INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Effects of spectral degradation on the cortical tracking of the speech envelope

Alexis Deighton MacIntyre, Tobias Goehring

Cambridge Hearing Group, MRC Cognition and Brain Sciences Unit, University of Cambridge, UK
AlexisDeighton.MacIntyre@mrc-cbu.cam.ac.uk, Tobias.Goehring@mrc-cbu.cam.ac.uk

Abstract

During speech listening, recurring patterns of neural activity become temporally coupled to stimulus features, such as the speech envelope. This cortical tracking can be measured using electroencephalography (EEG). Quantifying speech-brain coupling (e.g., as a correlation coefficient) sheds light on the neurobiological processes underlying perception and holds promise as an objective measure, particularly for clinical populations such as cochlear implant (CI) users. How spectral degradation associated with CI stimulation affects cortical tracking, however, remains unclear. In this EEG study, we simulate CI listening using vocoded speech with and without current spread, a realistic complication of CI stimulation. We find no effect of either vocoding or current spread on cortical tracking, despite differences in subjective reports of speech comprehension and implicit behavioural measures. We conclude that, when speech is intelligible, cortical tracking is robust to spectral degradation.
Index Terms: Speech perception, electroencephalography, speech entrainment, cochlear implants

1. Introduction

During continuous speech listening, patterns of neural activity within a listener's brain become temporally coupled to features of the acoustic stimulus. This form of coupling, sometimes termed cortical tracking, can be detected using neuroimaging techniques with good temporal resolution (e.g., electroencephalography; EEG). Multivariate linear regression is one technique used to characterise this tracking. Within this framework, the neural response can be estimated as a linear convolution of a given stimulus–for instance, slow amplitude fluctuations in the speech signal conveyed by the temporal envelope [1]. When the stimulus is used to predict the EEG signal, this is known as an *encoding* model. Conversely, a model can be trained to reconstruct the stimulus from neural data, which is the *decoding* model. Whereas encoding models shed some light on the topography and latency of stimulus-evoked neural processing, decoding models are less amenable to interpretation; however, they do offer a simple and intuitive way to quantify how well cortical tracking has been captured using EEG [2]. Namely, a linear correlation coefficient can be derived between the reconstructed stimulus and ground truth, thereby providing a measure of model accuracy. This value can be taken forward as a dependent variable to compare within or across participants and experimental conditions.

Capitalising on these and similar methods, research into continuous listening has broadened our understanding of many fundamental topics in auditory cognitive neuroscience, such as the nature of selective auditory attention, language acquisition, and audiovisual speech perception. Cortical tracking also holds promise for clinical applications as a potential objective measure of audiological function. For instance, higher correlations between the speech envelope and EEG recorded from typically hearing listeners predict better comprehension of speech in stationary noise [3]. In older adults with varying degrees of hearing loss, a higher magnitude of cortical tracking was also found to be correlated with speech understanding, particularly for individuals with more severe hearing loss [4, 5]. The underlying neural processes driving these correlations are not fully understood: Enhanced cortical tracking to the speech envelope may indicate increased listening effort; more attention directed towards this component of the signal; or the recruitment of additional or distributed neural resources to compensate for degraded auditory input and signal distortions. Nonetheless, converging evidence suggests that cortical speech tracking could serve as an objective measure of speech information transmission, which could be used to personalise medical interventions and improve speech perception outcomes.

