INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Congruent Audiovisual Speech Enhances Cortical Envelope Tracking during Auditory Selective Attention

Zhen Fu[1], *Jing Chen*[1]

[1]Department of Machine Intelligence, Speech and Hearing Research Center, and Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China
`fuzhen364@pku.edu.cn, chenj@cis.pku.edu.cn`

Abstract

Listeners usually have the ability to selectively attend to the target speech while ignoring competing sounds. The mechanism that top-down attention modulates the cortical envelope tracking to speech was proposed to account for this ability. Additional visual input, such as lipreading was considered beneficial for speech perception, especially in noise. However, the effect of audiovisual (AV) congruency on the dynamic properties of cortical envelope tracking activities was not discussed explicitly. And the involvement of cortical regions processing AV speech was unclear. To solve these issues, electroencephalography (EEG) was recorded while participants attending to one talker from a mixture for several AV conditions (audio-only, congruent and incongruent). Approaches of temporal response functions (TRFs) and inter-trial phase coherence (ITPC) analysis were utilized to index the cortical envelope tracking for each condition. Comparing with the audio-only condition, both indices were enhanced only for the congruent AV condition, and the enhancement was prominent over both the auditory and visual cortex. In addition, timings of different cortical regions involved in cortical envelope tracking activities were subject to stimulus modality. The present work provides new insight into the neural mechanisms of auditory selective attention when visual input is available.

Index Terms: audiovisual speech, auditory selective attention, EEG, temporal response function, inter-trial phase coherence

1. Introduction

In complex auditory scenes where several speakers are talking at the same time, a listener with normal hearing is usually able to selectively attend to the target speaker while ignoring interfering speakers. This phenomenon is well known as the "cocktail party problem" [1]. This phenomenon could be explained by the top-down attention modulation on cortical envelope tracking to speech [2, 3, 4]. Firstly, electroencephalography (EEG) and magnetoencephalography (MEG) studies exhibited that auditory neural activities entrain to the temporal envelope of speech [5]. Furthermore, during a selective attention task, neuronal excitability was modulated by the top-down attention, in which the cortical envelope tracking was enhanced for the attended and inhibited for the unattended speech [2, 3, 4]. In these studies, an approach called stimulus reconstruction was utilized to map the recorded neural responses backward to stimulus features (e.g., temporal envelope) via a set of reconstruction filters [6, 7]. Indexed by the correlation coefficient between the actual envelope (of both attended and unattended speech) and the reconstructed envelope, cortical envelope tracking was greater for the attended than the unattended speech.

During daily face-to-face conversations, auditory and visual cues are both available. Visual cues such as talker's articulatory movements are beneficial for speech perception, especially in noisy environments [8], which is attributed to the predictive and constraint effect of congruent visual input [9]. On one hand, congruent visual inputs, such as lip movements, precede and correlate with corresponding vocal signals [10], they are utilized to improve the prediction about the timing of the upcoming auditory input. On the other hand, lip movements indicating the place and manner of articulation serves to constrain the candidates of phoneme for the upcoming syllable [11].

Relative to the audio-only condition, the stimulus reconstruction accuracy was found improved by additional visual input only when it is congruent with the attended speech [12, 13]. However, the backward reconstruction filters are not neurophysiologically interpretable [14], therefore the properties of neural encoding remain unknown. To investigate the properties of cortical envelope tracking to audiovisual (AV) speech, the forward mapping (from speech envelope to neural responses) was utilized [15]. Temporal response functions (TRFs), defined as the channel-specific impulse responses of the forward mapping system, are estimated to reflect how, when and where cortical envelope tracking works [4, 5]. Measuring with MEG, TRFs peak amplitude of the attended speech was significantly greater for the congruent AV condition than the audio-only condition [15]. However, the reported TRFs waveforms were averaged across MEG channels and thus smeared both the temporal profile of TRFs and the spatial distribution of TRFs across cortex. Besides, the reported topographies of TRFs were averaged between the AV and audio-only condition. Therefore, the effect of AV congruency on the spatial-temporal distribution of TRFs still remains unclear. Spatio-frequency property was another aspect to explore the mechanism underlying cortical envelope tracking of AV speech. The channel-specific inter-trial phase coherence (ITPC) describes the phase consistency of the temporal pattern of neural responses to speech at each frequency [16, 17]. Averaged across representative channels, ITPC values were greater for the congruent AV condition than the audio-only condition [15]. Similarly, the spatial distribution of ITPC values were averaged across conditions so that the effect of AV congruency on the spatial-frequency distribution of ITPC remains unclear.

