Phase resetting in human auditory cortex to visual speech

Pierre Mégevand[1,2], Manuel R. Mercier[3,4,5], David M. Groppe[1,6], Elana Zion Golumbic[7], Nima Mesgarani[8], Michael S. Beauchamp[9], Charles E. Schroeder[10,11,12*], Ashesh D. Mehta[1,12*]

[1]Department of Neurosurgery, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, and Feinstein Institute for Medical Research, Manhasset, NY 11030, USA

[2]Department of Basic Neurosciences, Faculty of Medicine, University of Geneva, 1211 Geneva, Switzerland

[3]Department of Neurology, Montefiore Medical Center, Bronx, NY 10467, USA

[4]Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY 10461, USA

[5]Centre de Recherche Cerveau et Cognition, CNRS, UMR5549, 31052 Toulouse, France

[6]The Krembil Neuroscience Centre, University Health Network, Toronto, ON M5T 1M8, Canada

[7]The Gonda Brain Research Center, Bar Ilan University, Ramat Gan, 5290002, Israel

[8]Department of Electrical Engineering, Columbia University, New York, NY 10027, USA

[9]Department of Neurosurgery, Baylor College of Medicine, Houston, TX 77030, USA

[10]Nathan S. Kline Institute, Orangeburg, NY 10962, USA

[11]Department of Neurosurgery, Columbia University, New York, NY 10032, USA

[12]These senior authors contributed equally

*Correspondence: cs2388@columbia.edu; amehta@northwell.edu

ABSTRACT

Natural conversation is multisensory: when we can see the speaker's face, visual speech cues influence our perception of what is being said. The neuronal basis of this phenomenon remains unclear, though there is indication that neuronal oscillations—ongoing excitability fluctuations of neuronal populations in the brain—represent a potential mechanism. Investigating this question with intracranial recordings in humans, we show that some sites in auditory cortex track the temporal dynamics of unisensory visual speech using the phase of their slow oscillations and phase-related modulations in neuronal activity. This effect is asymmetric, as we find much less detectable tracking of auditory speech by visual cortex. Auditory cortex thus builds a representation of the speech stream's envelope based on visual speech alone, at least in part by resetting the phase of its ongoing oscillations. Phase reset amplifies the representation of the speech stream and organizes the information contained in neuronal activity patterns.

Keywords: neuronal oscillations, crossmodal stimuli, intracranial EEG, intertrial phase coherence, broadband high-frequency activity, phase-amplitude coupling

INTRODUCTION

While viewing one's interlocutor is not always necessary for speech perception, it significantly improves intelligibility under noisy conditions (Sumby and Pollack, 1954). Moreover, mismatched auditory and visual speech stimuli can induce striking perceptual illusions (McGurk and Macdonald, 1976). Despite the ubiquity and power of visual influences on speech perception, the underlying neuronal mechanisms remain an open question. The cerebral processing of auditory and visual speech converges in multisensory cortical areas, especially the superior temporal lobe (Miller and D'Esposito, 2005; Beauchamp, Nath and Pasalar, 2010). Crossmodal influences are also found in cortex traditionally considered to be unisensory; in particular, visual speech modulates the activity of auditory cortex (Calvert *et al.*, 1997; Lakatos *et al.*, 2007; Besle *et al.*, 2008; Kayser, Petkov and Logothetis, 2008).

The articulatory movements that constitute visual speech strongly correlate with the corresponding speech sounds (Chandrasekaran *et al.*, 2009; Schwartz and Savariaux, 2014) and predict them to some extent (Schroeder *et al.*, 2008; Arnal *et al.*, 2009; Arnal, Wyart and Giraud, 2011), suggesting that visual speech might serve as an alerting cue to auditory cortex, preparing the neural circuits to process the incoming speech sounds more efficiently. Our hypothesis is that this preparation occurs through a resetting of the phase of neuronal oscillations: through this phase reset, visual speech cues influence neuronal excitability in auditory cortex (Schroeder *et al.*, 2008).

This hypothesis rests on four lines of evidence. First, auditory speech is rhythmic, with syllables arriving at a relatively rapid rate (4-7 Hz) nested within the slower (1-3 Hz) rates of phrase and word production. These rhythmic features of speech are critical for it to be intelligible (Shannon *et al.*, 1995; Greenberg *et al.*, 2003). Second, auditory cortex

synchronizes its oscillations to the rhythm of heard speech, and the magnitude of this synchronization correlates with the intelligibility of speech (Ahissar *et al.*, 2001; Luo and Poeppel, 2007; Vander Ghinst *et al.*, 2016). Third, neuronal oscillations correspond to momentary changes in neuronal excitability, so that the response of sensory cortex depends on the phase of its oscillations upon stimulus arrival (Lakatos *et al.*, 2005; Whittingstall and Logothetis, 2009). Fourth, even at the level of primary sensory cortex, oscillations can be phase-reset by stimuli from other modalities, and this crossmodal reset influences the processing of incoming stimuli from the preferred modality (Lakatos *et al.*, 2007; Kayser, Petkov and Logothetis, 2008).

Human EEG and MEG studies of cerebral responses to continuous, naturalistic audiovisual speech have provided observations consistent with the phase reset hypothesis. These studies showed that the visual component of audiovisual speech influences cerebral activity, mostly in visual areas, but also in superior temporal, inferior frontal and premotor cortex (Luo, Liu and Poeppel, 2010; Park *et al.*, 2016, 2018; Giordano *et al.*, 2017) and detail patterns of interactions between brain regions that are critical components of the neural underpinnings of audiovisual speech perception (Park *et al.*, 2016, 2018). The reconstruction of the temporal dynamics of speech is more accurate from whole-brain responses to audiovisual speech than to unisensory auditory speech, and it is also possible to perform this reconstruction from whole-brain responses to unisensory visual speech (Crosse, Butler and Lalor, 2015; Crosse, Di Liberto and Lalor, 2016; O'Sullivan *et al.*, 2016). However, these non-invasive studies are not able to resolve the neuronal processes underlying the crossmodal influence of visual speech on auditory cortex, as we explain below.

The phase reset hypothesis predicts that visual speech entrains oscillations in auditory cortex, and that this entrainment occurs through phase reset rather than a succession of crossmodal sensory-evoked responses (Schroeder *et al.*, 2008). Pure phase reset is characterized by a realignment of the phase of ongoing oscillations without any increase in their power, and does not entail an increase in local neuronal activity; by contrast, evoked responses are defined by a time-locked increase in oscillatory power and a consequent increase in phase concentration measures, and imply an increase in local neuronal activity (Makeig *et al.*, 2004; Shah *et al.*, 2004; Lakatos *et al.*, 2007). In order to distinguish phase reset from evoked responses, one must be able to resolve the local neuronal activity of a given cortical area well enough to dissociate a momentary increase in phase alignment from any coincident increase in spectral power. No non-invasive neurophysiological study to date meets that standard; invasive techniques are required for that level of granularity.

Human intracranial EEG (iEEG) studies have shown that the posterior superior temporal gyrus responds to both the auditory and visual components of audiovisual speech (Ozker *et al.*, 2017; Micheli *et al.*, 2018), but did not focus on the underlying neuronal mechanisms. In a monkey study that focused on a voice-sensitive area in the anterior temporal lobe, audiovisual vocalization stimuli that began with a visual-only component produced an initial crossmodal evoked response. The timing of that evoked response with respect to the onset of the auditory vocalization, and thus the phase of visual-evoked activity in auditory cortex, determined whether neurons increased or decreased their firing rate in response to the incoming auditory stimulus (Perrodin et al., 2015). That study did not demonstrate that multisensory integration of audiovisual stimuli occurs through phase reset in auditory cortex, though, precisely because of the initial crossmodal response evoked by the visual stimulus; i.e., any phase concentration detected at this point could be an artifact of a “phasic” increase in power.

Here, we used iEEG to probe the mechanistic basis for the effect of visual speech on auditory processing. We verify the predictions that unisensory visual speech resets the phase of neuronal oscillations in auditory cortex, and that the phase alignment of auditory cortex to audiovisual speech is stronger than that to auditory speech alone. Building upon the strong correlation between the temporal dynamics of visual and auditory speech, we also successfully reconstruct these temporal dynamics from the response of auditory cortex to unisensory visual speech. Finally, we show that visual cortex does not track the temporal dynamics of heard speech to the same extent as the auditory cortex does visual speech, confirming the direction specificity of the phase reset hypothesis.

