelligibility and comprehension of speech presented in babble noise and that particularly individuals with high working memory capacity benefitted. Furthermore, we found that the visual enhancement effect was accompanied by increased cortical tracking of the speech envelope particularly in individuals who were low performers in the audio-only condition (i.e., $< 70\%$ correct) leading to improved speech in noise comprehension in an audiovisual setting. Further, these low auditory-only performers showed deficient neural speech tracking in a right auditory topographical cluster, which improved significantly when visual speech cues were presented leading to more bilateral processing of speech. Overall, our results show that older adults with hearing impairment who have difficulties with speech in noise perception use visual speech cues to improve performance which is reflected in more bilateral processing of speech.

Key points

- Visual speech cues can enhance intelligibility and comprehension of speech in noise.
- This visual enhancement effect is accompanied b increased cortical tracking of the speech envelope.
- Increased cortical tracking is reflected in more bilateral processing of speech.

Introduction

Difficulty understanding speech, particularly in challenging listening situations, is a common occurrence with ageing (Zekveld et al., 2011). An explanation for speech-in-noise (SiN) perceptual difficulties lies in the highly prevalent hearing loss in older adults, characterized by elevated pure-tone thresholds in the high-frequency range (Gates & Mills, 2005). Pure-tone hearing loss is typically driven by damage to the hair cells of the inner ear and accompanied by reduced synaptic connections between cochlear hair cells (Bowl & Dawson, 2019; Dubno et al., 2013) The resulting damage to the nerve fibers promotes inadequate encoding of the acoustic signal, which compromises speech perception (Liberman & Kujawa, 2017; Schneider & Pichora-Fuller, 2001).

Nevertheless, pure-tone hearing loss does not seem to fully account for SiN perceptual difficulties. This is indicated by individuals who have matching audiograms and who still vary significantly in their SiN perceptual performance (Anderson et al., 2011; Vermiglio et al., 2012) .Given such inter-individual variability in SiN perceptual abilities in older adults with similar audiograms, varying cognitive capacity should also be considered as a contributing factor (Humes et al., 2012). Cognition, particularly working memory, appears to partly bridge the gaps in auditory input created by reduced audibility, supporting sensory processing and improving reduced spectro-temporal precision (Anderson et al., 2013; Pichora-Fuller & Souza, 2003; Wong et al., 2009, 2010) . This rationale aligns with a theory of an integrated perceptual-cognitive system (Schneider & Pichora-Fuller, 2000), which assumes that a shared pool of resources is available for cognitive as well as perceptual demands.

Also, on the neural level there is individual variability on how speech in noise is being processed in auditory-related areas. In general, while listening to natural continuous speech in quiet or in noise, a synchronization between low frequency activity in the auditory cortex and temporal regularities of the speech signal can be observed. Such phase-locking of the neural response, particularly to the amplitude envelope, is often referred to as neural tracking of speech (Luo & Poeppel, 2007). This phenomenon is considered to be based on the rhythmicity of spoken language, whereby speech tracking is thought to facilitate comprehension by segmenting the continuous auditory signal into smaller, discrete acoustic units that provide input for further decoding steps (Giraud & Poeppel, 2012; Poeppel & Assaneo, 2020). In a recent study by Schmitt and colleagues (2022), it was shown that neural speech tracking serves as a neural basis for successful speech comprehension. Schmitt and colleagues (2022) and other studies have found a positive relationship between neural speech tracking and speech comprehension in both normal and hearing-impaired adults (Decruy et al., 2020; Gillis et al., 2021; Kurthen et al., 2021) . This positive relationship between speech tracking and comprehension highlights the importance of considering neural processing of speech in the context of SiN perceptual difficulties. The authors also found that a higher degree of pure-tone hearing loss was associated with a linear increase of tracking (quantified by greater cross-correlation coefficients), which is in line with several other studies (Decruy et al., 2020; Fuglsang et al., 2020; Gillis et al., 2021; Mirkovic et al., 2019), suggesting that the brain might compensate for lower intelligibility of speech (due to hearing impairment) by more precise neural tracking of speech.

There is still debate about the neurofunctional localization and lateralization of continuous speech processing in the brain. Poeppel (2003) suggests that the initial representation of speech is bilateral in auditory-related areas (i.e., in the superior temporal cortex Hickok & Poeppel, 2007) but is processed asymmetrically over time. Information from short temporal chunks (20 to 40 ms) is processed preferably in the left auditory regions and longer temporal windows (150 to 250 ms) in the right auditory regions. However, this asymmetric sampling in time (AST) hypothesis relates to younger individuals with age-typical auditory processing. Giroud and colleagues (2019) extend the AST model through the perspective of ageing. Following Poeppel and colleagues (2007), they argue that older individuals involve bilateral auditory regions to counteract age-related neurostructural decline (Giroud et al., 2019) and maintain sensitivity to different speech cues and temporal windows.

Previous studies on the influence of pure-tone hearing loss, cognition, and neural processing of speech on SiN perception or comprehension have used classical laboratory settings presenting only auditory speech information. In a natural conversation, however, the listeners often find themselves in an audio-visual setting, where the speaker's face and mouth movements are visible. Accordingly, there should be more emphasis on audio-visual speech and how it affects SiN perception and comprehension, particularly in older hearing-impaired individuals. While it is undisputed that visual speech cues improve speech perception, interestingly, the existing literature on individual differences in visual speech signal benefits for SiN comprehension is somewhat divided. Several studies suggest that older adults with high-frequency hearing loss in particular benefit from visual speech cues and show improved SiN performance (Altieri & Hudock, 2014; Hallam & Corney, 2014; Lidestam et al., 2014; Winneke & Phillips, 2011), while other studies clearly find a benefit to speech perception from audio-visual speech presentation, but independent of age and the degree of hearing loss and with considerable individual variability in the extent of this benefit (Başkent & Bazo, 2011; Rosemann & Thiel, 2018; Sommers et al., 2005; Tye-Murray et al., 2007). At the neural level, visual speech cues appear to enhance neural tracking of the speech envelope (Aller et al., 2022; Crosse et al., 2016; Micheli et al., 2018; Park et al., 2016) and restore early cortical tracking of speech presented in noise, complementing impaired auditory input (Atilgan et al., 2018; Crosse et al., 2015; Zion Golumbic et al., 2013). This audio-visual enhancement of neural speech tracking has also been observed in older individuals (Puschmann et al., 2019), although the relationship to SiN perception and comprehension has not yet been sufficiently established.

Thus, the aim of this study is to investigate the extent to which visual speech cues facilitate speech perception in noise in a large sample of older adults (N=67) with varying degrees of pure-tone hearing loss and cognitive capacity. More specifically, the relationship between neural speech tracking and SiN intelligibility and comprehension performance will be investigated, in auditory as well as audio-visual speech presentation. We hypothesize that audio-visual speech presentation is associated with improved SiN intelligibility and comprehension in older individuals with hearing impairment compared to auditory-only speech exposure. In addition, we hypothesize that cognitive capacity explains some of the individual variability in SiN intelligibility and comprehension. Furthermore, we hypothesize that the neural response is altered in the audio-visual condition such that neural speech tracking is greater in the audio-visual condition than in the auditory-only condition. Lastly, we assume that increased neural speech tracking is associated with better SiN intelligibility and comprehension.

Materials and methods

Sample

The study included 67 older participants ($M_{\text{Age}} = 72$, $\text{Range}_{\text{Age}} = 64\text{-}80$, $SD_{\text{Age}} = 4.3$, male = 41) who were in good health and did not show any cognitive impairments (Montreal Cognitive Assessment > 26 points; Nasreddine et al., 2005), pre-existing neurological or psychiatric conditions, speech or language disorders, neither were they professional musicians. The participants were all native speakers of Swiss German and had not learned a second language before age 7 years. Additionally, pure-tone hearing loss did not exceed 60 dB HL for octave frequencies between 0.5 and 8 kHz ($M_{\text{PTA}} = 42.58$, $\text{Range}_{\text{PTA}} = 31.85$ - 57, $SD_{\text{PTA}} = 6.33$) and pure-tone averages (PTA) were nearly symmetrical for both ears (< 15 dB interaural threshold difference). The sample was split into two groups, with 34 participants having experience using hearing aids (HA) for at least 12 months and 33 having never used hearing aids (nHA). Both groups were included in the study to represent a broad range of older individuals with hearing loss, and there were no significant differences found between the two groups in cognitive abilities, age, or hearing loss (table 1). All participants provided written informed consent and were compensated for their participation. The study was conducted ethically, in compliance with the Declaration of Helsinki and approved by the local ethics committee (Cantonal Ethics Committee Zurich, application no. 2017-00284).

Table 1. Comparability between individuals with and without hearing aids regarding age, hearing loss, and working memory capacity.

Assistance	HA (n = 34)	HA (n = 34)	nHA (n = 33)	nHA (n = 33)	$t(65)$	p
	M	*SE*	*M*	*SE*		
Age	72.97	4.15	71.12	4.42	1.765	.082
PTA	43.22	6.75	42.13	6.28	0.686	.495
WMC	49.35	8.79	49.33	8.76	0.009	.993

Audiometry

To determine participants' hearing loss, pure-tone thresholds were measured using a MATLAB-based probe-detection paradigm that has been described in detail in previous studies (Giroud et al., 2018; Lecluyse & Meddis, 2009). Stimuli were controlled via a sound card (RME Babyface Pro, RME, Haimhausen, Germany) and delivered through a linear frequency response loudspeaker (8030B Studio Monitor, Genelec, Iisalmi, Finland), with participants seated in an electrically shielded soundproof booth. After the measurement, the PTA was calculated by averaging the individual thresholds over the frequencies of 0.5, 1, 2, 4, and 8 kHz. Audiograms are visualized in Figure 1.

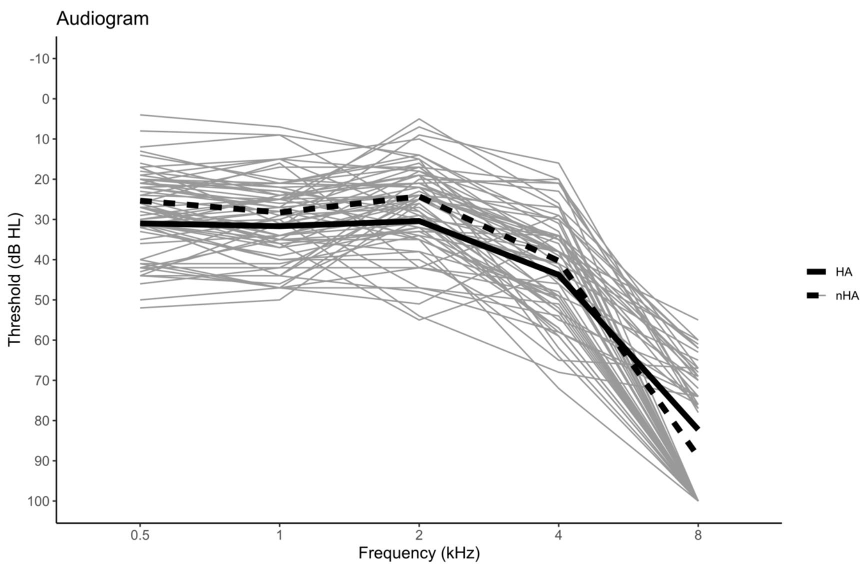

Figure 1. **Pure-tone audiometry** . The audiogram depicts individual pure-tone thresholds at frequencies between 0.5 and 8 kHz. There is no systematic difference between hearing-aid users (HA; group average shown in solid line) and non-hearing-aid user (nHA; group average depicted in dashed line).