As a complement to the previous study which focused on the backward stimulus reconstruction [12], the present study aimed to investigate the effect of AV congruency on the dynamic properties of cortical envelope tracking activities and the association of cortical regions in AV speech processing. Approaches of TRFs and ITPC analysis were utilized. From the temporal perspective, TRFs reflected how, when and where cortex responded to temporal envelope of speech. From the frequency perspective, ITPC exhibited the phase coherence of neural responses within the frequency range corresponding to speech temporal envelope.

http://dx.doi.org/10.21437/Interspeech.2020-1957

2. Methods

2.1. Subjects

The subject set is the same as the previous work [12]. Sixteen participants (7 females, age range: 22–27 years) with normal hearing participated in the experiment. All subjects were Mandarin-native speakers, had normal or corrected-to-normal vision, and reported no medical history of brain injury or cognitive deficits. Subjects were given informed consent in accordance with a protocol approved by the Peking University Institutional Review Board.

2.2. Stimuli and experimental procedure

The AV stimuli were explicitly described in a previous report [12]. Concisely, there were 20 speech mixtures, with each lasting for one minute and consisting of speech segment of one female and one male talker (Mandarin native speakers). Accordingly, there were 40 recorded videos shooting talkers' mouth movements while they narrating the speech.

Experiment was conducted in an acoustically and electrically shielded booth. Before the stimulus presentation, subjects were cued to attend to a certain talker of the mixture. Audio stimuli were presented diotically through an ER-3A earphone (Etymotic Research) at 65 dB SPL. Video stimuli were presented on a displayer 1 meter in front of the subject. During the stimulus presentation, four conditions were established (Table 1). For audio-only condition (i.e., A), subjects were instructed to maintain visual fixation on the white crosshair. For AV conditions (i.e., AVa, AVu and AVi), subjects were instructed to maintain visual fixation on talker's mouth. For each subject, five mixtures out of twenty were assigned to each condition, and each mixture was presented twice with different attended talker. Thus, each subject undertook 40 trials in total (2 attended talkers × 5 speech mixtures × 4 conditions).

2.3. EEG data acquisition and preprocessing

The EEG dataset was the same as previously reported in [12]. In brief, Continuous EEG data was acquired by a NeuroScan SynAmps2 system (Compumedics), using a 64-channel electrode cap with a 10/20 layout. Besides 62 scalp electrodes, 1 reference (nose-tip), 1 ground (forehead midline) and 4 bipolar electrooculograms electrodes were also used. EEG recordings were bandpass filtered (0.15–100 Hz), sampled (500 Hz) and stored for offline analysis. A trigger indicating the beginning of each trial was recorded along with the EEG.

Details about the preprocessing were described previously [12]. Noisy EEG channels were recalculated by interpolation, EEG data were then re-referenced to a common average, downsampled (64) Hz and baseline corrected. Since low-frequency ($<$ 8 Hz) neural activities were reported phase-locked to speech temporal envelopes [16, 18], EEG data were also bandpass filtered (2–8 Hz). Besides, electroocular artifacts were removed using independent component analysis [19]. All the procedures were performed using the EEGLAB toolbox [20, 21] with MATLAB (Mathworks).

Table 1: *Experimental conditions and stimuli contents.*

Cond.	**Stimuli contents**	
	Audio	Video
A	Speech mixture of a male and a female talker	Black screen with white crosshair
AVa		Congruent attended talker
AVu		Congruent unattended talker
AVi		Incongruent talker

2.4. Speech temporal envelope calculation

The waveform of each clean speech was first passed through 8 bandpass filters to account for the bandpass characteristic of the basilar membrane. Center frequencies of the filters were uniformly spaced between 150 Hz and 8000 Hz on the equivalent rectangular bandwidth scale [22]. Hilbert transform was then performed on the output of each filter. The analytical envelope signals obtained from the Hilbert transform were further raised to the power of 0.3 to account for the nonlinear compressive characteristic of the cochlea [23], and lowpass filtered ($<$ 8 Hz). Subsequently, the envelopes of all bands were averaged.

