Phase resetting in human auditory cortex to visual speech

Pierre Mégevand[1,2,13], 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

[13]Lead contact: pierre.megevand@unige.ch

*Correspondence: amehta@northwell.edu

SUMMARY

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 firing. This effect appears 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 firing 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 mechanism remains an open question. The cerebral processing of auditory and visual speech converges in multisensory cortical areas, especially the superior temporal sulcus (Beauchamp et al., 2010; Miller and D'Esposito, 2005). Crossmodal influences are also found in cortex traditionally considered to be unisensory; in particular, visual speech modulates the activity of auditory cortex (Besle et al., 2008; Calvert et al., 1997).

The articulatory movements that constitute visual speech tend to precede the production of speech sounds (Chandrasekaran et al., 2009; Schwartz and Savariaux, 2014), suggesting that visual speech might serve as an alerting cue to auditory cortex, preparing the neural circuits to process the incoming auditory speech more efficiently. Our hypothesis is that this preparation occurs through a resetting of the phase of neuronal oscillations in auditory cortex (ten Oever and Sack, 2015; Perrodin et al., 2015; Schroeder et al., 2008). This hypothesis rests on four findings.

First, auditory speech is rhythmic, with syllables arriving at a regular rate (4-7 Hz) nested within the slower rise and fall of prosody (1-3 Hz). These rhythmic features of speech are critical for it to be intelligible (Greenberg et al., 2003; Shannon et al., 1995). 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; Zion Golumbic et al., 2013). 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 (Kayser et al., 2008; Lakatos et al., 2007; Mercier et al., 2013, 2015).

The phase reset hypothesis makes a number of predictions. At the most basic level, we expect unisensory visual speech to reset the phase of oscillations in auditory cortex. In turn, this phase reset should influence neuronal firing through the coupling between oscillatory phase and neuronal excitability. Through this phase reset, auditory cortex should be able to build a representation of the temporal dynamics of speech sounds from visual speech cues alone. Finally, because visual speech tends to occur before auditory speech, we expect visual speech to phase-reset auditory cortex, but not the reverse.

RESULTS

Phase reset of delta-band 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 as they attended to segments of natural speech. 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 activity related to neuronal firing (as indexed by the amplitude of broadband high-frequency activity, BHA; (Lachaux et al., 2012; Ray et al., 2008)) in response to auditory speech (see Figure 1A to 1C for an example). To determine how visual speech influences activity in auditory cortex, we

computed BHA as well as power and intertrial coherence (ITC, a measure of the phase alignment) in the delta (1-3 Hz), theta (4-7 Hz) and alpha (8-12 Hz) oscillatory frequency bands.

Figures 1A to 1D show data for a representative electrode in auditory cortex that displayed phase alignment of its delta oscillations, but no neuronal firing increase, to unisensory visual speech. Across all electrodes and participants, there were two types of responses. On the one hand, 7 of 25 auditory electrodes displayed increased BHA responses to visual speech, whereas others showed decreases (Figure 1E). On the other, some electrodes displayed robust alignment of the phase of their slow oscillations in response to visual speech, as demonstrated by an ITC increase in the delta (Figure 1F; 6 of 25 electrodes) and, to a lesser extent, theta bands (data not shown; 3 of 25 electrodes). The amplitudes of BHA increase and delta ITC did not correlate over electrodes (Figure 1G), and the auditory electrodes that displayed increased neuronal firing or delta phase alignment were not the same (compare Figures 1E and 1F, and see also Figure S1). Furthermore, delta power did not increase, but in fact decreased, in those electrodes that displayed phase alignment (data not shown). These results support the view that the delta phase alignment to visual speech observed in some portions of auditory cortex is mediated by rapid, repetitive phase resetting of ongoing oscillations rather than by a succession of crossmodal sensory-evoked responses (Schroeder et al., 2008; Shah et al., 2004).

