

Visual speech cues recruit neural oscillations to optimise auditory perception: Ways forward for research on human communication



Benedikt Zoefel

Centre de Recherche Cerveau et Cognition (CerCo), CNRS UMR 5549, CHU Purpan, Pavillon Baudot, 31052, Toulouse, France

ARTICLE INFO

Keywords:

Speech perception
Prediction
Neural oscillations
Transcranial alternating current stimulation
Neural entrainment
Cross-modality

ABSTRACT

In pandemic times, when visual speech cues are masked, it becomes particularly evident how much we rely on them to communicate. Recent research points to a key role of neural oscillations for cross-modal predictions during speech perception. This article bridges several fields of research – neural oscillations, cross-modal speech perception and brain stimulation – to propose ways forward for research on human communication. Future research can test: (1) whether “speech is special” for oscillatory processes underlying cross-modal predictions; (2) whether “visual control” of oscillatory processes in the auditory system is strongest in moments of reduced acoustic regularity; and (3) whether providing information to the brain via electric stimulation can overcome deficits associated with cross-modal information processing in certain pathological conditions.

The human brain cannot be maximally responsive to all input it receives. Consequently, it has developed various mechanisms to efficiently allocate its attentional and perceptual resources to when they are needed. Such mechanisms, however, work only well if cues are available that predict the timing (and identity) of upcoming information. Such cues are not restricted to the modality of the information that is to be predicted (Bauer et al., 2020). This fact has become painfully evident in the current pandemic: If the mouth is covered during speaking, comprehension often declines. This – quite literal – masking of visual cues is particularly problematic for those who crucially rely on them to comprehend speech, such as hearing-impaired individuals or those living in a country whose language they do not fully speak.

Which neural processes does the brain employ for such cross-modal predictions during speech perception? A new study by Biau and colleagues (in press) points to a crucial role of neural oscillations. Such oscillations reflect regular fluctuations of neural excitability (Buzsáki and Draguhn, 2004) and, consequently, sensitivity to incoming information. Biau et al. provide important evidence that these oscillations are modulated by visual speech cues (the speaker's lip movements) so that moments of highest auditory sensitivity coincide with the expected timing of acoustic information (Fig. 1A).

In an elegant experimental paradigm, participants were exposed to silent movies of a person speaking and were asked to detect a target tone embedded in noise. As typical for human speech, coherence between lip movements and the (non-presented) acoustic envelope peaked in the theta-range (~4–8 Hz). As expected, participants' electroencephalogram

(EEG) over visual brain regions followed the theta-rhythm of the lip movements. Importantly however, this visual rhythm also modulated their ability to detect the auditory target. Connectivity analyses revealed theta oscillations in visual sources preceding – and potentially driving – those in auditory sources. These findings are exciting as they suggest that visual information functionally modulates auditory perception during speech processing (Fig. 1A). But results go even further. Longer exposure to silent movies increased the visually-induced rhythmic modulation of auditory detection, and decreased the lag between visual and auditory sources. Such a result is important as it speaks against a simple relay of visual information to auditory regions, and points to a neural process that takes some time to adapt to the timing of stimulus input, such as a neural oscillator synchronising to external rhythms (Pikovsky et al., 2001).

The study by Biau et al. does not only bring us closer to understanding multi-sensory interactions during speech perception, it also opens up new questions for future work. From the brain's perspective, human speech might seem special as visual cues robustly precede auditory information (Chandrasekaran et al., 2009). Does this mean that the reported cross-modal effect is tailored to speech, or can vision's role of orchestrating audition's oscillations be extended to multi-modal information processing in general? On the one hand, neural oscillations synchronise more reliably to speech than to other sounds that are matched in their broadband acoustic envelope (Peelle et al., 2013). A recent study (Hauswald et al., 2018), demonstrating vision's ability to extract phonological information from lip movements, suggests that such speech-specific processes might exist in the visual system as well. On the

E-mail address: benedikt.zoefel@cnrs.fr.

<https://doi.org/10.1016/j.crneur.2021.100015>

Received 3 June 2021; Received in revised form 21 June 2021; Accepted 21 June 2021

2665-945X/© 2021 The Author. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

