scientific reports

OPEN

Vibrotactile speech cues are associated with enhanced auditory processing in middle and superior temporal gyri

Alina Schulte[1,2,6]✉, Jeremy Marozeau[3,4], Andrej Kral[1,5] & Hamish Innes-Brown[2,6]

Combined auditory and tactile stimuli have been found to enhance speech-in-noise perception both in individuals with normal hearing and in those with hearing loss. While behavioral benefits of audio-tactile enhancements in speech understanding have been repeatedly demonstrated, the impact of vibrotactile cues on cortical auditory speech processing remains unknown. Using functional near-infrared spectroscopy (fNIRS) with a dense montage setup, we first identified a region-of-interest highly sensitive to auditory-only speech-in-quiet. In the same region, we then assessed the change in activity ('audio-tactile gains') when presenting speech-in-noise together with a single-channel vibratory signal to the fingertip, congruent with the speech envelope's rate of change. In data from 21 participants with normal hearing, audio-tactile speech elicited on average 20% greater hemodynamic oxygenation changes than auditory-only speech-in-noise within bilateral middle and superior temporal gyri. However, audio-tactile gains did not exceed the sum of the unisensory responses, providing no conclusive evidence of true multisensory integration. Our results support a metamodal theory for the processing of temporal speech features in the middle and superior temporal gyri, providing the first evidence of audio-tactile speech processing in auditory areas using fNIRS. Top-down modulations from somatosensory areas or attention networks likely contributed to the observed audio-tactile gains through temporal entrainment with the speech envelope's rate of change. Further research is needed to understand the neural responses in concordance with their behavioral relevance for speech perception, offering future directions for developing tactile aids for individuals with hearing impairments.

Keywords Functional near-infrared spectroscopy, Multisensory processing, Multimodal speech, Auditory processing, Audio-tactile perception

Sensory information from non-auditory modalities, such as visual cues from a speaker's face, plays a significant role in auditory speech processing[1]. This becomes particularly important in noisy listening environments and for individuals with hearing loss, where auditory cues are less reliable. Multimodal speech information enhances speech perception and improves intelligibility[2,3] by providing additional a priori information in a Bayesian sense[4]. The broadband envelope has been identified as a crucial speech feature for successful speech understanding[5–8] and is also reflected in lip and mouth movements[9,10]. Consequently, slow temporal fluctuations of speech appear to enhance intelligibility even when presented non-auditorily[11]. When delivered as vibrotactile cues to the skin, such as on the fingertip or wrist, congruent envelope cues have been found to improve understanding of speech in noise in people with normal hearing thresholds as well as cochlear implant users[6,12–20]. These findings suggest novel rehabilitation opportunities for individuals with hearing loss, underscoring the need for a deeper understanding of the neural encoding and integration of tactile stimuli into speech processing.

[1]Department of Experimental Otology of the Clinics of Otolaryngology, Institute for AudioNeuroTechnology (VIANNA), Hannover Medical School, Hannover, Germany. [2]Eriksholm Research Center, Oticon A/S, Snekkersten, Denmark. [3]Music and Cochlear Implants Lab, Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark. [4]Department of Basic Neurosciences, University of Geneva, Geneva, Switzerland. [5]Australian Hearing Hub, School of Medicine and Health Sciences, Macquarie University, Sydney, Australia. [6]Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark. ✉email: aicu@eriksholm.com

Audio-visual[21–23] and visual-only speech processing[24–26] have both been localized to auditory areas of the brain. These findings support an amodal perspective of speech processing in temporal areas[27], showing that separate unisensory processing is not necessarily required prior to convergence. Integration of multimodal speech cues likely occurs at multiple processing stages, including areas beyond auditory cortices[28,29]—a hypothesis further supported by recent evidence from a large-scale meta-analysis[30]. Several brain regions have been linked to speech perception, and theories on the contribution of, for example, the motor system, have been extensively discussed[31,32]. Recent frameworks also suggest a role for the somatosensory system in the neural processing of speech[33], offering a promising perspective for the use of vibrotactile speech information.

Different psychophysiological methods have been used to examine how vibrotactile speech support affects auditory speech processing. Electroencephalography (EEG) research has shown that vibrotactile cues aligned with the syllabic rate of speech entrain theta-range oscillatory neural activity[34], and that phase alignment of amplitude modulations between auditory noise and electrotactile carrier signals is critical for enhancing auditory steady-state responses[35]. Multisensory gain has also been observed in temporal response functions (using forward modelling of EEG) for synchronously combined audio-tactile speech stimuli, likely originating in auditory cortices[18]. A follow-up study using similar vibrotactile pulses at 5 Hz found that responses were sustained even beyond the stimulation period[36]. Together, these findings suggest that tactile input supports the segmentation of auditory speech into syllables, possibly mediated by a phase-resetting mechanism[37]. Moreover, a functional magnetic resonance imaging (fMRI) study demonstrated activation of superior temporal gyri (STG) in response to continuous tactile speech stimuli, but not to token-based stimuli[38]. These results support the hypothesis that vibrotactile enhancement engages auditory cortical regions and highlight the importance of preserving the temporal structure of speech in the tactile signal to elicit auditory-like processing. While EEG offers high temporal resolution for assessing the impact of tactile stimuli on oscillatory dynamics, its spatial resolution is limited and relies on source localization algorithms to estimate the origin of neural signals. In contrast, fMRI provides high spatial resolution but requires a costly, stationary setup with a loud acoustic environment that poses challenges for auditory experiments. Both methods are prone to artifacts from hearing devices, rendering them suboptimal for studying individuals with hearing loss—those who may benefit most from vibrotactile speech support. To overcome these limitations, the present study employs functional near-infrared spectroscopy (fNIRS) to investigate whether previously reported tactile enhancements of auditory speech processing are reflected in amplitude differences in evoked cortical hemodynamic responses. While fNIRS represents a compromise in spatial and temporal resolution, it provides sufficient sensitivity to detect speech-evoked cortical responses from optodes positioned over temporal regions, rather than reflecting summed or distant sources as in EEG. Additionally, it is quiet, easy to set up, and compatible with hearing aids and cochlear implants, without causing interfering artifacts[39].

fNIRS as a method for studying cortical speech-related hemodynamics

fNIRS is increasingly popular in hearing and language research[40–43]. It measures relative concentration changes in oxygenated (HbO) and deoxygenated (HbR) hemoglobin in superficial cortical tissue, which indirectly reflect neural activity through neurovascular coupling. fNIRS enables the identification of cortical sources of neural activation, making it well-suited to pinpoint superficial cortical sites involved in auditory speech processing. Numerous fNIRS studies have reported bilateral temporal and left inferior frontal cortical activations associated with different aspects of speech processing[44–51], and have identified differences in activation patterns between participants with natural hearing and those with cochlear implants[52–54]. Cross-modal activations have been observed across different participant groups[55–57] and this cross-modal activation has been linked to speech perception outcomes in cochlear implant users; however, the direction of these associations has varied[58,59]. Evidence of multisensory integration of speech stimuli using fNIRS remains elusive[60,61].

Not all studies have consistently observed robust auditory activations. Factors such as participants' vigilance, vascular confounds (e.g. blood-stealing[62]) or the choice of baseline[48] have been proposed as possible explanations. Auditory fNIRS responses predominantly reflect activity from the auditory belt and association cortices, as the primary auditory cortex lies deep within the lateral sulcus and is reached with less than 1% specificity[63]. The isolation of neural auditory activity is further complicated by significant contributions from physiological sources, including activity from the muscle temporalis and hemodynamic changes in extracranial vessels in temporal regions.

As a result, auditory fNIRS experiments require careful consideration of experimental design and analysis methods[64]. For instance, signal enhancement algorithms have been found essential for data cleaning and extracting auditory responses[65]. Additionally, approaches to maximize individual responses by determining each participant's best-responding channels prior to group-level analyses have been applied[57,58].

Aims of the current study

The central aim of the present study was to examine whether vibrotactile speech cues influence auditory cortical responses to speech, as measured with fNIRS. To address this question, we adopted a two-step approach.

We first identified fNIRS channels that consistently responded to auditory speech-in-quiet stimuli. This step enabled the definition of a data-driven region-of-interest (ROI), which served as the basis for the subsequent analysis in the present study and will be applicable in future studies using a similar montage setup. Second, we compared hemodynamic responses to audio-tactile speech-in-noise with those to auditory speech-in-noise within the previously identified auditory ROI. This comparison aimed to determine whether vibrotactile cues enhance auditory processing. Additionally, we evaluated whether the observed audio-tactile responses exceeded the linear sum of the unisensory responses, a conventional criterion for identifying multisensory integration[66].

Material and methods

Participants

Twenty-one normal-hearing adults (8 females, 13 males; mean age = 35 years, SD = 9) were included in the analysis. Data from four additional participants were excluded due to insufficient data quality, as more than 50% of their fNIRS channels had to be rejected based on the criteria described below (see in "Preprocessing" section). All participants were recruited from among colleagues at Oticon or the Technical University of Denmark. They were right-handed, had normal hearing, and were highly proficient in English, but not native speakers. Previous research has reported the largest multisensory benefits in participant groups who are non-native speakers of the language used for stimulus presentation[16,17,67–69]. Ethical approval was obtained from the Danish Science-Ethics Committee (reference H-16036391). The experiment was conducted in accordance with the respective guidelines and regulations, and all participants provided written informed consent.

Equipment

The study was conducted in a sound-treated booth at Eriksholm Research Center, Denmark. Auditory stimuli were presented diotically via Etymotic ER-2 insert earphones (Etymotic Research, Illinois, USA). Tactile stimuli were delivered through a Bruel and Kjær minishaker type 4810 equipped with an accelerometer (B&K 4533) featuring a circular contact area of 12 mm in diameter, positioned under the participant's right index fingertip. Stimulus files were converted by an RME Fireface UCX audio interface, with one output channel for each signal (auditory, tactile). These outputs were connected to separate amplifiers, which drove the earphones and minishaker, respectively (Fig. 1a).

Acceleration of the tactile probe was measured from the accelerometer, which also served as the contact surface for stimulus presentation. The accelerometer was connected to a B&K 1704 CCLD conditioner, which in turn was connected to the input of the RME Fireface soundcard. Sound intensities were measured using a B&K sound level meter type 2250, along with an ear simulator type 4157, and the external ear accessory DB 2012.

Cortical oxygenation changes were recorded using an fNIRS system (NIRx NirScout) with 16 LED sources and 16 avalanche photodiode detectors. Measurements targeted temporal, inferior-frontal and parietal regions (see Fig. 3 for the probe layout). Near-infrared light at 760 nm and 850 nm wavelengths was sampled at a rate of 3.9063 Hz using the NIRStar 15.3 acquisition software (NIRx Medizintechnik GmbH, Berlin, Germany).

Stimuli

The stimuli consisted of English HINT sentences presented in nine different conditions, of which only five were relevant for the present investigation: *auditory speech-in-quiet*, *auditory speech-in-noise*, *audio-tactile speech-in-noise*, *tactile speech,* and a silent *control* condition. Each condition comprised 20 trial repetitions with silent inter-stimulus intervals. A sixth condition (*auditory noise-in-quiet*) was disregarded in the present analysis. In addition to these six conditions with silent inter-stimulus intervals, three further conditions involving continuous background noise were recorded but excluded from analysis as they were not relevant for the research questions of this study. These were: S*peech-in-continuous background noise*, *tactile speech-in-continuous background noise* and a second control condition with *continuous background* noise only.