Cochlear implant (CI) users are one population that could benefit from such an objective measure. A CI is a hearing device that includes an array of electrodes implanted within the cochlea of someone with severe to profound hearing loss. These electrodes directly stimulate the auditory nerve, thereby replacing the function of damaged sensory hair cells, which would otherwise convert physical force from sound waves into electrical impulses. Although CI users often have good speech reception in quiet [6], background noise and compounding factors, such as reverberation, can pose a serious challenge to communication [7]. These difficulties partially arise from spectral degradation due to the limited number of spectral channels (12-24 electrodes) available, electrical current spreading within the cochlea fluid, and compromised function of neural receptors in the auditory nerve. Such factors effectively reduce the number of independent channels of information within the perceived speech signal and are detrimental to speech processing due to spectral and temporal degradation effects [8, 9, 10]. It remains an active area of research to improve spectral resolution with CIs [11], and successful interventions for some individuals may exacerbate the problem for others. Measurements of a CI user's spectral resolution typically included behavioural sound discrimination [12, 13] and/or speech perception experiments [14, 15]. However, such measurements require a minimum level of performance, may vary with listening experience, often require acclimatization or training, and can be too time-consuming for clinical use. The ability to objectively quantify the degree to which an acoustic stimulus is neurally represented by CI users could, therefore, provide valuable insights for clinicians and

This research was funded by Wellcome Trust ISSF (ADM) and CDA Fellowship MR/T03095X/1 from the Medical Research Council UK (TG).

 10.21437/Interspeech.2023-1776

help ensure that the device is optimally configured to each patient. Indeed, cortical tracking has been successfully measured in CI listeners and found to correlate with behavioural reports of speech intelligibility and to be capable of decoding selective auditory attention [16, 17, 18]. These are promising results, however, electrical stimulation generated by CIs interferes with EEG recordings and may confound the correlations obtained between neural responses and speech stimuli. Previous studies implemented measures to avoid this contamination, for example by inserting periodic gaps for EEG recording in the CI stimulation pattern, but strategies to deal with CI stimulation artefacts remain an active area of investigation [19]. Instead, tone or noise vocoders can be used to simulate CI speech perception in typically-hearing individuals. This has also been a popular method to investigate the cortical tracking of speech with poor to zero intelligibility [20, 21]. Although the perceptual quality of vocoded speech may not reflect the percept of CI listeners, key advantages for this approach are the avoidance of electrical stimulation artefacts with full control over spectral resolution and degradation, as well as the availability of an ideal control condition in form of the acoustic, non-vocoded stimuli. In this study, a vocoder was therefore used to simulate the effects of CI processing and their effects on the cortical tracking of speech in typically hearing listeners.

We aimed to control for and investigate the effects of spectral degradation relevant to CI speech perception on cortical tracking and the linear decoding of speech envelope information. For translation to clinical populations, we employed a simulation of CI speech perception as well as an easy behavioural task designed to sustain auditory attention across listening conditions. The current study is presented in three sections: Section 2 describes the methodologies for the experimental design, stimuli processing, and behavioural task, as well as data acquisition and model training. Section 3 presents the results and statistical analyses performed. Section 4 discusses the findings and puts them into context.

2. Methodology

2.1. Participants

Twenty adults (14 female, 7 male) between the ages of 18-35 (Mean age 23.75 years, SD 5.00) were recruited from the university community. The participants spoke English (including varieties of British, North American, South African, and Nigerian English) as a primary language. All participants reported typical hearing and no history of speech or language disabilities. Participants gave informed consent before taking part and were compensated for their time. The study was approved by the local ethical review committee.