2.5. TRFs estimation

In the present study, temporal envelope of the actually attended speech was used to estimate the TRFs. For speech temporal envelope $s(t)$ sampled at discrete time t ($t = 1, \cdots, T$), and the corresponding EEG $r(t, n)$ recorded at channel n ($n = 1, \cdots, N$), suppose that a set of spatio-temporal filters (i.e., the channel-specific TRFs) $w(\tau, n)$ could map from $s(t)$ to $r(t, n)$ in a convolutional way, as in equation 1:

$$r(t, n) = \sum_{\tau} w(\tau, n) s(t - \tau) + \epsilon(t, n), \tag{1}$$

where $\epsilon(t, n)$ represents the residual responses not explained by the model. The TRFs coefficients were calculated using the reverse correlation method with the ridge regression to solve the ill-posed problems and overfitting, as in equation 2:

$$w = (S^T S + \lambda M)^{-1} S^T r, \tag{2}$$

where S is a matrix consists of lagged time series of the speech envelope $s(t)$, λ is the ridge parameter and M is a quadratic penalty (details were referred to [7]). Time lags ranging from -50 ms pre-stimulus to 450 ms post-stimulus were used. For each trial, once TRFs were calculated, the predicted response $\hat{r}(t, n)$ was calculated by equation 1. Afterward, for each EEG channel n , the correlation coefficient (Pearson's r) between $\hat{r}(t, n)$ and $r(t, n)$ was calculated as the index of prediction accuracy.

For TRFs calculation of each condition, a leave-one-out cross-validation procedure was adopted to optimize the ridge parameter λ, whereby EEG data of each trial was predicted using the TRFs obtained by averaging the TRFs calculated on every other trials. Such procedure was repeated for ridge parameters with the range of $2^1, 2^3, \cdots, 2^{19}$. The optimal λ was determined by maximizing the prediction accuracy averaged across trials and subjects. For each condition, TRFs corresponding to the optimal ridge parameter were averaged across subjects and trials so that they were not biased toward either trial or subject. In the present study, TRFs analysis was implemented based on the mTRF toolbox [7], and the optimal λ was determined as 2^7 for all conditions.

2.6. ITPC analysis

For a stimulus s (e.g., attending to a certain speech in a certain condition), suppose there are K_s within-stimulus trials. For each EEG channel of each trial k within the K_s trials, to estimate the momentary phase $\varphi_k(t, f)$, the Morlet wavelet decomposition was performed in logarithmic steps between 1 and 20