Phase-amplitude coupling links slow oscillations to neuronal firing

Neuronal oscillations reflect momentary fluctuations in neuronal excitability through phase-amplitude coupling (PAC; (Buzsáki and Draguhn, 2004; Canolty and Knight, 2010; Lakatos et al., 2005; Whittingstall and Logothetis, 2009)). 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 delta-band PAC in response to visual speech (delta band: 22 of 25 electrodes, Figure 2C; theta band: 24 of 25 electrodes, data not shown). The magnitude of PAC correlated with the magnitude of delta phase alignment, but not with the BHA response (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 firing rate of auditory cortex, it shapes the temporal dynamics of auditory cortical activity at frequencies that are relevant for the processing of auditory speech (Giraud and Poeppel, 2012; Schroeder et al., 2008).

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 sounds from unisensory visual speech 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 (a feature of auditory speech) 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 4 of 25 electrodes (Figure 3C), even in the complete absence of any auditory input. Importantly, these 4 electrodes were among those that exhibited delta phase-locking to visual speech (compare Figures 3C and 1F, and see also Figure S1). As a control, we failed to reconstruct the speech envelope from visual cortex responses to auditory speech, despite the fact that auditory stimuli were presented in that case.

These results indicate that some sites in auditory cortex build a faithful representation of perceived speech based on visual speech cues alone. As has been shown before (Zion Golumbic et al., 2013), this representation complements and enriches that built from auditory speech, thus facilitating the attentional selection of the speech stream (Arnal and Giraud, 2012; Schroeder and Lakatos, 2009) as well as its parsing into phonetically and linguistically relevant building blocks (Giraud and Poeppel, 2012; Schroeder et al., 2008).

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 et al., 1993).

No evidence of phase alignment to auditory speech in visual cortex

Because the articulatory movements of speech tend to precede phonation (Chandrasekaran et al., 2009; Schroeder et al., 2008; Schwartz and Savariaux, 2014), the phase-reset hypothesis makes a site- and direction-specific prediction regarding phase reset of auditory cortex oscillations by visual speech. 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). Visual cortex exhibited relatively robust increases in neuronal firing in response to auditory speech (Figure 4A), but little to no detectable phase alignment of its slow oscillations (Figure

3B). 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.

DISCUSSION

We used iEEG recordings in humans to show that auditory cortex tracks the temporal dynamics of unisensory visual speech using the phase of its low-frequency oscillations. This phase alignment in turn determines systematic, stimulus-locked variations in neuronal firing, as indexed by fluctuations in broadband high-frequency activity. 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 a succession of evoked responses. Indeed, phase concentration is accompanied by a power decrease, rather than the power increase that accompanies 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; Mercier et al., 2015; Schepers et al., 2014). 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 signals 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 the intensity and spatial pattern of neuronal responses (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 signals, potentially helping to parse the continuous speech stream into linguistically relevant processing units such as syllables (Giraud and Poeppel, 2012; Zion Golumbic et al., 2012). As attention strongly reinforces the tracking of a specific speech stream (Mesgarani and Chang, 2012; Zion Golumbic et al., 2013), phase resetting will tend to amplify an attended speech stream above background noise, increasing its perceptual salience. These findings carry therapeutic implications, by highlighting the need to consider oscillatory phase in targeting potential neuromodulation therapy to enhance communication.

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 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., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M.M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc. Natl. Acad. Sci. U. S. A. *98*, 13367–13372.

Arnal, L.H., and Giraud, A.L. (2012). Cortical oscillations and sensory predictions. Trends Cogn. Sci. *16*, 390–398.

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. J. Neurosci. *30*, 2414–2417.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. *57*, 289–300.

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., and Giard, M.-H. (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans. J. Neurosci. *28*, 14301–14310.

Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. Science *304*, 1926–1929.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, P.W., Iversen, S.D., and David, A.S. (1997). Activation of auditory cortex during silent lipreading. Science *276*, 593–596.

Canolty, R.T., and Knight, R.T. (2010). The functional role of cross-frequency coupling. Trends Cogn. Sci. *14*, 506–515.

Canolty, R.T., Edwards, E., Dalal, S.S., Soltani, M., Nagarajan, S.S., Kirsch, H.E., Berger, M.S., Barbaro, N.M., and Knight, R.T. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. Science *313*, 1626–1628.

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. J. Neurophysiol. *60*, 1653–1677.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A.A. (2009). The natural statistics of audiovisual speech. PLoS Comput. Biol. *5*.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage *31*, 968–980.