Stimuli and experimental set-up

The presented material consisted of neutral, natural, continuous sentences (European Union shipping regulations), recorded by a trained female speaker. The sentences were presented in two different presentation conditions, one being purely auditory multi-talker babble noise (AB) and the other audio-visual multi-talker babble noise (AVB), where congruent lip movements of the speaker and the lower half of her face were visible. In each condition, a total of 30 sentences were presented at 70 dB, with babble noise consisting of a total of 8 randomly superimposed sentences, with silent gaps edited out. The sequence of conditions and the order of sentences were randomized across participants. To measure auditory perception at the behavioral level, an intelligibility task was presented after each sentence. Participants were presented with a short sound snippet (300 ms) and had to decide whether it was part of the sentence they had just heard. After every fifth sentence, a comprehension question (four-alternative forced choice) was also presented, which referred to the previous sentence and the participants were asked to indicate the correct answer on the keyboard. At the beginning, a practice round was conducted to make sure that the participants had understood the task. The conditions are illustrated in figure 2.

Figure 2. **Illustration of the stimuli presentation** . The two conditions differed in that audio-visual-babble (AVB) included a video sequence of the mouth and jaw movement whereas the audio-babble (AB) condition only contained auditory stimuli. Five sentences were presented, with an intelligibility question after each sentence. After every fifth sentence, a comprehension question was asked. There was a total of 30 items per condition.

Working-memory capacity

A computerized N-Back task was used, which is part of the Test of Attentional Performance assessment battery (Leclercq & Zimmermann, 2004). The N-Back paradigm visually presented digits that appeared in rapid succession on the screen. Participants were asked to press a key if the displayed digit had also been displayed three items/digits prior to the current one. Notably, participants did not have to press a key if the presented item did not match with the one 3 items prior. The completion of the assessment took a total of about 5 minutes. Instructions were given in standard German. T-scored N-Back accuracy was used as a primary indicator for WMC.

EEG recording and pre-processing

A continuous EEG was recorded with an 128 Ag/AgCl scalp electrode cap (BioSemi ActiveTwo, Amsterdam, The Netherlands) at a sampling rate of 512 Hz, while an online filter between 0.1 and 100 Hz was applied. Using the Fieldtrip toolbox (Oostenveld et al., 2011) , in MATLAB (R2021a, Mathworks), data were cut into trials of presented sentence lengths, which were re-referenced to Cz. A bandpass filter between 0.1 and 30 Hz (twopass Butterworth, 3rd order, Hamming window), a bandstop filter between 49-51 Hz (to minimize electrical interference artefacts) as well as visual inspection for bad segments was applied, before independent component analysis (Jung et al., 2000) was used to localize and remove artefacts (saccadic eye movements, eye blink and cardiac activity). Interpolation of noisy channels was done by applying spherical spline interpolation and average re-referencing. After resampling to 128 Hz and bandpass filtering between 2 and 8 Hz, data were baseline corrected and cut into segments of 3-5 seconds, excluding noise onset.

Envelope extraction

With the use of a gammatone filterbank, which is designed to mimic the compression response of the inner ear (Biesmans et al., 2017), speech envelopes were extracted. In this process, the raw acoustic signals are passed through a filter bank of 24 bandpass filters with an equivalent rectangular bandwidth of 1 and centre frequencies of 100 Hz to 4 kHz. The result of each filter was full-wave rectified and power-law compressed (i.e., the absolute value was raised to a power of 0.6). The generated sub-band envelopes were then combined to an average envelope and finally resampled, band-pass filtered between 2 and 8 Hz (optimal range for temporal modulations in the auditory system according to Poeppel & Assaneo, 2020) and cut to 3-5 seconds, matching the EEG signal.

Neural speech tracking

The synchronization between auditory cortex activity and speech envelope was quantified by cross-correlation. The cross-correlation function expresses the similarity between two signals with respect to a time lag, with values towards +-1 indicating perfect correlation and values towards 0 indicating no correlation. Cross-correlation coefficients were calculated for each trial of both conditions and the corresponding speech envelopes. We decided on three topographical clusters containing a right- (1-A26, 1-A27, 1-A28, 1-A29, 1-B6, 1-B7, 1-B8, 1-B9, 1-.B10, 1-B11, 1-B12) and a left-hemispheric temporo-parieto-occipital (1-A9, 1-A10, 1-A11, 1-A12, 1-A13, 1-A14, 1-A15, 1-A16, 1-D30, 1-D31, 1-D32) as well as a fronto-central (1-B31, 1-B32, 1-C1, 1-C2, 1-C3, 1-C4, 1-C11, 1-C12, 1-C13, 1-C20, 1-C21, 1-C22, 1-C23, 1-C24, 1-C25, 1-C26, 1-C27, 1-C32, 1-D1, 1-D2, 1-D3, 1-D4, 1-D12, 1-D13, 1-D18, 1-D19) cluster. The clusters were defined by the topography and time course of the grand average cross-correlation signal of both conditions and adapted from the approach of Schmitt and colleagues (2022) (Figure 4B). For a control condition, randomly selected speech envelopes were correlated with the EEG signal and compared with the experimental conditions. The comparison between both experimental conditions and the control condition was done for time lags ranging from 0 to 300 milliseconds since this time window contains two noticeable peaks in the cross-correlation function (Braiman et al., 2018; Horton et al., 2013; Zoefel & VanRullen, 2016). The average cross-correlation function of each condition and cluster can be found in figure 4A with significant time lags highlighted. These time windows provide the base for all further analyses related to neural speech tracking. Statistical tests were performed in R, version 4.1.2 (R Core Team, 2020).

Statistical analysis

Statistical analyses: Behavioral performance

To estimate the individual difference between conditions and the dependency between behavioral performance in the intelligibility and comprehension task and expected enhancement through the presentation of additional visual speech cues, a generalized linear mixed model (GLMM) was used. Two binomial models (for intelligibility and comprehension response) with a logistic link function from the *lme4* package (Bates et al., 2014) were estimated. Each model contained fixed effects of condition (categorical variable with two levels: audio-babble (AB), audio-visual babble (AVB)) and cross-correlation (continuous variable), an interaction between condition and cross-correlation, and the covariates cluster, working memory, age, and PTA. All continuous variables were z-standardized, and the binary response was coded with 1 for correct and 0 for incorrect answer. Additionally, each model contained by-subject and by-item random effects. A model with maximum random effect structure was estimated (Barr et al., 2013), while iteratively adjusting the structure to avoid overparameterization, non-convergence and a singular fit. The full model had the following specifications (using the formula notation in R):

*Response ~ condition * cross-correlation + cluster + working memory + age + pta + (condition | subject) + (condition * cross-correlation | item)*

Since interactions were included in the model, orthogonal sum-to-zero coding was used. The main effects were accordingly estimated at the grand-mean level and interpreted as such. Statistical interference was determined using likelihood ratio tests, comparing an encompassing model with a reduced model. The final model identified by this step-down process was fitted using restricted maximum likelihood. Degrees of freedom for t -tests and associated p -values were estimated using Satterthwaite's method for approximation.

Statistical analyses: Neural speech tracking

To estimate how the speech presentation may affect neural speech tracking and whether this relationship differs across clusters, a linear mixed model (LMM) with cross-correlation as a continuous outcome variable was used. The model contained fixed effects of condition and cluster and their interaction. Working-memory, age and PTA were included as covariates (z-standardized) and by-subject and by-item random effects were included. The maximum model had the following specifications (using the R formula notation):

*Cross-correlation ~ condition * cluster + working memory + age + pta + (condition*cluster | subject) + (condition * cluster | item)*

Since interactions were included in the model, orthogonal sum-to-zero coding was again used (Singmann & Kellen, 2019). Statistical interference was determined using likelihood ratio tests, comparing an encompassing model with a reduced model. The final model identified by this step-down process was fitted using restricted maximum likelihood. Degrees of freedom for t -tests and associated p -values were estimated using Satterthwaite's method for approximation.

Results

Behavioral data: Speech intelligibility and speech comprehension

Throughout initial visualization of the distribution of the behavioral data for both tasks (see Figure 3), we noticed a dichotomous distribution. When inspecting the performance in AB (without visual speech cues), a strong and a weak performance group (further referred to as “low” (purple) and “high” (orange) sub-sample) clearly emerged, which also differed regarding their change in performance when visual speech cues were additionally presented. Using a median split, the cut-off between the groups was 70% performance level (for both intelligibility- and comprehension performance). In other words, participants who answered less than 70% correctly in the audio-babble condition showed improvement in performance with visual speech cues, while participants who answered 70% or more correctly in the audio-babble condition showed no improvement with additional visual information potentially because they did not show signs of speech perception problems in AB. This observation of a dichotomous distribution and associated differential change in performance prompted further interest in possible distinguishing factors for these groups. As a result, we decided to analyze the groups separately and focus first on the “low” performing sub-sample for subsequent analyses, to examine a representative sample in which there was a statistical possibility of improvement in speech perception using additional visual speech information. The “high” performing sub-sample was analyzed separately as the results would have been biased due to the ceiling effect and thus no longer representative. Additional post-hoc analyses were then conducted that included the entire sample and focused specifically on the group differences between the “low” and “high” performing sub-samples, no longer focusing on potential performance change via visual speech cues, but rather on what causes these initial differences in the purely auditory condition. A visualization of the performance level in the comprehension and intelligibility task can be found in Figure 3.

