NeuroImage 202 (2019) 116134

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Audio-tactile enhancement of cortical speech-envelope tracking

Lars Riecke [a,*,1], Sophia Snipes [a,b,1], Sander van Bree [a,c], Amanda Kaas [a], Lars Hausfeld [a]

[a] Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Oxfordlaan 55, 6229, EV Maastricht, the Netherlands
[b] Child Development Center, University Children's Hospital Zurich, University of Zurich, Steinwiesstrasse 75, 8032, Zurich, Switzerland
[c] MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, CB2 7EF, United Kingdom

ARTICLE INFO

Keywords:
Electroencephalography
Multisensory
Speech envelope
Speech tracking
Tactile

ABSTRACT

Viewing a speaker's lip movements can improve the brain's ability to 'track' the amplitude envelope of the auditory speech signal and facilitate intelligibility. Whether such neurobehavioral benefits can also arise from tactually sensing the speech envelope on the skin is unclear. We hypothesized that tactile speech envelopes can improve neural tracking of auditory speech and thereby facilitate intelligibility. To test this, we applied continuous auditory speech and vibrotactile speech-envelope-shaped stimulation at various asynchronies to the ears and index fingers of normally-hearing human listeners while simultaneously assessing speech-recognition performance and cortical speech-envelope tracking with electroencephalography. Results indicate that tactile speech-shaped envelopes improve the cortical tracking, but not intelligibility, of degraded auditory speech. The cortical speech-tracking benefit occurs for tactile input leading the auditory input by 100 m s or less, emerges in the EEG during an early time window (~0–150 m s), and in particular involves cortical activity in the delta (1–4 Hz) range. These characteristics hint at a predictive mechanism for multisensory integration of complex slow time-varying inputs that might play a role in tactile speech communication.

1. Introduction

Comprehension of auditory speech is an integral part of social life and communication. Unfortunately, the intelligibility of auditory speech can be hampered by hearing impairments or noisy environments. Under these adverse conditions, intelligibility benefits from speech-relevant information in other non-auditory modalities. Viewing a talker's articulatory movements improves intelligibility (Sumby and Pollack, 1954), even if the visual information leads the acoustic input by more than ~200 m s (Grant and Greenberg, 2001). It is less known that also tactile information can facilitate speech perception. Audio-tactile research has shown that normally-hearing, non-trained listeners can exploit tactile temporal information (e.g., air puffs or skin stretches applied to the hand or neck) to detect or identify auditory syllables near threshold (Blamey et al., 1989; Fowler and Dekle, 1991; Gick et al., 2008, 2010; Gick and Derrick, 2009; Ito et al., 2009; Sato et al., 2010; Tjan et al., 2014). Some trained hearing-impaired listeners even benefit from tactile speech aids such as spatiotemporal displays resembling multi-channel tactile vocoders (Kirman, 1973; Weisenberger and Miller, 1987; Working Group on Communication Aids for the Hearing-Impaired, 1991; Rizza et al., 2018) or the Tadoma method, in which speech is received by placing a hand on the talker's face (Reed et al., 1982). Beyond these perceptual benefits, the underlying neural mechanisms are still poorly understood. Neural integration of basic audio-tactile non-speech stimuli occurs in the auditory cortex (Foxe et al., 2002; Fu et al., 2003; Brosch, 2005; Kayser et al., 2005; Caetano and Jousmäki, 2006; Schürmann et al., 2006; Lakatos et al., 2007) in a rapid (within 200 m s), automatic fashion (Foxe et al., 2000; Lütkenhöner et al., 2002; Gobbelé et al., 2003; Murray et al., 2004; Butler et al., 2011), possibly via intracerebral auditory-somatosensory anatomical connections (Hackett et al., 2007; Ro et al., 2013). It is still unknown whether these basic findings (review: Soto-Faraco and Deco, 2009) generalize to more complex signals such as natural speech.

We hypothesized that a mechanism for audio-tactile speech integration is the tracking of slow fluctuations (<20 Hz) in the speech-signal amplitude (i.e. the speech envelope) present also in the tactile input (reviews: Peelle and Davis, 2012; Zion Golumbic et al., 2012; Ding and Simon, 2014). This speech-envelope tracking is thought to reflect the accuracy with which the cortex encodes temporal speech features. Cortical speech-envelope tracking and speech intelligibility mutually influence each other. Increasingly degraded auditory speech results in increasingly reduced cortical speech-envelope tracking (Peelle et al.,

* Corresponding author.
E-mail address: l.riecke@maastrichtuniversity.nl (L. Riecke).
[1] Equal contribution.

https://doi.org/10.1016/j.neuroimage.2019.116134
Received 18 April 2019; Received in revised form 7 August 2019; Accepted 26 August 2019
Available online 27 August 2019

2013; Kong et al., 2015), while lip reading enhances cortical speech-envelope tracking (Zion Golumbic et al., 2013; Crosse et al., 2015, 2016b; Park et al., 2016). Conversely, impairing cortical speech-envelope tracking with transcranial electric current stimulation results in impaired speech intelligibility (Riecke et al., 2018). Based on these neural findings and the behavioral audio-tactile findings above, we reasoned that tactile speech envelopes can enhance auditory speech intelligibility by enhancing speech-envelope tracking in the cortex.

To test this idea, we simultaneously presented degraded (envelope-reduced) continuous auditory speech and speech envelope-shaped vibrotactile stimulation to normally-hearing listeners, while assessing the cortical tracking and intelligibility of the speech (using electroencephalography [EEG] and a speech-recognition task, respectively). To extract and characterize the hypothesized audio-tactile speech-envelope integration, we included unisensory stimulation conditions and manipulated the asynchrony of the audio-tactile stimuli. We predicted that adding synchronous tactile speech envelope to the auditory speech would have a supra-additive effect on cortical speech-envelope tracking and listeners' speech-recognition performance.

2. Materials and Methods

2.1. Participants

Eighteen healthy, normally hearing volunteers (ages: 18–45 years, 11 females, 16 right-handed) participated in the study. They gave their written informed consent and received study credits or monetary reward for their participation. The experimental procedure was approved by the local research ethics committee (*Ethical Review Committee Psychology and Neuroscience,* Maastricht University, #165_04_04_2016).

2.2. Stimuli

2.2.1. Auditory stimuli

Auditory stimuli consisted of sequences of meaningful everyday sentences (e.g., "Hij deed zijn ogen open" [*He opened his eyes*]) recorded from various male and female Dutch speakers (Oostdijk, 2000; Versfeld et al., 2000). An excerpt from an exemplary sentence-recording waveform is shown in Fig. 1A (top).

Sentences from the corpus by Oostdijk (2000) were selected with the following criteria: (1) the transcription only contained the lowercase Dutch alphabet, thus excluding questions, exclamations, subordinates, hyphenations, proper names, and foreign characters; (2) sentences starting with "en", "maar", "daar", "er", "met" (*and, but, there, with*), or containing words with a frequency rank lower than 30,000 (determined by the corpus by Oostdijk (2000)) were excluded; (3) only audio files with a duration between 1.7 and 5s were kept. This resulted in a set of 10,500 sentences. Lastly, a Dutch native speaker manually selected 1,247 sentences based on the following criteria: (1) the audio was clear; (2) there was no surprising, political, or emotional content; (3) there were no difficult or archaic expressions. Sentences from the corpus by Versfeld et al. (2000) consisted of eight or nine syllables, distributed over four to nine words, with a maximum of three syllables per word. From a joint pool of 2,261 sentences, 1,770 were randomly selected for the experiment, half of them spoken by a male and the other half spoken by a female. The recordings were equated for root-mean square level and appended to each other (inter-sentence interval: 400 m s) to create sequences alternating male and female speakers (Fig. 1B top, gray waveform). The final sentence of each sequence was always from Versfeld et al. (2000) and was used to assess intelligibility (see Task and Experimental Design). The number of sentences per sequence varied pseudo-randomly between one and seven, resulting in an overall sequence duration between 1.8s and 20.4s. The majority of sequences (70%) were at least 16s long to facilitate the EEG speech-tracking analysis. The remaining shorter sequences were added to ensure that participants could not reliably predict the onset of the test sentence and to encourage participants to pay attention to all sentences.