Stimuli were grouped in eight experimental blocks, with optional breaks between blocks. Blocks 3 and 6 included the conditions with continuous background noise, presented in random order. The remaining blocks comprised the six conditions with silent background between stimuli, also played in random order. During all experimental blocks, a silent movie [Koyaanisqatsi (Godfrey Reggio[70])], unrelated to the speech stimuli, was shown continuously on the computer screen to keep participants engaged. Including breaks, the listening task lasted approximately 70 min.

Auditory stimuli

Speech stimuli consisted of three concatenated sentences, resulting in trial durations between 5.56 and 5.7 s. In the *auditory speech-in-noise* and *audio-tactile speech-in-noise* conditions, noise started playing 0.5 s before speech onset and continued until the end of the speech signal. For the remaining conditions, 0.25 s pauses were inserted between sentences to achieve comparable stimulus lengths.

For the two speech-in-noise conditions, speech-shaped noise with an identical long-term average spectrum as the speech material was generated and presented at a fixed signal-to-noise ratio of − 2 dB for all participants. Given that participants were not native English speakers, and based on previous investigations showing little variance in speech reception thresholds for normal-hearing participants, it can be assumed that the presented speech stimuli were intelligible at approximately 50–70%. All auditory stimuli were equalized to have the same root-mean-square and presented at a level of 73.8 dB SPL LAF.

Tactile stimulus

The *tactile speech* stimulus was based on the sentence envelope's rate of change, which modulated a single carrier signal. As a consequence, the peaks of the tactile signal align with syllable onsets rather than the centers of syllables, as would be the case when using the original broadband envelope (Fig. 1c). This tactile stimulus differs from our previous study[19], in which we used a 4-band tactile stimulus that was summed into a single output, based on Fletcher et al.[13]. The change to a simpler, single-carrier tactile signal cueing the envelope's peak rate was inspired by an EEG study revealing that neurons in the superior temporal gyrus (STG) encode amplitude changes rather than absolute amplitude of speech broadband envelopes[71]. In a pilot experiment with seven participants, no differences in effectiveness between the old and the new tactile speech stimuli on speech enhancement were identified. Nonetheless, the finding that very sparse tactile signals aligned with phonemes elicited similar enhancement effects in another study[18] further supported our decision to switch to a simpler stimulus.

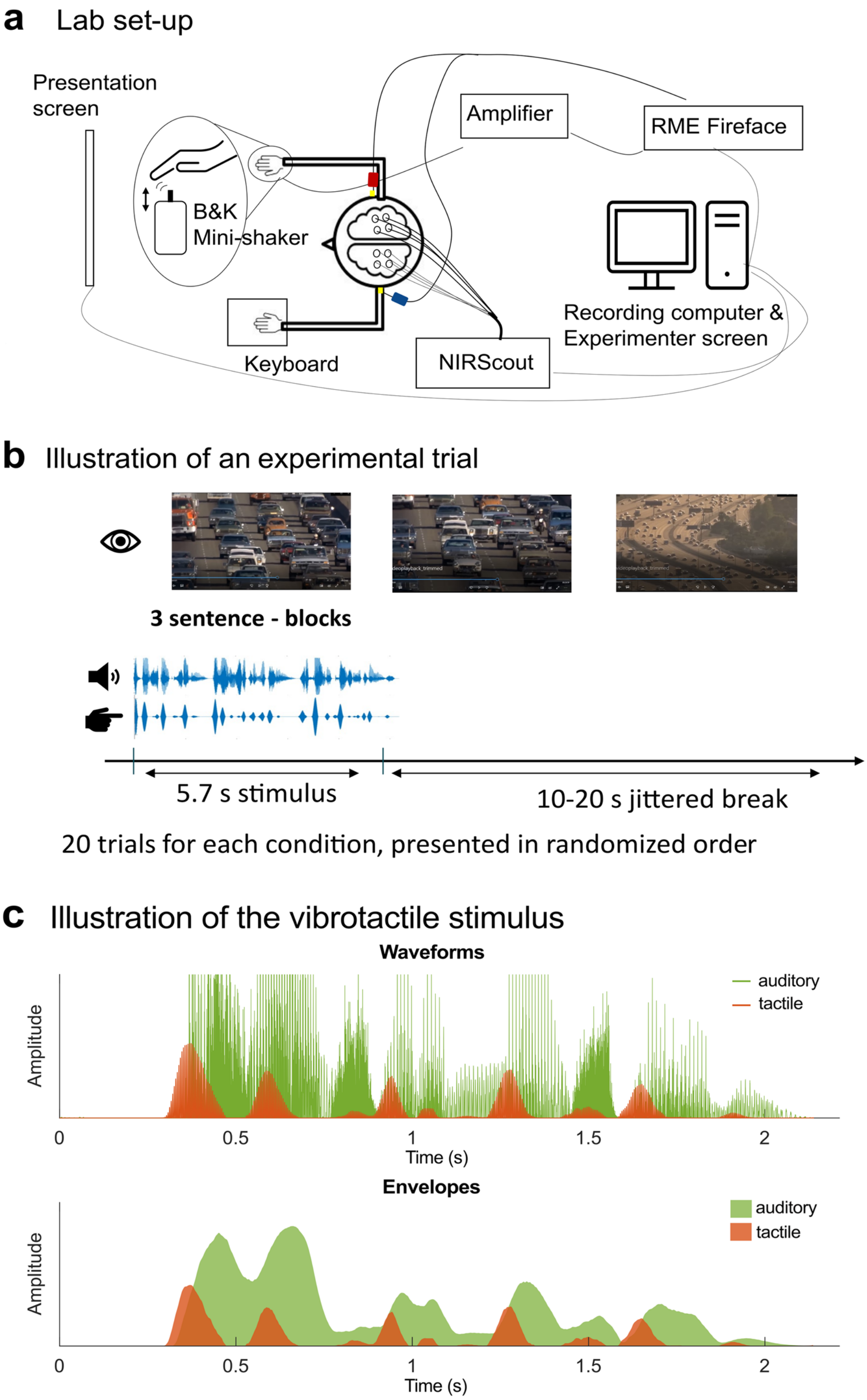

Fig. 1. Experimental setup and paradigm. (**a**) Schematic illustration of a participant in the lab receiving auditory stimuli through insert earphones and vibrotactile stimulation on the right index fingertip. (**b**) One trial of the experimental procedure including auditory, tactile, and visual stimulus presentations. (**c**) Auditory and vibrotactile signals corresponding to the sentence “The house had a nice garden”. The upper panel shows rectified waveforms of the auditory and tactile stimuli. The tactile stimulus (in orange) is a 230 Hz carrier modulated by the rate of change of the auditory envelope. The bottom panel displays area-filled envelopes of the auditory and tactile speech signals. Note that in the experiment, three sentences were always presented consecutively, forming one stimulus block, as illustrated in (**b**).

To generate the tactile envelope signal we followed the procedure from MacIntyre et al.[72], using their env3-function to extract the envelope[73], followed by taking its first derivative, smoothing, rectifying and rescaling. The resulting envelope rate of change was then used to modulate a carrier signal of 230 Hz—a frequency to which the Pacinian corpuscles in glabrous skin are highly sensitive[74]. Tactile stimuli were presented with an average displacement of 3.7 µm to the right fingertip (dominant hand for all participants). We used a stronger presentation intensity compared to our previous investigation[19], in which detection threshold estimates were around 0.1 µm for a 230 Hz carrier signal and stimulus presentation was at 1.4 µm displacement on average. Hence, stimulus intensity was most likely well above threshold in the current study as was confirmed subjectively by all participants during the experiment.

All stimuli were generated prior to data collection using MATLAB (Version R2020a, Natick, Massachusetts, The MathWorks Inc.) and presented during the experiment using PsychoPy3[75]. Event triggers were sent from PsychoPy to NirStar using LabStreamingLayer[76] at the start of stimulus presentation. The tactile and auditory signals were presented simultaneously, as they were stored in the same two-channel file that was sent to the respective playback devices (earphones and minishaker).

Experimental design and procedure

At the beginning, participants were informed about the experimental procedure and provided written informed consent. If no audiogram had been conducted within the last 12 months, pure-tone audiometry at 0.25, 0.5, 1, 2, 4 and 8 kHz was performed to ensure participants could be classified as normal-hearing based, on a mean threshold below 20 dB HL. Subsequently, participants were equipped with the fNIRS cap and instructed to avoid head movements and contractions of facial muscles. A passive speech perception paradigm followed, during which participants were instructed to attend to the stimuli, although no active task had to be performed. Between experimental blocks, participants were asked how well they could understand the speech-in-noise stimuli to remind them to direct attention to the stimuli. Trials of the experimental conditions occurred in random order, interleaved with a jittered break of 10–20 s. The trial order and stimuli were identical across all participants (see Fig. 1b for an example trial).

fNIRS montage

A suitable optode placement targeting speech-relevant areas was created using NIRSite 2021.4 (NIRx Medizintechnik GmbH, Berlin, Germany). Source-detector pairs were formed based on specificity information from the fNIRS Optodes' Location Decider (fOLD) toolbox[63], using the Juelich brain parcellation atlas. Optodes covered auditory and speech processing areas in the temporal lobes, as well as the left inferior frontal gyrus and somatosensory cortex. Most fNIRS studies consider EEG 10–10 locations as possible positions for fNIRS optodes. As one of the study's aims was to localize auditory responses as precisely as possible, spatial resolution was increased by adding channels between the 10–10 locations. Note that the additional optode positions are not equivalent to the 10–5 system, in which positions are defined at 5% steps from nasion to inion. Instead, the additional optodes remained spaced at 10% intervals from nasion to inion, while being placed at 5% distances between existing optodes along the preauricular-to-preauricular axis. The denser grid enabled the addition of diagonal channel connections to the montage, which would otherwise exceed a suitable optode distance. The final montage included a total of 49 long channels with source-detector distances ranging from 2.1 to 4.3 cm. In addition, eight short-channel detectors were included in the montage at a distance of 0.8 cm from their respective sources to capture scalp hemodynamics for signal correction (see Fig. 3).

Data analysis

Data were analyzed using the Python-based toolbox MNE-NIRS, including MNE[77], MNE-NIRS (version 1.5.1[47]) Nilearn[78] and statsmodels[79]. Cortical activation was inferred from the response magnitudes of HbO signals, estimated using a waveform averaging analysis and a general linear model (GLM). While GLMs address the statistical properties of fNIRS data more appropriately and are generally expected to yield more accurate results, waveform visualizations offer a more intuitive and conventional representation of the data, facilitating interpretation and understanding. Rather than comparing the two analysis methods directly, we interpret converging results from both as an indicator of robust auditory responses, with cross-validation between approaches increasing the credibility of our findings.