RESULTS

Phase reset of low-frequency oscillations in auditory cortex in response to visual speech

We recorded intracranial EEG (iEEG) signals from electrodes implanted in the brain of six human participants undergoing invasive electrophysiological monitoring for epilepsy. Patients attended to clips of a speaker telling a short story (7-11 seconds long), presented in the auditory (soundtrack with black screen) and visual (silent movie) modalities. iEEG electrodes were considered to be in auditory cortex (25 electrodes over 5 participants) if they fulfilled both an anatomical criterion (location in the superior temporal lobe) and a physiological criterion: increase in local neuronal activity (as indexed by the amplitude of broadband high-frequency activity, BHA; (Ray *et al.*, 2008)) in response to auditory speech. To determine how visual speech influences activity in auditory cortex, we computed BHA as well as power and intertrial coherence (ITC, a measure of phase alignment) in the delta (1-3 Hz), theta (4-7 Hz) and alpha (8-12 Hz) oscillatory frequency bands. We focused our analyses on a time period starting 1 s

after stimulus onset and ending before stimulus offset so as to avoid any potential contamination by onset- and offset-related responses.

Figures 1A to 1D show data for a representative electrode in auditory cortex that displayed a sustained alignment in the phase of its delta-band oscillations in response to unisensory visual speech (Figure 1C). If this phase alignment were caused by sensory-evoked responses, increases in delta power and local neuronal activity would be expected (Makeig *et al.*, 2004; Shah *et al.*, 2004; Lakatos *et al.*, 2007). In fact, delta power decreased (see Figure 1C), and local neuronal activity did not increase (Figure 1D). Thus, this combination of observations points towards phase reset of ongoing neuronal oscillations as the more likely mechanism.

This was verified across electrodes and participants: several auditory cortex electrodes displayed robust phase alignment of their slow oscillations in response to visual speech, as demonstrated by a significant ITC increase in the delta and, to a lesser extent, theta bands (Figure 1E). By contrast, local neuronal activity (Figure 1F) and low-frequency power (Figure 1G) tended to decrease in a majority of auditory electrodes in response to visual speech. Importantly, there was no correlation between the intensity of delta phase alignment and neuronal activity, and delta phase alignment correlated inversely with delta power (Figure 1H). Taken together, these results support the view that the low-frequency phase alignment to visual speech observed in some portions of auditory cortex is mediated by rapid, repetitive crossmodal phase resetting of ongoing neuronal oscillations rather than by a succession of sensory-evoked responses (Schroeder *et al.*, 2008).

To examine the consequences of this phase reset, we contrasted the phase alignment of auditory cortex oscillations to auditory and audiovisual speech stimuli. We found that delta-band ITC

was higher in response to audiovisual speech than to unisensory auditory speech (Figure 1I), which confirms that the addition of visual information indeed boosts the continuous tracking of speech by auditory cortex (Schroeder *et al.*, 2008; E. Zion Golumbic *et al.*, 2013).

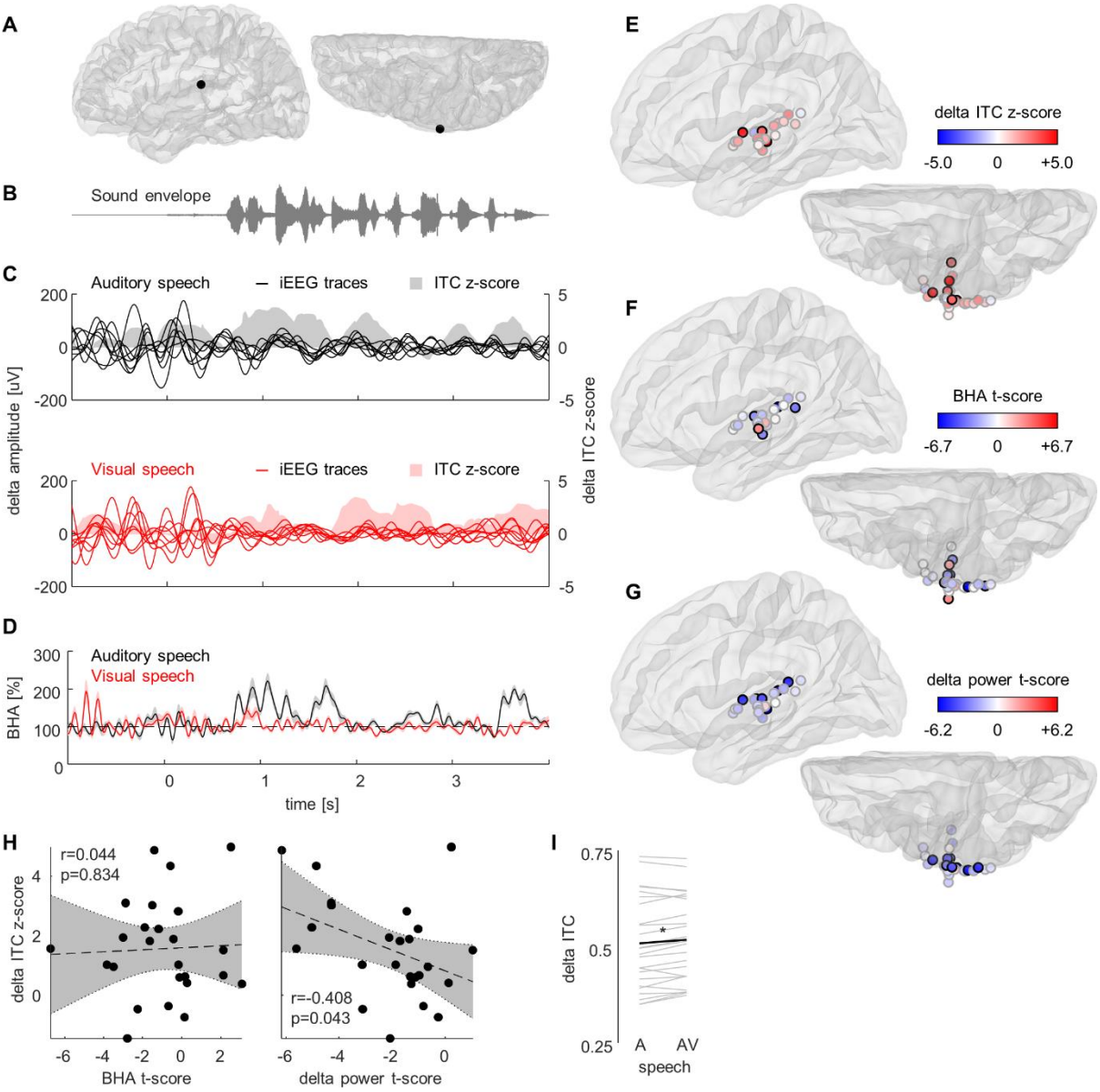

Figure 1. Responses of auditory cortex to visual speech.

A. A representative auditory electrode located in the left superior temporal gyrus (lateral and superior views of the participant's left cerebral hemisphere). **B.** Sound wave of an example speech stimulus. **C.** The representative electrode displayed increased broadband high-frequency activity (BHA, indexing local neuronal activity) to auditory speech (soundtrack only: black trace), but not visual speech (silent movie: red trace; mean and standard error on the mean of BHA responses to 8 repetitions of the same stimulus). **D.** Single-trial delta-band intracranial

EEG (iEEG) traces, superimposed on top of each other, in response to 8 repetitions of the same auditory (top, black traces) and visual speech stimulus (bottom, red traces). Phase alignment is evident as the simultaneous occurrence of peaks and troughs in the waveforms after stimulus onset, and is quantified by the intertrial coherence (ITC, expressed as a z-score: shaded areas). Power decrease is evident as a reduction in the amplitude of oscillations during stimulus presentation compared to the pre-stimulus baseline. **E.** Delta-band ITC for all auditory electrodes in response to visual speech is color-coded on lateral and superior views of a template left cerebral hemisphere (25 electrodes over 5 participants). 6/25 electrodes displayed significant phase alignment of delta oscillations (circled in black; $p=6.87*10^{-7}$). 3 of these 6 electrodes also displayed significant theta-band phase alignment (not shown; $p=5.01*10^{-4}$). Z-scores and significance testing were computed through a permutation test, with FDR correction for multiple comparisons across electrodes and frequency bands, with a family-wise error rate set at 0.05. Electrodes originally located in the right hemisphere were projected to the left one for display. **F.** BHA for all auditory electrodes in response to visual speech. 2/25 electrodes displayed a significant increase in neuronal activity, whereas activity decreased in 6/25. T-scores and significance testing were computed through a t-test relative to the pre-stimulus baseline, with false discovery rate (FDR) correction for multiple comparisons. **G.** Delta-band power for all auditory electrodes in response to visual speech. 8/25 electrodes displayed a significant decrease in delta power, whereas none showed an increase. **H.** Left panel: the intensity of phase alignment (delta ITC z-score) did not correlate with the intensity of local neuronal activity (BHA t-score). Right panel: phase alignment correlated negatively with delta power (delta power t-score). **I.** Delta-band ITC in auditory cortex electrodes was higher in response to audiovisual (AV) speech than to unisensory auditory (A) speech (paired t-test: $t(24)=2.702$, $p=0.037$, corrected for multiple comparisons over frequency bands).