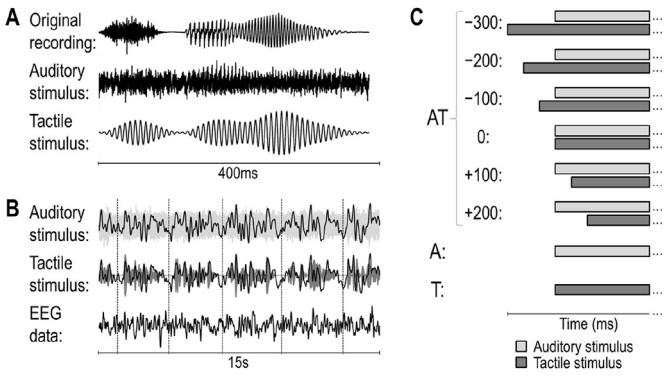

Fig. 1. Audio-Tactile Stimuli and Experimental Design.
A. The top waveform shows an original excerpt (the Dutch word "zwemmen") from an exemplary sentence recording. The middle waveform shows the same recording after removal of the speech envelope; this degraded version was presented as the auditory stimulus. The bottom row shows a sinusoidal carrier modulated by the removed speech envelope; this version was presented as the tactile stimulus.
B. Data are shown from a single trial in the synchronous audio-tactile condition (condition 0; see panel C). The waveforms illustrate the final 15s of a sentence sequence that was presented as auditory stimulus (light gray band) and tactile stimulus (dark gray band), and a participant's evoked EEG activity (channel Cz, lowpass-filtered below 20 Hz). The speech envelope that was reconstructed from the EEG is superimposed in black on the stimulus waveforms (top and middle). The vertical lines indicate the sentence structure. Speech intelligibility was assessed with the last sentence of the sequence.
C. The rows of this diagram illustrate the different experimental conditions. Rectangles represent the *initial portions* of auditory (light gray) and tactile (dark gray) stimulation intervals in a given trial. The design included audio-tactile (AT) conditions, a purely auditory (A) condition, and a purely tactile (T) condition. Additionally in condition AT, the delay of the tactile stimulation ('tactile lag') was experimentally manipulated in six steps (−300, −200, …, +200 m s) to identify a time window of audio-tactile speech integration.

To promote multisensory enhancement according to the principle of inverse effectiveness (Alex Meredith and Stein, 1986; Crosse et al., 2016b), adverse listening conditions were created by degrading the auditory speech. This was done by reducing the speech envelope, which was achieved by passing the concatenated original audio recordings through a 30-channel vocoder, extracting the envelope from each channel signal, and removing its low-frequency (<16 Hz) portion. The high-frequency (>64 Hz) portion was retained and its amplitude was scaled to define the amount of speech degradation (for details, see Riecke et al., 2018). The amount of degradation was adjusted individually to fix speech intelligibility across participants in the EEG experiment (see Procedure). The degradation procedure resulted in a reduction of the temporal structure, as can be seen from the exemplary auditory stimulus shown in Fig. 1B (top, gray waveform).

2.2.2. Tactile stimuli

To promote multisensory speech integration, the tactile stimuli were designed to carry information that was complementary to the information carried by the underspecified auditory speech. This was achieved by shaping the tactile stimuli exactly as the envelopes removed from the original auditory speech. This was done by squaring the low-frequency portion of the channel envelopes (see Auditory Stimuli), summing across channels, normalizing the resulting composite speech envelope, and multiplying it with a 128-Hz sinusoidal carrier. The carrier served to drive Pacinian mechanoreceptors, which are sensitive to vibrations between 60 and 400 Hz (Kandel et al., 2000). An exemplary tactile stimulus is shown in Fig. 1A (bottom) and 1B (middle). The normalized composite speech envelope was used for all speech-tracking analyses (see below, section Assessment of Speech-Envelope Tracking).

2.2.3. Audio-tactile stimulus presentation

Auditory and tactile stimuli were digitally generated using a sampling rate of 16 kHz and then converted collectively to analog signals using a multi-channel D/A converter (National Instruments). Stimulus timing was controlled using Datastreamer software (ten Oever et al., 2016). Auditory stimuli were presented diotically at 68 dB SPL via insert earphones (EARTone 3A) attenuating ambient noise by 30 dB or more. Tactile stimuli were presented bimanually via a multi-channel stimulator (bandwidth: 1–500 Hz; Mini PTS, Dance Design, UK) and piezoelectric transducers attached to the participants' left and right index fingers with Velcro. The transducer houses a vertically moving disk (diameter: 6 mm) centered in a static aperture (diameter: 8 mm). The maximum mechanical power delivered to the skin was 75 mW (corresponding to a disk movement within the range of ±0.5 mm) on each trial.

2.3. EEG recording

EEG was recorded with 62 scalp electrodes positioned according to a modified 10–20 system (Easycap montage 11) and a reference electrode above the left mastoid, using BrainAmp amplifiers (Brain Products, Munich, Germany) decoupled from the stimulation system. Electrooculography was recorded with additional electrodes below and next to the left eye. Inter-electrode impedances were kept below 5kΩ. The EEG recordings were bandpass-filtered (cutoffs: 0.01 and 200 Hz, analog filter) and digitized with a sampling rate of 500 Hz.

2.4. Task and Experimental Design

Our two measures of interest were the intelligibility of the auditory speech and the cortical tracking of the speech envelope. Speech intelligibility was measured using a speech-recognition task requiring participants to listen to the sentence sequence and verbally repeat as many words of the final sentence as possible. Trials consisted of a stimulation interval containing the sentence sequence and a subsequent response interval (duration: 5s, cued by a color change of a fixation cross) during which participants' verbal responses were recorded. Participants were instructed to keep still and avoid eye movements while they received stimulation. They were further informed that stimuli would be auditory, tactile, or both, and that some of the latter stimuli would be asynchronous. Cortical speech-envelope tracking was measured by applying a system-identification approach to the speech envelopes and the recorded EEG data (Crosse et al., 2015). More specifically, regularized linear regression and backward modelling were used to determine the accuracy with which the speech envelope could be reconstructed from the multi-channel EEG data (for details, see section Data Analysis below).

The within-subject experimental design is illustrated in Fig. 1C. It included audio-tactile stimulation conditions (AT), a purely auditory stimulation condition (A), and a purely tactile stimulation condition (T), which collectively served to extract audio-tactile speech integration. In condition AT, the tactile stimulation was delayed by −300, −200, −100, 0, +100, or +200 in ms relative to the auditory stimulation (negative values indicate tactile leading). This experimental manipulation of the 'tactile lag' served to identify the time window for audio-tactile speech integration.

Each run of the experiment consisted of on average 42 trials from all conditions and lasted in total 14 min. Each sentence was presented only once per participant, except for the two unisensory conditions (A, T), in which the same sentence sequences were presented for each modality. The assignment of sentences to conditions and the order of trials within runs were individually randomized.

2.5. Procedure

The experimental procedure involved the following steps: first, participants were screened for hearing impairments using a questionnaire or pure-tone audiometry (hearing threshold <25 dB HL at 0.25, 0.5, 1, 2, 4, or 6 kHz). Second, they were familiarized with the degraded auditory stimuli for ~30min while the EEG cap was applied. Third, they were seated in a sound-attenuated, electrically-shielded chamber isolated from the experimenter, and their auditory speech-recognition threshold was measured. The threshold was defined by fitting the data obtained on 110 trials of five degradation levels with a psychometric function and identifying the amount of degradation yielding an estimated starting performance level of 40%. Finally, the amount of auditory speech degradation was fixed to the individually identified threshold and seven to ten (average: 9.4) runs of the experiment were conducted while EEG was simultaneously recorded, with breaks between runs. Due to the extended duration of the experiment, participants could end the experiment at their own discretion.

2.6. Data Analysis

Of the 395 trials that were on average presented, only trials lasting 16s or longer (in total ~272 trials, ~35 per condition) were retained. This allowed for an EEG epoch duration that was fixed and sufficiently long for analyzing speech tracking across the long range of tactile lags (see below, Assessment of Speech-Envelope Tracking). As mentioned previously (see Auditory Stimuli), shorter trials only served to keep participants' attention constant within trials. Participants' verbal responses were scored by a blinded native Dutch speaker (author S.B.).