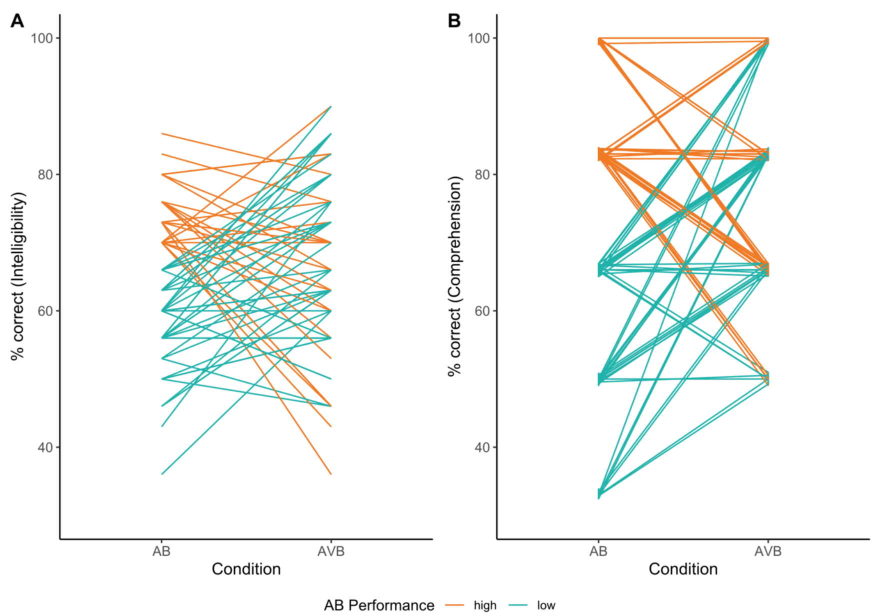

Figure 3. **Segregation of the sample into high and low performers.** A: Individual differences of intelligibility performance between audio-babble and audio-visual babble with color-codes for high and low performers in auditory babble. B: Individual differences of comprehension performance between audio-babble and audio-visual babble. Since there were only 6 comprehension questions per condition, there are 6 different percentile ranks in this figure. For a clear visual representation, the actual percentile ranks were jittered slightly across the participants.

Significant time windows for neural speech tracking

The cortical representation of speech was quantified by the cross-correlation between EEG time series and the amplitude envelope of the speech stimuli. As in previous work (Crosse et al., 2016; Puschmann et al., 2019; Zion Golumbic et al., 2013), the grand average cross-correlation functions showed a prominent peak at ~ 100 ms and a later one at around ~ 250 ms with an inversed polarity across the scalp (Figure 4). Pairwise comparisons between the cross-correlation functions and the control condition revealed several significant time lags (Bonferroni corrected; Figure 4A). In the left temporo-parieto-occipital electrode cluster significant lags were found from 62 to 164 ms for AB and from 55 to 150 ms for AVB. In the right temporo-parieto-occipital cluster, lags from 55 to 164 ms appeared to be significant for AB, whereat lags from 54 to 156 ms were significant for AVB. In the fronto-central cluster, time lags from 180 to 296 ms were significant for AB whereat lags ranging from 180 to 288 ms were significant for AVB.

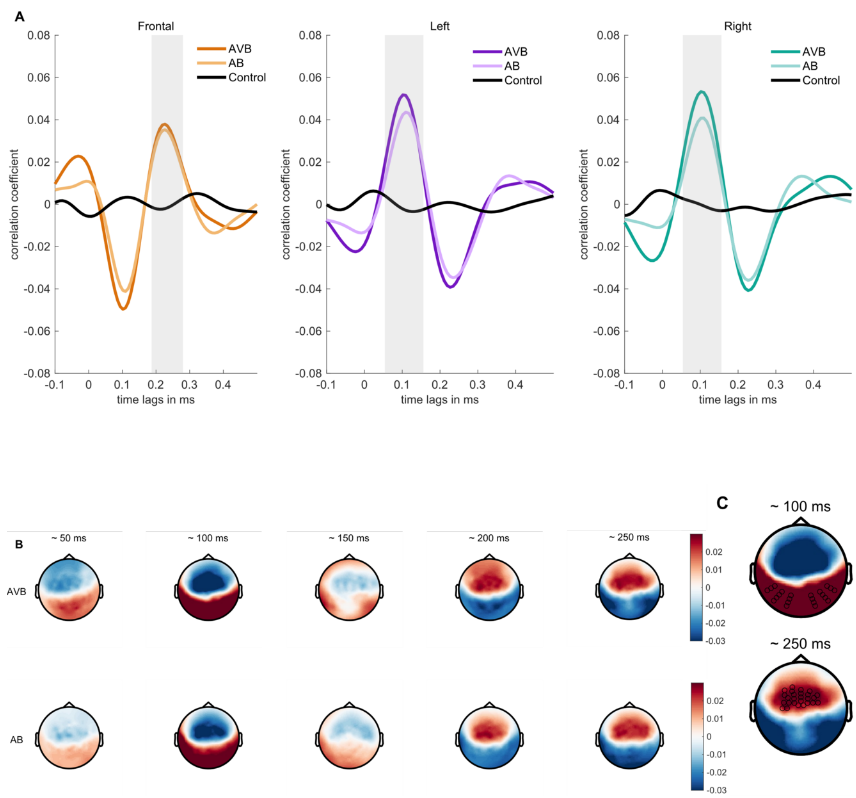

Figure 4: **Topographic distribution and time course of neural speech tracking** . A: Grand mean cross correlation functions of the frontal, the left and the right cluster. Significant time windows are marked as bars over the function. B: Topographic distribution and time course of the grand average cross correlation in both listening conditions from approximately 50 to 250 ms. C: Topographical distribution of the grand-average cross-correlation at the peaks at approximately 100 and 250 ms. Selected electrode clusters are marked with "*". Warm colors denote positive- and cool colors negative correlations.

Behavioral models for low performer sub-sample

The behavioral performances in an intelligibility and comprehension task were investigated with the reduced sample consisting of the "low" performers. Intelligibility (N = 39, $M_{\text{Age}} = 70.79$, $\text{Range}_{\text{Age}} = 64\text{–}80$, $SD_{\text{Age}} = 12.4$, $M_{\text{PTA}} = 41.87$, $\text{Range}_{\text{PTA}} = 31.9\text{–}54.25$, $SD_{\text{PTA}} = 9.47$) scores were significantly above chance in both conditions (One-sample t -tests against chance (50%): AB, $t(38) = 4.68$, $p < .001$; AVB, $t(38) =$

9.08, p < .001) with a mean intelligibility performance of 56.83% (± 1.46 % *SEM*) in AB and 67.77 % (± 1.96 % *SEM*) in AVB.

Comprehension (N = 39, M_{Age} = 71.67, $\text{Range}_{\text{Age}}$ = 64 - 80, SD_{Age} = 4.35, M_{PTA} = 42.9, $\text{Range}_{\text{PTA}}$ 31.8 – 55.25, SD_{PTA} = 6.68) scores were also significantly above chance in both conditions (One-sample t -tests against chance (50%): AB, t (38) = 1.75, p = 0.04, AVB, t (31) = 9.09, p < .001) with a mean comprehension performance of 53.92% (± 2.83% *SEM*) in AB and 73.43% (± 1.73 % *SEM*) in AVB. The focus was on determining whether there was a visual enhancement effect in performance and if this change in performance was associated with a change in neural speech tracking.

To estimate the statistical effect of condition (AB vs. AVB) on behavioral performance, two GLMMs were fitted. For the intelligibility task, the main effect of condition ($\Delta\chi^2(1)$ = 13.302, p< .001) was significant. The odds of answering correctly appeared to be significantly higher (on average 13%) in AVB compared to AB (AB - AVB: ΔOR = 0.58, *SE* = 0.15, p < .001). Additionally, WMC appeared to be a significant covariate ($\Delta\chi^2(1)$ = 9.060, p = .003), while the covariates cluster, age, and PTA seem not to be significantly associated with intelligibility performance (Table 2). Nevertheless, there seems to be an expected trend for PTA regarding intelligibility performance. With enhancing PTA the probability of answering correctly in an intelligibility task is declining, although not significantly. As for the relationship between performance and cross-correlation there was no significant effect revealed in the analysis ($\Delta\chi^2(1)$ = 0.303, p = .582). Furthermore, a likelihood ratio test indicated no statistically significant effect for the two-way interaction between condition and cross-correlation ($\Delta\chi^2(1)$ = 0.379, p = .538). The relationship between the covariates and the response as well as the main effects are visualized in figure 5.

Table 2. Estimated dependency between condition, cross-correlation, and intelligibility performance. Model configuration: Intelligibility response ~ condition + cross-correlation + cluster + working memory + age + PTA + (condition | subject) + (condition | trial).

Predictors	Estimate	*SE*	Odds Ratio	p
Intercept	0.63	0.08	1.87	<.001
Condition (AVB)	0.303	0.08	1.34	<.001
Cross-correlation	0.03	0.03	1.02	.235
Cluster (left)	-0.01	0.04	1.00	.956
Cluster (right)	-0.01	0.04	1.00	.965
Working memory	0.14	0.04	1.15	.001
Age	-0.03	0.05	0.97	.467
PTA	-0.08	0.05	0.93	.101

Regarding the comprehension task, the analysis revealed a significant main effect of condition ($\Delta\chi^2(1)$ = 20.001, p < .001). The odds of answering correctly appeared to be significantly higher (on average 24%) in AVB compared to AB (AB – AVB: ΔOR = 0.607, *SE* = 0.154, p < .001). There was no significant effect found for cross-correlation ($\Delta\chi^2(1)$ = 0.089, p = .765) and no statistically significant effect for the two-way interaction between condition and cross-correlation ($\Delta\chi^2(1)$ = 0.426, p = .514). Furthermore, the covariates cluster, WMC, age, and PTA seem not to be significantly associated with comprehension performance (Table 3). The relationship between the covariates and the response as well as the main effects are visualized in figure 5.

Table 3. Estimated dependency between condition, cross-correlation, and comprehension performance. Model configuration: Comprehension response ~ condition + cross-correlation + cluster + working memory + age + PTA + (condition | subject) + (condition | trial).