Phase-amplitude coupling links slow oscillations to local neuronal activity

Neuronal oscillations reflect momentary fluctuations in neuronal excitability through phase-amplitude coupling (PAC; (Buzsáki and Draguhn, 2004; Lakatos *et al.*, 2005; Whittingstall and Logothetis, 2009; Canolty and Knight, 2010)). We looked for evidence of PAC in auditory cortex during the perception of visual speech (see Figure 2A and 2B for an example) by computing the modulation index (MI; (Tort *et al.*, 2010)) between slow oscillations and BHA. Across participants, most auditory electrodes displayed significant PAC in response to visual speech (Figure 2C). The magnitude of PAC in auditory cortex correlated with the magnitude

of phase alignment, but not with the intensity of local neuronal activity (Figure 2D). The combination of phase alignment of auditory cortex to visual speech in the delta and theta bands with evidence of phase-amplitude coupling at these frequencies suggests that, even though visual speech does not increase the overall rate of neuronal activity in auditory cortex, it shapes the temporal dynamics of auditory cortical activity at frequencies that are relevant for the processing of auditory speech.

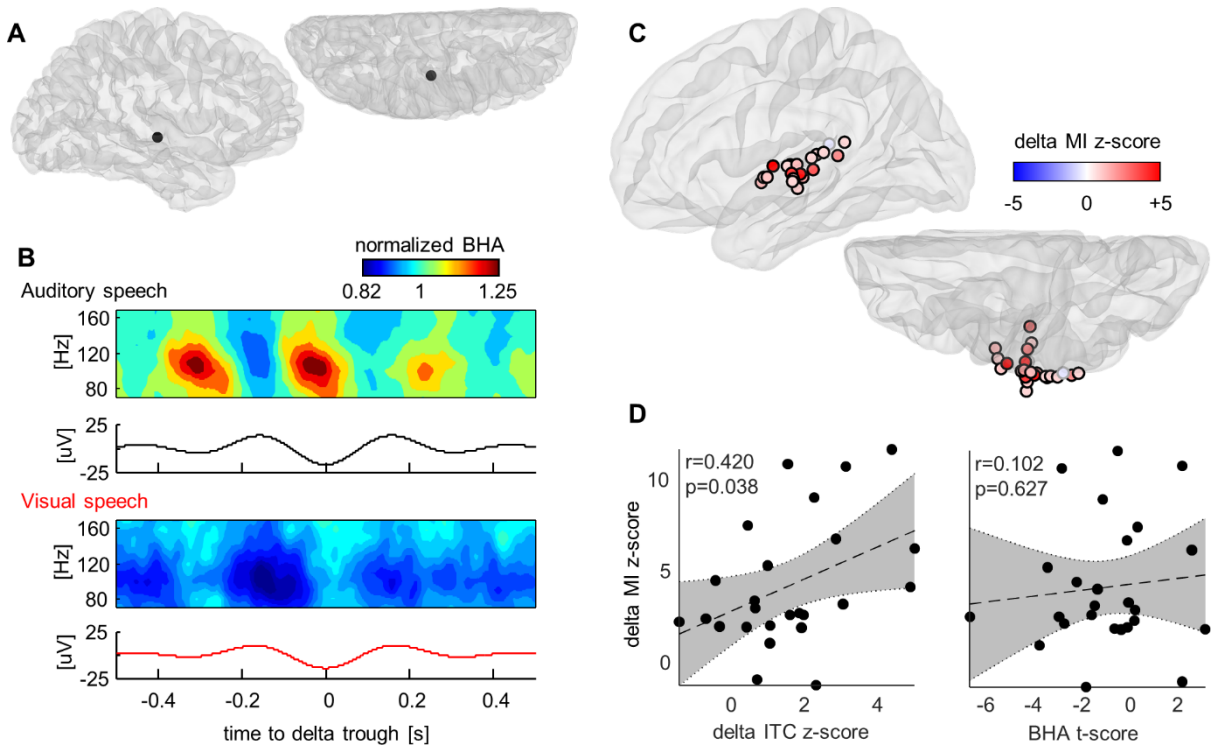

Figure 2. Phase-amplitude coupling to visual speech in auditory cortex.

A. A representative auditory electrode located in the superior temporal gyrus (lateral and superior views of the participant's right cerebral hemisphere). **B.** Displaying BHA (shown as time-frequency representations) locked to the trough of the simultaneous delta oscillations (shown as solid traces) illustrates phase-amplitude coupling (PAC), i.e. the systematic relationship between the phase of the slow oscillation and the amplitude of local neuronal activity, at the representative electrode (Canolty *et al.*, 2009)) in response to auditory (upper plot, black trace) and, to a lesser extent, visual speech (lower plot, red trace). **C.** Delta modulation index (MI) z-score for all auditory electrodes in response to visual speech. MI quantifies the magnitude of PAC. 22/25 electrodes displayed significant PAC in the delta band (circled in black; $p<<10^{-9}$), and 24/25 in the theta band (not shown). Z-scores

and significance testing were computed through a permutation test, FDR-corrected for multiple comparisons. **D.** The amplitude of PAC (delta MI z-score) correlated with the amount of delta phase alignment to visual speech (delta ITC z-score, left plot), but not with the intensity of local neuronal activity to visual speech (BHA t-score, right plot).

Auditory cortex represents the temporal dynamics of speech sounds from visual cues

Our observation of phase-locking of auditory cortex to visual speech, combined with the established correlation between parameters of visual speech such as the area of mouth opening and the envelope of speech sounds (Chandrasekaran *et al.*, 2009), suggests that auditory cortex might be able to build a relatively detailed representation of the temporal dynamics of speech from unisensory visual inputs. In order to probe this representation, we applied a technique of stimulus reconstruction (Mesgarani *et al.*, 2009) in an attempt to reconstitute the speech envelope from the responses of auditory cortex to visual speech alone (see Figure 3A and 3B for an example). Over participants, we found that reconstruction performed significantly above chance in a subset of auditory electrodes (Figure 3C), even in the complete absence of any auditory input. Importantly, these electrodes were among those that exhibited delta phase-locking to visual speech (compare Figures 3C and 1E, and see also Figure S1). As a control, we failed to reconstruct the speech envelope from visual cortex responses to auditory speech (not shown), despite the fact that auditory stimuli were in fact presented in that case. These results indicate that some portions of auditory cortex build a faithful representation of perceived speech based on visual speech cues alone. As has been shown before (E. M. Zion Golumbic *et al.*, 2013), this representation complements and enriches that built from auditory speech, thus facilitating the attentional selection of the speech stream, as well as its parsing into phonetically and linguistically relevant building blocks (Schroeder *et al.*, 2008; Schroeder and Lakatos, 2009; Arnal and Giraud, 2012; Giraud and Poeppel, 2012).

Since each stimulus was presented multiple times, it could be that stimulus representation in auditory cortex became more faithful over repetitions, as participants associated visual gestures and speech sounds in the stimulus set. To probe this, we reconstructed the speech envelope from auditory cortex responses to visual speech separately for each repetition of the stimuli. We did not find any tendency for reconstruction accuracy to improve over stimulus repetitions (Figure 3D). Nevertheless, it is very likely that repeated exposure will strengthen the associations between visual and auditory speech tokens, as suggested by the literature on speechreading training (Massaro, Cohen and Gesi, 1993).

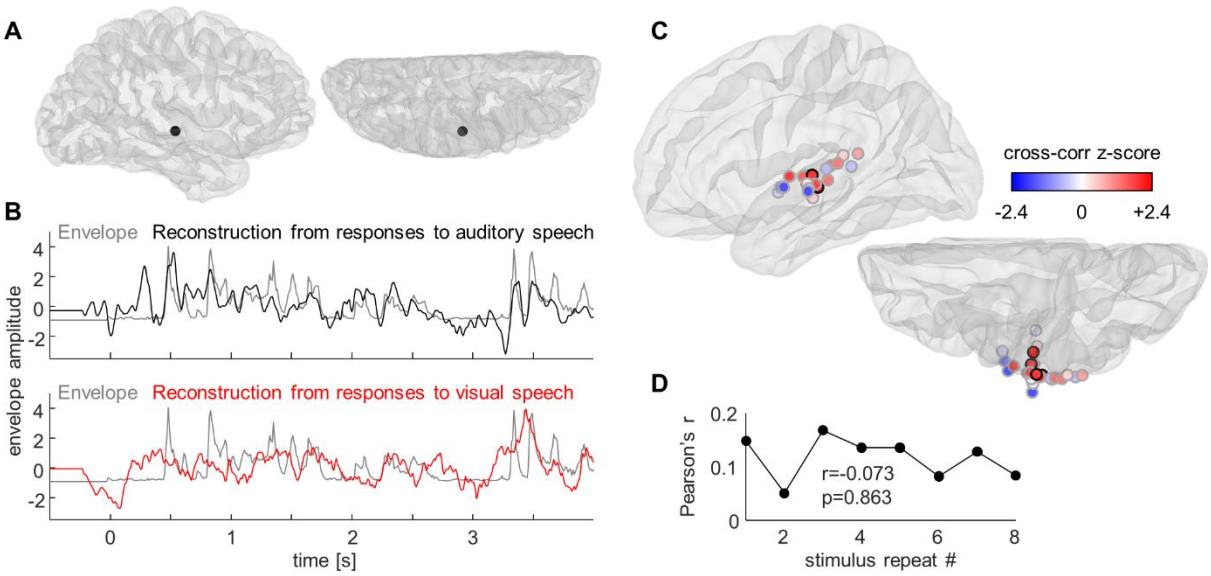

Figure 3. Reconstructing the speech envelope from auditory cortex responses to visual speech.