2.6.1. Assessment of Speech intelligibility

Speech-recognition performance was computed as the percentage of correctly recognized words pooled over all trials per condition. The order in which words were reported was disregarded. Reports of words not present in the sentence were excluded from the analysis.

2.6.2. EEG data preprocessing

EEG data were preprocessed using EEGLAB 14.1.2 (Delorme and Makeig, 2004). The preprocessing involved bandpass-filtering between 0.5 and 45 Hz (FIR filter with zero phase shift, filter order: 3300), reduction of artifacts (Artifact Subspace Reconstruction method; Chang et al., 2018), and re-referencing to an average reference. For speech-envelope tracking analysis (see next section), the preprocessed EEG data and the associated speech envelopes (see Tactile Stimuli) were further lowpass-filtered below 20 Hz, resampled at 64 Hz, and z-scored after epoching each trial from −15.3s to −0.3s relative to the end of the stimulation interval, to exclude potential responses to onsets or offsets of asynchronous stimuli.

2.6.3. Assessment of Speech-envelope tracking

Cortical tracking of the original (non-degraded) speech envelope carried by the tactile stimuli was assessed using a stimulus-reconstruction approach and a leave-one-trial-out cross-validation procedure as implemented in the mTRF toolbox (Crosse et al., 2016a). First, a multivariate linear function ('decoder') that optimally mapped the measured spatio-temporal EEG-response pattern onto the associated speech envelope (see Tactile Stimuli) was determined for all but one trial of a given condition. The decoder was estimated at 64 time lags (referred to here as 'cortical lags' to distinguish them from the tactile lags) between −300 m s and +700 m s representing the timing of cortical activity relative to the continuous speech envelope. This relatively long range was chosen to take into account the long range of tactile lags. The regularization parameter λ was optimized (range: 10^0,10^1, …, 10^5) in a nested leave-one-trial-out cross-validation procedure using the same subset of trials. Second, the average of the optimal single-trial decoders (i.e., decoders with regularizations yielding maximum performance in the nested loop) was used to reconstruct a speech envelope from the EEG-response pattern measured on the left-out trial (e.g., see Fig. 1B top and middle, black waveform). Third, the speech-envelope reconstruction accuracy (i.e., the performance of the decoder) was assessed as the linear correlation (Pearson's R) between the reconstructed speech envelope and the

actual speech envelope (see Tactile Stimuli) associated with the left-out trial. Fourth, the first three steps were iterated leaving out a different trial each time. Fifth, the single-trial accuracies obtained on all iterations were Fisher-transformed and averaged. The resulting average accuracy R_z, which was taken to quantify the accuracy of speech tracking in the given condition, was statistically compared to the empirical chance level. The latter was estimated separately for each condition by iteratively shuffling (500 iterations) the presented speech envelopes across trials within the condition and repeating all the steps described above on each iteration (Hausfeld et al., 2018).

2.6.4. Assessment of audio-tactile speech integration

Audio-tactile speech integration was assessed using a supra-additivity criterion (Stein and Meredith, 1993; Stevenson et al., 2014). According to this criterion, multisensory integration can be inferred if the response pattern to a multisensory stimulus differs from the sum of the response patterns to the unisensory stimuli (e.g., AT ≠ A + T). We applied this criterion in our analysis of speech intelligibility as follows: Listeners' speech-recognition performance in condition A was subtracted from that in condition AT to define a Multisensory Integration Index (MSI, in units of percentage points [pp]). Note that performance in condition T was not assessed in our study because the purely tactile speech was unrecognizable. MSI values above zero were interpreted as audio-tactile speech integration. For reference the theoretical upper limit of MSI was estimated based on previous speech-recognition data obtained in a purely auditory condition where the broad-band speech envelope was presented as a component of the auditory, not tactile, stimuli (for details, see Fig. S3 in Riecke et al., 2018).

Application of the supra-additivity criterion in the speech-envelope tracking analysis was done similarly to Crosse et al. (2015): First, for each unisensory condition a unisensory speech-envelope decoder was determined (see above, Assessment of Speech-Envelope Tracking) for each value of λ (10^0,10^1, …, 10^5).

Second, for each possible λ combination, the unweighted sum of the two unisensory decoders (decoder A and decoder T) was computed to define a reference decoder A + T. This summed reference decoder served as an estimate of the purely linear superposition of synchronous unisensory (auditory and tactile) cortical response patterns, resembling the case that no multisensory integration occurs.

Third, for each λ combination, the accuracy with which the reference decoder A + T could reconstruct the speech envelope in the multisensory condition was assessed. The maximum (λ combination-specific) accuracy was taken to define a baseline, which was used to extract supra-additive effects (see next two steps below). Initial analyses using lag-specific reference decoders yielded relatively low baseline values for all asynchronous conditions, leading to inflated supra-additive effects (for details, see Supplementary Material). We therefore chose to use a non-lagged (lag-unspecific) reference decoder A + T (derived from the fully synchronous condition 0), which provided the most conservative baseline for all analyses.

Fourth, a neural MSI (ΔR_z, a differential correlation coefficient) was defined for each multisensory condition. This was done by subtracting the baseline (i.e., the accuracy with which the reference decoder A + T could reconstruct the speech envelope in condition 0) from the reconstruction accuracy that was observed in the conventional speech-tracking analysis of the given multisensory condition (decoder AT; see above, Assessment of Speech-Envelope Tracking).

Finally, the supra-additivity criterion was applied. That is, as for the speech-recognition data, we interpreted values of MSI above zero as audio-tactile speech integration (Crosse et al., 2015).

2.6.5. Analysis of spectral, temporal, and spatial characteristics of audio-tactile speech-envelope integration

The contribution of cortical oscillatory activity in different frequency bands to audio-tactile speech-envelope integration was assessed by bandpass-filtering the preprocessed EEG data and the associated speech envelopes into the delta range (1–4 Hz), theta range (4–8 Hz), or alpha range (8–15 Hz). To assess the timing of audio-tactile speech-envelope integration in the cortex, the stimulus-reconstruction approach was applied separately at each individual cortical lag instead of jointly to all these lags, yielding a reconstruction accuracy for each lag. Finally, to assess the spatial distribution of audio-tactile speech-envelope integration across the scalp, a forward modelling approach (i.e., an EEG-prediction approach analogous to the stimulus-reconstruction approach above) was applied separately to each EEG channel. The accuracy with which this model could predict the EEG channel data from the presented speech envelopes was assessed using linear correlation. This yielded a spatial map of correlation coefficients, which was averaged across participants. The similarity of the average maps obtained from different conditions or decoders was assessed using linear correlation of the vectorized average maps. The resulting correlation coefficients between maps were qualitatively compared.

2.6.6. Statistical testing

Participants' individual measures ([differential] speech-recognition performance and speech-envelope reconstruction accuracy) were submitted to second-level (random-effects) group analyses using parametric statistical tests (t-tests, ANOVAs) for repeated measures. Violation of sphericity was compensated by using Greenhouse-Geisser correction. A significance criterion $\alpha = 0.05$ was used. Type-I error probabilities (P values) inflated by multiple comparisons were corrected by controlling the false-discovery rate (Benjamini and Hochberg, 1995). For analysis of lag-specific reconstruction accuracies, non-parametric statistics (based on 500 permutations) and a multiple comparison correction based on a lag-cluster size criterion were used (Maris and Oostenveld, 2007). Reported summary statistics are mean and s.e.m. across all participants.