Preprocessing

For each of the two wavelengths, the recorded traces of raw light intensities were converted to optical density (i.e., transformation to a logarithmic scale with zero mean). To identify channels with insufficient data quality, that are unlikely to reflect cerebral oxygenation changes, we considered scalp coupling indices (SCI) and peak power (cross-correlation of wavelengths 1 and 2 and its power spectral density within the cardiac range as, described by Pollonini et al.[80]), as well as each channel's standard deviation (SD). Gaps larger than 30 s between segments from − 5 to 15 s around stimulus onset were discarded for this step, ensuring data quality assessment was based on experimental blocks only (i.e., not including break periods of the experiment). Channels were excluded from analysis if they met one or more of the following criteria: SCI < 0.7, more than 50% of 10 s segments with a peak power < 0.1, or a SD > 0.2 (equivalent to a coefficient of variation of 20%). These criteria led to the exclusion of, on average, 2.2 out of 49 long channels (SD = 3.6 channels) and 1.4 out of 8 short channels (SD = 2.2 channels) per participant (see Fig. S1). After channel rejection, temporal derivative distribution repair[81] was applied to diminish motion artifacts. Corrected optical density data were converted to hemoglobin concentration changes by applying the modified Beer–Lambert law with a partial pathlength factor of 0.1. The hemoglobin data were low-pass filtered using a 5th order IIR Butterworth filter with a cut-off frequency of 0.25 Hz, and then high-pass filtered by application of the same filter with a high-pass cut-off frequency of 0.005 Hz.

Finally, signal quality was further enhanced using the correlation-based signal improvement method[82]. This technique must be treated with caution, as it transforms the data substantially and relies on the assumptions that HbO and HbR signals are positively correlated during motion artifacts and (perfectly) negatively correlated otherwise—premises that do not fully hold. For instance, a slightly delayed dip in the HbR signal relative to the HbO peak is expected and these relationships are likely to vary across individuals and brain regions. Correlation-based signal improvement was included in our preprocessing pipeline only after evaluating its effect on the contrast-to-noise ratio (CNR, defined as in Zhou et al.[83]) of waveform averages, comparing all stimulus conditions combined versus no stimulus (control conditions). Note that this response was not of interest for our main analysis. Comparing the CNR before and after applying correlation-based signal improvement, we found that the method significantly enhanced detection of the hemodynamic response to stimuli without altering the response to the control stimuli (see Fig. S2). The algorithm computes a new HbO signal based on both original chromophore signals, whereas the new HbR signal is entirely derived from the corrected HbO. Consequently, the corrected HbR signal does not contain any independent information after this processing step. For visualization purposes, both chromophores are displayed in waveform averages throughout the manuscript, however, statistical analyses and interpretations are based solely on the corrected HbO signal.

To remove contributions from scalp hemodynamics, we performed a separate GLM using all available short-channel traces as regressors (up to 8 HbO and 8 HbR traces). In one participant, this step was omitted because no short channels with sufficient data quality were available. Previous studies have demonstrated the effectiveness of including either all available short channels or their principal components in the model[84,85] and recommend performing short-channel correction in a separate step to reduce collinearity between regressors[85,86]. We found that this approach considerably increased the CNR compared to the alternative method implemented in MNE-NIRS (i.e., subtracting a scaled version of the nearest short channel based on Fabbri et al.[87], Saager and Berger[87] and Scholkmann et al.[88], see Fig. S2).

Single participant feature extraction: GLM and waveform averaging

GLM analysis Beta-values, reflecting the amplitude of the modelled hemodynamic response, were estimated for each participant. Regressors in the design matrix were generated by convolving a Glover hemodynamic response function (HRF)[89] with a boxcar function of 3.4 s duration. This deviation from the actual stimulus duration was applied following Luke et al.[47], who found that a boxcar duration 1.67 times shorter than the actual stimulus length best captured the hemodynamic responses in a similar in a passive auditory listening paradigm. Additional cosine drift regressors were included to account for drifts not removed by the prior high-pass filtering. The high-pass cut-off frequency for the drift model was defined based on the longest inter-stimulus interval between two conditions of interest. An autoregressive model of order 5 was used to estimate beta-values for each channel, weighting each regressor to explain the maximum variance in the data.

Waveform averaging analysis As a second measure, stimulus-evoked neural responses were estimated for each participant using a waveform averaging analysis. Channels were segmented into epochs from − 5 to 20 s relative to stimulus onset and baseline-corrected by subtracting the mean of the pre-stimulus period. Additionally, a linear detrend was applied to each epoch. Epochs with HbO peak-to-peak amplitudes exceeding 45 Δμmol/l were rejected. On average, this threshold led to the exclusion of 0.6% of epochs per participant (SD = 14.5%, median = 4.3%).

To reduce the waveform to a single value suitable for statistical analysis, the mean HbO signal between 5 and 8 s post-stimulus onset was calculated (referred to as the waveform mean amplitude). This time window corresponds to the expected peak of the hemodynamic response (around 6.5 s after stimulus onset, considering a Glover HRF convolved with a boxcar function of 3.4 s length).

Generation of data-driven auditory region-of-interest

To generate a ROI containing channels sensitive to auditory speech processing, single-participant results for each channel in the *auditory speech-in-quiet* condition were analyzed. For each participant, the ten channels showing the highest HbO responses were identified. Then, for each channel, the number of participants for which it appeared among the top ten was counted. This process was performed separately for both waveform mean amplitudes and beta-values. The final data-driven ROI was defined as the intersection of the top-ranked channels from the two top ten lists. This procedure revealed six left and right temporal channels which were defined as the *auditory-ROI* (see Fig. 3). To compute the ROI response, amplitude estimates (waveform mean amplitudes and beta-values, respectively) were averaged with equal weighting.

Group level analysis

Single participant results were summarized using linear mixed-effect models (LMMs) to determine group HbO responses for individual condition activations, as well as to test our hypothesis of potentially stronger responses for *audio-tactile speech-in-noise* compared to *auditory speech-in-noise* within the identified *auditory-ROI*.

Separate models were conducted for beta-values and waveform mean amplitudes. Within each model, multiple comparisons were accounted for using the Benjamini–Hochberg procedure to control the false discovery rate (FDR) as implemented in the statsmodels package[79]. This method uses an accepted false discovery rate of 0.05 to determine a *p* value threshold for rejecting hypotheses, based on the rank and total number of comparisons (i.e., the number of conditions multiplied by the number of channels or ROIs). We considered the agreement of results from both GLM and waveform-averaging analyses as an indicator of the robustness of the overall HbO results and their relevance for interpretation.

Condition activations To initially display the data, HbO responses for each condition were compared to baseline with an interaction of channels and condition as main effect, participants as a random effect and a suppressed intercept (in Roger–Wilkinson notation: HbO response ~ − 1 + Condition:channel + (1|Participant)). To obtain ROI results instead of single-channel activations, the model was repeated with Condition as the main effect, using each participant's auditory-ROI activation as input instead of single channel results.

Contrasts To test for the presence of audio-tactile gains in the *auditory-ROI*, an LMM with *Condition* as the main effect and *Participants* as a random effect was conducted, comparing *audio-tactile speech-in-noise* against *auditory speech-in-noise* in the *auditory-ROI* (Δ HbO response ~ Condition + 1|Participant). By not suppressing the intercept, the first level of the Condition factor (auditory speech-in-noise) serves as a reference category, and the resulting coefficient for the audio-tactile condition represents the contrast between the two conditions. If significant audio-tactile gains were present, their effect size was compared to each participant's sum of the unimodal activations to assess multisensory integration: an additional LMM, including the conditions *audio-tactile speech-in-noise* and an auxiliary condition comprising the sum of the two unisensory activations (*auditory speech-in-noise* and *tactile speech*), was conducted.

Results

Condition activations

Single channel results

Group results for single-channel activations across task conditions revealed significant positive HbO responses predominantly in temporal areas for all conditions except for the *control* (Fig. 2). Activation patterns were overall similar for beta-values and waveform mean amplitudes, with significant activations observed in auditory regions. However, they differed in somatosensory channels: beta-values revealed significant activations in channels over the left primary somatosensory cortex across all stimulus conditions (most pronounced for *audio-tactile speech-in-noise*), while waveform mean amplitudes indicated left primary somatosensory cortex activation selectively for *tactile speech* and *audio-tactile speech-in-noise*.

ROI results

Based on consistent *auditory speech-in-quiet* responses across participants, an *auditory-ROI* was defined, comprising six channels positioned over the left and right middle and superior temporal gyri (Fig. 3, Table 1).

In line with the activation patterns observed in single channel results (Fig. 2), the ROI group results also showed significant responses for all task conditions except the *control* (Fig. 4a,b; Table 2). *Tactile speech* reached significance only in the waveform averaging analysis (Table 2). Overall, waveform mean amplitudes and beta-values appeared consistent, with similar distributions across participants (Fig. 4b).

Contrasts

Audio-tactile gains were assessed by contrasting the *audio-tactile speech-in-noise* and *auditory speech-in-noise* conditions. *Audio-tactile speech-in-noise* elicited significantly larger HbO responses than *auditory speech-in-noise* in both the GLM analysis (beta difference = 0.6 Δμmol/l, $p = 0.037$) and the waveform averaging analysis (waveform mean amplitude difference = 0.7 Δμmol/l, $p = 0.045$), corresponding to an increase in HbO response of 19% and 23%, respectively. However, this effect did not exceed an additive criterion for multisensory integration. The summed responses of the *auditory speech-in-noise* and *tactile speech* conditions resulted on average in 0.28 Δμmol/l larger activations for waveform mean amplitudes ($p = 0.01$) and 0.7 Δμmol/l larger activations for beta-values ($p = 0.06$) compared to the *audio-tactile speech-in-noise* condition. These results indicate a sub-additive to additive enhancement effect at the group level. Individual audio-tactile gains from waveform averaging and GLM analyses are shown in Fig. 4c.

Discussion

In the present study, we demonstrated that previously described audio-tactile advantages in speech understanding[19] can be attributed to the influence of tactile stimuli on auditory neuronal processing. Using a dense optode montage, we first identified channels with the strongest responses to auditory sentences presented in quiet. The identified ROI comprised the middle and superior temporal gyri, in line with well-established literature on brain regions involved in speech processing[47,56,90,91]. We then found that vibrotactile speech cues enhanced cortical auditory responses in these channels, providing the first evidence of tactile enhancement effects on auditory speech processing using fNIRS. While *audio-tactile speech-in-noise* elicited larger responses than *auditory-only speech-in-noise*, the observed audio-tactile gains did not exceed the sum of unisensory responses, a criterion traditionally used to determine multisensory integration. Our result may reflect a near-saturation of auditory hemodynamic responses under unimodal stimulation, limiting the extent to which they can be further augmented by multimodal speech inputs.

Individual differences in audio-tactile gains

The variability of individual audio-tactile gains as shown in Fig. 4c, underscores that despite the overall positive mean audio-tactile gains (sub-additive based on waveform averaging results, additive based on GLM results), not all participants exhibited stronger cortical oxygenation in the auditory ROI for audio-tactile compared to auditory-only speech.