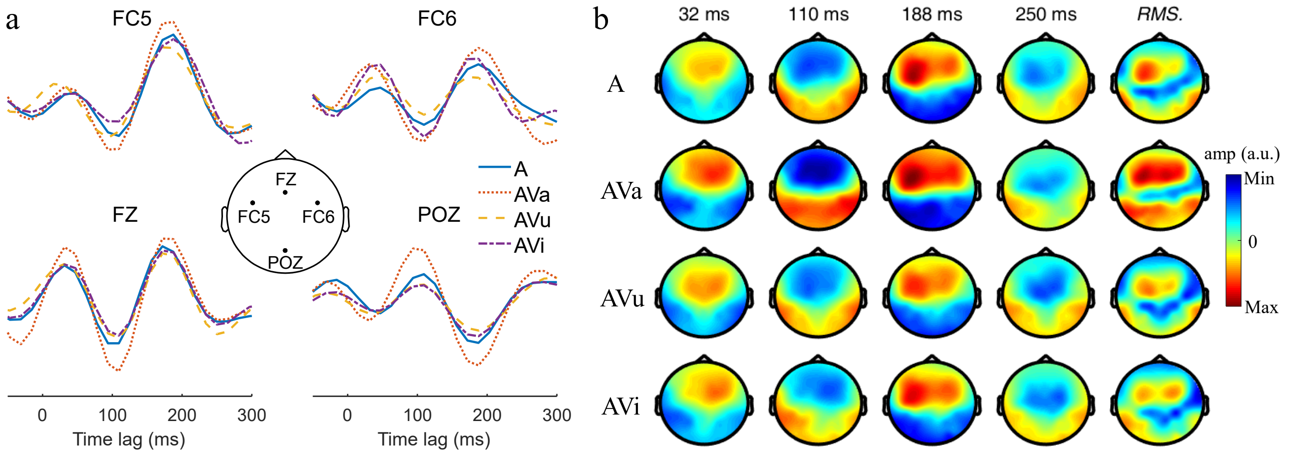

Figure 1: *TRFs of different conditions. (a) TRFs at several representative electrodes. (b) Topographies of TRFs at representative time lags and RMS values of the TRFs between 0 and 250 ms.*

Hz with wavelet cycles increasing from 4 to 10. Subsequently, the ITPC at each time-frequency bin was calculated by averaging phase angles across K_s trials, as in equation 3:

$$ITPC_{t,f,s} = \frac{1}{K_s}|\sum_{k=1}^{K_s} e^{i\cdot\varphi_k(t,f)}|. \tag{3}$$

This procedure was repeated for 160 stimuli (2 attended talkers × 20 speech mixtures × 4 conditions). For each condition, the within-stimuli ITPC was obtained by averaging $ITPC_{t,f,s}$ across time course and stimuli. As the control, the across-stimuli ITPC was obtained using the same approach except that K_s trials corresponded to different stimuli selected randomly. The difference between within-stimuli and across-stimuli ITPC was finally reported.

3. Results

3.1. TRFs estimation

To examine the effect of AV congruency on cortical processing of speech, TRFs of each condition were estimated. Figure 1(a) shows the estimated TRFs at several electrodes covering fontal (FZ), fronto-temporal (FC5, FC6) and partieto-occipital (POZ) regions. The patterns are similar to the results reported in [13, 24]. Generally, for all conditions, TRFs had distinct fluctuations across time that peaks or valleys appear at certain time lags. For each electrode, the fluctuations were greater for condition AVa than other conditions. For each condition, timings of the fluctuation were dependent of the cortical region. Topographies of TRFs were further demonstrated in Figure 1(b), to manifest the effect of AV congruency on spatio-temporal pattern of TRFs. Among time lags at which TRFs fluctuated the most in Figure 1(a), TRFs amplitudes of the AVa condition were found the greatest among all conditions. Specifically, at an early stage of about 32 ms post-stimulus, TRFs had a prominent peak over the frontal region (associated with cognitive processing) for all conditions. At about 110 ms post-stimulus, for all conditions, clear valleys and peaks appeared over the frontal and parietal regions (associated with attentional control), respectively. Specially, TRFs had a prominent peak over the occipital region only for condition AVa, indicating the involvement of the visual cortex. At a later stage of about 188 ms post-stimulus, TRFs had distinct peaks over the frontal-temporal regions bilaterally for all conditions, indicating the involvement of the auditory cortex. The result that the left temporal region had greater TRFs than the right one was likely for the reason that the left hemisphere was critical for speech recognition [25]. For other time lags, TRFs patterns were similar among conditions. The root-mean-square (RMS) value of TRFs was further used to reflect the processing strength across all time lags. For all conditions, temporal region had greater amplitude over other regions, while the occipital processing was only evident for condition AVa. The general processing strength was the greatest for condition AVa. And it was slightly greater for condition A than condition AVu or AVi, which was in line with the notion that incongruent visual inputs would degrade speech perception [12].

Overall, TRFs analysis revealed that cortical regions were associated with AV speech processing at different timings, and the processing was subject to stimulus modality. Primarily, occipital region (about 110 ms) was involved in AV speech processing early than temporal region (about 188 ms).