Predictors	Estimate	*SE*	Odds Ratio	p
Intercept	1.68	0.24	4.53	<.001

Predictors	Estimate	*SE*	Odds Ratio	*p*
Condition (AVB)	0.94	0.19	2.61	<.001
Cross-correlation	-0.02	0.04	0.96	.763
Cluster (left)	-0.01	0.04	0.99	.878
Cluster (right)	0.01	0.04	1.00	.924
WMC	0.38	0.19	1.45	.051
Age	0.11	0.20	1.12	.578
PTA	0.16	0.18	1.18	.387

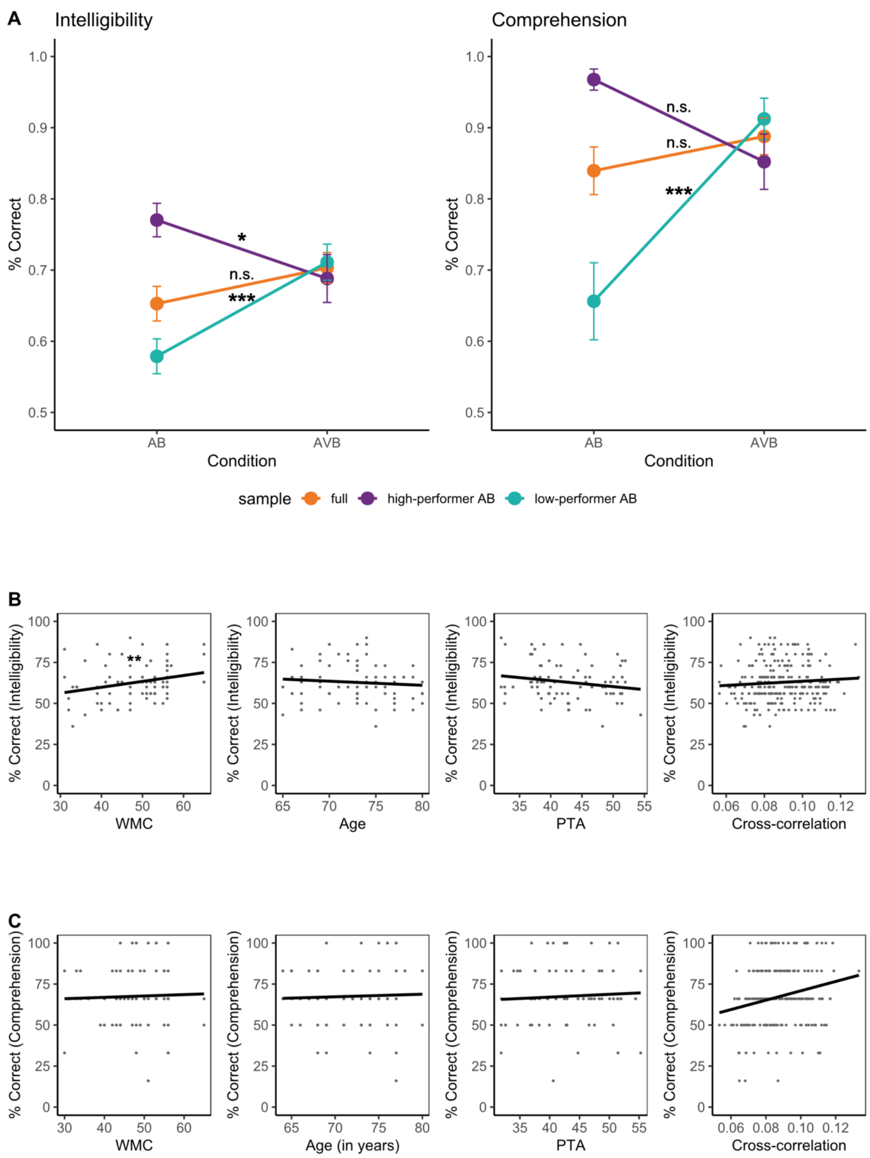

Figure 5. **Change of SiN intelligibility and comprehension performance through the presentation of additional visual speech cues** . A: In both, the intelligibility task and the comprehension task, there is a significant increase in SiN perception under AV speech presentation for the low-performer sub-sample, but

no significant change in performance for the full sample, nor for the high-performer sub-sample. B: In the low performing group, representative of the covariates, WMC is significantly associated with intelligibility performance, while age, hearing loss and neural speech tracking are not. C: None of the covariates seem to be significantly associated with comprehension performance.

Which factors account for SiN performance differences in AB?

As a further step, we analyzed the data for the full sample (N = 68,M_{Age} = 72.06, $\text{Range}_{\text{Age}}$ = 64–80,SD_{Age} = 4.36, M_{PTA} = 42.68, $\text{Range}_{\text{PTA}}$ = 31.85–57,SD_{PTA} = 6.49) including both performance states ("low" and "high"). The inclusion of the full sample allows us to further identify factors that contribute to the dichotomous distribution in the audio-babble performance. Since the high performer group suffered from a ceiling effect, we decided to only analyze the data from the audio-babble performance and exclude AV condition as a predictor. Including this condition would have produced non-representative results, since the participants who could not statistically improve by the additional visual speech cues more would either show stagnation or decline in their performance in the audio-visual babble condition. Intelligibility as well as comprehension scores were again significantly above chance in the full sample (one-sample t -tests against chance (50%) for intelligibility: AB, t (64) = 9.06, $p<$ .001; For comprehension: AB, t (64) = 6.88, $p<$ .001) with a mean intelligibility performance of 66.25% (± 1.79 % *SEM*) in AB, and a mean comprehension performance of 67.48% (± 2.54% *SEM*) in AB. Again, GLMMs were fitted, this time containing the full sample and specifying cross-correlation and WMC as main effects (since those two variables were of primary interest to explain variability in SiN performance in AB). For the intelligibility task a statistically significant effect for the two-way interaction between cross-correlation and working memory ($\Delta\chi^2(1) = 10.129$, $p=$ .002) was found. Visualized in Figure 6, the higher the neural tracking response was, the more likely a listener was to be correct in the intelligibility task, but only in participants with high WMC. With lower WMC this relationship diminished. A unit change in cross-correlation (i.e., +1 standard deviation) is associated with 13% increase in the odds of answering correctly in an intelligibility task with WMC held at 0 (e.g., the mean WMC). In a participant with higher WMC (i.e., 1 standard deviation above the mean) a unit change in cross-correlation is associated with 24% increase in in the odds of answering correctly (1.13^1 * 1.10), whereas in a participant with lower WMC (i.e., 1 standard deviation below the mean) a unit change in cross correlation is associated with 3% decrease in the odds of answering correctly (1.13^{-1}*1.10).

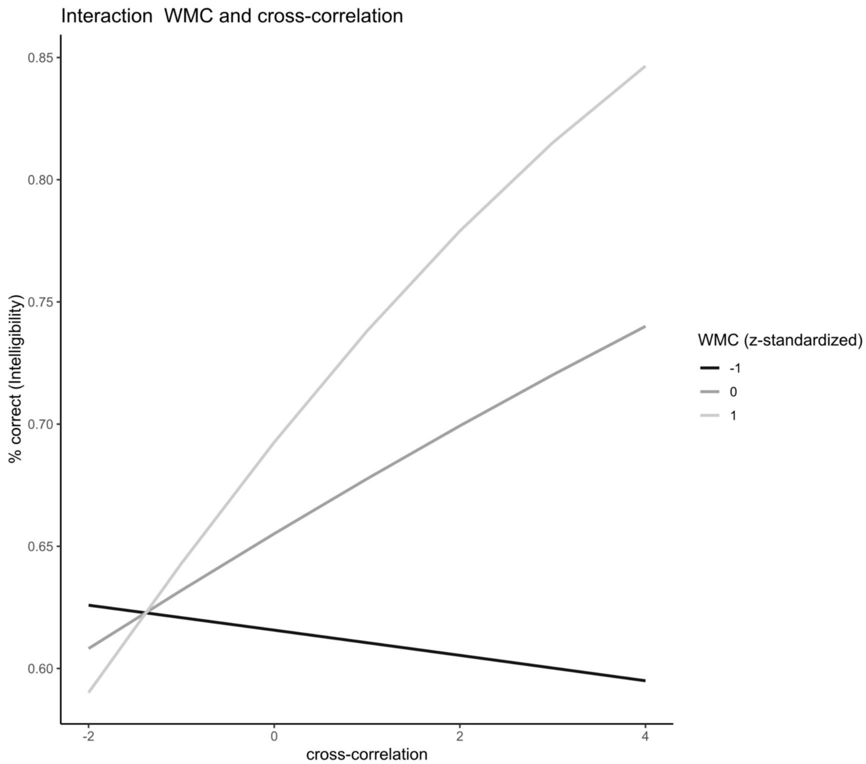

Figure 6. **Interaction between WMC and cross-correlation in the intelligibility task** . The higher the neural speech tracking response, the more likely it is to be correct in the intelligibility task, but only for participants with high WMC, with lower WMC this relationship seems to diminish.

Regarding the main effects, the analysis revealed no significant relationship between cross-correlation and performance ($\Delta\chi^2(1) = 2.471$,$p = .116$), whereas working-memory appeared to be significant($\Delta\chi^2(1) = 5.216$, $p = .022$). Furthermore, no significant associations were found for the covariates cluster, age, and PTA. The results of the additional analysis in the full sample for the audio-babble condition in the intelligibility task are summarized in table 4.

Table 4. Estimated dependency between cross-correlation, working-memory, and intelligibility performance for the full sample in audio-babble. Model configuration: Intelligibility response ˜ cross-correlation * working memory + cluster + PTA + age + (cross-correlation | subject) + (cross-correlation | trial).

Predictors	Estimate	*SE*	Odds Ratio	*p*
Intercept	0.63	0.11	1.88	<.001
Cross-correlation	0.10	0.06	1.10	.115
Working memory	0.10	0.06	1.10	.015
Cluster (left)	-0.01	-0.08	0.99	.926
Cluster (right)	0.01	0.12	1.01	.889
PTA	-0.05	0.05	0.95	.407
Age	-0.09	0.05	0.91	.107
xcorr:WM	0.12	0.04	1.13	.002

Regarding the comprehension task, the model for the full sample in AB revealed a significant main effect of working-memory ($\Delta\chi^2(1) = 7.491, p = .006$) but no significant main effect of cross-correlation ($\Delta\chi^2(1) = 0.022$, $p = .882$), indicating that more working-memory capacity is associated with better comprehension performance. There was no statistically significant interaction between cross-correlation and working memory ($\Delta\chi^2(1) = 0.001$, $p = .977$) and no significant association between the covariates cluster, and age (table 5)

Table 5. Estimated dependency between cross-correlation, working-memory, and comprehension performance for the full sample in audio-babble. Model configuration: Comprehension response ˜ cross-correlation + working memory + cluster + PTA + age + (cross-correlation | subject) + (cross-correlation | trial).