Considering that all participants were normal-hearing, individual factors other than hearing status must have influenced the cortical processing of audio-tactile speech stimuli. Various factors such as tactile sensitivity, neural connectivity, prior experience with audio-tactile stimuli (e.g. from playing an instrument) or psychological

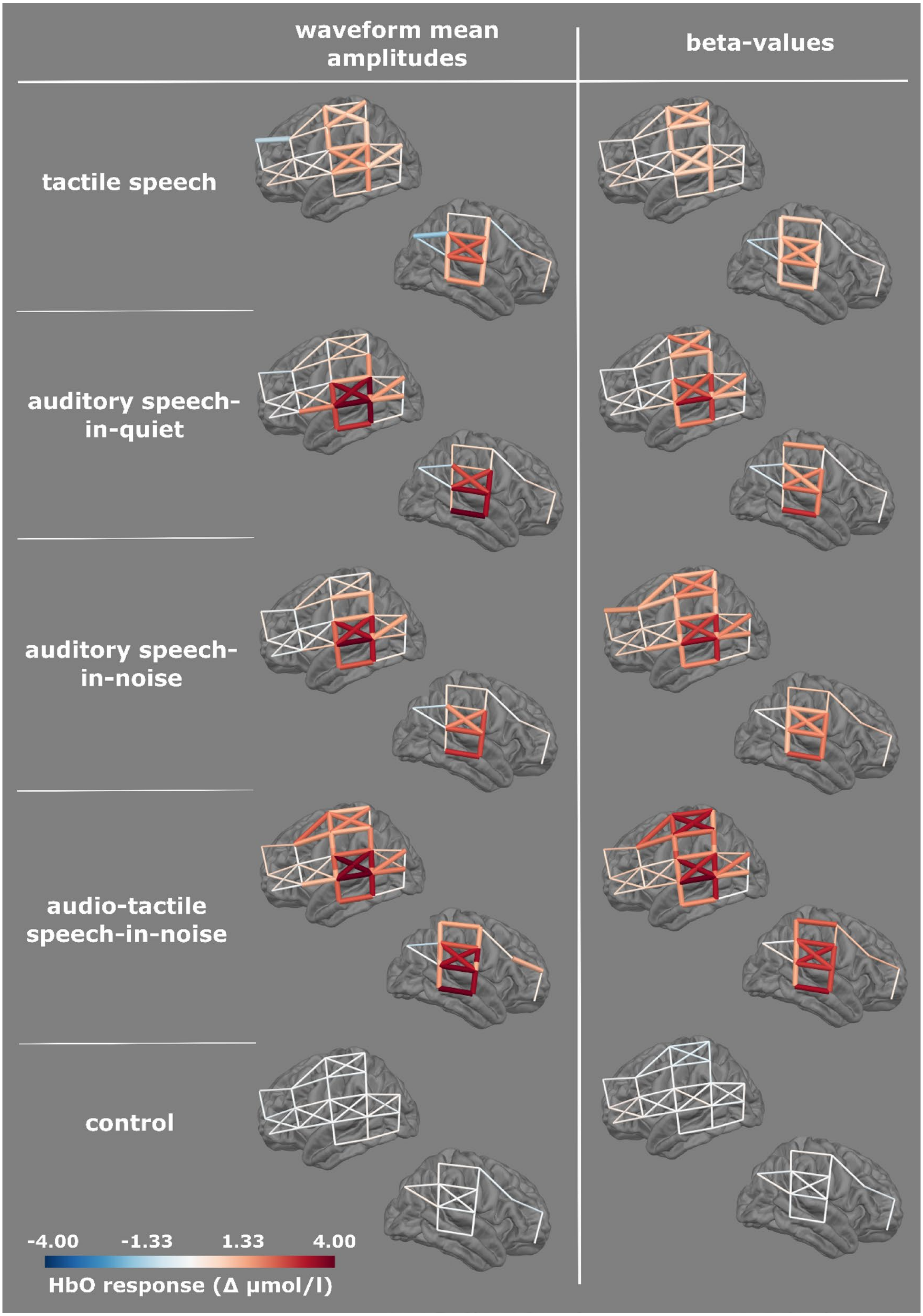

Fig. 2. Single channel group results. HbO waveform mean amplitudes and beta-values for all experimental conditions are displayed in lateral views of the left and right hemisphere. Responses are displayed as colored tubes, in the respective position of the montage. Thicker tubes represent significant channels after FDR correction.

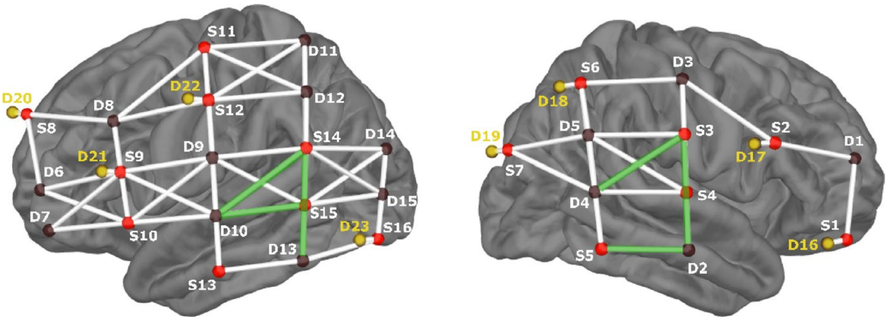

Fig. 3. Montage with auditory-ROI. Lateral views of left and right hemispheres. Red circles indicate the positions of sources, and brown circles the positions of detectors. Short-channel detectors are highlighted in yellow. Channels between optodes are displayed in white, except for those forming the auditory ROI, which are displayed in green.

Region	Channels	EEG locations	Specificity
Data-driven *auditory*-ROI	S3-D2	C6-T8	54% rSTG, 31% rMTG
	S5-D2	TP8-T8	63% rMTG, 35% rIFG
	S3-D4	C6-TP8h	rMTG, rSTG
	S14-D13	CP5-TP7	72% lMTG, 17% lSTG
	S14-D10	CP5-T7h	*lMTG, lSTG*
	S15-D10	TP7h-T7h	*lMTG, lSTG*

Table 1. Auditory-ROI: specificity of underlying targeted brain regions. Specificity information on cortical structures targeted byc each channel of the auditory-ROI is provided when available, based on the fOLD toolbox[63]. The montage included additional channels beyond the 10–10 system for which no specificity data are available. *STG* superior temporal gyrus, *MTG* middle temporal gyrus, *r* right, *l* left.

aspects such as motivation or expectation may contribute to the magnitude of audio-tactile gains. It is likely that the amount of attention directed to the tactile stimulus significantly influenced the observed audio-tactile gains, which was not assessed in the present study. Thus, the extent to which participants focused on the concurrent vibrotactile input was uncontrolled and may have varied. Other studies have ensured attention to both stimulus modalities by asking participants to detect vibrotactile patterns and answer comprehension questions[18], but have not tested whether performance in detecting tactile patterns correlates with auditory-tactile speech understanding or hemodynamic audio-tactile gains.

Moreover, it should be noted that waveform averaging and GLM analysis yielded highly comparable audio-tactile gains at the group level in this study (Fig. 4b). However, in only 10 out of the 21 participants did both methods show the same direction of effect. The fact that half of the participants did not show a consistent direction of effect across both methods suggests that the method chosen can significantly impact the results. Other studies confirm this replication issue for fNIRS data analyzed with different approaches[92]. Generally, the GLM approach is considered more robust to noise and false positives, as unexplained variance is accounted for by a noise term.

Considerations of the experimental paradigm

This experiment was designed to identify cortical responses to auditory, vibrotactile and audio-tactile speech stimuli using fNIRS. To avoid distortion of the responses of interest, we aimed to minimize speaking artifacts, and therefore, did not record data on the participants' speech understanding performance. Nevertheless, there is a tradeoff between active and passive task paradigms. Active tasks allow for monitoring of participants' attention and tend to elicit stronger responses[93], as has been shown for both, listening tasks and also somatosensory perception[94]. However, they often create artificial scenarios and require additional cognitive processes, such as memory consolidation, recall, and motor responses (e.g., button presses), which can introduce false positives due to event-locked motion artifacts. Passive listening paradigms more closely mirror natural listening conditions. They are preferred for objective assessments because they can be applied to a broader range of participants, including those who may not be able to provide active responses.

Here participants were instructed to listen to the speech carefully to understand what was being said, but we cannot rule out that the choice of a passive paradigm may have compromised participant engagement and overall effect sizes. Overall, the trade-off between active and passive task paradigms needs to be carefully evaluated with respect to the research questions being addressed and the intended use of the results, e.g. in future clinical applications targeting specific patient groups. Including an active paradigm and additional tactile conditions

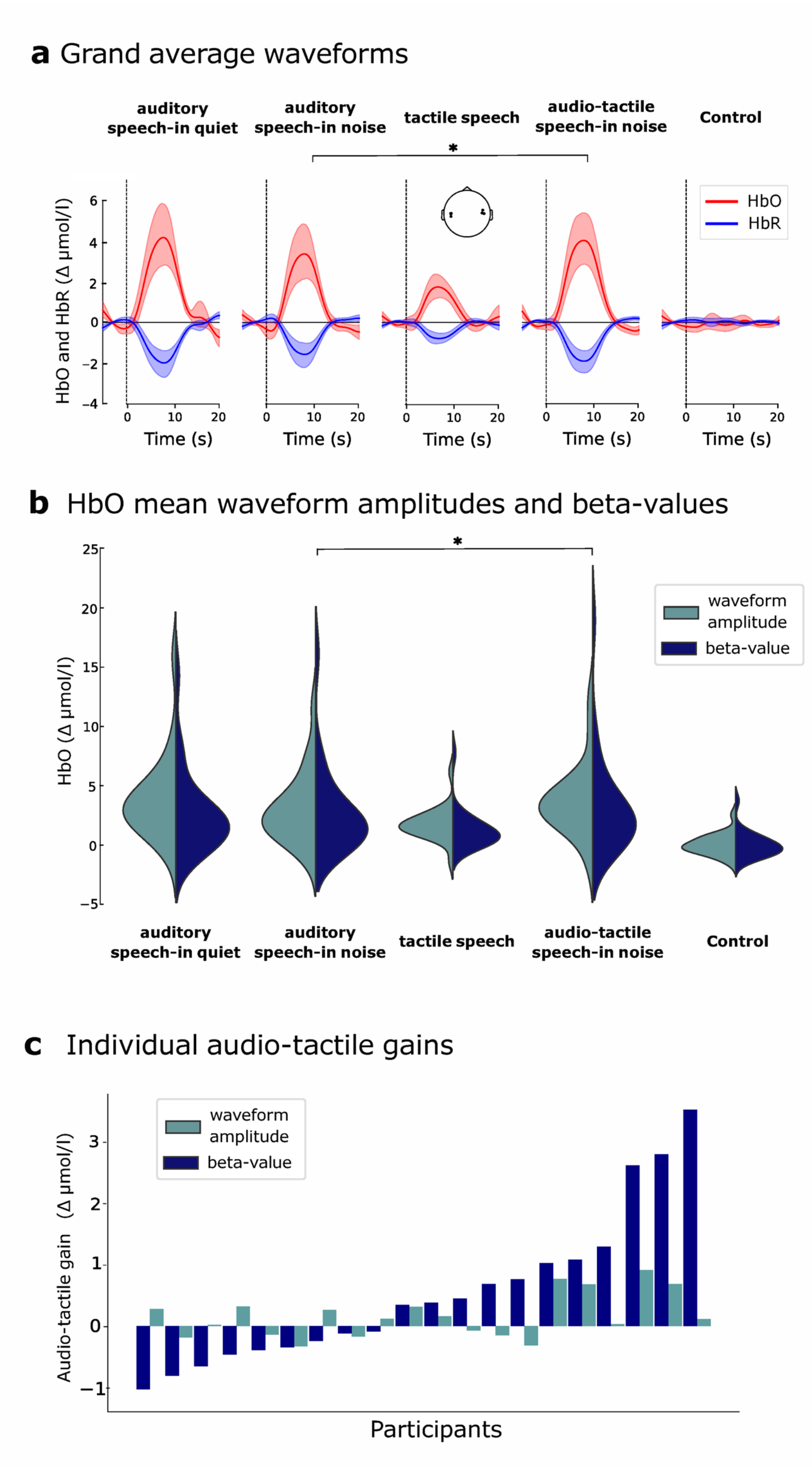

Fig. 4. ROI results. (**a**) Grand average waveforms for each task condition in the auditory-ROI. HbO traces are displayed in red, HbR traces in blue, with shaded areas representing 95% confidence intervals. Both chromophores are displayed for illustrative purposes, but only HbO was analyzed statistically. (**b**) Distributions of HbO waveform mean amplitudes and HbO beta-values in the auditory-ROI are shown for each task condition using split violin plots. (**c**) Individual audio-tactile gains in the auditory-ROI. Each participant's audio-tactile gain, as derived from the GLM and the waveform averaging analysis is shown in a bar plot, sorted by size based on the GLM-derive.