3. Results

3.1. Speech intelligibility

Fig. 2A shows results from analysis of participants' speech-recognition performance. They recognized on average 61.7 ± 1.9% of the words correctly (condition AT: 61.9%, condition A: 60.7%), which was higher than the 40% criterion we used to define threshold. Comparison of the first vs. second half of trials revealed no significant difference for any condition in the EEG experiment (all $t_{17}<2.09$, $P>0.37$), however there was a significant improvement from the first to second half of the threshold procedure ($t_{17} = -3.07$, $P = 0.0035$). This indicates that the aforementioned above-threshold performance arose from a learning effect during threshold estimation, not the main EEG

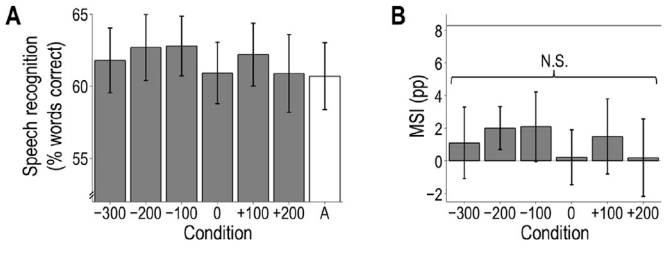

Fig. 2. Speech-Intelligibility Results.
A. Average speech-recognition performance (mean ± s.e.m. across all participants) is shown for each experimental condition (except for condition T, which involved no audible speech). Condition labels are defined in Fig. 1C.
B. The average multisensory integration index (mean ± s.e.m. across all participants; obtained by subtracting condition A from condition AT) is shown for each audio-tactile condition. Although the index was positive in all conditions, it was not significantly above zero or different across tactile lags, providing no evidence for audio-tactile integration in the context of auditory speech intelligibility. The horizontal gray line represents the theoretical upper limit of the index (see Materials and Methods). N.S., non-significant.

experiment. Fig. 2B shows the MSI for speech intelligibility (see Assessment of Multisensory Integration) as a function of the tactile lag. Positive values of this index were observed at all lags and the maximum value occurred when the tactile input led the auditory speech by 100 m s (condition −100: 2.1pp, which is 6.2pp below the theoretical upper limit of MSI). However, contrary to our prediction of a supra-additive effect, the index was not significantly above zero for any tactile lag (all $t_{17}<1.49$, $P>0.46$) or averaged across these lags ($t_{17}=0.70$, $P=0.25$). It also did not vary significantly across tactile lags (no main effect of tactile lag: $F_{4.1,69.8}=0.39$, $P=0.82$). Thus, these results provide no strong evidence for audio-tactile integration in the context of auditory speech intelligibility.

3.2. Cortical speech-envelope tracking

Fig. 3A shows results from the analysis of cortical speech-envelope tracking. The accuracy of speech-envelope reconstruction was on average 0.23 ± 0.02 (mean ± s.e.m. non-normalized correlation coefficient R across all participants) and significantly above chance in all conditions (all $t_{17}>8.67$, $P<10^{-7}$). It did not vary significantly across tactile lags (no main effect of tactile lag: $F_{3.7,62.1}=1.90$, $P=0.13$). However, it was significantly higher in the purely tactile condition relative to the purely auditory condition ($t_{17}=3.30$, $P=0.004$), reflecting the fact that speech envelope was intact in the tactile stimuli and reduced in the auditory stimuli. Moreover, reconstruction accuracy showed a positive correlation with listeners' speech-recognition performance ($R=0.51$, $P=0.022$, averaged across conditions), as shown in Fig. 4A.

Fig. 3B shows the neural MSI as a function of the tactile lag. Positive values were observed at all lags and the maximum value occurred in the synchronous condition (condition 0: MSI = 0.025). Consistent with our prediction of a supra-additive effect, the neural MSI averaged across tactile lags was significantly above zero ($t_{17}=2.39$, $P=0.028$). This effect was significant when the tactile input led the auditory speech by 100 m s or less (condition −100: $t_{17}=2.71$, $P=0.045$; condition 0: $t_{17}=5.03$, $P=0.00061$), but not in the other conditions (all $t_{17}<2.21$, $P>0.082$). Moreover, neural MSI across participants did not correlate

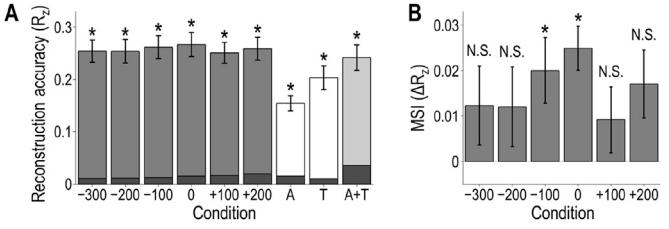

Fig. 3. Cortical Speech-Envelope Tracking Results.
A. The bars show average speech-envelope reconstruction accuracy (mean ± s.e.m. across all participants) for each experimental condition (dark gray: audio-tactile conditions, white: unisensory conditions) and a baseline obtained with reference decoder A + T (light gray: condition A + T; see Materials and Methods). The baseline resembled the case that synchronous unisensory cortical responses add up in a purely linear (non-integrative) manner. The shaded area at the bottom denotes the upper limit of the 95% confidence interval for the average empirical chance level (mean across all participants) for each condition. Reconstruction accuracy was significantly above chance in every condition. *, $P<0.05$. See also Supplementary Fig. 1.
B. The bar plot shows the average multisensory integration index (mean ± s.e.m. across all participants) for each audio-tactile condition. This index was obtained by subtracting the baseline (panel A, light gray bar) from the multisensory conditions (panel A, dark gray bars) to extract supra-additive effects, which were taken to index audio-tactile speech-envelope integration. The plot shows that the index was significantly above zero specifically when the tactile input led the auditory speech by 100 m s or less (conditions −100 and 0), providing evidence for a time window of audio-tactile integration in the context of cortical speech-envelope tracking, given our definitions. N.S., non-significant.

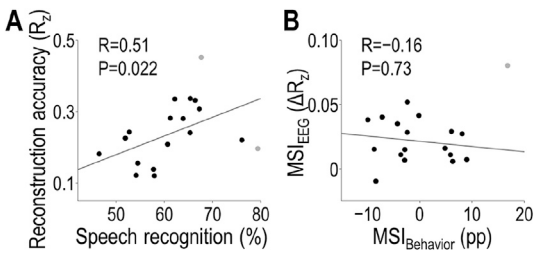

Fig. 4. Link between Speech Intelligibility and Cortical Speech-Envelope Tracking.
A. The scatterplot indicates a positive linear association between speech-recognition performance and speech-envelope reconstruction accuracy (both averaged across conditions). Each dot represents an individual participant. Light gray dots represent outliers (participants exhibiting values more than two standard deviations away from the mean) that were excluded from this specific analysis.
B. The scatterplot shows no significant linear association between the behavioral multisensory integration index (based on speech intelligibility) and the neural multisensory integration index (based on cortical speech-envelope tracking) in the synchronous condition (condition 0).

significantly with the behavioral MSI (condition −100: $R=0.025$, $P=0.73$; condition 0: $R=-0.16$, $P=0.73$) as shown in Fig. 4B. Overall, these results provide evidence for audio-tactile integration in the context of cortical speech-envelope tracking. They further suggest that this process integrates audio-tactile input within a time window from around −100 m s to 0 m s (relative to the auditory input).

3.3. Characteristics of Cortical Audio-tactile speech-envelope integration

The top row of Fig. 5 summarizes results from exploratory spectral, temporal, and spatial analyses of the neural MSI in condition 0, which showed the strongest cortical audio-tactile speech-envelope integration above (see Fig. 3B).

Fig. 5A (top row) shows the neural MSI for each frequency band (delta, theta, alpha) of cortical activity and the speech envelope. Similar to the broadband results above, the index was significantly above zero in every band (all $t_{17}>2.78$, $P<0.0064$). This effect was strongest in the delta band, although it did not vary significantly across bands (no main effect of frequency band: $F_{1,17}=1.0$, $P=0.33$). Fig. 5B (top row) shows the neural MSI as a function of the cortical lag, a measure representing the timing of cortical activity relative to the continuous speech envelope. The index was significantly above zero at cortical lags between −20 m s and 170 m s (all $t_{17}>2.11$, $P<0.05$). Fig. 5C (top row) shows the spatial distribution of the neural MSI assessed with an EEG-prediction approach across the scalp. The index was above zero for 68% of electrodes and largest over the middle and lateral central scalp (locations near CPz, C3 and C4).

The lower rows of Fig. 5 show measures that underlie the multisensory-integration indices reported above, i.e., measures associated with decoders and forward models 0, A + T, A, and T. The oscillatory band and cortical lags that contributed most strongly to the integration of audio-tactile speech envelopes (Fig. 5A, B top row) were observed to contribute strongly to the reconstruction of speech envelopes in the unisensory conditions as well (Fig. 5A, B bottom rows). The pattern of scalp locations that strongly contributed to EEG prediction was observed to be highly similar across condition 0, the reference condition A + T, and the purely tactile condition. The average pattern contributing to audio-tactile speech-envelope integration was qualitatively more similar to the average pattern observed in the purely auditory condition (Fig. 5C, correlation of maps for MSI [first row] and forward model A [fourth row]: $R=0.41$) than to the average pattern observed in the purely tactile condition (cf. Fig. 5C, correlation of maps for MSI [first row] and forward model T [fifth row]: $R=0.02$).