A. A representative auditory electrode located in the superior temporal gyrus (lateral and superior views of the participant's right cerebral hemisphere). **B.** Examples of actual (grey traces) and reconstructed speech envelopes from auditory cortex responses in the representative electrode to auditory (black trace) and visual speech (red trace). **C.** The accuracy of speech envelope reconstruction from auditory cortex responses to visual speech, quantified by the z-score of the cross-correlation between the reconstructed and actual speech envelope, is color-coded on lateral and superior views of a template left cerebral hemisphere. Reconstruction was significantly more accurate than chance in 4/25 electrodes (circled in black; $p=5.45*10^{-5}$). Z-scores and significance testing were computed through a permutation test, FDR-corrected for multiple comparisons. **D.** Using the brain responses from the 4 significant electrodes of Figure 3C as inputs, stimulus reconstruction was performed separately for each

repetition of the stimuli. The accuracy of speech envelope reconstruction did not increase as a function of stimulus repeat number (t-test on the slope of the regression line).

Little evidence of phase alignment to auditory speech in visual cortex

The phase-reset hypothesis makes a site- and direction-specific prediction regarding phase reset of auditory cortex oscillations by visual speech gestures. To test this prediction, we examined the responses of visual cortex to auditory speech (28 electrodes over 4 participants, selected according to anatomical and physiological criteria: location in the occipital lobe and increased BHA to visual speech). There was little detectable phase alignment of slow oscillations in visual cortex to auditory speech (Figure 4). This observation fits with the notion that the phase-resetting effect of visual speech on auditory cortex is specific, and is not merely due to an indiscriminate phase-reset of oscillations in sensory cortex by crossmodal stimuli.

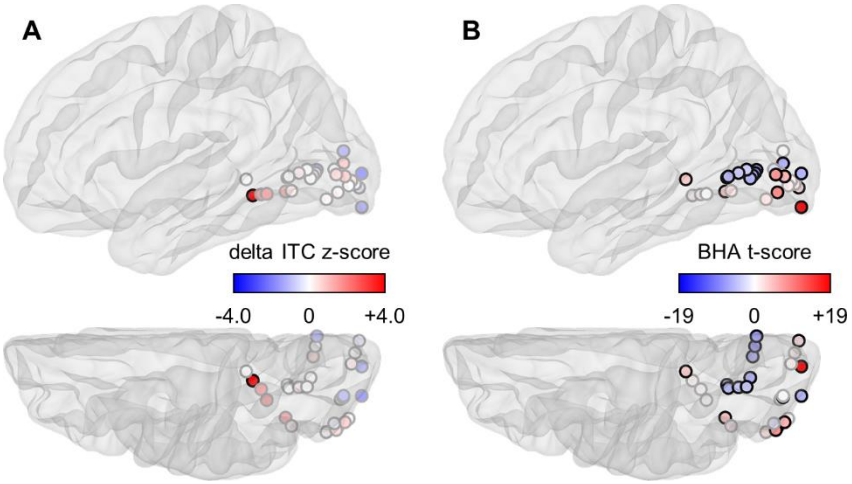

Figure 4. Little detectable phase alignment to auditory speech in visual cortex.

A. Delta ITC z-score for all visual electrodes in response to auditory speech is color-coded on lateral and superior views of a template left cerebral hemisphere (28 electrodes over 4 participants). 1/28 electrode displayed significant phase alignment (circled in black; p=0.05). Z-scores and significance testing were computed through a permutation test, FDR-corrected for multiple comparisons. **B.** BHA t-score for all visual electrodes in response to auditory speech stimuli. 7/28 electrodes displayed a significant increase in neuronal activity, whereas 10/28

displayed a decrease (circled in black). T-scores and significance testing were computed through a t-test relative to the pre-stimulus baseline, FDR-corrected for multiple comparisons.

DISCUSSION

It is widely observed that both phase-entrained low-frequency activity and fluctuations in broadband high-frequency activity in auditory cortex track the temporal dynamics of unisensory auditory speech. It is also widely held that visual speech gestures enhance intelligibility by facilitating auditory cortical entrainment to the speech stream (Crosse, Butler and Lalor, 2015; Crosse, Di Liberto and Lalor, 2016; Park *et al.*, 2016, 2018; Giordano *et al.*, 2017). Here, we used iEEG recordings for a more direct examination of the neurophysiological mechanisms underlying visual enhancement of auditory cortical speech processing. Our findings significantly elaborate the mechanistic description of crossmodal stimulus processing as a critical contribution to speech perception under complex and noisy natural conditions. Three aspects of the findings are novel and fundamentally important.

First, the low-frequency tracking reflects a pattern of phase resetting linked to the succession of visual cues, rather than simply a succession of evoked responses. Indeed, phase concentration is accompanied by an amplitude decrease, rather than the amplitude increase that defines evoked responses (Makeig *et al.*, 2004; Shah *et al.*, 2004). Interestingly, this may help to explain the paradoxical observation that, despite the general perceptual amplification that attends audiovisual speech, neurophysiological responses to multisensory audiovisual stimuli in both auditory and visual cortex are generally smaller than those to the preferred-modality stimulus alone (Besle *et al.*, 2008; Schepers, Yoshor and Beauchamp, 2014; Mercier *et al.*, 2015). While the physiological mechanisms of the low-frequency power decrease are not yet clear, our findings represent an unequivocal demonstration of cross-modal phase-reset in speech perception, and they strongly support the hypothesis that oscillatory phase reset is a

mechanism by which visual speech cues influence the processing of speech sounds by auditory cortex (Schroeder *et al.*, 2008).

Second, auditory cortical responses to visual speech in isolation reflect stimulus-specific features of the visual speech cues. As such, they suggest a key role for oscillatory phase as a neuronal coding mechanism—along with the intensity and spatial pattern of neuronal firing (Kayser *et al.*, 2009)—underlying specific aspects of audiovisual speech integration such as the categorical perception of syllables (ten Oever and Sack, 2015). Such a prediction could be tested in future studies by investigating how conflicting auditory and visual speech cues hijack spike-phase coding to cause perceptual illusions (McGurk and Macdonald, 1976).

Finally, the pattern of rapid quasi-rhythmic phase resetting we observe has strong implications for the mechanistic understanding of speech processing in general. Indeed, this phase resetting aligns the ambient excitability fluctuations in auditory cortex with the incoming sensory stimuli, potentially helping to parse the continuous speech stream into linguistically relevant processing units such as syllables (Schroeder *et al.*, 2008; Giraud and Poeppel, 2012; Zion Golumbic, Poeppel and Schroeder, 2012). As attention strongly reinforces the tracking of a specific speech stream (Mesgarani and Chang, 2012; E. M. Zion Golumbic *et al.*, 2013; O'Sullivan *et al.*, 2015), phase resetting will tend to amplify an attended speech stream above background noise, increasing its perceptual salience.

It is clear that visual enhancement of speech takes place within the context of strong top-down influences from frontal and parietal regions that support the processing of distinct linguistic features (Park *et al.*, 2016, 2018; Di Liberto *et al.*, 2018; Keitel, Gross and Kayser, 2018). It also appears that low-frequency oscillations relevant to speech perception can themselves be

modulated by transcranial electrical stimulation (Zoefel, Archer-Boyd and Davis, 2018). Our findings highlight the need to consider oscillatory phase in targeting potential neuromodulation therapy to enhance communication.

METHODS

Experimental design

Participants

Six patients (3 women, age range 21-52 years old) suffering from drug-resistant focal epilepsy and undergoing video-intracranial EEG (iEEG) monitoring at North Shore University Hospital (Manhasset, NY 11030, USA) participated in the experiments. All participants were native speakers of English. The participants provided written informed consent under the guidelines of the Declaration of Helsinki, as monitored by the Feinstein Institute for Medical Research's institutional review board.