Dykstra, A.R., Chan, A.M., Quinn, B.T., Zepeda, R., Keller, C.J., Cormier, J., Madsen, J.R., Eskandar, E.N., and Cash, S.S. (2012). Individualized localization and cortical surface-based registration of intracranial electrodes. Neuroimage *59*, 3563–3570.

Fischl, B. (2012). FreeSurfer. Neuroimage *62*, 774–781.

Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. *15*, 511–517.

Grandchamp, R., and Delorme, A. (2011). Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials. Front. Psychol. *2*, 1–14.

Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal properties of spontaneous speech - A syllable-centric perspective. J. Phon. *31*, 465–485.

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*, 1711–1725.

Groppe, D.M., Bickel, S., Dykstra, A.R., Wang, X., Mégevand, P., Mercier, M.R., Lado, F.A., Mehta, A.D., and Honey, C.J. (2017). iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. J. Neurosci. Methods *281*, 40–48.

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., and Smith, S.M. (2012). FSL. Neuroimage *62*, 782–790.

Kayser, C., Petkov, C.I., and Logothetis, N.K. (2008). Visual modulation of neurons in auditory cortex. Cereb. Cortex *18*, 1560–1574.

Kayser, C., Montemurro, M.A., Logothetis, N.K., and Panzeri, S. (2009). Spike-Phase Coding Boosts and Stabilizes Information Carried by Spatial and Temporal Spike Patterns. Neuron *61*, 597–608.

Khalighinejad, B., Nagamine, T., Mehta, A., and Mesgarani, N. (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.

Lachaux, J.-P., Axmacher, N., Mormann, F., Halgren, E., and Crone, N.E. (2012). High-frequency neural activity and human cognition: Past, present and possible future of intracranial EEG research. Prog. Neurobiol. *98*, 279–301.

Lakatos, P., Shah, A.S., Knuth, K.H., Ulbert, I., Karmos, G., and Schroeder, C.E. (2005). An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. J. Neurophysiol. *94*, 1904–1911.

Lakatos, P., Chen, C.M., O'Connell, M.N., Mills, A., and Schroeder, C.E. (2007). Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. Neuron *53*, 279–292.

Luo, H., and Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. Neuron *54*, 1001–1010.

Makeig, S., Makeig, S., Debener, S., Debener, S., Onton, J., Onton, J., Delorme, A., and Delorme, A. (2004). Mining event-related brain dynamics. Trends Cogn. Sci. *8*, 204–210.

Massaro, D.W., Cohen, M.M., and Gesi, A.T. (1993). Long-term training, transfer, and retention in learning to lipread. Percept. Psychophys. *53*, 549–562.

McGurk, H., and Macdonald, J. (1976). Hearing lips and seeing voices. Nature *264*, 691–811.

Mercier, M.R., Foxe, J.J., Fiebelkorn, I.C., Butler, J.S., Schwartz, T.H., and Molholm, S.

(2013). Auditory-driven phase reset in visual cortex: Human electrocorticography reveals mechanisms of early multisensory integration. Neuroimage *79*, 19–29.

Mercier, M.R., Molholm, S., Fiebelkorn, I.C., Butler, J.S., Schwartz, T.H., and Foxe, J.J. (2015). Neuro-oscillatory phase alignment drives speeded multisensory response times: an electro-corticographic investigation. J. Neurosci. *35*, 8546–8557.

Mesgarani, N., and Chang, E.F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. Nature *485*, 233–236.

Mesgarani, N., Shamma, S.A., David, S. V, and Fritz, J.B. (2009). Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. J. Neurophysiol. *102*, 3329–3339.

Miller, L.M., and D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J. Neurosci. *25*, 5884–5893.

ten Oever, S., and Sack, A.T. (2015). Oscillatory phase shapes syllable perception. Proc. Natl. Acad. Sci. *112*, 15833–15837.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. *2011*.

Papademetris, X., Jackowski, M.P., Rajeevan, N., DiStasio, M., Okuda, H., Constable, R.T., and Staib, L.H. (2006). BioImage Suite: An integrated medical image analysis suite: An update. Insight J. *2006*, 209.

Pasley, B.N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S.A., Crone, N.E., Knight, R.T., and Chang, E.F. (2012). Reconstructing speech from human auditory cortex. PLoS Biol. *10*.