In sum, these observations show that the strongest contributions to

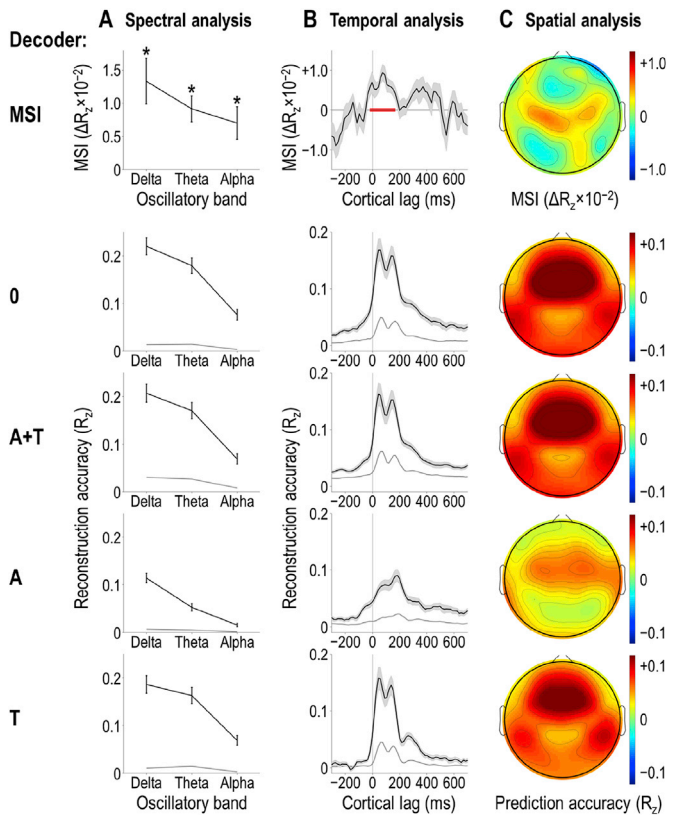

Fig. 5. Characteristics of Cortical Audio-Tactile Speech-Envelope Integration. The top row illustrates results from the analysis of the multisensory integration index in the synchronous condition (condition 0), which showed the strongest cortical audio-tactile speech-envelope integration above (see Fig. 3B). The lower rows show results from the analysis of reference measures associated with decoders or forward models 0, A + T (baseline), A, and T (from top to bottom).
A. The average multisensory integration index is shown as a function of the oscillatory frequency band (top). The strongest audio-tactile speech-envelope integration was observed in the delta band. The plots below show the average reconstruction accuracy per band achieved with the decoder indicated on the left. Similar to the multisensory integration index, these measures also revealed a peak in the delta band. The light gray line at the bottom denotes the upper limit of the 95% confidence interval for the average empirical chance level (mean across all participants) determined through permutation testing. Error bars represent s.e.m. across all participants. *, $P < 0.05$.
B. The average multisensory integration index is shown as a function of the cortical lag (top). The index was significantly above zero specifically between −20 and 170 m s (relative to the timing of the continuous speech envelope, vertical line), indicating that audio-tactile speech-envelope integration emerged in the cortex within this time window (see red dash, $P < 0.05$). The plots below show the time-resolved average reconstruction accuracy achieved with the decoder indicated on the left. These measures revealed peaks in a similar time window as the multisensory integration index. The light gray waveform at the bottom denotes the upper limit of the 95% confidence interval for the average empirical chance level (mean across all participants). Shaded areas represent s.e.m. across all participants.
C. The top scalp map shows the average multisensory integration index obtained with an EEG-prediction approach for each EEG channel. Warmer colors represent more positive indices, i.e., stronger audio-tactile speech-envelope integration. The largest indices were observed over the middle and lateral central scalp (locations near CPz, C3 and C4). The plots below show scalp maps of the average prediction accuracies achieved with the forward model indicated on the left. The average map resembling audio-tactile speech-envelope integration (top row) showed a qualitatively higher correlation with the average map obtained with the forward model A (fourth row) than with the average map obtained with the forward model T (fifth row).

the observed cortical audio-tactile speech-envelope integration emerged from cortical activity in the delta band at ~0–150 m s relative to the speech envelope. They further indicate that the integration arose from neural generators that were strongly involved in the tracking of the speech envelope in auditory input rather than tactile input.

4. Discussion

We tested whether cortical speech-envelope tracking is a mechanism for audio-tactile speech integration. We presented continuous tactile speech envelopes and degraded auditory speech and assessed cortical speech-envelope tracking and intelligibility based on decoders' speech-envelope reconstruction accuracy and listeners' speech-recognition performance, respectively. We assumed that supra-additivity in these measures reflects multisensory integration. Under this assumption, we found integration in the cortical tracking of auditory speech and tactile speech envelopes leading by 100 m s or less. This audio-tactile speech-envelope integration particularly involves cortical delta activity occurring at ~0–150 m s from auditory stimulus onset. We found no evidence for audio-tactile integration in speech intelligibility. In sum, tactile speech-shaped stimulation can enhance the cortical encoding of degraded auditory speech input within a ~100-ms window; however this benefit appears to be insufficient for improving intelligibility.

4.1. Tactile speech-shaped stimulation enhances cortical speech-envelope tracking

The supra-additive effect on cortical auditory speech tracking (Fig. 3B) indicates that temporal encoding of ongoing auditory speech can be increased by applying speech-shaped stimulation to the skin. Given our null result on speech intelligibility, the opposite might also be possible, i.e., that auditory stimulation enhanced temporal encoding of tactile input. However, this interpretation appears less plausible. First, the effect occurs when tactile input leads auditory input, but not vice versa, thus implying multisensory integration especially when tactile input can predict auditory input, but not vice versa. Second, our tactile stimuli were probably more effective in modulating the perceived intensity of the auditory stimuli than vice versa (Yau et al., 2010). Third, as further discussed below, our spatial results suggest that the observed supra-additivity affected neural generators of auditory, rather than tactile, speech-envelope tracking.

Our neural finding agrees with corresponding audio-visual results (Crosse et al., 2015, 2016b), suggesting that cortical speech integration in the audio-tactile and audio-visual modalities could rely on similar mechanisms. It diverges from null results of an EEG study that tested audio-tactile stimulus-envelope integration based on cortical steady-state responses (Budd and Timora, 2013). The difference in outcome may be ascribed to the shorter non-speech stimuli and non-auditory task in that study. The audio-tactile stimulus-envelope integration observed here may be specific to attentive speech listening, an idea that may be tested in future studies comparing speech stimuli versus non-speech equivalents. Indeed, cortical speech-envelope tracking reflects not merely a passive acoustic input-following response, but also top-down modulation related to intelligibility (Peelle et al., 2013; Ding et al., 2015; Steinmetzger and Rosen, 2017; but see Zoefel and VanRullen, 2016), linguistic knowledge (Di Liberto et al., 2018; but see Millman et al., 2015), and attention to speech (Makov et al., 2017; Ding et al., 2018).

4.2. Time window for cortical audio-tactile speech-envelope integration

We observed cortical audio-tactile speech-envelope integration when tactile input led auditory speech by 100 m s or was synchronous with it

(Fig. 3B). This is consistent with animal findings showing stronger integration of synchronous vs. asynchronous basic audio-tactile stimuli in auditory cortex (Kayser et al., 2005; Lakatos et al., 2007) and extends these findings to human listeners and continuous speech-shaped stimuli. It is further consistent with audio-visual speech findings showing more accurate cortical speech-envelope tracking when the visual and auditory speech are temporally congruent (Luo et al., 2010; Crosse et al., 2015). More generally, our findings are consistent with the view that multisensory integration depends on cross-modal temporal coherence (Stein and Meredith, 1993).