Predictors	Estimate	*SE*	Odds Ratio	*p*
Intercept	1.66	0.25	5.29	<.001
Cross-correlation	-0.001	0.05	0.99	.881
Working memory	0.65	0.23	1.93	.006
Cluster (left)	-0.01	0.05	0.99	.993
Cluster (right)	0.01	0.05	1.01	.943
PTA	0.08	0.22	1.08	.730
Age	0.14	0.23	1.15	.544

Neural data: neural speech tracking across clusters, conditions, and tasks

Neural speech tracking as a function of condition

The above-mentioned time windows (figure 4A) served as base to extract trial-level data by averaging cross-correlation coefficients for each trial over all significant time windows within each condition and cluster. To estimate the effect of conditions on neural speech tracking, a LMM was fitted. Additionally, we analyzed differences between different electrode clusters regarding neural speech tracking. This first model was fitted with the low-performer sub-sample to investigate an unbiased relationship between cortical envelope tracking and the speech presentation condition, as with the full sample a ceiling effect might have biased this relationship. Our analysis revealed a significant interaction between condition and cluster ($\Delta\chi^2(2) = 35.468$, $p<.001$). Post hoc pairwise comparisons were computed, revealing a significant increase in the right cluster compared to the left cluster (right − left: $\Delta\beta = 0.010$, $SE = 0.002$, $t = 4.41$, $p< .001$) as well as to the frontal cluster (right − fronto: $\Delta\beta = 0.014$, $SE = 0.002$, $t = 5.99$, $p < .001$) in AVB, whereas the left and frontal cluster did not show a significant increase in comparison (left − fronto: $\Delta\beta = 0.004$, $SE = 0.002, t = 1.56$, $p = .625$). Neural speech tracking increased significantly more under audio-visual speech cues in the right parieto-temporo occipital cluster compared to the other clusters. The model additionally revealed a significant main effect of cluster ($\Delta\chi^2(2) = 8.146$, $p = .017$) as well as a significant main effect for condition ($\Delta\chi^2(1) = 6.978$, $p = .008$), indicating a higher tracking response in AVB compared to AB. Individuals who have difficulty perceiving SiN in a purely auditory condition show a significant increase in speech envelope tracking by the use of additional visual speech cues, especially in the right temporo-occipital cluster which parallels with behavioral improvement in intelligibility and comprehension. The distribution of neural speech tracking across the clusters in AB and AVB for the low-performer sub-sample are visualized in figure 7A and 8A (accompanied by the distribution of neural speech tracking for the high-performer sub-sample in figure 7B and 8B). The results from the model are summarized in table 6.

Table 6. Estimated dependency between condition, clusters, and neural speech tracking. Model configuration: Cross-correlation ˜ condition * cluster + age + PTA + working memory + (condition | subject) + (condition | trial).

Predictors	Estimate	*SE*	*df*	*t*	*p*
Intercept	0.0919	0.0026	51.46	37.48	<.001

Predictors	Estimate	*SE*	*df*	*t*	*p*
Condition (AVB)	0.0034	0.0013	30.21	2.71	.011
Cluster (right)	0.0026	0.0009	4830.34	2.75	.006
Cluster (left)	-0.0003	0.0009	4807.65	-0.37	.715
Age	0.0004	0.0020	49.31	0.19	.853
PTA	-0.0010	0.0020	44.69	-0.61	.613
Working Memory	0.0012	0.0021	35.94	0.56	.579
Condition(AVB):Cluster(right)	0.0055	0.0009	4830.90	5.86	<.001
Condition (AVB):Cluster (left)	-0.0019	0.0009	4808.23	-1.97	.049

Neural speech tracking in full sample in AB condition

Similarly, for neural speech tracking, an additional model was calculated, including the complete sample, but only focusing on differences in the audio-babble condition. This additional model contained performance state as a categorical predictor (2 levels: "low", "high") to explicitly investigate if there were initial differences in the neural response between those two sub-groups. The model revealed a significant two-way interaction between performance state and cluster ($\Delta\chi^2(2) = 6.289$, $p = .043$) showing significantly less tracking in the right cluster compared to the other clusters for the low performer group, whereas the high performer group did not show such a significant difference in tracking between the cluster in the purely auditory condition. Post hoc pairwise comparisons were computed, revealing significantly less speech tracking for the low performer group in the right cluster compared to the frontal cluster (right-frontal: $\Delta\beta = -0.007$, $SE = 0.002$, $t = -3.24$, $p = .015$) but not significantly less tracking in the right cluster compared to the left cluster (right − left: $\Delta\beta = -0.005$, $SE = 0.002$, $t = -2.52$, $p = .113$) nor the left cluster compared to the frontal cluster (left − fronto: $\Delta\beta = -0.001$, $SE = 0.002$, $t = -0.777$, $p = .972$). Furthermore, no significant main effects of performance state ($\Delta\chi^2(1) = 0.877$, $p = .349$) nor cluster ($\Delta\chi^2(2) = 5.513$, $p = .063$) were found. The covariate working memory, age and PTA showed no significant association with neural speech tracking as well (table 7). These results indicate that there seems to be a difference regarding the neural response between both sub-groups in the purely auditory setting, but only regarding the right cluster. This relationship is also shown in figure 7.

Table 7. Estimated dependency between performance state, clusters, and neural speech tracking for the full sample in audio-babble. Model configuration: Cross-correlation ~ performance state * cluster + PTA + age + working memory + (1 | subject) + (performance state | trial).

Predictors	Estimate	*SE*	*df*	*t*	*p*
Intercept	0.0881	0.0018	73.05	47.78	<.001
Performance(low)	-0.0017	0.0018	67.68	-0.95	.347
Cluster (right)	-0.0018	0.0009	4127.06	-2.01	.078
Cluster (fronto)	0.0014	0.0009	4131.63	1.59	.112
PTA	0.0017	0.0015	67.92	1.11	.274
Age	0.0001	0.0016	67.82	0.04	.968
Working Memory	0.0010	0.0016	66.71	0.63	.529
Performance(low):Cluster(right)	-0.0023	0.0009	4127.01	-2.56	.013
Performance(low):Cluster(fronto)	0.0013	0.0009	4131.09	1.94	.137

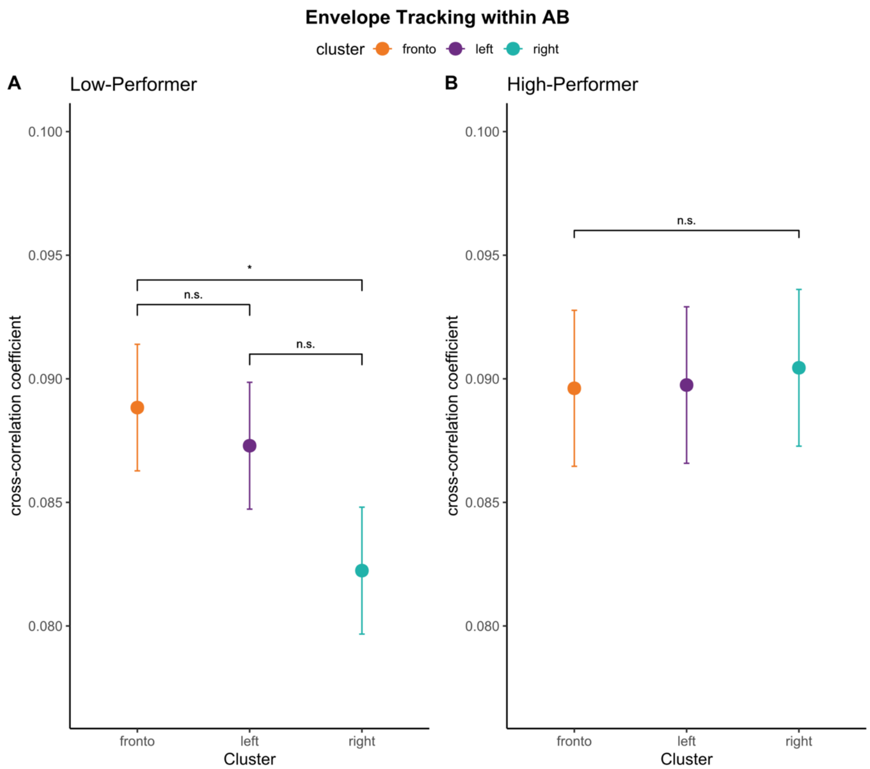

Figure 7. **Differences in neural speech tracking in the audio-babble condition between the two sub-samples** . A: There is no significant difference in the neural response between the frontal and the left nor the left and the right cluster, but a significant difference between the right and the frontal cluster for the low-performer sub-sample. B: There is no significant difference in envelope tracking between any of the three clusters for the high-performer sub-sample in the auditory-babble condition.

For the sake of completeness, another model was conducted, including both conditions (AB and AVB). The decision for this additional analysis resulted from the interest in how the tracking response changes over visual speech cues. The model revealed a significant three-way interaction between condition, performance state and cluster ($\Delta\chi^2(2) = 12.941$, $p = .002$) indicating a significant increase in tracking in the right cluster for the low-performer group in the AVB condition compared to the other clusters, whereas no such effect was found for the high-performer group. Post hoc pairwise comparisons were computed, revealing significantly more speech tracking for the low-performer group in the right cluster in AVB compared to the frontal cluster (right-frontal: $\Delta\beta = 0.014$, $SE = 0.002$, $t = 5.89$, $p < .001$) as well as significantly more tracking in the right cluster compared to the left cluster (right − left: $\Delta\beta = 0.010$,$SE = 0.002$, $t = 4.31$, $p = .001$). The results are visualized in figure 8. Ultimately, however, these results must be interpreted with caution, as the neural response of the “high” performer group in the audio-visual condition is not fully representative due to the ceiling effect and the neural response of the “low” performer group was taken in comparison to the “high” performer group. Nevertheless, the trajectory of the neural response is clearly investigated in the initial model by analyzing the low performer group, which is representative in terms of age and sensory characteristics.

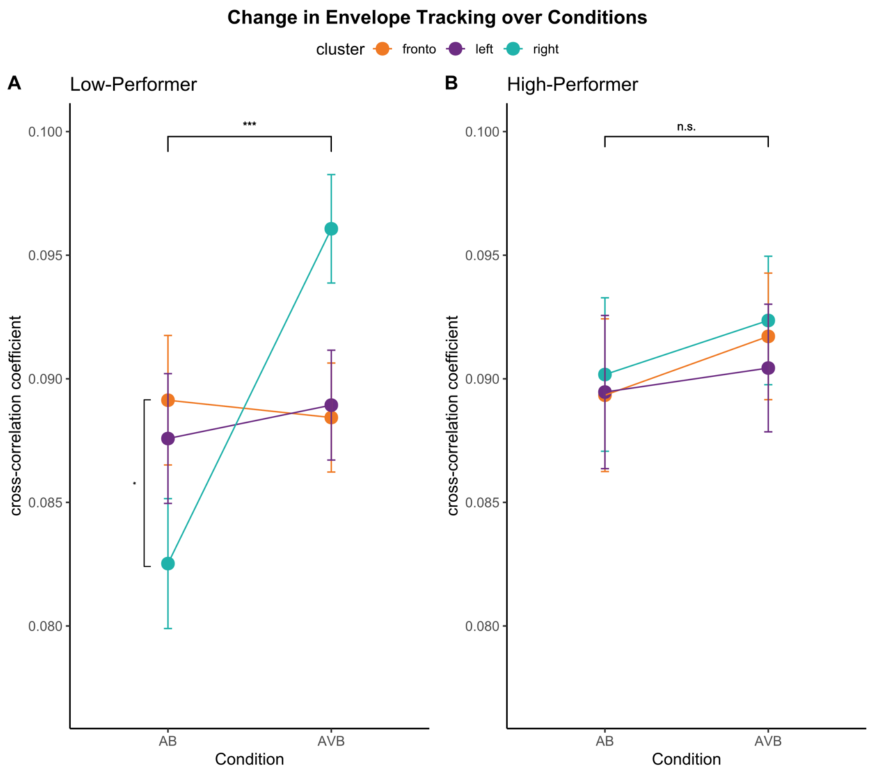

Figure 8. **Change in neural speech tracking through audio-visual speech presentation in the full sample with respect to performance state** . A: There is a significant three-way interaction between condition, cluster and performance state indicating a significant increase in neural tracking in audio-visual speech presentation for the low performer group in the right temporo-occipital cluster. B: There is no significant interaction between the condition, cluster, and performance state for the high performer group.