Condition	Beta-values (*p* value)	Waveform mean amplitude (*p* value)
Auditory speech-in-quiet	2.6 (< 0.001)	3.8 (< 0.001)
Auditory speech-in-noise	2.7 (< 0.001)	3.0 (< 0.001)
Tactile speech	1.2 (0.09)	1.7 (0.002)
Audio-tactile speech-in-noise	3.2 (< 0.001)	3.7 (< 0.001)
Control	− 0.05 (0.9)	− 0.4 (0.9)

Table 2. Beta-values and waveform mean amplitudes for single condition activations in the auditory-ROI. Values are given in Δμmol/l with the respective FDR corrected *p* value in brackets.

would allow assessment of whether our results mirror behavioral enhancements effect found for congruent but not incongruent[16,17] or neutral[19] tactile speech stimuli.

It should be noted that we tested for automatically evoked audio-tactile effects in cortical processing, as participants had no prior experience with tactile speech stimuli. The inclusion of a familiarization or training phase prior to the fNIRS recordings might have resulted in larger effects.

Unlike in our previous experiment where we used sound field presentation[19] and could easily align the spatial direction of the auditory and tactile stimuli, spatial congruency was not explicitly addressed in the present study. Here, auditory stimuli were delivered through insert earphones, which did not create a binaural experience of the sound coming from the same direction as the tactile input, presented to the right index fingertip. Spatial congruency for multisensory integration has been attributed less importance than temporal congruency, but it still might have lowered the likelihood of integration of both signals[95,96].

Activation of auditory areas by tactile stimulation

The described audio-tactile gain suggests that the vibrotactile speech stimulus activated brain regions typically identified as auditory. That tactile stimuli activate auditory areas is conceivable via known cortico-cortical[97,98] and thalamocortical[99] connections, and has been repeatedly documented in the literature[100–102]. Multisensory processing in auditory areas was found particularly relevant for processing temporal stimulus features[103,104]. For example, Bolognini et al.[105] demonstrated a causal role of the contralateral superior temporal gyrus for a tactile temporal discrimination task. With respect to vibrotactile speech stimuli, trained perception of sixteen-channel vocoded vibrotactile words presented on the forearm, produced similar fMRI responses in temporal lobes as auditory speech[38]. Remarkably, this result was present only for the vocoded vibrotactile stimulus that preserved temporal speech cues, while a token-based vibrotactile signal that led to similar behavioral performance in tactile word recognition did not show the same neural responses and changes in auditory-somatosensory connectivity. These results demonstrate a general ability of speech-like processing of a vibrotactile stimulus conveying temporal envelope information and suggest metamodal engagement of auditory areas[106,107]. Rather than responding selectively to auditory speech stimuli, temporal cortices seem to be sensitive to temporal structure and may be modulated also through heteromodal inputs[4].

To ensure that noise from the shaker was not audible and did not evoke an auditory response, additional investigations were conducted (see Supplementary material). The results indicate that an actual auditory response in the tactile conditions is highly unlikely and cannot explain the observed audio-tactile gains in the auditory ROI.

While tactile processing in auditory regions remains well conceivable, the limited spatial accuracy of fNIRS complicates the separation of secondary somatosensory from auditory activation. Vibrotactile cues presented to the right fingertip are known to be processed in the left primary somatosensory cortex, from where input is projected further to bilateral secondary somatosensory cortices, located in the parietal operculum, which lies directly opposite to the primary auditory cortex in Heschl's gyrus as part of the temporal operculum[108]. Neuroimaging studies have verified activations following this somatosensory pathway[109]. In accordance, our single channel results of the *tactile speech* condition show responses in left somatosensory and bilateral temporal areas (Fig. 2). Whereas primary somatosensory cortex activity can be mapped to activation in distinct channels in postcentral gyrus (e.g. S11-D9, S11-D12, S11-D11, S6-D3, see Fig. 3) responses in channels located over superior temporal gyri could be attributed to both auditory and secondary somatosensory areas.

Nonetheless, it is unlikely that responses in several channels over temporal cortices could be due solely to secondary somatosensory cortex activity. Even the most inferior channel S5-D2 located over right middle temporal gyrus showed activation in response to tactile speech alone (Fig. 2), pointing towards activation of speech-sensitive areas by the tactile stimulus rather than a secondary somatosensory cortex contribution, even when *tactile speech* was presented without auditory input.

To further verify whether the obtained audio-tactile gain could be interpreted as enhancement in auditory areas, we conducted the same procedure used to determine the *auditory-ROI*, using the *tactile speech* condition. This way, we obtained a frequency count for channels responding most often among participants to *tactile speech*, allowing us to explore whether different channels emerged with the highest preference for *tactile speech* compared to *auditory speech-in-quiet* stimuli. Contralateral (left) primary somatosensory cortex activation and bilateral activation in secondary somatosensory cortices are expected in response to right fingertip stimulation[110–112]. We identified a group of channels comprising five channels, distributed across left primary somatosensory cortex and right superior temporal gyrus, that responded most consistently to *tactile speech* in our participant group (S12-D11, S12-D12, S4-D5, S4-D4, S3-D4; see Fig. 5). This pattern may contrast with the results shown in Fig. 2, where bilateral activation of secondary somatosensory areas was observed in response to tactile speech.

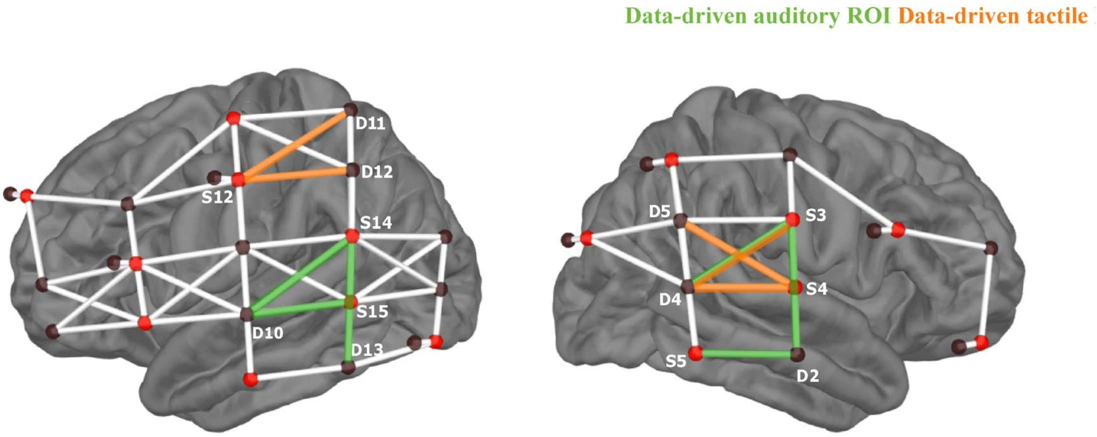

Fig. 5. Auditory-ROI and tactile-ROI. Both ROIs were defined in a data-driven approach based on channels responding most consistent with the largest HbO amplitudes across the participant group. With the exception of one channel (S3-D4), both ROIs revealed distinct channel locations that can be attributed to somatosensory vs. auditory processing. S = Source, D = Detector.

However, activation magnitude in the left-hemispheric channels was lower (Fig. 2), which likely explains why these channels did not emerge among the most responsive channels in this follow-up analysis (Fig. 5). The right-lateralized responses observed could be due to an optode placement that was more sensitive to the right secondary somatosensory cortex, despite similar levels of bilateral neural activation.

Alternatively, there might be a genuine right-hemispheric lateralization in the processing of vibrotactile speech stimuli. This question remains open for future research.

Importantly, within the identified *tactile ROI*, only channel S3-D4 overlapped with the earlier defined *auditory-ROI*, indicating that auditory-sensitive areas are not the same as tactile-sensitive areas. This follow-up analysis thus supports the interpretation that the observed audio-tactile gain reflects a genuine enhancement effect in auditory regions with minimal to no contribution from purely somatosensory processing areas.

Modulations by projections from other areas or the inherent multimodality of the auditory-ROI may underlie this effect. The latter is often referred to as "piggybacking", where somatosensory input is "piggybacked" onto auditory processing, assuming a fundamentally modality-neutral nature of speech areas[113]. This hypothesis was supported by recent observations demonstrating a dominance of audio-tactile multisensory neurons across all cortical layers, justifying the designation of temporal cortices as inherently multisensory[114]. However, similar cross-modal activations have been found to be driven by altered sensory excitability in deep anesthetic states[115], making this finding questionable. Possible bottom-up versus top-down modulations contributing to the effects will be discussed in the next section.

Vibrotactile influences on auditory cortical activations are likely mediated by top-down modulations

The presented findings speak in favor of an actual influence of the vibrotactile stimulus on processing in auditory regions. However, audio-tactile gains did not exceed an additive criterion. Often, true multisensory integration is interpreted as bottom-up sensory interaction, inferred if the response magnitude in the multisensory condition significantly deviates from the unisensory activation, e.g. significantly exceeds or suppresses the sum of unisensory responses[116]. While originally defined for spike counts of single neurons in feline superior colliculus, multisensory integration principles have been adopted also to fMRI and fNIRS measurements[61,117–119]. We anticipated audio-tactile gains (instead of suppressions), because previous fMRI and fNIRS investigations found enhancements for audio-visual speech and audio-tactile non-speech stimuli[121]. Similar to our findings, other fNIRS investigations that assessed MSI effects neither observed multisensory integration effects based on superadditivity nor the principle of inverse effectiveness[56,60,61].

Inferring multisensory integration as defined for firing rates of single neurons from fNIRS measurements is generally problematic because they constitute only indirect measures of a large number of neural populations. Instead of spike counts on an ordinal scale, relative oxygenation changes influenced by the metabolism of millions of neurons are measured on an interval scale with a normalized baseline. Consequently, linear summation may not be directly applicable. Even when cortical activations exceed a certain multisensory integration criterion (whether maximum or additive), the measured response could arise from metabolic changes driven by multisensory neuronal populations as well as co-activated unisensory neurons[118,122]. Hence, multisensory thresholds may be considered as a loose indicator for multisensory integration rather than its proof. Its absence as observed in our case, can either be attributed to a saturation effect or argues against true bottom-up integration effects. The latter case would suggest a stronger contribution of top-down influences in driving tactile activations in auditory areas. These are likely modulations from somatosensory regions via cortico-cortical connections[98] or top-down attentional influences[123]. It is known that feedback projections from higher order associative areas dominate outer cortical layers[124–126] and that the concentration of these feedback connections is reflected in cortical thickness[127]. fNIRS, measuring from superficial cortical areas may thus record activation related to feedback projections with higher sensitivity than those elicited by feedforward projections. Hence, it might also be that afferent auditory and tactile information is integrated in deeper cortical layers and was not detected.