3.2. ITPC pattern

To examine the effect of AV congruency on phase consistency of neural responses to speech, ITPC of each condition was calculated. Figure 2(a) shows the ITPC values of two representative electrodes, at which the differences of ITPC between condition AVa and A were relatively greater. For all conditions, high ITPC values were found for frequencies between 1 and 8 Hz, which is similar to the results in [15]. And within this frequency range, ITPC values were greater for condition AVa than other conditions. Topographies of ITPC values at representative frequencies are shown in Figure 2(b). It demonstrated that ITPC values were greater for condition AVa than other conditions almost for all frequencies and EEG channels, and the increment mainly located over the temporal and occipital region. Unlike TRFs analysis, the differences of ITPC values among condition A, AVu and AVi were little. This might imply that incongruent visual inputs degrade the amplitude instead of the phase of the cortical envelope tracking activities.

These results indicated that the phase of neural activity faithfully tracked the presented speech token in the frequency range corresponding to speech temporal envelope.

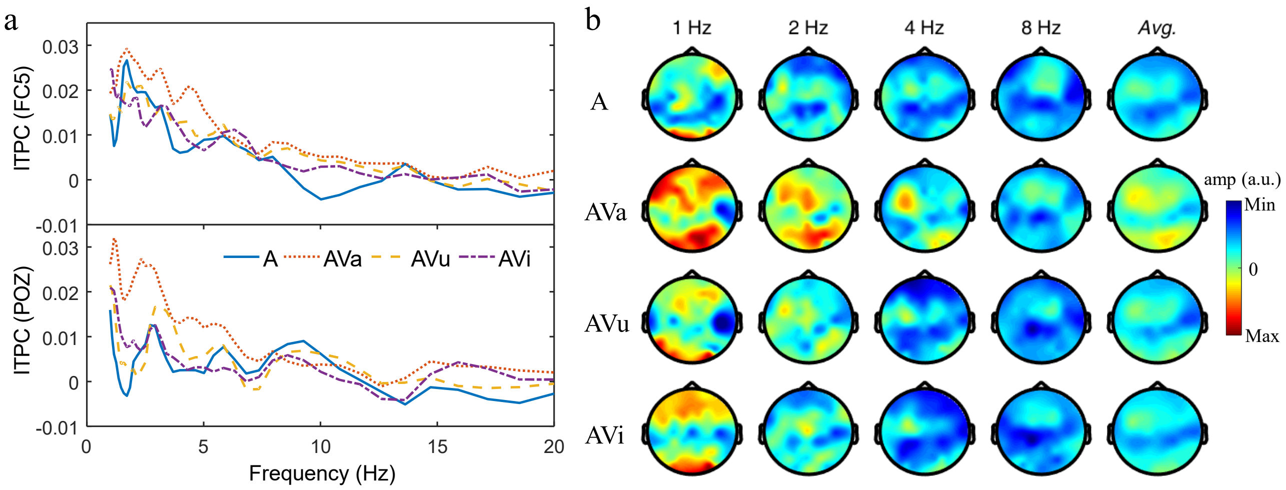

Figure 2: *ITPC of different conditions. (a) ITPC at fronto-temporal electrode FC5 and partieto-occipital electrode POZ. (b) Topographies of ITPC at representative frequencies and averaged ITPC values between 1 and 8 Hz.*

4. Discussion

4.1. Effect of AV congruency on cortical envelope tracking

The primary finding of the present study is that, in a two-talker scenario, additional visual input could improve the cortical envelope tracking (indexed by TRFs and ITPC) only when it is congruent with the attended speech. Since the involvement of the occipital region (about 110 ms) was found early than the temporal region (about 188 ms), the present study conforms to the conclusion that congruent visual inputs play a predictive role in assisting speech perception [9]. Specifically, listener's attentional resources would be directed to an appropriate timing that the to-be-attended speech is expected to arrive. Consequently, high neuronal excitability would align with the arrival of the attended speech, and thus the cortical envelope tracking activities are enhanced. The differential timings between these two regions also accord with the long-term crossmodal integration model of AV speech processing, in which visual input was considered restoring early tracking of the acoustic signal in noise [26].

It has been reported that the top-down attention could increase the phase coherence between distinct neuronal clusters tuned to the same target, which facilitates the formation of attended speech stream and thereby helps distinguishing between the target speech and the background [27]. The present study further indicates that such a synchronization effect can be enhanced by congruent AV inputs, and the effect is evident over both the auditory and visual cortex. This cross-modal synchronization is consistent with the previous finding, in which the visual cortex is involved in the processing of temporal envelope of speech [28], due to the correlation between congruent visual (e.g., lip movements) and auditory (e.g., temporal envelope) inputs [10].