Discussion

Visual speech cues facilitate SiN perception and comprehension

We hypothesized that audio-visual speech presentation is associated with improved SiN intelligibility and comprehension compared to auditory-only speech exposure in older individuals with hearing impairment. Indeed, our analysis revealed that audio-visual speech presentation significantly improves both SiN perception and comprehension (by 13 and 24 % respectively on average). It is important to note that this visual enhancement effect occurred in the auditory low performing sub-sample, but not in individuals who performed well (above 70% correct in the auditory-only condition). This result shows that audio-visual enhancement is particularly important for individuals who have auditory SiN perception difficulties. The high-performer sub-sample already showed a very good performance in the purely auditory condition and consequently no further increase through visual speech cues. Pure-tone hearing loss and age do not seem to have a direct impact on SiN perception and comprehension in our data, which is not surprising since our sample explicitly focuses on older, hearing-impaired individuals and thus does not include a normal hearing or younger age group (($M_{\text{Age}} = 72$, $\text{Range}_{\text{Age}} = 64$ –80,$SD_{\text{Age}} = 4.3$, $M_{\text{PTA}} = 42.68$, $\text{Range}_{\text{PTA}} = 31.85$–57, $SD_{\text{PTA}} = 6.49$). Nevertheless, there is a trend in our data indicating that SiN perception (measured in the intelligibility task) but not SiN comprehension is negatively affected by increased pure-tone hearing loss and age.

In line with our results, several studies show this visual enhancement effect in relation to SiN comprehension (O'Neill, 1954; Sumby and Pollack, 1954; Helfer and Freyman, 2005; Ross et al., 2007 which is evident in older populations with normal as well as impaired (Avivi-Reich et al., 2018; Tye-Murray et al., 2007, 2016).A common element of this work is that the visual enhancement effect is often characterized by inter-individual variability, as it is in our case. Tye-Murray and colleagues (2016) argue that this variability is partly explained by age-related unimodal degradation, while Avivi-Reich and colleagues (2018) find no age-related effect on AV-benefit. Our data do not allow for a conclusive answer to this either, as it explicitly concerns older individuals with hearing loss. However, in the context of this population, neither age nor hearing loss seem to explain this variability sufficiently. In contrast to the studies mentioned above, we extend the argument by also considering cognitive capacity, especially working memory.

Previous work (Anderson et al., 2013; Peelle, 2018; Puschmann et al., 2019; Zekveld et al., 2011) suggests that more working memory capacity (WMC) is associated with better SiN perception. It is argued that hearing-impaired individuals rely more on cognitive resources to perceive SiN (Campbell & Sharma, 2013; Peelle & Wingfield, 2016). Puschmann and colleagues (2019) speculate that hearing-impaired individuals devote a considerable amount of their cognitive resources to SiN perception and therefore show reduced behavioral performance in a purely auditory condition compared to normal-hearing individuals. Audiovisual cues facilitate speech perception and thus a redistribution of cognitive resources to the behavioral task can occur, explaining differences in AV-benefit. When applied to our sample, it is arguable that differences in SiN perception performance in the auditory-only condition are partly explained by individual cognitive capacity. This assumption is, to some extent, supported by our data. WMC seems to be associated with better SiN perception for the low performance sub-sample (indicated by a significant association between WMC and intelligibility performance). Considering the whole sample in the purely auditory condition (i.e., both high and low performers), WMC also seems to characterize the two subgroups at least in part. With respect to the Intelligibility task, the higher the speech tracking response, the better the intelligibility performance, but only for individuals with high WMC. With decreasing WMC this relationship diminishes. It could therefore be argued that AV-benefit differences in our data can be partly explained by the fact that individuals who show low SiN performance particularly benefit from AV-input when sufficient WMC is present. This finding is important in terms of developing preventive interventions such as training, and suggests that a combined approach of auditory, audio-visual, and cognitive elements may be particularly helpful for people with hearing loss.

We further hypothesized that audio-visual speech presentation is associated with increased neural tracking and better SiN perception as well as comprehension. We investigated this question by studying cortical speech envelope tracking in auditory only as well as audio-visual speech presentation and applied an intelligibility and a comprehension task. We found that the high-performing and low-performing subsamples differ in their tracking response in the purely auditory condition. Individuals who have difficulty perceiving and comprehending speech in noise also show significantly reduced neural tracking in the right temporo-occipital cluster compared to the left and frontal clusters, while individuals who do not have such difficulty (and are matched regarding age and PTA) show no significant differences in tracking between their clusters. Furthermore, our data show a significant increase in neural speech envelope tracking by audio-visual speech cues, a result shown in a variety of previous studies (Besle et al., 2004; Bishop & Miller, 2009; Callan et al., 2003; Crosse et al., 2015; McGettigan et al., 2012; Puschmann et al., 2019; Sekiyama et al., 2003). Audio-visual speech presentation appears to enhance neural responses, particularly in challenging listening situations (Crosse et al., 2015; Zion Golumbic et al., 2013), and this dynamic is prevalent in both younger and older individuals.

In line with our results, Puschmann and colleagues (2019) have already demonstrated that cortical speech tracking and task performance were increased in audio-visual speech, but only with respect to a SiN perception task using a target-word detection paradigm. We extend these findings to a higher-level language task, namely a comprehension task. In addition, we argue that more speech envelope tracking partially explains auditory-only SiN performance variability, which can be explained through a central compensatory gain mechanism (Chambers et al., 2016) . This central compensatory hypothesis states that increased cortical activation serves to restore the degraded peripheral input from noise as well as hearing loss in order to understand speech. The absence of this gain, reflected in the reduced activity in the low-performer subsample, could therefore also explain the lower SiN performance in the auditory-only condition.

Another possible explanation could be that increased neural processing of speech may reflect lip-movement induced entrainment. Articulatory movement of the mouth and jaw, which are correlated with the temporal regularity of the speech envelope (Chandrasekaran et al., 2009; Grant & Seitz, 2000), appear to modulate auditory cortex activity and enhance the phase-alignment of auditory cortical oscillations to speech input (Besle et al., 2008; Lakatos et al., 2008; Micheli et al., 2018; Park et al., 2016). This explanation also seems to be relevant in our case, although it should be noted that we did not have a purely visual condition for comparison, which makes it difficult to disentangle this dynamic. Nevertheless, it can be assumed that the presentation of congruent lip movement is involved in the audio-visual reinforcement of neural tracking. One mechanism of action could be that visual articulatory information precedes auditory information (Chandrasekaran et al., 2009; Grant & Seitz, 2000), speculating that visual input facilitates speech perception by helping to predict the timing of upcoming auditory input. This assumption is also supported by Schwartz and colleagues (2004), who showed that intelligibility of auditory syllables is enhanced when parallel visual input contains predictive information, even when the visual content does not contain information regarding the identity of the syllables (Kim & Davis, 2003). It has also been shown at the neurophysiological level that the early neural response to audio-visual syllable presentation is proportionally enhanced by predictive content of the visual input (Arnal et al., 2009, 2011). From this, we argue that audio-visual enhancement arises, among other things, from the fact that auditory input using congruent visual cues leads to increased sensitivity and enhanced sensory processing in the auditory cortex. This assumption is also supported by the fact that this significant increase in neural tracking was only found in the early time window (75 ms–125 ms), which suggests that visual input in particular increases sensory processing. This association between audio-visual enhancement and early sensory processing in the auditory cortex was also found in Puschmann and colleagues (2019).

A final important point to discuss in our data is the lateralization of neural speech tracking as a function of performance sub-samples. The low-performer subsample shows a difference regarding their tracking response, with significantly reduced tracking in the right temporo-occipital cluster compared to the other clusters in the auditory-only condition, and a significant increase in this cluster through audio-visual speech cues paralleling improving in behavioral responses to the tasks. At the same time, the high-performer subsample does not show this deviation in speech tracking across the different cluster. The AST hypothesis (Poeppel, 2003) states that the initial representation of speech is bilateral but becomes asymmetric in later processing stages. The hypothesis refers to young adults with age-typical auditory performance. Giroud and colleagues (2019) extended this hypothesis through the perspective of ageing, arguing that age-related structural degradation is counteracted by increased bilateral processing of speech. In the context of our study, we can extend this argument by including the perspective of hearing loss. In addition to the age of our population, our sample is composed of individuals with pure-tone hearing loss. We speculate that the difference in SiN performance in the pure auditory condition is partly due to the fact that the low-performing sub-sample is not able to recruit both auditory areas (left as well as right) to overcome partly age-related and partly sensory-related decline in auditory networks, resulting in reduced SiN performance. By adding visual speech cues, which demonstrates an additional source of information, our data suggests that older individuals with auditory processing deficits benefit from visual input by means of more bilateral auditory processing. This argument is particularly important as it provides information regarding possible training designs. If audio-visual speech cues stimulate bilateral processing, and if bilateral processing generally leads to better encoding of speech in noise in older individuals with hearing impairment, this reinforces the already known potential of visual speech presentation. Furthermore, brain stimulation protocols for older adults with hearing impairment could test targeting specifically right auditory areas to improve auditory processing in aging. Following this line of reasoning, it would be essential to test whether this stimulation caused by visual speech presentation shows long-term effects and leads to an altered activation pattern for SiN perception, independent of visual speech cues. Furthermore, and especially in the context of the results presented here, it also seems important that future studies include cognitive capacity in this context. The extent to which external audio-visual input affects and facilitates neural speech processing in older individuals with hearing impairment also seems to depend on the individual's cognitive capacity.

Conclusion

In summary, audio-visual speech presentation seems to facilitate both SiN perception and comprehension, especially for individuals who have auditory SiN perception difficulties. It appears, that individual working memory capacity plays a role in this association and suggests that AV input in particular is beneficial when sufficient WMC is present. Our work also demonstrates that speech tracking increases, particularly in the right auditory clusters leading to more bilateral processing, under audio-visual speech presentation and is associated with improved SiN comprehension. In addition, we argue that higher speech tracking at least partially explains the interindividual differences in the benefit of AV speech presentation, possibly involving a central compensatory gain mechanism.