Perrodin, C., Kayser, C., Logothetis, N.K., and Petkov, C.I. (2015). Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-

sensitive cortex. Proc. Natl. Acad. Sci. *112*, 273–278.

Ray, S., Crone, N.E., Niebur, E., Franaszczuk, P.J., and Hsiao, S.S. (2008). Neural correlates of high-gamma oscillations (60-200 Hz) in macaque local field potentials and their potential implications in electrocorticography. J. Neurosci. *28*, 11526–11536.

Schaer, M., Bach Cuadra, M., Tamarit, L., Lazeyras, F., Eliez, S., and Thiran, J.P. (2008). A Surface-based approach to quantify local cortical gyrification. IEEE Trans. Med. Imaging *27*, 161–170.

Schepers, I.M., Yoshor, D., and Beauchamp, M.S. (2014). Electrocorticography Reveals Enhanced Visual Cortex Responses to Visual Speech. Cereb. Cortex 4103–4110.

Schroeder, C.E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. *32*, 9–18.

Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. Trends Cogn. Sci. *12*, 106–113.

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 Comput. Biol. *10*.

Shah, A.S., Bressler, S.L., Knuth, K.H., Ding, M., Mehta, A.D., Ulbert, I., and Schroeder, C.E. (2004). Neural Dynamics and the Fundamental Mechanisms of Event-related Brain Potentials. Cereb. Cortex *14*, 476–483.

Shannon, R. V, Zeng, F.G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. Science *270*, 303–304.

Sumby, W.H., and Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. J. Acoust. Soc. Am. *26*, 212–215.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., and Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. J. Neurosci. *16*, 4240–

4249.

Tort, A.B.L., Komorowski, R., Eichenbaum, H., and Kopell, N. (2010). Measuring Phase-Amplitude Coupling Between Neuronal Oscillations of Different Frequencies. J. Neurophysiol. *104*, 1195–1210.

Whittingstall, K., and Logothetis, N.K. (2009). Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. Neuron *64*, 281–289.

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 Lang. *122*, 151–161.

Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. Neuron *77*, 980–991.