The identified time window for cortical audio-tactile speech-envelope integration matches the time window for unisensory tactile integration (i.e., temporal integration of consecutive events), which spans ~75 m s (Yamashiro et al., 2011). It partially differs from behavioral estimates based on audio-tactile syllable identification (−50 to +200 m s; Gick et al., 2010) and audio-tactile tone detection (−200 to 0 m s; Wilson et al., 2009). Our observation that the tactile input must lead the auditory speech by 100 m s or less suggests that tactile events slightly preceding peaks in the auditory speech signal are most effective in aiding cortical speech-envelope tracking. This hints at a predictive mechanism, as has been proposed for audio-visual speech integration (Peelle and Sommers, 2015). It is possible that tactile temporal cues prepare the cortex for upcoming peaks in the auditory speech signal, possibly by resetting the phase of auditory cortical oscillations (Schroeder et al., 2008). Indeed, tactile stimuli can phase reset ongoing oscillations in monkey auditory cortex and thereby modulate responses to upcoming auditory events (Lakatos et al., 2007).

4.3. Characteristics of Cortical Audio-tactile speech-envelope integration

To better understand the basis of the observed cortical audio-tactile speech-envelope integration, we explored its relation to the observed cortical speech-envelope tracking in the spectral, temporal, and spatial domain. We observed that audio-tactile speech-envelope integration involves significant contributions from cortical activity in the delta, theta, and alpha band (Fig. 5A top). The delta band showed the strongest contribution, which is consistent with electrophysiological findings on audio-tactile integration in monkey auditory cortex (Lakatos et al., 2007) and the general notion that neural delta oscillations integrate cross-modal information over a ~125–250 m s time window (Schroeder et al., 2008). It also partially matches with the contribution of cortical activity in the 2–6 Hz range to audio-visual speech-envelope integration (Crosse et al., 2015, 2016b). In continuous speech, cortical delta activity tracks linguistic events occurring at the corresponding time scale (Ding et al., 2015); for the Dutch sentences used here, this corresponds approximately to individual words (Verhoeven et al., 2004). Whether tactile speech-shaped stimulation enhanced speech encoding especially at the word level remains unclear in our study, given the null result on speech intelligibility. The contribution of delta-band activity to audio-tactile speech-envelope integration probably originates from its involvement in speech tracking that we observed in the unisensory conditions (Fig. 5B bottom).

We further observed that audio-tactile speech-envelope integration involves cortical activity occurring at ~0–150 m s (Fig. 5B top). This effect started to build up already 20 m s before the audio-tactile input, further supporting the notion that the observed audio-tactile speech-envelope integration involves sensory predictions. Perhaps unsurprisingly, cortical activity in this latency range also contributed strongly to speech-envelope tracking in the unisensory conditions (Fig. 5B bottom), indicating that audio-tactile speech-envelope integration and speech-envelope tracking arise within a common time window. The observed short latency agrees with basic findings showing rapid integration of audio-tactile non-speech stimuli in early sensory cortex (see Introduction) and extends them to continuous speech-shaped stimuli. It further partially matches audio-visual findings showing cortical speech-envelope integration at 50 m s (Zion Golumbic et al., 2013), 140 m s and 220 m s (Crosse et al., 2015). The observations of similar latencies and time scales for cortical speech-envelope integration in audio-tactile and audio-visual modalities suggest that these modalities rely on similar mechanisms. Somatosensory-evoked responses peak at ~100 m s near the secondary somatosensory cortex (review: Yamashiro et al., 2011). We observed strong cortical audio-tactile speech-envelope integration over the middle and lateral central scalp, locations that were coupled more strongly with the tracking of speech envelope in auditory than tactile input (Fig. 5C). This suggests that audio-tactile speech-envelope integration primarily affects neural generators in the auditory rather than somatosensory cortex.

4.4. Does cortical audio-tactile speech-envelope integration enhance the intelligibility of auditory speech?

Contrary to our neural results, we found no evidence for audio-tactile integration in auditory speech intelligibility (Fig. 2B). While our data match previous results showing a positive link between cortical speech-envelope tracking and intelligibility (Fig. 4A), they provide no evidence for a benefit of audio-tactile speech-envelope integration for intelligibility (Fig. 4B). Our failure to observe behavioral effects possibly reflects a methodological limitation: our participants were familiarized only with the auditory, not the tactile, stimuli. Training on tactile speech stimuli could have been beneficial, considering that auditory speech comprehension is less naturally coupled with tactile than visual information. Moreover, our stimuli resembled an adverse listening condition in which tactile stimuli and degraded auditory stimuli were deliberately mostly uncorrelated (Supplementary Fig. 1C) and therefore carried complementary information rather than redundant information. More correlated stimuli carrying overlapping audio-tactile speech information could have been beneficial, as this might reflect a more natural condition for multisensory integration (Campbell, 2008). In addition, our intelligibility measure captured only the final sentence of each sequence, whereas our cortical speech-tracking measure captured the entire sequence. These three aspects likely reduced the detectability of a putative link between tactile enhancement of intelligibility and speech tracking.

Alternatively, tactile enhancement of cortical speech-envelope tracking might not suffice to enhance auditory speech intelligibility. It is possible that neural oscillatory phase resetting, while beneficial for simple temporal predictions, does not confer enough of a benefit in improving more complex linguistic processing. We did not assess oscillatory phase resetting in our study, but the aforementioned interpretation is supported by audio-visual results showing little or no benefit for auditory speech intelligibility from purely temporal, non-articulatory visual speech cues, e.g., visual shapes whose size co-varies with the speech envelope (Summerfield, 1979; Schwartz et al., 2004). Together, these results suggest that tactile or visual speech-shaped stimulation can inform the auditory system about *when* to listen (Tjan et al., 2014), but this 'bottom-up' temporal information appears to be insufficient for understanding *what* is being said. The latter may require additional, non-auditory post-perceptual processes (Rizza et al., 2018) that rely on the availability of articulatory information in the tactile input and the listener's linguistic knowledge (Weisenberger and Miller, 1987). Indeed, benefits of basic tactile stimuli for auditory speech perception have been reported so far only in the context of speech detection and syllable identification, not sentence intelligibility (Blamey et al., 1989). Moreover, the Tadoma method relies to a certain extent on articulatory cues (Tan et al., 1989) and successful use of it or other tactile speech aids benefits from prior training (Reed et al., 1985; Working Group on Communication Aids for the Hearing-Impaired, 1991).

4.5. Cortical tracking of inaudible speech-shaped stimuli and envelope-reduced speech

We observed that unisensory tactile speech-shaped stimulation and

unisensory degraded auditory speech are sufficient to elicit cortical speech-envelope tracking (Fig. 3A). The tactile envelope-tracking result extends similar results from silent lip-reading (Crosse et al., 2015; Park et al., 2016; O'Sullivan et al., 2017) to the tactile modality in the absence of intelligibility. Moreover, it extends tactile EEG results based on isochronous tone sequences and steady-state responses (Snyder, 1992; Tobimatsu et al., 1999) to more complex, speech-shaped stimuli. The auditory result agrees with previous EEG results (Zoefel and VanRullen, 2016) and underscores that cortical speech-envelope tracking reflects more than a passive acoustic input-following response. Although our speech-degradation procedure substantially reduced amplitude envelopes and their recovery in the peripheral auditory system (Ghitza, 2001; Gilbert and Lorenzi, 2006), it possibly preserved temporally correlated higher-order speech features, which may have contributed to the observed envelope tracking (Zeng et al., 2004).

5. Conclusion

Our study provides insights into how the human brain integrates continuous auditory and tactile speech-shaped input under adverse listening conditions. Together with related audio-visual studies, it suggests a cortical mechanism for multisensory speech-envelope integration that operates at a slow (delta) time scale in a rapid, anticipatory, and facilitative manner. This may be relevant for understanding the neural basis for tactile speech communication (e.g., Tadoma method). Moreover, the possibility to improve the temporal encoding of auditory speech with simultaneous tactile stimulation within less than 150 m s may be utilized to alleviate temporal encoding deficits, e.g., in dyslexia (Goswami, 2011).

Conflicts of interest

The authors declare that no conflict of interest exists.

Acknowledgement

This work was funded by the Netherlands Organization for Scientific Research (Veni grant 451-17-033 to L.H.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.116134.

References

Alex Meredith, M., Stein, B.E., 1986. Spatial factors determine the activity of multisensory neurons in cat superior colliculus. Brain Res. 365, 350–354.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300. https://www.jstor.org/stable/2346101.