Limitations

Given the focus on neuro-functional correlates, more specifically neural speech tracking, in the context of audio-visual speech presentation, the main limitation is that no visual-only control condition was included. Disentangling the interplay between auditory and visual input is potentially important in that it provides insight into the supra-additive effects of both streams. In particular, given the discussed potential role of lip-movement induced entrainment and the associated amplification of the neural response could further provide insight into which factors favor audio-visual speech processing. However, due to the rather long paradigm and the concern that an additional condition would have led to biased results, due to participant fatigue, this was not done. Nevertheless, it seems to be a promising aspect for future work on audio-visual speech presentation and its benefits for older persons with hearing impairment.

Data and code availability statement

The auditory and audio-visual stimuli as well as the behavioral and neural data used in this study are publicly available in the study's Open Science Framework repository (https://osf.io/sb4mu/). No part of the study procedures and analyses was preregistered prior to the research being conducted.

Authors contribution

Vanessa Frei: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – Original Draft, Visualization; Raffael Schmitt: Conceptualization, Writing-Review & Editing; Martin Meyer: Conceptualization, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition; Nathalie Giroud: Conceptualization, Methodology, Validation, Writing - Review & Editing, Supervision, Funding acquisition.

Acknowledgments

This work was supported by grants from the Swiss National Science Foundation (grant no. 105319_169964 to MM and grant no. PR00P1_185715 to NG).

Figures legend

Figure 1. **Pure-tone audiometry** . The audiogram depicts individual pure-tone thresholds at frequencies between 0.5 and 8 kHz. There is no systematic difference between hearing-aid users (HA; group average shown in solid line) and non-hearing-aid user (nHA; group average depicted in dashed line).

Figure 2. **Illustration of the stimuli presentation** . The two conditions differed in that audio-visual-babble (AVB) included a video sequence of the mouth and jaw movement whereas the audio-babble (AB) condition only contained auditory stimuli. Five sentences were presented, with an intelligibility question after each sentence. After every fifth sentence, a comprehension question was asked. There was a total of 30 items per condition.

Figure 3. **Segregation of the sample into high and low performers.** A: Individual differences of intelligibility performance between audio-babble and audio-visual babble with color-codes for high and low performers in auditory babble. B: Individual differences of comprehension performance between audio-babble and audio-visual babble. Since there were only 6 comprehension questions per condition, there are 6 different

percentile ranks in this figure. For a clear visual representation, the actual percentile ranks were jittered slightly across the participants.

Figure 4: **Topographic distribution and time course of neural speech tracking** . A: Grand mean cross correlation functions of the frontal, the left, and the right cluster. Significant time windows are marked as bars over the function. B: Topographic distribution and time course of the grand average cross correlation in both listening conditions from approximately 50 to 250 ms. C: Topographical distribution of the grand-average cross-correlation at the peaks at approximately 100 and 250 ms. Selected electrode clusters are marked with "*". Warm colors denote positive- and cool colors negative correlations.

Figure 5. **Change of SiN intelligibility and comprehension performance through the presentation of additional visual speech cues** . A: In both, the intelligibility task and the comprehension task, there is a significant increase in SiN perception under AV speech presentation for the low-performer sub-sample, but no significant change in performance for the full sample, nor for the high-performer sub-sample. B: In the low performing group, representative of the covariates, WMC is significantly associated with intelligibility performance, while age, hearing loss and neural speech tracking are not. C: None of the covariates seem to be significantly associated with comprehension performance.

Figure 6. **Interaction between WMC and cross-correlation in the intelligibility task** . The higher the neural speech tracking response, the more likely it is to be correct in the intelligibility task, but only for participants with high WMC, with lower WMC this relationship seems to diminish.

Figure 7. **Differences in neural speech tracking in the audio-babble condition between the two sub-samples** . A: There is no significant difference in the neural response between the frontal and the left nor the left and the right cluster, but a significant difference between the right and the frontal cluster for the low-performer sub-sample. B: There is no significant difference in envelope tracking between any of the three clusters for the high-performer sub-sample in the auditory-babble condition.

Figure 8. **Change in neural speech tracking through audio-visual speech presentation in the full sample with respect to performance state** . A: There is a significant three-way interaction between condition, cluster and performance state indicating a significant increase in neural tracking in audio-visual speech presentation for the low performer group in the right temporo-occipital cluster. B: There is no significant interaction between the condition, cluster, and performance state for the high performer group.

Tables legend

Table 1. Comparability between individuals with and without hearing aids regarding age, hearing loss, and working memory capacity.

Table 2. Estimated dependency between condition, cross-correlation, and intelligibility performance. Model configuration: Intelligibility response ˜ condition + cross-correlation + cluster + working memory + age + PTA + (condition | subject) + (condition | trial).

Table 3. Estimated dependency between condition, cross-correlation, and comprehension performance. Model configuration: Comprehension response ˜ condition + cross-correlation + cluster + working memory + age + PTA + (condition | subject) + (condition | trial).

Table 4. Estimated dependency between cross-correlation, working-memory, and intelligibility performance for the full sample in audio-babble. Model configuration: Intelligibility response ˜ cross-correlation * working memory + cluster + PTA + age + (cross-correlation | subject) + (cross-correlation | trial).

Table 5. Estimated dependency between cross-correlation, working-memory, and comprehension performance for the full sample in audio-babble. Model configuration: Comprehension response ˜ cross-correlation + working memory + cluster + PTA + age + (cross-correlation | subject) + (cross-correlation | trial).

Table 6. Estimated dependency between condition, clusters, and neural speech tracking. Model configuration: Cross-correlation ˜ condition * cluster + age + PTA + working memory + (condition | subject) +

(condition | trial).

Table 7. Estimated dependency between performance state, clusters, and neural speech tracking for the full sample in audio-babble. Model configuration: Cross-correlation ~ performance state * cluster + PTA + age + working memory + (1 | subject) + (performance state | trial).

References

Aller, M., Økland, H. S., MacGregor, L. J., Blank, H., & Davis, M. H. (2022). Differential Auditory and Visual Phase-Locking Are Observed during Audio-Visual Benefit and Silent Lip-Reading for Speech Perception. *Journal of Neuroscience* , *42* (31), 6108–6120. https://doi.org/10.1523/JNEUROSCI.2476-21.2022

Altieri, N., & Hudock, D. (2014). Hearing impairment and audiovisual speech integration ability: A case study report. *Frontiers in Psychology* , *5* . https://doi.org/10.3389/fpsyg.2014.00678

Anderson, S., Parbery-Clark, A., Yi, H.-G., & Kraus, N. (2011). A Neural Basis of Speech-in-Noise Perception in Older Adults. *Ear and hearing* , *32* (6), 750–757. https://doi.org/10.1097/AUD.0b013e31822229d3

Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hearing Research* , *300* , 18–32. https://doi.org/10.1016/j.heares.2013.03.006

Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A.-L. (2009). Dual Neural Routing of Visual Facilitation in Speech Processing.*Journal of Neuroscience* , *29* (43), 13445–13453. https://doi.org/10.1523/JNEUROSCI.3194-09.2009

Arnal, L. H., Wyart, V., & Giraud, A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech.*Nature Neuroscience* , *14* (6), Art. 6. https://doi.org/10.1038/nn.2810

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., & Bizley, J. K. (2018). Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding. *Neuron* , *97* (3), 640-655.e4. https://doi.org/10.1016/j.neuron.2017.12.034

Avivi-Reich, M., Puka, K., & Schneider, B. A. (2018). Do age and linguistic background alter the audiovisual advantage when listening to speech in the presence of energetic and informational masking?*Attention, Perception, & Psychophysics* , *80* (1), 242–261. https://doi.org/10.3758/s13414-017-1423-5

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal.*Journal of Memory and Language* , *68* (3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Başkent, D., & Bazo, D. (2011). Audiovisual asynchrony detection and speech intelligibility in noise with moderate to severe sensorineural hearing impairment. *Ear and Hearing* , *32* (5), 582–592. https://doi.org/10.1097/AUD.0b013e31820fca23

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting Linear Mixed-Effects Models using lme4* (arXiv:1406.5823). arXiv. https://doi.org/10.48550/arXiv.1406.5823

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., & Giard, M.-H. (2008). Visual Activation and Audiovisual Interactions in the Auditory Cortex during Speech Perception: Intracranial Recordings in Humans. *Journal of Neuroscience* , *28* (52), 14301–14310. https://doi.org/10.1523/JNEUROSCI.2875-08.2008

Besle, J., Fort, A., Delpuech, C., & Giard, M.-H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex.*European Journal of Neuroscience* , *20* (8), 2225–2234. https://doi.org/10.1111/j.1460-9568.2004.03670.x

Biesmans, W., Das, N., Francart, T., & Bertrand, A. (2017). Auditory-Inspired Speech Envelope Extraction Methods for Improved EEG-Based Auditory Attention Detection in a Cocktail Party Scenario.*IEEE Transactions on Neural Systems and Rehabilitation Engineering* , *25* (5), 402–412. https://doi.org/10.1109/TNSRE.2016.2571900

Bishop, C. W., & Miller, L. M. (2009). A Multisensory Cortical Network for Understanding Speech in Noise. *Journal of Cognitive Neuroscience* , *21* (9), 1790–1804. https://doi.org/10.1162/jocn.2009.21118

Bowl, M. R., & Dawson, S. J. (2019). Age-Related Hearing Loss.*Cold Spring Harbor Perspectives in Medicine* , *9* (8), a033217. https://doi.org/10.1101/cshperspect.a033217

Braiman, C., Fridman, E. A., Conte, M. M., Voss, H. U., Reichenbach, C. S., Reichenbach, T., & Schiff, N. D. (2018). Cortical Response to the Natural Speech Envelope Correlates with Neuroimaging Evidence of Cognition in Severe Brain Injury. *Current Biology* , *28* (23), 3833-3839.e3. https://doi.org/10.1016/j.cub.2018.10.057

Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *NeuroReport* ,*14* (17), 2213.