A bottom-up processing of vibrotactile speech in auditory areas may not be far-fetched, as the same tactile stimulus could be perceived by the auditory system if it was transmitted through air and picked up by the ear instead of presented to the skin. Early metamodal processing might be more natural for auditory and tactile speech stimuli than for e.g. visual facial speech cues which are transmitted through light and constitute a different form of energy at a completely different timescale. Thus, processing of tactile stimuli in an auditory fashion is conceivable from the lowest levels, while visual speech may only mimic auditory speech processing at higher hierarchical levels, such as when reaching phoneme (and viseme) representations.

For non-speech audio-tactile stimuli, bottom-up sensory integration has been localized in superior temporal gyrus, contralateral to the side of tactile stimulation presentation[128]. We did not contrast left and right-hemispheric effects in our study, but measured multisensory gains in a bilateral auditory ROI, which may have blurred a potentially stronger right-hemispheric integration effect.

Considering single channel activations elicited by separate conditions (Fig. 2), a more pronounced right-hemispheric temporal activation elicited by *tactile speech* and a left-hemispheric dominance in *auditory-only* conditions seemed to disappear for *audio-tactile speech*. While often the left hemisphere appears to be dominant in speech processing[129,130], particularly slow features of speech, such as syllabic processing have been attributed to right auditory areas[131,132]. Note that lateralization effects may differ between individuals (e.g. dependent on sex[133,134], which was neglected in our analysis. It was surprising that only right and not bilateral STG channels were identified in the data-driven tactile ROI as expected for vibrotactile stimulation[94,135]. This may be due to a right-hemispheric dominance of SII processes[136] or a montage placement that measured with higher specificity from right than left SII. Alternatively, the responses to tactile speech may reflect not purely secondary somatosensory cortex activation, but partly right-hemispheric speech envelope processing in auditory areas triggered by the tactile stimulus alone. A right-hemispheric lateralization has also been described for theta oscillatory activity associated with the integration of syllabic units[137].

That tactile envelope cues aid in parsing a continuous speech signal into words and syllables has been hypothesized as a mechanism behind benefits in speech understanding as reported in our and other previous studies[12,14,17–19]. Top-down connections from secondary somatosensory or higher order associative areas may mediate the enhancements in speech understanding, particularly in normal-hearing participants (review in Kral and Sharma[4]). The temporally aligned vibrotactile input may entrain attention in a temporally tuned manner towards the incoming speech stimulus, thereby facilitating linguistic parsing. Various literature indicates that such a process could be influenced by top-down projections[138–140] from, among others, prefrontal areas[141–144] and be accompanied by a synchronization of low-frequency cortical oscillations as has been reported for top-down attentional mechanisms[141,145], in auditory processing[146] and audio-tactile integration[18,36].

Next to top-down influence by temporal modulations, prior knowledge of the fact that vibrotactile cues provide speech information may have consciously or unconsciously led to an interpretation of the tactile stimulation as being auditory. Exposure to preceding auditory speech stimuli, paired with the congruent *tactile speech* stimulus may have caused a priming effect, so that participants anticipated or imagined a corresponding auditory signal. The imagination of a sound percept can elicit comparable activations to real stimulation, as shown for musical auditory imagination[147,148]. In that scenario, vibrotactile speech would not be processed in auditory areas by a bottom-up process but would indirectly trigger (for tactile input only) or reinforce (for audio-tactile input) an auditory percept via mental imagery that was reflected in enhanced cortical oxygenation changes in auditory areas by top-down connections from frontal cortices. Indeed, some participants reported after the experiment that they imagined what kind of sentences the signals could have been related to, supporting an auditory interpretation of tactile cues by top-down and context-dependent influence which likely contributed to the observed audio-tactile gains.

Multisensory gains are generally assumed to be driven by a combination of top-down and bottom-up processing. We assume that also here, both early vibrotactile interactions in auditory areas, along with attentional modulations that inform about syllable onsets in a top-down manner, are mechanisms that contributed to the obtained effect. Possibly, the audio-tactile gain was further reinforced by conscious or unconscious interpretations of the vibrotactile stimulus as auditory speech.

Conclusion

The present study showed that a subtle, congruent vibrotactile signal presented to one fingertip elicited measurable cortical gains in auditory areas compared to the auditory signal alone. The observed effects may be explained by attentional modulations enhancing perceptual focus on temporally relevant speech segments, modulations from somatosensory cortices, and learned associations of tactile cues with auditory speech sounds.

Outlook

Understanding the neural basis of audio-tactile speech enhancements will help to design more effective tactile stimuli and advance the development of tactile aids. These could be particularly supportive for hard of hearing or deaf individuals who are unable to access cochlear implants or experience unsatisfactory outcomes of their hearing treatment. Further applications are conceivable during language acquisition[149] and cochlear implant rehabilitation, second language learning, or for individuals with dyslexia[150].

Data availability

The dataset generated in this study is not publicly available due to data protection issues but can be made available from the corresponding author on reasonable request.

Received: 8 December 2024; Accepted: 17 June 2025
Published online: 12 July 2025

References

1. Sumby, W. H. & Pollack, I. Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* **26**, 212–215 (1954).
2. Grant, K. W. & Seitz, P.-F. The use of visible speech cues for improving auditory detection of spoken sentences. *J. Acoust. Soc. Am.* **108**, 1197–1208 (2000).
3. Erber, N. P. Auditory-visual perception of speech. *J. Speech Hear. Disord.* **40**, 481–492 (1975).
4. Kral, A. & Sharma, A. Crossmodal plasticity in hearing loss. *Trends Neurosci.* **46**, 377–393 (2023).
5. Peelle, J. E. & Davis, M. H. Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* **3**, 320 (2012).
6. Oh, Y., Kalpin, N., Hunter, J. & Schwalm, M. The impact of temporally coherent visual and vibrotactile cues on speech recognition in noise. *JASA Express Lett.* **3**, 25203 (2023).
7. Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J. & Ekelid, M. Speech recognition with primarily temporal cues. *Science* **270**, 303–304 (1995).
8. Fogerty, D. Acoustic predictors of intelligibility for segmentally interrupted speech: Temporal envelope, voicing, and duration. *J. Speech Lang. Hear. Res.* **56**, 1402–1408 (2013).
9. Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A. & Ghazanfar, A. A. The natural statistics of audiovisual speech. *PLoS Comput. Biol.* **5**, e1000436 (2009).
10. Poeppel, D. & Assaneo, M. F. Speech rhythms and their neural foundations. *Nat. Rev. Neurosci.* **21**, 322–334 (2020).
11. Calvert, G. A. et al. Activation of auditory cortex during silent lipreading. *Science* **276**, 593–596 (1997).
12. Răutu, I. S., De Tiège, X., Jousmäki, V., Bourguignon, M. & Bertels, J. Speech-derived haptic stimulation enhances speech recognition in a multi-talker background. *Sci. Rep.* **13**, 1–11 (2023).
13. Fletcher, M. D., Hadeedi, A., Goehring, T. & Mills, S. R. Electro-haptic enhancement of speech-in-noise performance in cochlear implant users. *Sci. Rep.* **9**, 11428 (2019).
14. Fletcher, M. D., Mills, S. R. & Goehring, T. Vibro-tactile enhancement of speech intelligibility in multi-talker noise for simulated cochlear implant listening. *Trends Hear.* **22**, 1–11 (2018).
15. Huang, J., Sheffield, B., Lin, P. & Zeng, F.-G. Electro-tactile stimulation enhances cochlear implant speech recognition in noise. *Sci. Rep.* **7**, 2196 (2017).
16. Cieśla, K. et al. Effects of training and using an audio-tactile sensory substitution device on speech-in-noise understanding. *Sci. Rep.* **12**, 1–16 (2022).
17. Cieśla, K. et al. Immediate improvement of speech-in-noise perception through multisensory stimulation via an auditory to tactile sensory substitution. *Restor. Neurol. Neurosci.* **37**, 155–166 (2019).
18. Guilleminot, P. & Reichenbach, T. Enhancement of speech-in-noise comprehension through vibrotactile stimulation at the syllabic rate. *Proc. Natl. Acad. Sci.* **119**, e2117000119 (2022).
19. Schulte, A. et al. Improved speech intelligibility in the presence of congruent vibrotactile speech input. *Sci. Rep.* **13**, 22657 (2023).
20. Oh, Y., Schwalm, M. & Kalpin, N. Multisensory benefits for speech recognition in noisy environments. *Front. Neurosci.* **16**, 1031424 (2022).
21. Möttönen, R., Krause, C. M., Tiippana, K. & Sams, M. Processing of changes in visual speech in the human auditory cortex. *Cogn. Brain Res.* **13**, 417–425 (2002).
22. Besle, J. et al. Visual activation and audiovisual interactions in the auditory cortex during speech perception: Intracranial recordings in humans. *J. Neurosci.* **28**, 14301–14310 (2008).
23. Okada, K., Venezia, J. H., Matchin, W., Saberi, K. & Hickok, G. An fMRI study of audiovisual speech perception reveals multisensory interactions in auditory cortex. *PLoS ONE* **8**, e68959 (2013).
24. Mégevand, P. et al. Crossmodal phase reset and evoked responses provide complementary mechanisms for the influence of visual speech in auditory cortex. *J. Neurosci.* **40**, 8530–8542 (2020).
25. Bröhl, F., Keitel, A. & Kayser, C. MEG activity in visual and auditory cortices represents acoustic speech-related information during silent lip reading. *eNeuro* **9**, 1–15 (2022).
26. Ruytjens, L., Albers, F., Van Dijk, P., Wit, H. & Willemsen, A. Activation in primary auditory cortex during silent lipreading is determined by sex. *Audiol. Neurotol.* **12**, 371–377 (2007).
27. Rosenblum, L. D. Speech perception as a multimodal phenomenon. *Curr. Dir. Psychol. Sci.* **17**, 405–409 (2008).
28. Campbell, R. The processing of audio-visual speech: Empirical and neural bases. *Philos. Trans. R. Soc. B Biol. Sci.* **363**, 1001–1010 (2007).
29. Peelle, J. E. & Sommers, M. S. Prediction and constraint in audiovisual speech perception. *Cortex* **68**, 169–181 (2015).
30. Gao, C. et al. Audiovisual integration in the human brain: A coordinate-based meta-analysis. *Cereb. Cortex* **33**, 5574–5584 (2023).
31. Liberman, A. M. & Mattingly, I. G. The motor theory of speech perception revised. *Cognition* **21**, 1–36 (1985).
32. Skipper, J. I., Devlin, J. T. & Lametti, D. R. The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain Lang.* **164**, 77–105 (2017).
33. Franken, M. K., Liu, B. C. & Ostry, D. J. Towards a somatosensory theory of speech perception. *J. Neurophysiol.* **128**, 1683–1695 (2022).
34. Riecke, L., Snipes, S., van Bree, S., Kaas, A. & Hausfeld, L. Audio-tactile enhancement of cortical speech-envelope tracking. *Neuroimage* **202**, 116134 (2019).
35. Fu, X. & Riecke, L. Effects of continuous tactile stimulation on auditory-evoked cortical responses depend on the audio-tactile phase. *Neuroimage* **274**, 120140 (2023).
36. Guilleminot, P., Graef, C., Butters, E. & Reichenbach, T. Audiotactile stimulation can improve syllable discrimination through multisensory integration in the theta frequency band. *J. Cogn. Neurosci.* **35**, 1760–1772 (2023).
37. Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A. & Schroeder, C. E. Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* **53**, 279–292 (2007).
38. Damera, S. R. et al. Metamodal coupling of vibrotactile and auditory speech processing systems through matched stimulus representations. *J. Neurosci.* **43**, 4984–4996 (2023).
39. Harrison, S. C., Lawrence, R., Hoare, D. J., Wiggins, I. M. & Hartley, D. E. H. Use of functional near-infrared spectroscopy to predict and measure cochlear implant outcomes: A scoping review. *Brain Sci.* **11**, 1439 (2021).
40. Saliba, J., Bortfeld, H., Levitin, D. J. & Oghalai, J. S. Functional near-infrared spectroscopy for neuroimaging in cochlear implant recipients. *Hear. Res.* **338**, 64 (2016).
41. Rossi, S., Telkemeyer, S., Wartenburger, I. & Obrig, H. Shedding light on words and sentences: Near-infrared spectroscopy in language research. *Brain Lang.* **121**, 152–163 (2012).
42. Basura, G. J., Hu, X. S., Juan, J. S., Tessier, A. M. & Kovelman, I. Human central auditory plasticity: A review of functional near-infrared spectroscopy (fNIRS) to measure cochlear implant performance and tinnitus perception. *Laryngosc. Investig. Otolaryngol.* **3**, 463–472 (2018).
43. Dieler, A. C., Tupak, S. V. & Fallgatter, A. J. Functional near-infrared spectroscopy for the assessment of speech related tasks. *Brain Lang.* **121**, 90–109 (2012).