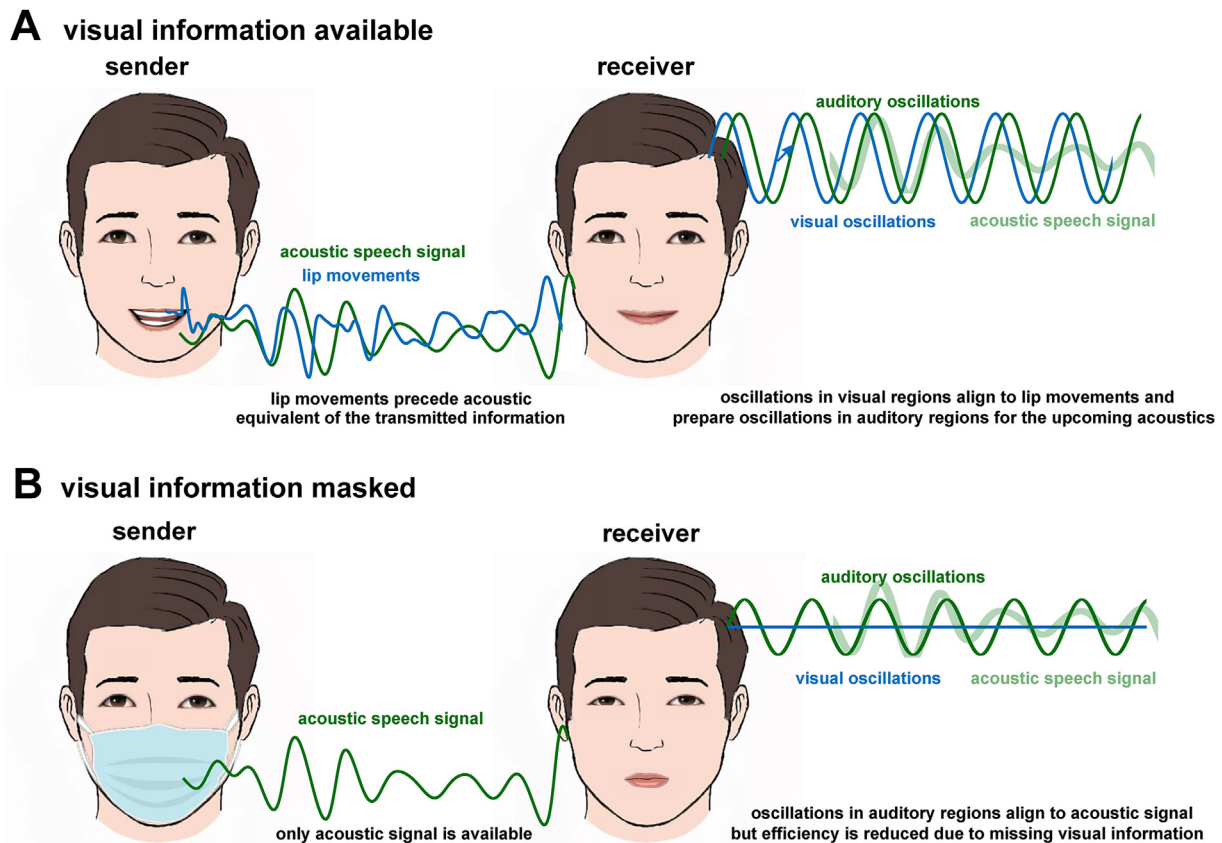


Fig. 1. A. Interpretation of experimental results reported by Biau et al. (in press). In typically produced human speech (left), lip movements (blue) precede the acoustic speech signal (green). Example signals have been band-pass filtered to illustrate fluctuations in the theta-range ($\sim 4\text{--}8$ Hz). In the receiver's brain (right), neural oscillations in visual regions (blue) align to lip movements and prepare oscillations in auditory regions (green) so that they become aligned with the acoustic signal (faint green). In the study described, the acoustic signal was not presented; instead, the detection of a target tone was found to be modulated by the theta phase of lip movements, demonstrating visual speech cues driving auditory perception. B. In scenarios in which such visual cues are unavailable, oscillations in auditory regions can still align to the speech signal. However, due to the lack of preparatory signals from visual regions, they are less efficient in adjusting to upcoming input, potentially leading to impaired speech perception.

other hand, cross-modal optimisation of auditory processing exists beyond the human species and seems to involve similar oscillatory processes (Atilgan et al., 2018). In humans, a delay of $\sim 30\text{--}100$ ms between visual and auditory input is optimal for perception in general, and not restricted to speech (Thorne and Debener, 2014). In the current study (Biau et al., in press), visual speech cues modulated the detection of tones, again indicating a more general (and not speech-specific) process. The question of whether “speech is special” for oscillatory processes underlying multi-modal predictions is still open, and answering it an exciting endeavour for the future.

It remains an open question how precisely oscillations in the visual and auditory systems interact during speech perception. Neural oscillations can align to the rhythm of speech even in the absence of visual cues (Fig. 1B; Peelle et al., 2013), and additional work is required to identify how vision can make this alignment more efficient. Human speech is not always rhythmic (Cummins, 2012) and neural oscillations might not align well in such moments of reduced rhythmicity and temporal predictability. Speculatively, oscillations in the auditory system might therefore most strongly rely on visual control when alignment to the acoustic speech signal is difficult, and when visual cues predict the timing of upcoming speech more faithfully than acoustic cues alone.