Campbell, J., & Sharma, A. (2013). Compensatory changes in cortical resource allocation in adults with hearing loss. *Frontiers in Systems Neuroscience* , *7* . https://www.frontiersin.org/articles/10.3389/fnsys.2013.00071

Chambers, A. R., Resnik, J., Yuan, Y., Whitton, J. P., Edge, A. S., Liberman, M. C., & Polley, D. B. (2016). Central Gain Restores Auditory Processing following Near-Complete Cochlear Denervation. *Neuron* ,*89* (4), 867–879. https://doi.org/10.1016/j.neuron.2015.12.041

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The Natural Statistics of Audiovisual Speech.*PLOS Computational Biology* , *5* (7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. *Journal of Neuroscience* , *35* (42), 14195–14204. https://doi.org/10.1523/JNEUROSCI.1829-15.2015

Crosse, M. J., Liberto, G. M. D., & 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. https://doi.org/10.1523/JNEUROSCI.1396-16.2016

Decruy, L., Vanthornhout, J., & Francart, T. (2020). Hearing impairment is associated with enhanced neural tracking of the speech envelope.*Hearing Research* , *393* , 107961. https://doi.org/10.1016/j.heares.2020.107961

Dubno, J. R., Eckert, M. A., Lee, F.-S., Matthews, L. J., & Schmiedt, R. A. (2013). Classifying Human Audiometric Phenotypes of Age-Related Hearing Loss from Animal Models. *Journal of the Association for Research in Otolaryngology* , *14* (5), 687–701. https://doi.org/10.1007/s10162-013-0396-x

Fuglsang, S. A., Märcher-Rørsted, J., Dau, T., & Hjortkjær, J. (2020). Effects of Sensorineural Hearing Loss on Cortical Synchronization to Competing Speech during Selective Attention. *Journal of Neuroscience* , *40* (12), 2562–2572. https://doi.org/10.1523/JNEUROSCI.1936-19.2020

Gates, G. A., & Mills, J. H. (2005). Presbycusis. *The Lancet* ,*366* (9491), 1111–1120. https://doi.org/10.1016/S0140-6736(05)67423-5

Gillis, M., Decruy, L., Vanthornhout, J., & Francart, T. (2021).*Hearing loss is associated with delayed neural responses to continuous speech* (S. 2021.01.21.427550). bioRxiv. https://doi.org/10.1101/2021.01.21.427550

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations.*Nature Neuroscience* , *15* (4), Art. 4. https://doi.org/10.1038/nn.3063

Giroud, N., Hirsiger, S., Muri, R., Kegel, A., Dillier, N., & Meyer, M. (2018). Neuroanatomical and resting state EEG power correlates of central hearing loss in older adults. *Brain Structure and Function* , *223* (1), 145–163. https://doi.org/10.1007/s00429-017-1477-0

Giroud, N., Keller, M., Hirsiger, S., Dellwo, V., & Meyer, M. (2019). Bridging the brain structure—Brain function gap in prosodic speech processing in older adults. *Neurobiology of Aging* , *80* , 116–126. https://doi.org/10.1016/j.neurobiolaging.2019.04.017

Grant, K. W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America* , *108* (3), 1197–1208. https://doi.org/10.1121/1.1288668

Hallam, R. S., & Corney, R. (2014). Conversation tactics in persons with normal hearing and hearing-impairment. *International Journal of Audiology* , *53* (3), 174–181. https://doi.org/10.3109/14992027.2013.852256

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience* , *8* (5), Art. 5. https://doi.org/10.1038/nrn2113

Horton, C., D'Zmura, M., & Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *Journal of Neurophysiology* , *109* (12), 3082–3093. https://doi.org/10.1152/jn.01026.2012

Humes, L. E., Dubno, J. R., Gordon-Salant, S., Lister, J. J., Cacace, A. T., Cruickshanks, K. J., Gates, G. A., Wilson, R. H., & Wingfield, A. (2012). Central presbycusis: A review and evaluation of the evidence. *Journal of the American Academy of Audiology* , *23* (8), 635–666. https://doi.org/10.3766/jaaa.23.8.5

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKEOWN, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* ,*37* (2), 163–178. https://doi.org/10.1111/1469-8986.3720163

Kim, J., & Davis, C. (2003). Task effects in masked cross-script translation and phonological priming. *Journal of Memory and Language* , *49* (4), 484–499. https://doi.org/10.1016/S0749-596X(03)00093-7

Kurthen, I., Galbier, J., Jagoda, L., Neuschwander, P., Giroud, N., & Meyer, M. (2021). Selective attention modulates neural envelope tracking of informationally masked speech in healthy older adults. *Human Brain Mapping* , *42* (10), 3042–3057. https://doi.org/10.1002/hbm.25415

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection. *Science* , *320* (5872), 110–113. https://doi.org/10.1126/science.1154735

Leclercq, M., & Zimmermann, P. (2004). *Applied Neuropsychology of Attention: Theory, Diagnosis and Rehabilitation* . Psychology Press.

Lecluyse, W., & Meddis, R. (2009). A simple single-interval adaptive procedure for estimating thresholds in normal and impaired listeners. *The Journal of the Acoustical Society of America* , *126* (5), 2570–2579. https://doi.org/10.1121/1.3238248

Liberman, M. C., & Kujawa, S. G. (2017). Cochlear synaptopathy in acquired sensorineural hearing loss: Manifestations and mechanisms. *Hearing Research* , *349* , 138–147. https://doi.org/10.1016/j.heares.2017.01.003

Lidestam, B., Moradi, S., Pettersson, R., & Ricklefs, T. (2014). Audiovisual training is better than auditory-only training for auditory-only speech-in-noise identification. *The Journal of the Acoustical Society of America* , *136* , EL142. https://doi.org/10.1121/1.4890200

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron* ,*54* (6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

McGettigan, C., Faulkner, A., Altarelli, I., Obleser, J., Baverstock, H., & Scott, S. K. (2012). Speech comprehension aided by multiple modalities: Behavioural and neural interactions. *Neuropsychologia* , *50* (5), 762–776. https://doi.org/10.1016/j.neuropsychologia.2012.01.010

Micheli, C., Schepers, I., Ozker, M., Yoshor, D., Beauchamp, M., & Rieger, J. (2018). Electrocorticography reveals continuous auditory and visual speech tracking in temporal and occipital cortex. *European Journal of Neuroscience* , *51* . https://doi.org/10.1111/ejn.13992

Mirkovic, B., Debener, S., Schmidt, J., Jaeger, M., & Neher, T. (2019). Effects of directional sound processing and listener's motivation on EEG responses to continuous noisy speech: Do normal-hearing and aided hearing-impaired listeners differ? *Hearing Research* , *377* , 260–270. https://doi.org/10.1016/j.heares.2019.04.005

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society* ,*53* (4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience* , *2011* , 1:1-1:9. https://doi.org/10.1155/2011/156869

Park, H., Kayser, C., Thut, G., & Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *ELife* , *5* , e14521. https://doi.org/10.7554/eLife.14521

Peelle, J. E. (2018). Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. *Ear and Hearing* , *39* (2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

Peelle, J. E., & Wingfield, A. (2016). The Neural Consequences of Age-Related Hearing Loss. *Trends in Neurosciences* , *39* (7), 486–497. https://doi.org/10.1016/j.tins.2016.05.001

Pichora-Fuller, M. K., & Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology* , *42* , 2S11-2S16.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication* , *41* (1), 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience* , *21* (6), Art. 6. https://doi.org/10.1038/s41583-020-0304-4

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2007). Speech perception at the interface of neurobiology and linguistics.*Philosophical Transactions of the Royal Society B: Biological Sciences* , *363* (1493), 1071–1086. https://doi.org/10.1098/rstb.2007.2160

Puschmann, S., Daeglau, M., Stropahl, M., Mirkovic, B., Rosemann, S., Thiel, C. M., & Debener, S. (2019). Hearing-impaired listeners show increased audiovisual benefit when listening to speech in noise.*NeuroImage* , *196* , 261–268. https://doi.org/10.1016/j.neuroimage.2019.04.017

Rosemann, S., & Thiel, C. M. (2018). Audio-visual speech processing in age-related hearing loss: Stronger integration and increased frontal lobe recruitment. *NeuroImage* , *175* , 425–437. https://doi.org/10.1016/j.neuroimage.2018.04.023

Schmitt, R., Meyer, M., & Giroud, N. (2022). Better speech-in-noise comprehension is associated with enhanced neural speech tracking in older adults with hearing impairment. *Cortex* , *151* , 133–146. https://doi.org/10.1016/j.cortex.2022.02.017

Schneider, B. A., & Pichora-Fuller, M. K. (2000). Implications of perceptual deterioration for cognitive aging research. In *The handbook of aging and cognition, 2nd ed* (S. 155–219). Lawrence Erlbaum Associates Publishers.

Schneider, B. A., & Pichora-Fuller, M. K. (2001). Age-Related Changes in Temporal Processing: Implications for Speech Perception.*Seminars in Hearing* , *22* (3), 227–240. https://doi.org/10.1055/s-2001-15628

Schwartz, J.-L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition* , *93* (2), B69–B78. https://doi.org/10.1016/j.cognition.2004.01.006

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research* , *47* (3), 277–287. https://doi.org/10.1016/S0168-0102(03)00214-1

Singmann, H., & Kellen, D. (2019). An Introduction to Mixed Models for Experimental Psychology. In *New Methods in Cognitive Psychology* . Routledge.

Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-Visual Speech Perception and Auditory-Visual Enhancement in Normal-Hearing Younger and Older Adults. *Ear and Hearing* , *26* (3), 263–275.

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007). Audiovisual Integration and Lipreading Abilities of Older Adults with Normal and Impaired Hearing. *Ear and Hearing* , *28* (5), 656–668. https://doi.org/10.1097/AUD.0b013e31812f7185

Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (2016). Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration. *Psychology and Aging* , *31* (4), 380–389. https://doi.org/10.1037/pag0000094

Vermiglio, A. J., Soli, S. D., Freed, D. J., & Fisher, L. M. (2012). The Relationship between High-Frequency Pure-Tone Hearing Loss, Hearing in Noise Test (HINT) Thresholds, and the Articulation Index. *Journal of the American Academy of Audiology* , *23* (10), 779–788. https://doi.org/10.3766/jaaa.23.10.4

Winneke, A. H., & Phillips, N. A. (2011). Does audiovisual speech offer a fountain of youth for old ears? An event-related brain potential study of age differences in audiovisual speech perception. *Psychology and Aging* , *26* (2), 427–438. https://doi.org/10.1037/a0021683

Wong, P. C. M., Ettlinger, M., Sheppard, J. P., Gunasekera, G. M., & Dhar, S. (2010). Neuroanatomical Characteristics and Speech Perception in Noise in Older Adults. *Ear and hearing* , *31* (4), 471–479. https://doi.org/10.1097/AUD.0b013e3181d709c2

Wong, P. C. M., Jin, J. X., Gunasekera, G. M., Abel, R., Lee, E. R., & Dhar, S. (2009). Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia* , *47* (3), 693–703. https://doi.org/10.1016/j.neuropsychologia.2008.11.032

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive Load During Speech Perception in Noise: The Influence of Age, Hearing Loss, and Cognition on the Pupil Response. *Ear and Hearing* , *32* (4), 498. https://doi.org/10.1097/AUD.0b013e31820512bb

Zion Golumbic, E. M., 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

Zoefel, B., & VanRullen, R. (2016). Eeg oscillations entrain their phase to high-level features of speech sound. *NeuroImage* , *124* , 16–23. https://doi.org/10.1016/j.neuroimage.2015.08.054