44. Wijayasiri, P., Hartley, D. E. H. & Wiggins, I. M. Brain activity underlying the recovery of meaning from degraded speech: A functional near-infrared spectroscopy (fNIRS) study. *Hear. Res.* **351**, 55–67 (2017).
45. Defenderfer, J., Kerr-German, A., Hedrick, M. & Buss, A. T. Investigating the role of temporal lobe activation in speech perception accuracy with normal hearing adults: An event-related fNIRS study. *Neuropsychologia* **106**, 31–41 (2017).
46. Zhou, X., Sobczak, G. S., McKay, C. M. & Litovsky, R. Y. Effects of degraded speech processing and binaural unmasking investigated using functional near-infrared spectroscopy (fNIRS). *PLoS ONE* **17**, e0267588 (2022).
47. Luke, R. et al. Analysis methods for measuring passive auditory fNIRS responses generated by a block-design paradigm. *Neurophotonics* **8**, 025008 (2021).
48. Mushtaq, F., Wiggins, I. M., Kitterick, P. T., Anderson, C. A. & Hartley, D. E. H. Evaluating time-reversed speech and signal-correlated noise as auditory baselines for isolating speech-specific processing using fNIRS. *PLoS ONE* **14**, e0219927 (2019).
49. Pollonini, L. et al. Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy. *Hear. Res.* **309**, 84–93 (2014).
50. Lawrence, R. J., Wiggins, I. M., Anderson, C. A., Davies-Thompson, J. & Hartley, D. E. H. Cortical correlates of speech intelligibility measured using functional near-infrared spectroscopy (fNIRS). *Hear. Res.* **370**, 53–64 (2018).
51. Zhang, M., Mary Ying, Y. L. & Ihlefeld, A. Spatial release from informational masking: Evidence from functional near infrared spectroscopy. *Trends Hear.* **22**, 1–12 (2018).
52. Olds, C. et al. Cortical activation patterns correlate with speech understanding after cochlear implantation. *Ear Hear.* **37**, e160–e172 (2016).
53. Sevy, A. B. G. et al. Neuroimaging with near-infrared spectroscopy demonstrates speech-evoked activity in the auditory cortex of deaf children following cochlear implantation. *Hear. Res.* **270**, 39–47 (2010).
54. Zhou, X. et al. Cortical speech processing in postlingually deaf adult cochlear implant users, as revealed by functional near-infrared spectroscopy. *Trends Hear.* **22**, 233121651878685 (2018).
55. Mushtaq, F., Wiggins, I. M., Kitterick, P. T., Anderson, C. A. & Hartley, D. E. H. The benefit of cross-modal reorganization on speech perception in pediatric cochlear implant recipients revealed using functional near-infrared spectroscopy. *Front. Hum. Neurosci.* **14**, 308 (2020).
56. Butera, I. M. et al. Functional localization of audiovisual speech using near infrared spectroscopy. *Brain Topogr.* **35**, 416 (2022).
57. Chen, L. C., Sandmann, P., Thorne, J. D., Herrmann, C. S. & Debener, S. Association of concurrent fNIRS and EEG signatures in response to auditory and visual stimuli. *Brain Topogr.* **28**, 710–725 (2015).
58. Chen, L.-C., Sandmann, P., Thorne, J. D., Bleichner, M. G. & Debener, S. Cross-modal functional reorganization of visual and auditory cortex in adult cochlear implant users identified with fNIRS. *Neural Plast.* **1**, 4382656 (2016).
59. Anderson, C. A., Wiggins, I. M., Kitterick, P. T. & Hartley, D. E. H. Adaptive benefit of cross-modal plasticity following cochlear implantation in deaf adults. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10256–10261 (2017).
60. Zhou, X., Innes-Brown, H. & McKay, C. Using fNIRS to study audio-visual speech integration in post-lingually deafened cochlear implant users. In *Proceedings of the International Symposium on Auditory and Audiological Research*, vol. 6, 55–62 (2017).
61. van de Rijt, L. P. H. et al. Temporal cortex activation to audiovisual speech in normal-hearing and cochlear implant users measured with functional near-infrared spectroscopy. *Front. Hum. Neurosci.* **10**, 48 (2016).
62. Steinmetzger, K., Shen, Z., Riedel, H. & Rupp, A. Auditory cortex activity measured using functional near-infrared spectroscopy (fNIRS) appears to be susceptible to masking by cortical blood stealing. *Hear. Res.* **396**, 108069 (2020).
63. Zimeo Morais, G. A., Balardin, J. B. & Sato, J. R. FNIRS optodes' location decider (fOLD): A toolbox for probe arrangement guided by brain regions-of-interest. *Sci. Rep.* **8**, 1–11 (2018).
64. van de Rijt, L. P. H. et al. Measuring cortical activity during auditory processing with functional near-infrared spectroscopy. *J. Hear. Sci.* **8**, 9–18 (2018).
65. Wiggins, I. M., Anderson, C. A., Kitterick, P. T. & Hartley, D. E. H. Speech-evoked activation in adult temporal cortex measured using functional near-infrared spectroscopy (fNIRS): Are the measurements reliable?. *Hear. Res.* **339**, 142–154 (2016).
66. Stein, B. E., Stanford, T. R., Ramachandran, R., Perrault, T. J. & Rowland, B. A. Challenges in quantifying multisensory integration: Alternative criteria, models, and inverse effectiveness. *Exp. Brain Res.* **198**, 113–126 (2009).
67. Fenwick, S. E., Davis, C., Best, C. T. & Tyler, M. D. The effect of modality and speaking style on the discrimination of non-native phonological and phonetic contrasts in noise. In *Auditory Visual Speech Processing* (2015).
68. Fuster-Duran, A. Perception of conflicting audio-visual speech: An examination across Spanish and German. *Speechreading by Humans and Machines: Models, Systems, and Applications.* 35–143 (1996).
69. Hazan, V. et al. The use of visual cues in the perception of non-native consonant contrasts. *J. Acoust. Soc. Am.* **119**, 1740–1751 (2006).
70. Godfrey Reggio. *Koyaanisqatsi* (1982).
71. Oganian, Y. & Chang, E. F. A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Sci. Adv.* **5**, eaay6279 (2019).
72. MacIntyre, A. D., Cai, C. Q. & Scott, S. K. Pushing the envelope: Evaluating speech rhythm with different envelope extraction techniques. *J. Acoust. Soc. Am.* **151**, 2002–2026 (2022).
73. MacIntyre, A. *AcousticLandmarks* (2021). Preprint at https://github.com/alexisdmacintyre/AcousticLandmarks.
74. Greenspan, J. D. & Bolanowski, S. J. The psychophysics of tactile perception and its peripheral physiological basis. In *Pain and Touch*, 25–103 (1996).
75. Peirce, J. et al. PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* **51**, 195–203 (2019).
76. Kothe, C. A. *LabStreamingLayer (Version 1.13)* (2014). Preprint at https://github.com/sccn/labstreaminglayer.
77. Gramfort, A. et al. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, 70133 (2013).
78. Abraham, A. et al. Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* **8**, 71792 (2014).
79. Seabold, S. & Perktold, J. Statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference* (2010).
80. Pollonini, L., Bortfeld, H. & Oghalai, J. S. PHOEBE: A method for real time mapping of optodes-scalp coupling in functional near-infrared spectroscopy. *Biomed. Opt. Express* **7**, 5104 (2016).
81. Fishburn, F. A., Ludlum, R. S., Vaidya, C. J. & Medvedev, A. V. Temporal derivative distribution repair (TDDR): A motion correction method for fNIRS. *Neuroimage* **184**, 171–179 (2019).
82. Cui, X., Bray, S. & Reiss, A. L. Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage* **49**, 3039 (2010).
83. Zhou, X., Sobczak, G., Colette, M. M. & Litovsky, R. Y. Comparing fNIRS signal qualities between approaches with and without short channels. *PLoS ONE* **15**, e0244186 (2020).
84. Santosa, H., Zhai, X., Fishburn, F., Sparto, P. J. & Huppert, T. J. Quantitative comparison of correction techniques for removing systemic physiological signal in functional near-infrared spectroscopy studies. *Neurophotonics* **7**, 035009 (2020).
85. Klein, F., Lührs, M., Benitez-Andonegui, A., Roehn, P. & Kranczioch, C. Performance comparison of systemic activity correction in functional near-infrared spectroscopy for methods with and without short distance channels. *Neurophotonics* **10**, 013503 (2022).
86. Cockx, H. et al. fNIRS is sensitive to leg activity in the primary motor cortex after systemic artifact correction. *Neuroimage* **269**, 119880 (2023).