If combined with recent advances in brain stimulation research, results obtained by Biau and colleagues could also help us design applications for people for whom visual cues are not available, or who struggle to process multi-sensory information, as reported for autism spectrum disorders (Gepner and Féron, 2009). If visual cues cannot be used to

optimise auditory processing in such conditions, then providing them to the brain via electric stimulation might overcome the resulting perceptual deficits. Due to its regularity, transcranial alternating current stimulation (tACS) is optimally suited to manipulate neural oscillations (Herrmann et al., 2013). Stimulating visual regions with theta-tACS might therefore lead to enhanced speech perception if applied at the “correct” phase lag relative to the acoustic speech envelope. Such optimal phase lags often vary across participants when tACS is used to boost synchronisation between oscillations and speech in auditory regions (van Bree et al., 2021), and similar inter-individual variability might be expected for tACS over visual regions. An interesting possibility that remains to be tested is that optimal phase lags for the two modalities are correlated *within* individuals, with vision “preferring” a shorter lag. Moreover, applying tACS at non-optimal phase lags should cancel potential benefits of visual speech cues, and reveal a causal role of cross-modal oscillatory processes for perception.

When visual speech cues are masked (Fig. 1B), it becomes particularly evident how much we rely on them to communicate. This phenomenon seems intimately linked to the brain's effort to optimise resources and adapt its sensitivity to the expected occurrence of information. Evidence accumulates for neural oscillations underlying such predictive processes, with vision preparing audition by adapting the oscillation's high-excitability phase to the expected timing of upcoming input. Further investigation of oscillatory processes during speech perception will reveal important insights into human communication and may lead to promising new developments to improve it.

CRediT authorship contribution statement

Benedikt Zoefel: Conceptualization, Visualization, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the CNRS. I thank Emmanuel Biau for providing example signals shown (in modified form) in Fig. 1.

Appendix A. Peer Review Overview and Supplementary data

A Peer Review Overview and (sometimes) Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.crneur.2021.100015>.

References

- Atilgan, H., Town, S.M., Wood, K.C., Jones, G.P., Maddox, R.K., Lee, A.K.C., Bizley, J.K., 2018. Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. *Neuron* 97, 640–655. <https://doi.org/10.1016/j.neuron.2017.12.034> e4.
- Bauer, A.-K.R., Debener, S., Nobre, A.C., 2020. Synchronisation of neural oscillations and cross-modal influences. *Trends Cognit. Sci.* 24, 481–495. <https://doi.org/10.1016/j.tics.2020.03.003>.
- Biau, E., Wang, D., Park, H., Jensen, O., Hanslmayr, S., (in press). Auditory detection is modulated by theta phase of silent lip movements. *Curr. Res. Neurobiol.*, (this issue).
- Buzsáki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. <https://doi.org/10.1126/science.1099745>.
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., Ghazanfar, A.A., 2009. The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5, e1000436 <https://doi.org/10.1371/journal.pcbi.1000436>.
- Cummins, F., 2012. Oscillators and syllables: a cautionary note. *Front. Psychol.* 3, 364. <https://doi.org/10.3389/fpsyg.2012.00364>.
- Gepner, B., Féron, F., 2009. Autism: a world changing too fast for a mis-wired brain? *Neurosci. Biobehav. Rev.* 33, 1227–1242. <https://doi.org/10.1016/j.neubiorev.2009.06.006>.
- Hauswald, A., Lithari, C., Collignon, O., Leonardelli, E., Weisz, N., 2018. A visual cortical network for deriving phonological information from intelligible lip movements. *Curr. Biol.* 28, 1453–1459. <https://doi.org/10.1016/j.cub.2018.03.044> e3.
- Herrmann, C.S., Rach, S., Neuling, T., Strüber, D., 2013. Transcranial alternating current stimulation: a review of the underlying mechanisms and modulation of cognitive processes. *Front. Hum. Neurosci.* 7 <https://doi.org/10.3389/fnhum.2013.00279>.
- Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebr. Cortex* 23, 1378–1387. <https://doi.org/10.1093/cercor/bhs118>.
- Pikovsky, A., Rosenblum, M., Kurths, J., 2001. Synchronization: A Universal Concept in Nonlinear Sciences, Cambridge Nonlinear Science Series. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511755743>.
- Thorne, J.D., Debener, S., 2014. Look now and hear what's coming: on the functional role of cross-modal phase reset. *Hear. Res.* 307, 144–152. <https://doi.org/10.1016/j.heares.2013.07.002>.
- van Bree, S., van, Sohoglu, E., Davis, M.H., Zoefel, B., 2021. Sustained neural rhythms reveal endogenous oscillations supporting speech perception. *PLoS Biol.* 19, e3001142 <https://doi.org/10.1371/journal.pbio.3001142>.