FIGURES

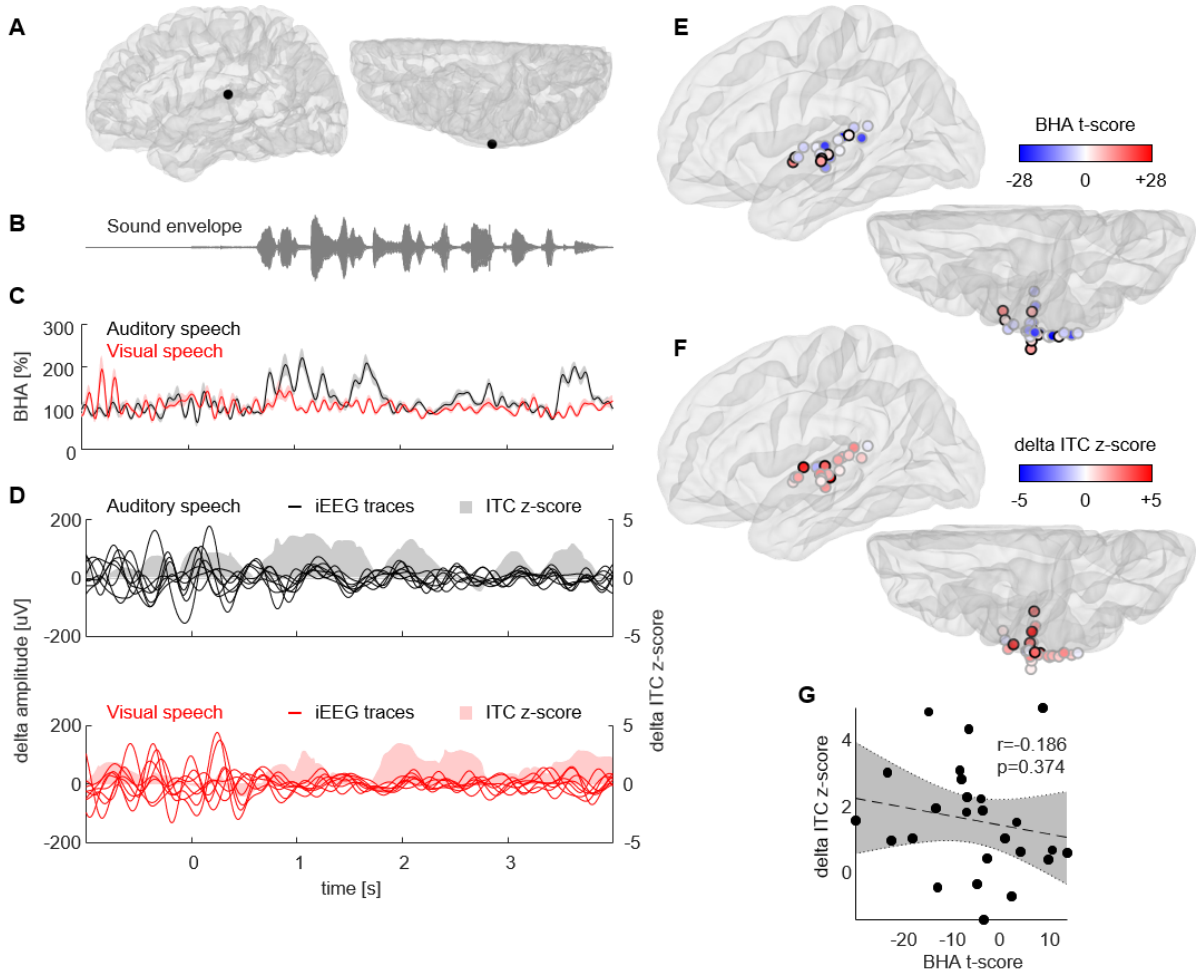

Figure 1. Responses of auditory cortex to visual speech.

A. A representative auditory electrode located in the superior temporal gyrus (lateral and superior views of the participant's left cerebral hemisphere). **B.** Sound wave of a representative speech stimulus. **C.** The representative electrode displayed increased broadband high-frequency activity (BHA, indexing neuronal firing) 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 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).

E. BHA for all auditory electrodes (25 electrodes over 5 participants) in response to visual speech is color-coded on lateral and superior views of a template left cerebral hemisphere. 7 of 25 electrodes displayed a significant increase in neuronal firing (circled in black; $p=7.6*10^{-8}$). 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 across electrodes 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.** Delta ITC for all auditory electrodes in response to visual speech. 6 of 25 electrodes displayed significant phase alignment (circled in black; $p=6.87*10^{-7}$). Z-scores and significance testing were computed through a permutation test, with FDR correction for multiple comparisons. **G.** The intensity of neuronal firing (BHA t-score) did not correlate with the amplitude of delta phase alignment to visual speech (delta ITC z-score).