87. Saager, R. B. & Berger, A. J. Direct characterization and removal of interfering absorption trends in two-layer turbid media. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **22**, 1874 (2005).
88. Scholkmann, F., Metz, A. J. & Wolf, M. Measuring tissue hemodynamics and oxygenation by continuous-wave functional near-infrared spectroscopy—How robust are the different calculation methods against movement artifacts?. *Physiol. Meas.* **35**, 717–734 (2014).
89. Glover, G. H. Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage* **9**, 416–429 (1999).
90. Correia, J. M., Jansma, B. M. B. & Bonte, M. Decoding articulatory features from fMRI responses in dorsal speech regions. *J. Neurosci.* **35**, 15015–15025 (2015).
91. Szycik, G. R., Tausche, P. & Münte, T. F. A novel approach to study audiovisual integration in speech perception: Localizer fMRI and sparse sampling. *Brain Res.* **1220**, 142–149 (2008).
92. Yücel, M. A. et al. The fNIRS Reproducibility Study Hub (FRESH): Exploring Variability and Enhancing Transparency in fNIRS Neuroimaging Research. *Commun. Biol.* (in press) (2025). Preprint at https://doi.org/10.31222/OSF.IO/PC6X8.
93. Vannest, J. J. et al. Comparison of fMRI data from passive listening and active-response story processing tasks in children. *J. Magn. Reson. Imaging* **29**, 971–976 (2009).
94. Nelson, A. J., Staines, W. R., Graham, S. J. & McIlroy, W. E. Activation in SI and SII; The influence of vibrotactile amplitude during passive and task-relevant stimulation. *Cogn. Brain Res.* **19**, 174–184 (2004).
95. Innes-Brown, H. & Crewther, D. The impact of spatial incongruence on an auditory-visual illusion. *PLoS ONE* **4**, e6450 (2009).
96. Spence, C. Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Ann. N. Y. Acad. Sci.* **1296**, 31–49 (2013).
97. Cappe, C. & Barone, P. Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *Eur. J. Neurosci.* **22**, 2886–2902 (2005).
98. Smiley, J. F. et al. Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J. Comp. Neurol.* **502**, 894–923 (2007).
99. Hackett, T. A. et al. Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *J. Comp. Neurol.* **502**, 924–952 (2007).
100. Schürmann, M., Caetano, G., Hlushchuk, Y., Jousmäki, V. & Hari, R. Touch activates human auditory cortex. *Neuroimage* **30**, 1325–1331 (2006).
101. Meredith, M. A. & Allman, B. L. Single-unit analysis of somatosensory processing in the core auditory cortex of hearing ferrets. *Eur. J. Neurosci.* **41**, 686–698 (2015).
102. Fu, K. M. G. et al. Auditory cortical neurons respond to somatosensory stimulation. *J. Neurosci.* **23**, 7510 (2003).
103. Occelli, V., Spence, C. & Zampini, M. Audiotactile interactions in temporal perception. *Psychon Bull. Rev.* **18**, 429–454 (2011).
104. Noesselt, T., Bergmann, D., Heinze, H. J., Münte, T. & Spence, C. Coding of multisensory temporal patterns in human superior temporal sulcus. *Front. Integr. Neurosci.* **6**, 30900 (2012).
105. Bolognini, N., Papagno, C., Moroni, D. & Maravita, A. Tactile temporal processing in the auditory cortex. *J. Cogn. Neurosci.* **22**, 1201–1211 (2010).
106. Pascual-Leone, A. & Hamilton, R. The metamodal organization of the brain. *Prog. Brain Res.* **134**, 427–445 (2001).
107. Heimler, B., Striem-Amit, E. & Amedi, A. Origins of task-specific sensory-independent organization in the visual and auditory brain: Neuroscience evidence, open questions and clinical implications. *Curr. Opin. Neurobiol.* **35**, 169–177 (2015).
108. Eickhoff, S. B., Grefkes, C., Zilles, K. & Fink, G. R. The somatotopic organization of cytoarchitectonic areas on the human parietal operculum. *Cereb. Cortex* **17**, 1800–1811 (2007).
109. Lamp, G. et al. Activation of bilateral secondary somatosensory cortex with right hand touch stimulation: A meta-analysis of functional neuroimaging studies. *Front. Neurol.* **10**, 1129 (2019).
110. Lin, Y. Y. & Forss, N. Functional characterization of human second somatosensory cortex by magnetoencephalography. *Behav. Brain Res.* **135**, 141–145 (2002).
111. Ruben, J. et al. Somatotopic organization of human secondary somatosensory cortex. *Cereb. Cortex* **11**, 463–473 (2001).
112. Schluppeck, D. & Francis, S. Secondary somatosensory cortex. In *Brain Mapping: An Encyclopedic Reference* Vol. 2 (ed. Toga, A. W.) 759–778 (Academic Press, 2015).
113. Rosenblum, L. D. Primacy of multimodal speech perception. In *The Handbook of Speech Perception*, 51–78 (2005).
114. Meredith, M. A. et al. What is a multisensory cortex? A laminar, connectional, and functional study of a ferret temporal cortical multisensory area. *J. Comp. Neurol.* **528**, 1864–1882 (2020).
115. Land, R., Engler, G., Kral, A. & Engel, A. K. Auditory evoked bursts in mouse visual cortex during isoflurane anesthesia. *PLoS ONE* **7**, e49855 (2012).
116. Stevenson, R. A. et al. Identifying and quantifying multisensory integration: A tutorial review. *Brain Topogr.* **27**, 707–730 (2014).
117. Noppeney, U. Characterization of multisensory integration with fMRI: Experimental design, statistical analysis, and interpretation. In *The Neural Bases of Multisensory Processes*, 233–252 (CRC Press, 2011).
118. James, T. W. & Stevenson, R. A. The use of fMRI to assess multisensory integration. In *The Neural Bases of Multisensory Processes*, 131–146 (CRC Press, 2011).
119. Beauchamp, M. S. Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics* **3**, 093–114 (2005).
120. Erickson, L. C. et al. Distinct cortical locations for integration of audiovisual speech and the McGurk effect. *Front. Psychol.* **5**, 83600 (2014).
121. Hoefer, M. et al. Tactile stimulation and hemispheric asymmetries modulate auditory perception and neural responses in primary auditory cortex. *Neuroimage* **79**, 371–382 (2013).
122. Goebel, R. & Van Atteveldt, N. Multisensory functional magnetic resonance imaging: A future perspective. *Exp. Brain Res.* **198**, 153–164 (2009).
123. Talsma, D., Senkowski, D., Soto-Faraco, S. & Woldorff, M. G. The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* **14**, 400–410 (2010).
124. Rockland, K. S. & Drash, G. W. Collateralized divergent feedback connections that target multiple cortical areas. *J. Comp. Neurol.* **373**, 529–548 (1996).
125. Shipp, S. Structure and function of the cerebral cortex. *Curr. Biol.* **17**, R443–R449 (2007).
126. Larkum, M. A cellular mechanism for cortical associations: An organizing principle for the cerebral cortex. *Trends Neurosci.* **36**, 141–151 (2013).
127. Wagstyl, K., Ronan, L., Goodyer, I. M. & Fletcher, P. C. Cortical thickness gradients in structural hierarchies. *Neuroimage* **111**, 241–250 (2015).
128. Murray, M. M. et al. Grabbing your ear: Rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb. Cortex* **15**, 963–974 (2005).
129. Narain, C. et al. Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb. Cortex* **13**, 1362–1368 (2003).
130. Peelle, J. E. The hemispheric lateralization of speech processing depends on what "speech" is: A hierarchical perspective. *Front. Hum. Neurosci.* **6**, 309 (2012).
131. Abrams, D. A., Nicol, T., Zecker, S. & Kraus, N. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J. Neurosci.* **28**, 3958 (2008).
132. Poeppel, D. The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* **41**, 245–255 (2003).

133. Ramos-Loyo, J., González-Garrido, A. A., Llamas-Alonso, L. A. & Sequeira, H. Sex differences in cognitive processing: An integrative review of electrophysiological findings. *Biol. Psychol.* **172**, 108370 (2022).
134. Koles, Z. J., Lind, J. C. & Flor-Henry, P. Gender differences in brain functional organization during verbal and spatial cognitive challenges. *Brain Topogr.* **23**, 199–204 (2010).
135. Wegner, K., Forss, N. & Salenius, S. Characteristics of the human contra-versus ipsilateral SII cortex. *Clin. Neurophysiol.* **111**, 894–900 (2000).
136. Tommerdahl, M. et al. Response of SII cortex to ipsilateral, contralateral and bilateral flutter stimulations in the cat. *BMC Neurosci.* **6**, 1–14 (2005).
137. Morillon, B., Liégeois-Chauvel, C., Arnal, L. H., Bénar, C. G. & Giraud, A. L. Asymmetric function of theta and gamma activity in syllable processing: An intra-cortical study. *Front. Psychol.* **3**, 248 (2012).
138. Zion Golumbic, E. M., Poeppel, D. & Schroeder, C. E. Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain Lang.* **122**, 151–161 (2012).
139. Kösem, A., Basirat, A., Azizi, L. & van Wassenhove, V. High-frequency neural activity predicts word parsing in ambiguous speech streams. *J. Neurophysiol.* **116**, 2497–2512 (2016).
140. Gwilliams, L., Marantz, A., Poeppel, D. & King, J. R. Top-down information shapes lexical processing when listening to continuous speech. *Lang. Cogn. Neurosci.* **39**(8), 1045–1058 (2023).
141. Friese, U. et al. Oscillatory brain activity during multisensory attention reflects activation, disinhibition, and cognitive control. *Sci. Rep.* **6**, 32775 (2016).
142. Bidet-Caulet, A., Bottemanne, L., Fonteneau, C., Giard, M. H. & Bertrand, O. Brain dynamics of distractibility: Interaction between top-down and bottom-up mechanisms of auditory attention. *Brain Topogr.* **28**, 423–436 (2015).
143. Büchel, C. & Friston, K. J. Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cereb. Cortex* **7**, 768–778 (1997).
144. Frith, C. & Dolan, R. J. Brain mechanisms associated with top-down processes in perception. *Philos. Trans. R. Soc. B Biol. Sci.* **352**, 1221 (1997).
145. Clayton, M. S., Yeung, N. & Cohen Kadosh, R. The roles of cortical oscillations in sustained attention. *Trends Cogn. Sci.* **19**, 188–195 (2015).
146. Yusuf, P. A., Hubka, P., Tillein, J., Vinck, M. & Kral, A. Deafness weakens interareal couplings in the auditory cortex. *Front. Neurosci.* **14**, 625721 (2021).
147. Zatorre, R. J. & Halpern, A. R. Mental concerts: Musical imagery and auditory cortex. *Neuron* **47**, 9–12 (2005).
148. Herholz, S. C., Halpern, A. R. & Zatorre, R. J. Neuronal correlates of perception, imagery, and memory for familiar tunes. *J. Cogn. Neurosci.* **24**, 1382–1397 (2012).
149. Ruiz-Stovel, V. D., González-Garrido, A. A., Gómez-Velázquez, F. R., Alvarado-Rodríguez, F. J. & Gallardo-Moreno, G. B. Quantitative EEG measures in profoundly deaf and normal hearing individuals while performing a vibrotactile temporal discrimination task. *Int. J. Psychophysiol.* **166**, 71–82 (2021).
150. Goswami, U. et al. Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proc. Natl. Acad. Sci.* **99**, 10911–10916 (2002).

Acknowledgements

The authors thank all test participants enrolled in this study for their time and efforts.

Author contributions

All authors contributed to conceiving the study and approved the final version of the manuscript. A.S. collected and analyzed the data, and wrote the original draft of the manuscript. H.I., J.M. and A.K. acquired funding, supervised all experimental phases and revised the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. Open Access funding enabled and organized by Projekt DEAL. This work was supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No 860755. Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-07718-8.

Correspondence and requests for materials should be addressed to A.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.