Blamey, P.J., Cowan, R., Alcantara, J.I., Whitford, L., Clark, G., 1989. Speech perception using combinations of auditory, visual, and tactile information. J. speech Rehabil. Res. Dev. 26, 15–24. http://hdl.handle.net/11343/27275.

Brosch, M., 2005. Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. J. Neurosci. 25, 6797–6806.

Budd, T.W., Timora, J.R., 2013. Steady state responses to temporally congruent and incongruent auditory and vibrotactile amplitude modulated stimulation. Int. J. Psychophysiol. 89, 419–432. https://doi.org/10.1016/j.ijpsycho.2013.06.001.

Butler, J.S., Molholm, S., Fiebelkorn, I.C., Mercier, M.R., Schwartz, T.H., Foxe, J.J., 2011. Common or redundant neural circuits for duration processing across audition and touch. J. Neurosci. 31, 3400–3406. https://doi.org/10.1523/JNEUROSCI.3296-10.2011.

Caetano, G., Jousmäki, V., 2006. Evidence of vibrotactile input to human auditory cortex. Neuroimage 29, 15–28. https://doi.org/10.1016/j.neuroimage.2005.07.023.

Campbell, R., 2008. The processing of audio-visual speech: empirical and neural bases. Philos. Trans. R. Soc. Biol. Sci. 363, 1001–1010.

Chang, C.Y., Hsu, S.H., Pion-Tonachini, L., Jung, T.P., 2018. Evaluation of artifact Subspace reconstruction for automatic EEG artifact removal. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 1242–1245, 2018.

Crosse, M.J., Butler, J.S., Lalor, E.C., 2015. Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. J. Neurosci. 35, 14195–14204. http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1829-15.2015.

Crosse, M.J., Di Liberto, G.M., Bednar, A., Lalor, E.C., 2016a. The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. Front. Hum. Neurosci. 10, 1–14. http://journal.frontiersin.org/article/10.3389/fnhum.2016.00604/full.

Crosse, M.J., Di Liberto, G.M., Lalor, E.C., 2016b. Eye can hear clearly now: inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. J. Neurosci. 36, 9888–9895. http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1396-16.2016.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Di Liberto, G.M., Lalor, E.C., Millman, R.E., 2018. Causal cortical dynamics of a predictive enhancement of speech intelligibility. Neuroimage 166, 247–258. https://doi.org/10.1016/j.neuroimage.2017.10.066.

Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2015. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19, 158–164. https://doi.org/10.1038/nn.4186.

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., Zhang, J., 2018. Attention is required for knowledge-based sequential grouping: insights from the integration of syllables into words. J. Neurosci. 38, 1178–1188. https://doi.org/10.1523/JNEUROSCI.2606-17.2017.

Ding, N., Simon, J.Z., 2014. Cortical entrainment to continuous speech: functional roles and interpretations. Front. Hum. Neurosci. 8, 1–7. http://journal.frontiersin.org/article/10.3389/fnhum.2014.00311/abstract.

Fowler, C.A., Dekle, D.J., 1991. Listening with eye and hand: cross-modal contributions to speech perception. J. Exp. Psychol. Hum. Percept. Perform. 17, 816–828. https://doi.org/10.1037/0096-1523.17.3.816.

Foxe, J.J., Morocz, I.A., Murray, M.M., Higgins, B.A., Javitt, D.C., Schroeder, C.E., 2000. Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Cogn. Brain Res. 10, 77–83. https://doi.org/10.1016/S0926-6410(00)00024-0.

Foxe, J.J., Wylie, G.R., Martinez, A., Schroeder, C.E., C.D, Guilfoyle, D., Ritter, W., Murray, M.M., Javitt, D.C., Guilfoyle, D., Ritter, W., Murray, M.M., 2002. Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. J. Neurophysiol. 88, 540–543. https://doi.org/10.1152/jn.2002.88.1.540.

Fu, K.-M.G., Johnston, T.A., Shah, A.S., Arnold, L., Smiley, J., Hackett, T.A., Garraghty, P.E., Schroeder, C.E., 2003. Auditory cortical neurons respond to somatosensory stimulation. J. Neurosci. 23, 7510–7515.

Ghitza, O., 2001. On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. J. Acoust. Soc. Am. 110, 1628–1640. http://asa.scitation.org/doi/10.1121/1.1396325.

Gick, B., Derrick, D., 2009. Aero-tactile integration in speech perception. Nature 462, 502–504.

Gick, B., Ikegami, Y., Derrick, D., 2010. The temporal window of audio-tactile integration in speech perception. J. Acoust. Soc. Am. 128, EL342–EL346. http://asa.scitation.org/doi/10.1121/1.3505759.

Gick, B., Jóhannsdóttir, K.M., Gibraiel, D., Mühlbauer, J., 2008. Tactile enhancement of auditory and visual speech perception in untrained perceivers. J. Acoust. Soc. Am. 123, EL72–EL76. http://asa.scitation.org/doi/10.1121/1.2884349.

Gilbert, G., Lorenzi, C., 2006. The ability of listeners to use recovered envelope cues from speech fine structure. J. Acoust. Soc. Am. 119, 2438–2444. http://asa.scitation.org/doi/10.1121/1.2173522.

Gobbelé, R., Schürmann, M., Forss, N., Juottonen, K., Buchner, H., Hari, R., 2003. Activation of the human posterior parietal and temporoparietal cortices during audiotactile interaction. Neuroimage 20, 503–511. https://doi.org/10.1016/S1053-8119(03)00312-4.

Goswami, U., 2011. A temporal sampling framework for developmental dyslexia. Trends Cogn. Sci. 15, 3–10. https://doi.org/10.1016/j.tics.2010.10.001.

Grant, K.W., Greenberg, S., 2001. Speech intelligibility derived from asynchronous processing of auditory-visual information. In: Proc Conf Audit Speech Process 2001, pp. 132–137. http://www.isca-speech.org/archive_open/avsp01/av01_132.html.

Hackett, T.A., Smiley, J.F., Ulbert, I., Karmos, G., Lakatos, P., De La Mothe, L.A., Schroeder, C.E., 2007. Sources of somatosensory input to the caudal belt areas of auditory cortex. Perception 36, 1419–1430. https://doi.org/10.1068/p5841.

Hausfeld, L., Riecke, L., Valente, G., Formisano, E., 2018. Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. Neuroimage 181, 617–626. https://doi.org/10.1016/j.neuroimage.2018.07.052.

Ito, T., Tiede, M., Ostry, D.J., 2009. Somatosensory function in speech perception. Proc. Natl. Acad. Sci. 106, 1245–1248. http://www.pnas.org/cgi/doi/10.1073/pnas.0810063106.

Kandel, E.R., Schwartz, J.H., Jessell, T.M., 2000. Principles of Neural Science.

Kayser, C., Petkov, C.I., Augath, M., Logothetis, N.K., 2005. Integration of touch and sound in auditory cortex. Neuron 48, 373–384. https://doi.org/10.1016/j.neuron.2005.09.018.

Kirman, J.H., 1973. Tactile communication of speech: a review and an analysis. Psychol. Bull. 80, 54–74. https://doi.org/10.1037/h0034630.

Kong, Y.-Y., Somarowthu, A., Ding, N., 2015. Effects of spectral degradation on attentional modulation of cortical auditory responses to continuous speech. J. Assoc. Res. Otolaryngol. 16, 783–796. http://link.springer.com/10.1007/s10162-015-0540-x.

Lakatos, P., Chen, C.M., O'Connell, M.N., Mills, A., Schroeder, C.E., 2007b. Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron 53, 279–292. https://doi.org/10.1016/j.neuron.2006.12.011.

Luo, H., Liu, Z., Poeppel, D., 2010. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. PLoS Biol. 8, 25–26. https://doi.org/10.1371/journal.pbio.1000445.

Lütkenhöner, B., Lammertmann, C., Simões, C., Hari, R., 2002. Magnetoencephalographic correlates of audiotactile interaction. Neuroimage 15, 509–522. https://doi.org/10.1006/nimg.2001.0991.

Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., Zion Golumbic, E., 2017. Sleep disrupts high-level speech parsing despite significant basic auditory processing. J. Neurosci. 37, 7772–7781. http://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.0168-17.2017.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024.

Millman, R.E., Johnson, S.R., Prendergast, G., 2015. The role of phase-locking to the temporal envelope of speech in auditory perception and speech intelligibility. J. Cogn. Neurosci. 27, 533–545. https://doi.org/10.1162/jocn_a_00719.