Stimuli and task

Stimuli (E. M. Zion Golumbic *et al.*, 2013) were presented at the bedside using a laptop computer and Presentation software (version 17.2, Neurobehavioral Systems, Inc., Berkeley, CA, http://www.neurobs.com). Trials started with a 1-s fixation cross on a black screen. The participants then viewed or heard video clips (7-11 seconds) of a speaker telling a short story. The clips were cut off to leave out the last word. A written word was then presented on the screen, and the participants had to select whether that word ended the story appropriately or not. There was no time limit for participants to indicate their answer; reaction time was not monitored. There were 2 speakers (one woman) telling 4 stories each (8 distinct stories); each story was presented once with one of 8 different ending words (4 appropriate), for a total of 64 trials. These were presented once in each of 3 sensory modalities: audiovisual (movie with

audio track), auditory (soundtrack with a fixation cross on a black screen), visual (silent movie). Trial order was randomized, with the constraint that the same story could not be presented twice in a row, regardless of modality.

Intracranial EEG recordings

iEEG electrode localization

The placement of iEEG electrodes (subdural and depth electrodes, Ad-Tech Medical, Racine, WI, and Integra LifeSciences, Plainsboro, NJ) was determined on clinical grounds, without reference to this study. The localization and display of iEEG electrodes was performed using iELVis (http://ielvis.pbworks.com) (Groppe *et al.*, 2017). For each participant, a post-implantation high-resolution CT scan was coregistered with a post-implantation 3D T1 1.5-tesla MRI scan and then with a pre-implantation 3D T1 3-tesla MRI scan via affine transforms with 6 degrees of freedom using the FMRIB Linear Image Registration Tool included in the FMRIB Software Library (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki) (Jenkinson *et al.*, 2012) or the bbregister tool included in FreeSurfer (https://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferWiki) (Fischl, 2012). Electrodes were localized manually on the CT scan using BioImage Suite (http://bioimagesuite.yale.edu/) (Papademetris *et al.*, 2006). The pre-implantation 3D T1 MRI scan was processed using FreeSurfer to segment the white matter, deep grey matter structures, and cortex, reconstruct the pial surface, approximate the leptomeningeal surface (Schaer *et al.*, 2008), and parcellate the neocortex according to gyral anatomy (Desikan *et al.*, 2006). In order to compensate for the brain shift that accompanies the insertion of subdural electrodes through a large craniotomy, subdural electrodes were projected back to the pre-implantation leptomeningeal surface (Dykstra *et al.*, 2012) using iELVis.

iEEG recording and preprocessing

Intracranial EEG signals were referenced to a vertex subdermal electrode, filtered and digitized (0.1 Hz high-pass filter, 200 Hz low-pass filter, 500-512 samples per second, XLTEK EMU128FS or Natus Neurolink IP 256 systems, Natus Medical, Inc., Pleasanton, CA). Analysis was performed offline using the FieldTrip toolbox (http://www.fieldtriptoolbox.org/) (Oostenveld *et al.*, 2011) and custom-made programs for MATLAB (The MathWorks Inc., Natick, MA, https://www.mathworks.com/products/matlab.html). 60-Hz line noise and its harmonics were filtered out using a discrete Fourier transform filter, iEEG electrodes contaminated with noise or abundant epileptiform activity were identified visually and rejected, and the remaining iEEG signals were re-referenced to average reference.

Time-frequency analysis of iEEG signals

Time-frequency analysis was performed using a Morlet wavelet transform. Wavelets (3 cycles) were centered every 10 ms from -2 to +10 s with respect to stimulus onset, and every 1 Hz from 1-8 Hz, every 2 Hz at 10 and 12 Hz, and every 10 Hz from 70-150 Hz. The complex number resulting from the wavelet transform was used to compute the power and phase of oscillations.

Power

Single-trial power was baseline-corrected by dividing it by the mean power over trials of the same modality during the -0.5 to -0.25-s baseline period preceding stimulus onset (Grandchamp and Delorme, 2011). Note that moving the baseline to earlier or later periods (-1.0 to -0.75 s, -0.75 to -0.5 s, or -0.25 to 0 s) did not significantly alter our observations. Power was averaged over canonical frequency bands (delta: 1-3 Hz, theta: 4-7 Hz, alpha: 8-12 Hz). Broadband high-frequency activity (BHA), which reflects local neuronal activity (Ray *et al.*,

2008), was computed by dividing single-trial power within each 10-Hz frequency band between 70 and 150 Hz by its own mean over the trial, baseline-correcting as described above, and then averaging over frequency bands. Single-trial power was then averaged over trials in each modality and over time from +1 to +7 seconds relative to stimulus onset; the first and last seconds were ignored to leave out onset and offset responses.

In order to assess whether auditory electrodes displayed changes in power in response to visual speech, a paired t-test for power during stimulus presentation compared to baseline was computed in each electrode, modality and frequency band. The corresponding p-values (two-tailed, because the null hypothesis was that power did not either increase or decrease from baseline) were corrected for multiple comparisons across electrodes using an FDR procedure with a family-wise error rate set at 0.05, implemented in the Mass Univariate ERP toolbox (Groppe, Urbach and Kutas, 2011). Note that the Benjamini-Hochberg FDR procedure maintains adequate control of the family-wise error rate even in the case of positive dependencies between the observed variables (Benjamini and Hochberg, 1995; Groppe, Urbach and Kutas, 2011). The same approach was used in visual electrodes to test for power changes in response to auditory speech.

Intertrial coherence

The intertrial coherence (ITC) quantifies the phase alignment of iEEG oscillations over trials and ranges from 0 to 1, 1 indicating perfect phase alignment (Tallon-Baudry *et al.*, 1996). ITC was computed as the mean resultant length of the phase angle of slow oscillations over the 8 trials where the same stimulus was presented. Single-stimulus ITC was then averaged over time from +1 to +7 seconds relative to stimulus onset and over stimuli within each modality.

In order to assess the hypothesis that auditory electrodes displayed increased ITC to visual speech, a permutation test was used to generate a surrogate distribution of ITC under the null hypothesis. For that purpose, 8 trials were selected at random from the 64 trials in each modality and the ITC over these 8 trials was computed in the same fashion as the observed ITC. The procedure was repeated 1000 times. Observed ITC values were converted to z-scores relative to the surrogate distribution. The corresponding p-values (considering the z-scores as one-tailed, because the null hypothesis was that observed ITC values are not higher than expected by chance) were corrected for multiple comparisons over electrodes and frequency bands (delta, theta and alpha) using an FDR procedure with a family-wise error rate set at 0.05. The same approach was used in visual electrodes to test for increased ITC to auditory speech.

Phase-amplitude coupling

Phase-amplitude coupling refers to the systematic relationship between the phase of a slow oscillation and the intensity of local neuronal activity, approximated by BHA (Canolty *et al.*, 2009). Phase-amplitude coupling was quantified by computing the modulation index (MI) (Tort *et al.*, 2010) using custom-made MATLAB code. MI relates to the Kullback-Leibler distance between the observed distribution of BHA values, binned as a function of slow oscillatory phase, and a uniform distribution. It ranges from 0 to 1, 0 indicating absolutely no phase-amplitude coupling. MI was computed for each trial from +1 to +7 seconds relative to stimulus onset, and was then averaged over stimuli within each modality.

In order to assess the hypothesis that auditory electrodes displayed significant phase-amplitude coupling during visual speech, a permutation test was used to generate a surrogate distribution of MI under the null hypothesis. For that purpose, the BHA and slow oscillatory phase of trials within each modality were paired at random and a surrogate value of MI was computed in the

same fashion as the observed MI. The procedure was repeated 1000 times. Observed MI values were converted to z-scores relative to the surrogate MI distribution. The corresponding p-values (one-tailed, because the null hypothesis was that observed MI values are not higher than expected by chance) were corrected for multiple comparisons over electrodes and frequency bands (delta, theta and alpha) using an FDR procedure with a family-wise error rate set at 0.05.