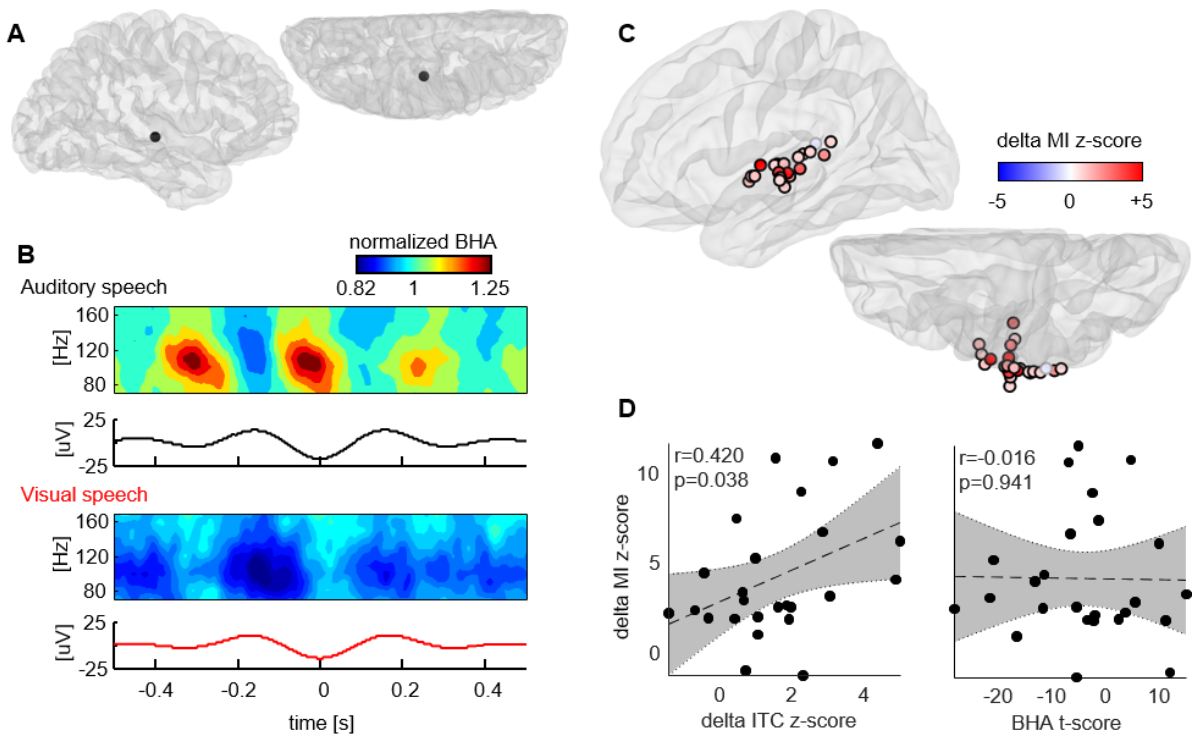

Figure 2. Phase-amplitude coupling to visual speech in auditory cortex.

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 neuronal firing, at the representative electrode (Canolty et al., 2006)) 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 of 25 electrodes displayed significant PAC (circled in black; $p<<10^{-9}$). 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 firing to visual speech (BHA t-score, right plot).

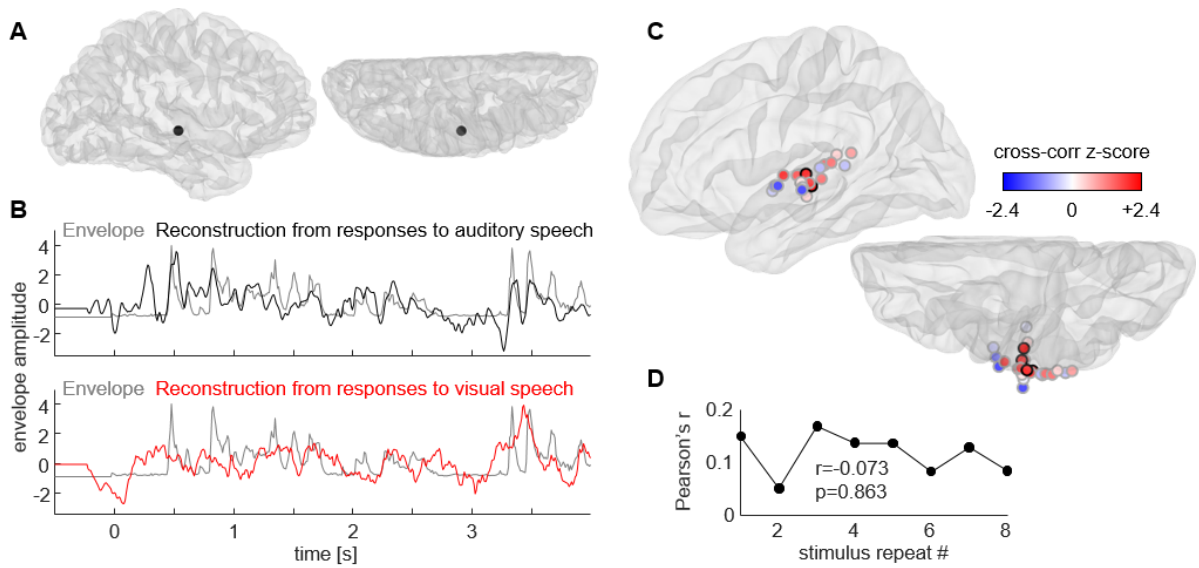

Figure 3. Reconstructing the speech envelope from auditory cortex responses to visual speech.

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. The 4 electrodes where reconstruction was significantly more accurate than chance are 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).