4.2. Similarities and differences to related studies

In the previous study using MEG [15], TRFs amplitudes and ITPC values were both greater for the congruent AV condition than the audio-only condition. However, the enhancement was only observed in the temporal region in [15], but it was found covering more cortical regions in the present work. One explanation for such a discrepancy is that the AV speeches were much shorter and presented repeatedly in [15] (about 10 s, 40 times) than the present work (60 s, 1 time). Specifically, since subjects could get familiar with the speech materials during the stimulus repetition, the selective attention task could rely on subject's ability to remember and predict the upcoming speech stream instead of processing stimulus features adequately. For this reason, visual cues would be less processed and high-level top-down modulation would be less applied, and thus the envelope tracking activities were weak for occipital, frontal and parietal regions in [15]. Another reason for this discrepancy could be that EEG was shown more sensitive to attention-related components amid neural activity than MEG [29]. Another difference between the two studies was that the present study further demonstrated the topographic distribution of time-varying TRFs and frequency-dependent ITPC for each AV condition to manifest the spatial characteristics of cortical envelope tracking activities clearly.

5. Conclusions

The present study examined whether additional visual input (i.e., lipreading) could enhance the cortical processing of speech. From perspectives of time (indexed by TRFs) and frequency (indexed by ITPC), the cortical envelope tracking to speech was found enhanced relative to the audio-only condition only for the congruent AV condition. Topographic analysis showed that the timings of cortical regions involved in AV speech processing were subject to stimulus modality. As a complement to related studies, the present work provides new insight into the neural mechanisms of cortical processing of speech and how attention shapes the internal representation of speech when visual input is also available. More sophisticated conclusion could be derived from more advanced neuroimaging techniques in future works.

6. Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61771023 and 11590773) and a research funding from SONOVA.

7. References

[1] E. C. Cherry, “Some experiments on the recognition of speech, with one and with two ears,” *Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, 1953.

[2] N. Mesgarani and E. F. Chang, “Selective cortical representation of attended speaker in multi-talker speech perception,” *Nature*, vol. 485, no. 7397, pp. 233–236, 2012.

[3] J. A. O’Sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor, “Attentional selection in a cocktail party environment can be decoded from single-trial EEG,” *Cerebral Cortex*, vol. 25, no. 7, pp. 1697–1706, 2015.

[4] N. Ding and J. Z. Simon, “Emergence of neural encoding of auditory objects while listening to competing speakers,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 29, pp. 11 854–11 859, 2012.

[5] E. C. Lalor and J. J. Foxe, “Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution,” *European Journal of Neuroscience*, vol. 31, no. 1, pp. 189–193, 2010.

[6] N. Mesgarani, S. David, J. Fritz, and S. A Shamma, “Influence of Context and Behavior on Stimulus Reconstruction From Neural Activity in Primary Auditory Cortex,” *Journal of neurophysiology*, vol. 102, pp. 3329–3339, 2009.

[7] M. J. Crosse, G. M. D. Liberto, A. Bednar, and E. C. Lalor, “The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli,” *Frontiers in Human Neuroscience*, vol. 10, p. 604, 2016.

[8] K. W. Grant and P.-F. Seitz, “The use of visible speech cues for improving auditory detection of spoken sentences,” *The Journal of the Acoustical Society of America*, vol. 108, no. 3, pp. 1197–1208, 2000.

[9] J. E. Peelle and M. S. Sommers, “Prediction and constraint in audiovisual speech perception,” *Cortex*, vol. 68, pp. 169–181, 2015.

[10] C. Chandrasekaran, A. Trubanova, S. Stillittano, A. Caplier, and A. A. Ghazanfar, “The natural statistics of audiovisual speech,” *PLoS computational biology*, vol. 5, no. 7, p. e1000436, 2009.

[11] N. Tye-Murray, M. S. Sommers, and B. Spehar, “Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing,” *Ear and Hearing*, vol. 28, no. 5, pp. 656–668, 2007.