Stimulus reconstruction

Because neuronal activity in auditory cortex reflects the dynamics of auditory stimuli, the spectro-temporal features of speech sounds can be reconstructed from the neural responses of auditory cortex (Mesgarani *et al.*, 2009; Pasley *et al.*, 2012). In order to determine whether cortex encodes features of speech stimuli that are detailed enough to allow their identification, the speech envelope was reconstructed from the iEEG responses using NAPLIB (Khalighinejad *et al.*, 2017). The rationale of reconstructing the speech envelope, a feature of auditory stimuli, from neural responses to unisensory visual speech is that the speech envelope correlates with the area of mouth opening (Chandrasekaran *et al.*, 2009; Schwartz and Savariaux, 2014). The broadband speech envelope was extracted by filtering the audio track of the video clips through a gammatone filter bank with 128 center frequencies equally spaced on the equivalent rectangle bandwidth-rate scale and ranging from 80 and 5000 Hz, approximating a cochlear filter (Carney and Yin, 1988); computing the power in each frequency band using a Hilbert transform; and averaging power over frequencies (University of Surrey's Institute of Sound Recording MATLAB Toolbox). The speech envelope was then reconstructed from the broadband iEEG signals (downsampled to 100 samples per second, then averaged over trials for each video clip in each modality) using optimal prior reconstruction, a linear mapping between the neural responses and the original stimulus (Mesgarani *et al.*, 2009). Lags of -200 to +200 ms between the speech envelope and the neural responses were allowed. The

reconstruction algorithm was first trained on 7 of the 8 stimuli in each modality, and then tested by reconstructing the speech envelope of the 8th stimulus from the corresponding neural responses. The procedure was repeated for all 8 video clips. In order to account for the varying lengths of the video clips, the speech envelopes were padded with zeros between -1 and +13 s relative to stimulus onset. The zero-lag cross-correlation between the actual and reconstructed speech envelopes (averaged over stimuli within each modality) was used as a metric for the accuracy of stimulus reconstruction.

In order to assess the hypothesis that the speech envelope can be reconstructed from the activity of auditory electrodes in response to visual speech, a permutation test was used to generate a surrogate distribution of cross-covariance under the null hypothesis. For that purpose, stimulus reconstruction was performed after the labels of the speech envelopes of each stimulus in each modality were shuffled, and surrogate values of the cross-covariance was computed in the same fashion as the observed cross-covariances. The procedure was repeated 1000 times. Observed cross-covariances were then converted to z-score relative to the surrogate distribution. The corresponding p-values (one-tailed, because the null hypothesis was that observed cross-covariance values are not higher than expected by chance) were corrected for multiple comparisons over electrodes using an FDR procedure with a family-wise error rate set at 0.05.

Electrode selection

The selection of electrodes for further analysis was based on a combination of anatomical and neurophysiological criteria. The anatomical criterion was based on the Desikan-Killiany parcellation of each participant's MRI (Desikan *et al.*, 2006). Within each participant, the selection of auditory electrodes started by identifying electrodes that lay in the superior temporal lobe (superior temporal gyrus, transverse temporal cortex, or banks of the superior

temporal sulcus of the Desikan-Killiany parcellation). Then, the BHA response of these electrodes to auditory speech was examined for a sustained increase (physiological criterion). BHA was averaged between +1 and +7 seconds relative to stimulus onset and compared to baseline using a two-tailed one-sample t-test. P-values were corrected for multiple comparisons over electrodes (2-8 per participant) using a false-discovery rate (FDR) procedure (Benjamini and Hochberg, 1995) with a family-wise error rate set at 0.05. 25 electrodes in 5 participants matched these criteria. Similarly, visual electrodes were defined as those electrodes that lay in the occipital lobe (lingual gyrus, pericalcarine cortex, cuneus, or lateral occipital cortex of the Desikan-Killiany parcellation) and that displayed a sustained BHA increase in response to visual speech. 28 electrodes in 4 participants matched these criteria.

Assessing the statistical significance of observed effects across all electrodes of interest

We computed the probability of observing a given number or more significant electrodes under the null hypothesis by simulating one billion null experiments and subjecting the simulated z-scores to the same FDR procedure as the observed data. The corresponding probabilities appear in the legends to the figures.

Data and software availability

Please see Table S1 for a list of materials and software used in this study. Data and custom-made software are available upon request from Pierre Mégevand (pierre.megevand@unige.ch).

ACKNOWLEDGEMENTS

We thank the patients for their participation; Erin Yeagle, Willie Walker Jr., the physicians, and other professionals of the Neurosurgery and Neurology departments of North Shore University Hospital for their assistance; Itzik Norman for help with brain surface

reconstruction; Bahar Khalighinejad for help with stimulus reconstruction. Part of the computations for this work were performed at the University of Geneva on the Baobab cluster. This work was supported by the Swiss National Science Foundation (grants 139829, 148388 and 167836 to PM), the NINDS (NS098976 to CES, MSB and ADM) and the Page and Otto Marx Jr. Foundation to ADM.

AUTHOR CONTRIBUTIONS

Conceptualization: PM, EZG, CES and ADM; Software: PM, MRM, DMG and NM; Formal Analysis: PM, MRM, DMG and NM; Investigation: PM and EZG; Writing – Original draft: PM; Writing – Review & Editing: PM, MSB, CES and ADM; Visualization: PM, MRM and DMG; Project Administration: PM, DMG and ADM; Funding Acquisition: PM, CES and ADM; Supervision: CES and ADM.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Ahissar, E. *et al.* (2001) ‘Speech comprehension is correlated with temporal response patterns recorded from auditory cortex.’, *Proceedings of the National Academy of Sciences of the United States of America*, 98(23), pp. 13367–72. doi: 10.1073/pnas.201400998.

Arnal, L. H. *et al.* (2009) ‘Dual Neural Routing of Visual Facilitation in Speech Processing’, *Journal of Neuroscience*, 29(43), pp. 13445–13453. doi: 10.1523/JNEUROSCI.3194-09.2009.

Arnal, L. H. and Giraud, A. L. (2012) ‘Cortical oscillations and sensory predictions’, *Trends in Cognitive Sciences*, 16(7), pp. 390–398. doi: 10.1016/j.tics.2012.05.003.

Arnal, L. H., Wyart, V. and Giraud, A.-L. (2011) ‘Transitions in neural oscillations reflect prediction errors generated in audiovisual speech.’, *Nature neuroscience*, 14(6), pp. 797–801. doi: 10.1038/nn.2810.

Beauchamp, M. S., Nath, A. R. and Pasalar, S. (2010) ‘fMRI-Guided Transcranial Magnetic Stimulation Reveals That the Superior Temporal Sulcus Is a Cortical Locus of the McGurk Effect’, *Journal of Neuroscience*, 30(7), pp. 2414–2417. doi: 10.1523/JNEUROSCI.4865-09.2010.

Benjamini, Y. and Hochberg, Y. (1995) ‘Controlling the false discovery rate: a practical and powerful approach to multiple testing’, *Journal of the Royal Statistical Society*. WileyRoyal Statistical Society, 57(1), pp. 289–300. doi: 10.2307/2346101.

Besle, J. *et al.* (2008) ‘Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans.’, *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28(52), pp. 14301–14310. doi: 10.1523/JNEUROSCI.2875-08.2008.

Buzsáki, G. and Draguhn, A. (2004) ‘Neuronal oscillations in cortical networks.’, *Science (New York, N.Y.)*, 304(5679), pp. 1926–9. doi: 10.1126/science.1099745.

Calvert, G. A. *et al.* (1997) ‘Activation of auditory cortex during silent lipreading.’, *Science*

(New York, N.Y.), 276(5312), pp. 593–6. doi: 10.1126/science.276.5312.593.

Canolty, R. T. *et al.* (2009) 'High gamma power is phase-locked to theta oscillations in human neocortex.', *Science (New York, N.Y.)*, 313(5793), pp. 1626–8. doi: 10.1126/science.1128115.

Canolty, R. T. and Knight, R. T. (2010) 'The functional role of cross-frequency coupling.', *Trends in cognitive sciences*. Elsevier Ltd, 14(11), pp. 506–15. doi: 10.1016/j.tics.2010.09.001.

Carney, L. H. and Yin, T. C. T. (1988) 'Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model.', *Journal of neurophysiology*, 60(5), pp. 1653–1677.

Chandrasekaran, C. *et al.* (2009) 'The natural statistics of audiovisual speech', *PLoS Computational Biology*, 5(7). doi: 10.1371/journal.pcbi.1000436.

Crosse, M. J., Butler, J. S. and Lalor, E. C. (2015) 'Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 35(42), pp. 14195–204. doi: 10.1523/JNEUROSCI.1829-15.2015.

Crosse, M. J., Di Liberto, G. M. and Lalor, E. C. (2016) 'Eye Can Hear Clearly Now: Inverse Effectiveness in Natural Audiovisual Speech Processing Relies on Long-Term Crossmodal Temporal Integration.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*. Society for Neuroscience, 36(38), pp. 9888–95. doi: 10.1523/JNEUROSCI.1396-16.2016.

Desikan, R. S. *et al.* (2006) 'An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest.', *NeuroImage*, 31(3), pp. 968–80. doi: 10.1016/j.neuroimage.2006.01.021.

Dykstra, A. R. *et al.* (2012) 'Individualized localization and cortical surface-based registration of intracranial electrodes', *NeuroImage*, 59(4), pp. 3563–3570. doi: 10.1016/j.neuroimage.2011.11.046.