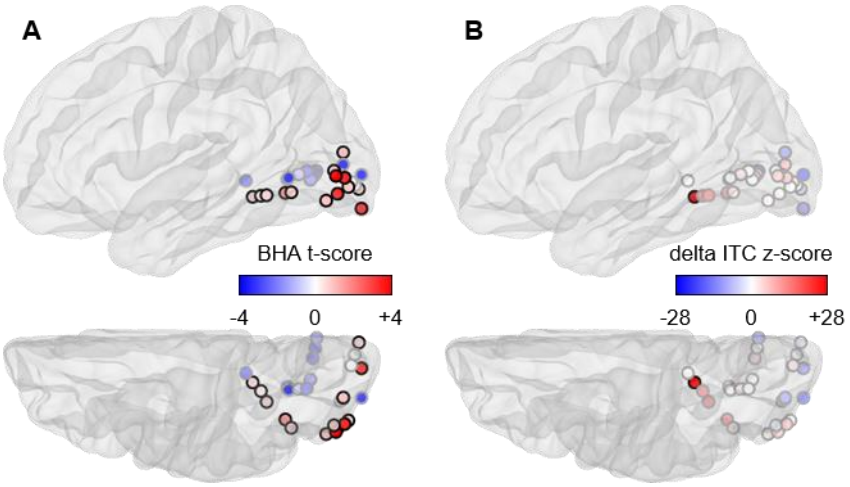

Figure 4. No detectable phase alignment to auditory speech in visual cortex.

A. BHA t-score for all visual electrodes (28 electrodes over 4 participants) in response to auditory speech stimuli is color-coded on lateral and superior views of a template left cerebral hemisphere. 14 of 28 electrodes displayed a significant increase in neuronal firing (circled in black; $p<10^{-9}$). T-scores and significance testing were computed through a t-test relative to the pre-stimulus baseline, FDR-corrected for multiple comparisons. **B.** Delta ITC z-score for all visual electrodes in response to auditory speech. Only 1 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.

METHODS

Contact for reagent and resource sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Pierre Mégevand (pierre.megevand@unige.ch).

Experimental model and subject details

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.

Method details

Stimuli and task

Stimuli (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 (8-12 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 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.

Intracranial EEG recordings 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

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 250 ms preceding stimulus onset (Grandchamp and Delorme, 2011). 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 approximates local neuronal firing (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.

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.

Phase-amplitude coupling

Phase-amplitude coupling refers to the systematic relationship between the phase of a slow oscillation and the intensity of neuronal firing, here approximated by BHA (Canolty et al., 2006). Phase-amplitude coupling was quantified by computing the modulation index (MI) (Tort et al., 2010). 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.

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. 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. 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.

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.

Quantification and statistical analysis

Statistical analysis for power

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. 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 et al.,

2011). The same approach was used in visual electrodes to test for power changes in response to auditory speech.

Statistical analysis for intertrial coherence

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.

Statistical testing for phase-amplitude coupling

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.

Statistical analysis for 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.

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

The EEG recording system and iEEG electrodes, as well as MATLAB and Presentation, are commercially available. The iELVis, FSL, FreeSurfer, BioImage Suite and FieldTrip software are open-source. The MATLAB function to compute FDR correction is part of the open-source Mass Univariate ERP toolbox (Groppe et al., 2011), available on GitHub: https://github.com/dmgroppe/Mass_Univariate_ERP_Toolbox. The custom-made MATLAB

function to compute MI is open-source, available on GitHub: https://github.com/pierremegevand/modulation_index. The MATLAB gammatone filter bank function is part of the University of Surrey's Institute of Sound Recording MATLAB Toolbox, available on GitHub: https://github.com/IoSR-Surrey/MatlabToolbox. The MATLAB function to perform stimulus reconstruction is part of NAPLIB (Khalighinejad et al., 2017), an open-source toolbox for real-time and offline neural acoustic processing, available on GitHub: https://github.com/Naplib/Naplib. Data and custom-made software produced for this study are available upon request from the Lead Contact, Pierre Mégevand (pierre.megevand@unige.ch).

Key Resources Table

See separate document.