Murray, M.M., Molholm, S., Michel, C.M., Heslenfeld, D.J., Ritter, W., Javitt, D.C., Schroeder, C.E., Foxe, J.J., 2004. Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cerebr. Cortex 15, 963–974.

O'Sullivan, A.E., Crosse, M.J., Di Liberto, G.M., Lalor, E.C., 2017. Visual cortical entrainment to motion and categorical speech features during silent lipreading. Front. Hum. Neurosci. 10, 1–11. http://journal.frontiersin.org/article/10.3389/fnhum.2016.00679/full.

Oostdijk, N.H.J., 2000. The spoken Dutch Corpus. Outline and first evaluation. In: Proceedings of Second International Conference on Language Resources and Evaluation (LREC), pp. 887–894. http://repository-acc.ubn.ru.nl/handle/123456789/76342.

Park, H., Kayser, C., Thut, G., Gross, J., 2016. Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. Elife 5, 1–17.

Peelle, J.E., Davis, M.H., 2012. Neural oscillations carry speech rhythm through to comprehension. Front. Psychol. 3, 1–17.

Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cerebr. Cortex 23, 1378–1387. https://doi.org/10.1093/cercor/bhs118.

Peelle, J.E., Sommers, M.S., 2015. Prediction and constraint in audiovisual speech perception. Cortex 68, 169–181. https://doi.org/10.1016/j.cortex.2015.03.006.

Reed, C.M., Durlach, N.I., Braida, L.D., 1982. Research on tactile communication of speech: a review. ASHA Monogr. 20, 1.

Reed, C.M., Rabinowitz, W.M., Durlach, N.I., Braida, L.D., Conway-Fithian, S., Schultz, M.C., 1985. Research on the Tadoma method of speech communication. J. Acoust. Soc. Am. 77, 247–257. http://asa.scitation.org/doi/10.1121/1.392266.

Riecke, L., Formisano, E., Sorger, B., Başkent, D., Gaudrain, E., 2018. Neural entrainment to speech modulates speech intelligibility. Curr. Biol. 28, 161–169. https://doi.org/10.1016/j.cub.2017.11.033.

Rizza, A., Terekhov, A.V., Montone, G., Olivetti-Belardinelli, M., O'Regan, J.K., 2018. Why early tactile speech aids may have failed: No perceptual integration of tactile and auditory signals. Front. Psychol. 9, 767. https://doi.org/10.3389/fpsyg.2018.00767.

Ro, T., Ellmore, T.M., Beauchamp, M.S., 2013. A neural link between feeling and hearing. Cerebr. Cortex 23, 1724–1730.

Sato, M., Cavé, C., Ménard, L., Brasseur, A., 2010. Auditory-tactile speech perception in congenitally blind and sighted adults. Neuropsychologia 48, 3683–3686. https://doi.org/10.1016/j.neuropsychologia.2010.08.017.

Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., Puce, A., 2008. Neuronal oscillations and visual amplification of speech. Trends Cogn. Sci. 12, 106–113. https://doi.org/10.1016/j.tics.2008.01.002.

Schürmann, M., Caetano, G., Hlushchuk, Y., Jousmäki, V., Hari, R., 2006. Touch activates human auditory cortex. Neuroimage 30, 1325–1331. https://doi.org/10.1016/j.neuroimage.2005.11.020.

Schwartz, J.L., Berthommier, F., Savariaux, C., 2004. Seeing to hear better: evidence for early audio-visual interactions in speech identification. Cognition 93, 69–78.

Snyder, A.Z., 1992. Steady-state vibration evoked potentials: description of technique and characterization of responses. Electroencephalogr. Clin. Neurophysiol. Evoked Potentials 84, 257–268.

Soto-Faraco, S., Deco, G., 2009. Multisensory contributions to the perception of vibrotactile events. Behav. Brain Res. 196, 145–154. https://doi.org/10.1016/j.bbr.2008.09.018.

Stein, B.E., Meredith, M.A., 1993. The Merging of the Senses. The MIT Press.

Steinmetzger, K., Rosen, S., 2017. Effects of acoustic periodicity and intelligibility on the neural oscillations in response to speech. Neuropsychologia 95, 173–181. https://doi.org/10.1016/j.neuropsychologia.2016.12.003.

Stevenson, R.A., Ghose, D., Fister, J.K., Sarko, D.K., Altieri, N.A., Nidiffer, A.R., Kurela, L.A.R., Siemann, J.K., James, T.W., Wallace, M.T., 2014. Identifying and quantifying multisensory integration: a tutorial review. Brain Topogr. 27, 707–730. https://doi.org/10.1523/JNEUROSCI.3615-13.2014.

Sumby, W.H., Pollack, I., 1954. Perceptual amplification of speech sounds by visual cues. J. Acoust. Soc. Am. 26, 212–215.

Summerfield, Q., 1979. Use of visual information for phonetic perception. Phonetica 36, 314–331. http://www.karger.com/doi/10.1159/000259969.

Tan, H.Z., Rabinowitz, W.M., Durlach, N.I., 1989. Analysis of a synthetic Tadoma system as a multidimensional tactile display. J. Acoust. Soc. Am. 86, 981–988.

ten Oever, S., de Graaf, T.A., Bonnemayer, C., Ronner, J., Sack, A.T., Riecke, L., 2016. Stimulus presentation at specific neuronal oscillatory phases experimentally controlled with tACS: implementation and applications. Front. Cell. Neurosci. 10, 1–8. http://journal.frontiersin.org/article/10.3389/fncel.2016.00240/full.

Tjan, B.S., Chao, E., Bernstein, L.E., 2014. A visual or tactile signal makes auditory speech detection more efficient by reducing uncertainty. Eur. J. Neurosci. 39, 1323–1331. http://doi.org/10.1111/ejn.12471.

Tobimatsu, S., Zhang, Y.M., Kato, M., 1999. Steady-state vibration somatosensory evoked potentials: physiological characteristics and tuning function. Clin. Neurophysiol. 110, 1953–1958.

Verhoeven, J., De Pauw, G., Kloots, H., 2004. Speech rate in a pluricentric language: a comparison between Dutch in Belgium and The Netherlands. Lang. Speech 47, 297–308.

Versfeld, N.J., Daalder, L., Festen, J.M., Houtgast, T., 2000. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. J. Acoust. Soc. Am. 107, 1671–1684. http://asa.scitation.org/doi/10.1121/1.428451.

Weisenberger, J.M., Miller, J.D., 1987. The role of tactile aids in providing information about acoustic stimuli. J. Acoust. Soc. Am. 82, 906–916. http://asa.scitation.org/doi/10.1121/1.395289.

Wilson, E.C., Reed, C.M., Braida, L.D., 2009. Integration of auditory and vibrotactile stimuli: effects of phase and stimulus-onset asynchrony. J. Acoust. Soc. Am. 126, 1960. http://scitation.aip.org/content/asa/journal/jasa/126/4/10.1121/1.3204305.

Working Group on Communication Aids for the Hearing-Impaired, 1991. Speech-perception aids for hearing-impaired people: current status and needed research. J. Acoust. Soc. Am. 90, 637–685.

Yamashiro, K., Inui, K., Otsuru, N., Urakawa, T., Kakigi, R., 2011. Temporal window of integration in the somatosensory modality: an MEG study. Clin. Neurophysiol. 122, 2276–2281. https://doi.org/10.1016/j.clinph.2011.03.028.

Yau, J.M., Weber, A.I., Bensmaia, S.J., 2010. Separate mechanisms for audio-tactile pitch and loudness interactions. Front. Psychol. 1, 160.

Zeng, F.-G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y.-Y., Chen, H., 2004. On the dichotomy in auditory perception between temporal envelope and fine structure cues (L). J. Acoust. Soc. Am. 116, 1351–1354. http://asa.scitation.org/doi/10.1121/1.1777938.

Zion Golumbic, E., Cogan, G.B., Schroeder, C.E., Poeppel, D., 2013. Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party. J. Neurosci. 33, 1417–1426. http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.3675-12.2013.

Zion Golumbic, E.M., Poeppel, D., Schroeder, C.E., 2012. Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. Brain Lang. 122, 151–161. https://doi.org/10.1016/j.bandl.2011.12.010.

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.