Fischl, B. (2012) ‘FreeSurfer’, *NeuroImage*, 62(2), pp. 774–781. doi: 10.1016/j.neuroimage.2012.01.021.

Vander Ghinst, M. *et al.* (2016) ‘Left Superior Temporal Gyrus Is Coupled to Attended Speech in a Cocktail-Party Auditory Scene’, *The Journal of Neuroscience*. Society for Neuroscience, 36(5), pp. 1596–1606. doi: 10.1523/JNEUROSCI.1730-15.2016.

Giordano, B. L. *et al.* (2017) ‘Contributions of local speech encoding and functional connectivity to audio-visual speech perception.’, *eLife*, 6. doi: 10.7554/eLife.24763.

Giraud, A.-L. and Poeppel, D. (2012) ‘Cortical oscillations and speech processing: emerging computational principles and operations.’, *Nature neuroscience*. Nature Publishing Group, 15(4), pp. 511–7. doi: 10.1038/nn.3063.

Grandchamp, R. and Delorme, A. (2011) ‘Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials’, *Frontiers in Psychology*, 2(SEP), pp. 1–14. doi: 10.3389/fpsyg.2011.00236.

Greenberg, S. *et al.* (2003) ‘Temporal properties of spontaneous speech - A syllable-centric perspective’, *Journal of Phonetics*, 31(3–4), pp. 465–485. doi: 10.1016/j.wocn.2003.09.005.

Groppe, D. M. *et al.* (2017) ‘iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data.’, *Journal of neuroscience methods*. Elsevier B.V., 281, pp. 40–48. doi: 10.1016/j.jneumeth.2017.01.022.

Groppe, D. M., Urbach, T. P. and Kutas, M. (2011) ‘Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review’, *Psychophysiology*, 48(12), pp. 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x.

Jenkinson, M. *et al.* (2012) ‘FSL.’, *NeuroImage*, 62(2), pp. 782–90. doi: 10.1016/j.neuroimage.2011.09.015.

Kayser, C. *et al.* (2009) ‘Spike-Phase Coding Boosts and Stabilizes Information Carried by Spatial and Temporal Spike Patterns’, *Neuron*. Elsevier Ltd, 61(4), pp. 597–608. doi:

10.1016/j.neuron.2009.01.008.

Kayser, C., Petkov, C. I. and Logothetis, N. K. (2008) ‘Visual modulation of neurons in auditory cortex.’, *Cerebral cortex (New York, N.Y. : 1991)*, 18(7), pp. 1560–74. doi: 10.1093/cercor/bhm187.

Keitel, A., Gross, J. and Kayser, C. (2018) ‘Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features’, *PLoS Biology*, 16(3), pp. 1–19. doi: 10.1371/journal.pbio.2004473.

Khalighinejad, B. *et al.* (2017) ‘NAPLib: An open source toolbox for real-time and offline Neural Acoustic Processing’, in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. IEEE, pp. 846–850. doi: 10.1109/ICASSP.2017.7952275.

Lakatos, P. *et al.* (2005) ‘An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex’, *Journal of Neurophysiology*, 94(3), pp. 1904–1911. doi: 10.1152/jn.00263.2005.

Lakatos, P. *et al.* (2007) ‘Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex’, *Neuron*, 53(2), pp. 279–292. doi: 10.1016/j.neuron.2006.12.011.

Di Liberto, G. M. *et al.* (2018) ‘Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia.’, *NeuroImage*. Elsevier Inc., 175, pp. 70–79. doi: 10.1016/j.neuroimage.2018.03.072.

Luo, H., Liu, Z. and Poeppel, D. (2010) ‘Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation’, *PLoS Biology*, 8(8), pp. 25–26. doi: 10.1371/journal.pbio.1000445.

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

Makeig, S. *et al.* (2004) ‘Mining event-related brain dynamics.’, *Trends in cognitive sciences*. Elsevier Current Trends, 8(5), pp. 204–10. doi: 10.1016/j.tics.2004.03.008.

Massaro, D. W., Cohen, M. M. and Gesi, A. T. (1993) ‘Long-term training, transfer, and retention in learning to lipread.’, *Perception & psychophysics*, 53(5), pp. 549–62. doi: 10.3758/BF03205203.

McGurk, H. and Macdonald, J. (1976) ‘Hearing lips and seeing voices.’, *Nature*, 264, pp. 691–811. doi: 10.1038/264746a0.

Mercier, M. R. *et al.* (2015) ‘Neuro-oscillatory phase alignment drives speeded multisensory response times: an electro-corticographic investigation.’, *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 35(22), pp. 8546–57. doi: 10.1523/JNEUROSCI.4527-14.2015.

Mesgarani, N. *et al.* (2009) ‘Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex.’, *Journal of Neurophysiology*, 102(6), pp. 3329–3339. doi: 10.1152/jn.91128.2008.

Mesgarani, N. and Chang, E. F. (2012) ‘Selective cortical representation of attended speaker in multi-talker speech perception.’, *Nature*. Nature Publishing Group, 485(7397), pp. 233–6. doi: 10.1038/nature11020.

Micheli, C. *et al.* (2018) ‘Electrocorticography reveals continuous auditory and visual speech tracking in temporal and occipital cortex.’, *The European journal of neuroscience*. John Wiley & Sons, Ltd (10.1111). doi: 10.1111/ejn.13992.

Miller, L. M. and D’Esposito, M. (2005) ‘Perceptual fusion and stimulus coincidence in the cross-modal integration of speech.’, *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 25(25), pp. 5884–93. doi: 10.1523/JNEUROSCI.0896-05.2005.

O’Sullivan, A. E. *et al.* (2016) ‘Visual Cortical Entrainment to Motion and Categorical Speech Features during Silent Lipreading.’, *Frontiers in human neuroscience*, 10(January), p. 679. doi:

10.3389/fnhum.2016.00679.

O'Sullivan, J. A. *et al.* (2015) 'Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG.', *Cerebral cortex (New York, N.Y. : 1991)*, 25(7), pp. 1697–706. doi: 10.1093/cercor/bht355.

ten Oever, S. and Sack, A. T. (2015) 'Oscillatory phase shapes syllable perception', *Proceedings of the National Academy of Sciences*, 112(52), pp. 15833–15837. doi: 10.1073/pnas.1517519112.

Oostenveld, R. *et al.* (2011) 'FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data', *Computational Intelligence and Neuroscience*, 2011. doi: 10.1155/2011/156869.

Ozker, M. *et al.* (2017) 'A Double Dissociation between Anterior and Posterior Superior Temporal Gyrus for Processing Audiovisual Speech Demonstrated by Electrocorticography.', *Journal of cognitive neuroscience*. MIT PressOne Rogers Street, Cambridge, MA 02142-1209USAjournals-info@mit.edu, 29(6), pp. 1044–1060. doi: 10.1162/jocn_a_01110.

Papademetris, X. *et al.* (2006) 'BioImage Suite: An integrated medical image analysis suite: An update.', *The insight journal*, 2006, p. 209. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25364771.

Park, H. *et al.* (2016) 'Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility', *eLife*. eLife Sciences Publications Limited, 5(MAY2016), p. e14521. doi: 10.7554/eLife.14521.

Park, H. *et al.* (2018) 'Representational interactions during audiovisual speech entrainment: Redundancy in left posterior superior temporal gyrus and synergy in left motor cortex.', *PLoS biology*. Edited by J. Bizley, 16(8), p. e2006558. doi: 10.1371/journal.pbio.2006558.

Pasley, B. N. *et al.* (2012) 'Reconstructing speech from human auditory cortex', *PLoS Biology*, 10(1). doi: 10.1371/journal.pbio.1001251.

Ray, S. *et al.* (2008) ‘Neural correlates of high-gamma oscillations (60-200 Hz) in macaque local field potentials and their potential implications in electrocorticography.’, *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28(45), pp. 11526–11536. doi: 10.1523/JNEUROSCI.2848-08.2008.

Schaer, M. *et al.* (2008) ‘A Surface-based approach to quantify local cortical gyrification’, *IEEE Transactions on Medical Imaging*, 27(2), pp. 161–170. doi: 10.1109/TMI.2007.903576.

Schepers, I. M., Yoshor, D. and Beauchamp, M. S. (2014) ‘Electrocorticography Reveals Enhanced Visual Cortex Responses to Visual Speech.’, *Cerebral cortex (New York, N.Y. : 1991)*, (November), pp. 4103–4110. doi: 10.1093/cercor/bhu127.

Schroeder, C. E. *et al.* (2008) ‘Neuronal oscillations and visual amplification of speech.’, *Trends in cognitive sciences*, 12(3), pp. 106–13. doi: 10.1016/j.tics.2008.01.002.