SUPPLEMENTAL INFORMATION

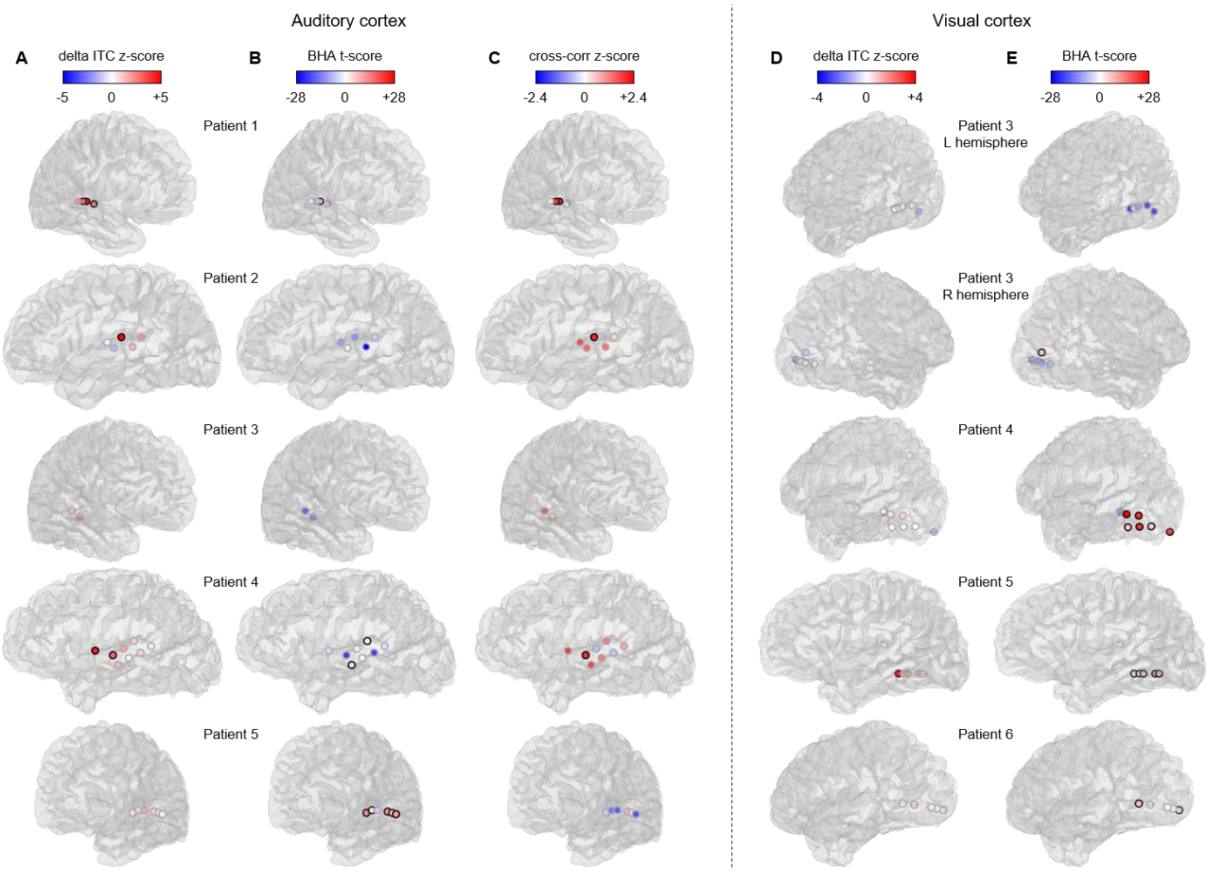

Figure S1. Related to Figures 1, 3 and 4. Results in individual participants.

This figure contains the individual patient versions of Figures 1E and 1F (auditory cortex, delta ITC z-score and BHA t-score), 3C (auditory cortex, cross-corr z-score), and 4 (visual cortex, delta ITC z-score and BHA t-score). **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.** Cross-correlation (cross-corr) z-score quantifies the accuracy of stimulus reconstruction from auditory cortex responses to visual speech. **D.** Delta ITC z-score indexes the magnitude of delta-band oscillatory phase alignment to auditory speech in visual cortex. **E.** BHA t-score quantifies neuronal activity changes in visual cortex in response to auditory speech.