[12] Z. Fu, X. Wu, and J. Chen, “Congruent audiovisual speech enhances auditory attention decoding with EEG,” *Journal of Neural Engineering*, vol. 16, no. 6, p. 064002, 2019.

[13] M. J. Crosse, J. S. Butler, and E. C. Lalor, “Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions,” *Journal of Neuroscience*, vol. 35, no. 42, pp. 14 195–14 204, 2015.

[14] S. Haufe, F. Meinecke, K. Görgen, S. Dahne, J.-D. Haynes, B. Blankertz, and F. Bießmann, “On the interpretation of weight vectors of linear models in multivariate neuroimaging,” *NeuroImage*, vol. 87, pp. 96–110, 2014.

[15] E. Zion-Golumbic, G. B. Cogan, C. E. Schroeder, and D. Poeppel, “Visual Input Enhances Selective Speech Envelope Tracking in Auditory Cortex at a Cocktail Party,” *Journal of Neuroscience*, vol. 33, no. 4, pp. 1417–1426, 2013.

[16] H. Luo and D. Poeppel, “Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex,” *Neuron*, vol. 54, no. 6, pp. 1001–1010, 2007.

[17] R. M. van Diepen and A. Mazaheri, “The Caveats of observing Inter-Trial Phase-Coherence in Cognitive Neuroscience,” *Scientific Reports*, vol. 8, no. 1, pp. 1–9, 2018.

[18] E. Zion-Golumbic, N. Ding, S. Bickel, P. Lakatos, C. Schevon, G. M McKhann, R. R Goodman, R. Emerson, A. Mehta, J. Simon, D. Poeppel, and C. E Schroeder, “Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a ”Cocktail Party”,” *Neuron*, vol. 77, no. 5, pp. 980–991, 2013.

[19] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, “Independent component analysis of electroencephalographic data,” *Advances in Neural Information Processing Systems*, vol. 8, no. 3, pp. 145–151, 1996.

[20] A. Delorme and S. Makeig, “EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis,” *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004, wOS:000189330500002.

[21] S. Makeig, T. Bell, T. Lee, T. Jung, and S. Enghoff, “EEGLAB: ICA toolbox for psychophysiological research,” 2000. [Online]. Available: www.sccn.ucsd.edu/eeglab

[22] B. R. Glasberg and B. C. Moore, “Derivation of auditory filter shapes from notched-noise data,” *Hearing Research*, vol. 47, no. 1-2, pp. 103–138, 1990.

[23] C. J. Plack, A. J. Oxenham, A. M. Simonson, C. G. O’Hanlon, V. Drga, and D. Arifianto, “Estimates of compression at low and high frequencies using masking additivity in normal and impaired ears,” *The Journal of the Acoustical Society of America*, vol. 123, no. 6, pp. 4321–4330, 2008.

[24] S. A. Fuglsang, T. Dau, and J. Hjortkjær, “Noise-robust cortical tracking of attended speech in real-world acoustic scenes,” *Neuroimage*, vol. 156, pp. 435–444, 2017.

[25] D. Kimura, “Functional Asymmetry of the Brain in Dichotic Listening,” *Cortex*, vol. 3, no. 2, pp. 163–178, 1967.

[26] M. J. Crosse, L. G. Di, and E. C. Lalor, “Eye can hear clearly now: inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration,” *Journal of Neuroscience*, vol. 36, no. 38, pp. 9888–9895, 2016.

[27] S. A. Shamma, M. Elhilali, and C. Micheyl, “Temporal coherence and attention in auditory scene analysis,” *Trends in Neurosciences*, vol. 34, no. 3, pp. 114–123, 2011.

[28] M. P. Broderick, A. J. Anderson, G. M. Di Liberto, M. J. Crosse, and E. C. Lalor, “Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech,” *Current Biology*, vol. 28, no. 5, pp. 803–809.e3, 2018.

[29] S. Kähkönen, J. Ahveninen, I. P. Jääskeläinen, S. Kaakkola, R. Näätänen, J. Huttunen, and E. Pekkonen, “Effects of haloperidol on selective attention: a combined whole-head MEG and high-resolution EEG study,” *Neuropsychopharmacology*, vol. 25, no. 4, pp. 498–504, 2001.