Schroeder, C. E. and Lakatos, P. (2009) ‘Low-frequency neuronal oscillations as instruments of sensory selection’, *Trends in Neurosciences*. Elsevier, 32(1), pp. 9–18. doi: 10.1016/j.tins.2008.09.012.

Schwartz, J. L. and Savariaux, C. (2014) ‘No, There Is No 150 ms Lead of Visual Speech on Auditory Speech, but a Range of Audiovisual Asynchronies Varying from Small Audio Lead to Large Audio Lag’, *PLoS Computational Biology*, 10(7). doi: 10.1371/journal.pcbi.1003743.

Shah, A. S. *et al.* (2004) ‘Neural Dynamics and the Fundamental Mechanisms of Event-related Brain Potentials’, *Cerebral Cortex*, 14(5), pp. 476–483. doi: 10.1093/cercor/bhh009.

Shannon, R. V *et al.* (1995) ‘Speech recognition with primarily temporal cues.’, *Science (New York, N.Y.)*, 270(5234), pp. 303–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7569981.

Sumby, W. H. and Pollack, I. (1954) ‘Visual Contribution to Speech Intelligibility in Noise’, *The Journal of the Acoustical Society of America*, 26(2), pp. 212–215. doi: 10.1121/1.1907309.

Tallon-Baudry, C. *et al.* (1996) ‘Stimulus specificity of phase-locked and non-phase-locked 40

Hz visual responses in human.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 16(13), pp. 4240–9. doi: 10.1016/j.neuropsychologia.2011.02.038.

Tort, A. B. L. *et al.* (2010) 'Measuring Phase-Amplitude Coupling Between Neuronal Oscillations of Different Frequencies', *Journal of Neurophysiology*, 104(2), pp. 1195–1210. doi: 10.1152/jn.00106.2010.

Whittingstall, K. and Logothetis, N. K. (2009) 'Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex.', *Neuron*. Elsevier Ltd, 64(2), pp. 281–9. doi: 10.1016/j.neuron.2009.08.016.

Zion Golumbic, E. *et al.* (2013) 'Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party".', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33(4), pp. 1417–26. doi: 10.1523/JNEUROSCI.3675-12.2013.

Zion Golumbic, E. M. *et al.* (2013) 'Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party".', *Neuron*. Elsevier Inc., 77(5), pp. 980–91. doi: 10.1016/j.neuron.2012.12.037.

Zion Golumbic, E. M., Poeppel, D. and Schroeder, C. E. (2012) 'Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective', *Brain and Language*. Elsevier Inc., 122(3), pp. 151–161. doi: 10.1016/j.bandl.2011.12.010.

Zoefel, B., Archer-Boyd, A. and Davis, M. H. (2018) 'Phase Entrainment of Brain Oscillations Causally Modulates Neural Responses to Intelligible Speech.', *Current biology : CB*. Elsevier Ltd., 28(3), p. 401–408.e5. doi: 10.1016/j.cub.2017.11.071.

SUPPLEMENTAL INFORMATION

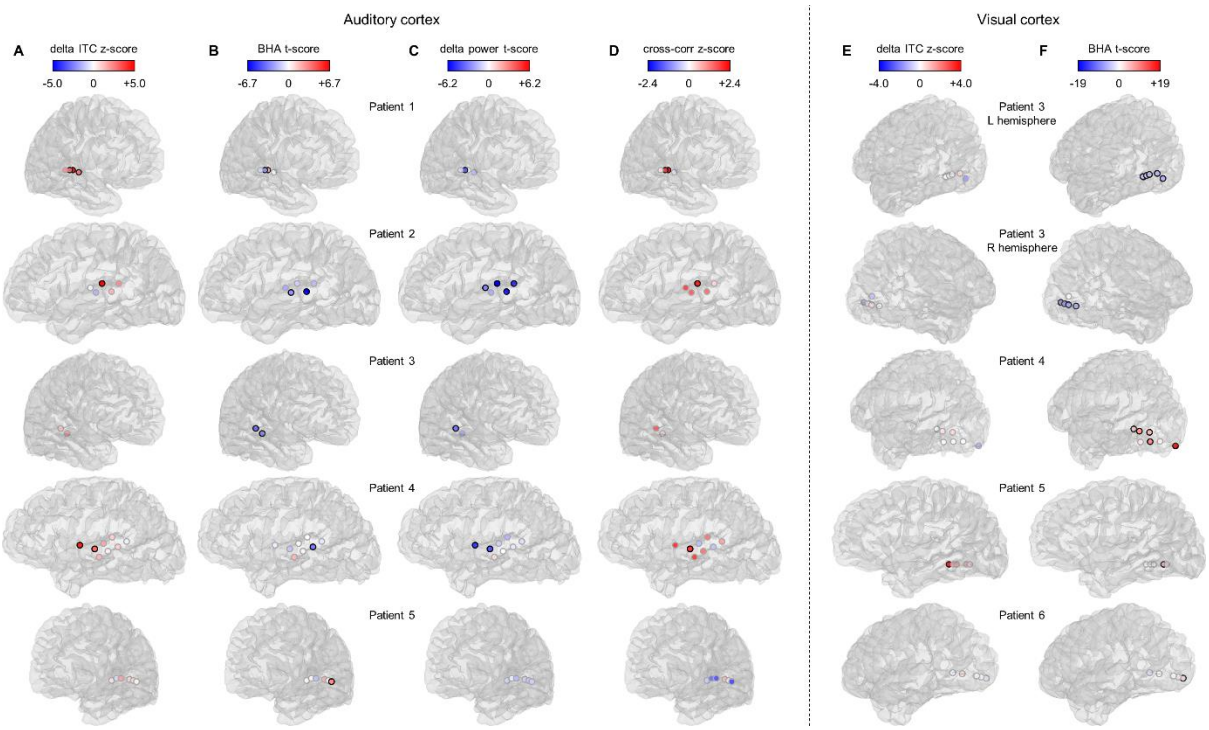

Figure S1. Results in individual participants.

A. Delta intertrial coherence (ITC) z-score indexes the magnitude of phase alignment of delta-band oscillations in auditory cortex in response to visual speech. **B.** Broadband high-frequency activity (BHA) t-score indexes changes in neuronal activity (with respect to the pre-stimulus baseline) in auditory cortex in response to visual speech. **C.** Delta power t-score quantifies changes in delta power in auditory cortex in response to visual speech. **D.** Cross-correlation (cross-corr) z-score quantifies the accuracy of stimulus reconstruction from auditory cortex responses to visual speech. **E.** Delta ITC z-score indexes the magnitude of delta-band oscillatory phase alignment to auditory speech in visual cortex. **F.** BHA t-score quantifies neuronal activity changes in visual cortex in response to auditory speech. Electrodes circled in black showed significant results (FDR-corrected across all electrodes and frequency bands).

Table S1. Materials and software used

RESOURCE	SOURCE	IDENTIFIER
Software		
Presentation	Neurobehavioral Systems, Inc.	RRID: SCR_002521; http://www.neurobs.com/
MATLAB	The MathWorks, Inc.	RRID: SCR_001622; https://www.mathworks.com/products/matlab.html
iELVis	(Groppe *et al.*, 2017)	RRID: SCR_016109; http://ielvis.pbworks.com/
FSL	(Jenkinson *et al.*, 2012)	RRID: SCR_002823; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki
FreeSurfer	(Fischl, 2012)	RRID: SCR_001847; https://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferWiki
BioImage Suite	(Papademetris *et al.*, 2006)	RRID: SCR_002986; http://bioimagesuite.yale.edu/
FieldTrip	(Oostenveld *et al.*, 2011)	RRID: SCR_004849; http://www.fieldtriptoolbox.org/
Mass Univariate ERP Toolbox	(Groppe, Urbach and Kutas, 2011)	RRID: SCR_016108; https://github.com/dmgroppe/Mass_Univariate_ERP_Toolbox
ModulationIndex	(Tort *et al.*, 2010)	N/A; https://github.com/pierremegevand/modulation_index
Gammatone filter bank	University of Surrey's Institute of Sound Recording	N/A; https://github.com/IoSR-Surrey/MatlabToolbox
Stimulus reconstruction	(Mesgarani *et al.*, 2009); (Khalighinejad *et al.*, 2017)	N/A; https://github.com/Naplib/Naplib
Materials		
EEG recording system	Natus Medical, Inc.	Product ID: XLTEK EMU128FS, Natus Neurolink IP 256
Intracranial EEG electrodes	Ad-Tech Medical	Cat# FG64C-SP10X-0C6, TS08R-SP10X-000, SD08R-SP05X-000
Intracranial EEG electrodes	Integra LifeSciences	Cat# AU8X8P4, AU1X8P, 15819D508

RRID: Resource Identification Initiative (https://scicrunch.org/resources)