2.2. Experimental design

2.2.1. Stimuli and procedure

The acoustic stimuli consisted of a recording of a Sherlock Holmes story by Arthur Conan Doyle (duration 35 minutes and 26 seconds) read aloud by a male speaker with a North American English accent. The audio was filtered to match the international long-term spectrum of speech [22] and intensity was normalised by RMS using a sliding window and perceptually corrected where necessary. The story was then divided into 6 blocks (Mean duration 5 min. 54 s, SD 1.63 s). Each listening condition (i.e., acoustic, vocoded, or vocoded with simulated current spread) was assigned to 2 blocks, the order of which in relation to the story was counter-balanced across participants. To simulate CI processing, we employed the SPIRAL vocoder [9], which emulates the limited number of electrode channels available to CI listeners as well as the deleterious effects of current spread and neural degeneration that can result in additional spectral degradation of the perceived signals. Behavioural speech reception data suggest that this method more accurately reflects the perceptual challenges typical of CI listening and associated listening performance than simulations using traditional noise or tone vocoders. Here, we used 16 analysis filter bands to produce both vocoded speech conditions (Fig. 1). In the current spread simulating condition, we introduced a decay slope of -16 dB/octave as proposed in [9]. The experiment took place in a sound attenuating, electrically shielded room. Speech stimuli were sampled at 44100 Hz and presented binaurally through Etymotic Research ER-2 insert earphones (Etymotic Research, Elk Grove Village, IL, USA) at a comfortable volume. Participants were instructed to avoid excessive blinking or extraneous movement, and to visually focus on a fixation marker centred in the presentation computer screen. After each block ended, participants rated the preceding section on a scale from $1-7$ for *i)* how well they found they could follow what was happening and *ii)* how engaged they felt by the story content, before beginning a self-paced rest period.

2.2.2. Behavioural task

To sustain auditory attention across listening conditions, participants were given a behavioural target detection task. The targets were repeated phrases, such that the audio playback would repeat a particular phrase one additional time after first presentation, before continuing with the story. Repetitions occurred pseudo-randomly once every $\sim$45 s (Mean repeated phrase duration 2.06 s, SD 0.04 s). There were a total of 8 repeated phrases per block. Before beginning the experiment, participants were given a practice session with feedback for accuracy and reaction time to ensure the task was understood. Target detection was recorded via button press using a custom-made USB button box that the participant held in one hand on their lap to avoid unnecessary movement.

2.3. Data acquisition and preprocessing

EEG was recorded using a BioSemi Active Two system (Biosemi, Amsterdam, The Netherlands) with 64-channels at a sampling rate of 2048 Hz. The acoustic stimuli and EEG triggers were controlled through an RME Fireface UCX (RME, Haimhausen, Germany) external soundcard. Preprocessing was performed in MATLAB using functions from the Fieldtrip [23] and noisetools [24] toolboxes. The data were first offline re-referenced using the channel average and resampled to 256 Hz. Line noise was removed using zapline-plus [25], and the EEG was high-pass filtered at 0.5 Hz and subsequently low-pass filtered at 40 Hz using 4th-order Butterworth filters. Bad channels were visually identified and interpolated after removing eye-blink artifacts from the remaining channels using independent component analysis (FastICA algorithm). Finally, the first and final 5 seconds of each block were removed to avoid filtering-related artifacts. The acoustic stimulus envelope was extracted as described in [26, 27]. Briefly, we filtered the speech signal within critical bands based on the Bark scale, square-rectified the signal within each filter band, and averaged across bands. This technique is well-suited to capture the timing of vowel onsets, which are argued to be particularly important for speech perception [26]. Finally, both the EEG and acoustic envelope

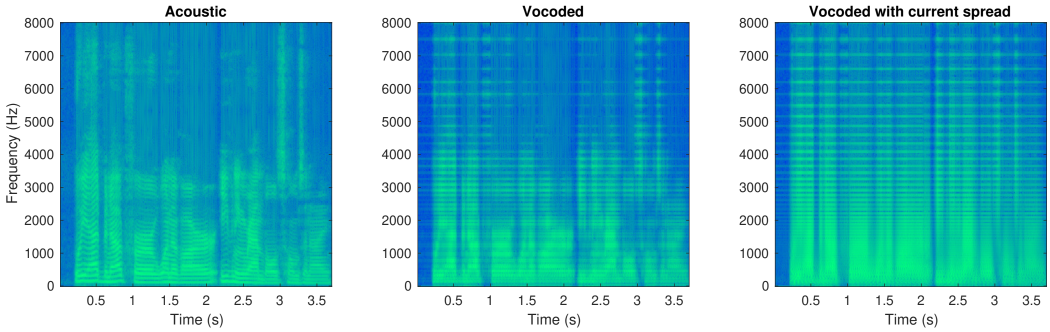

Figure 1: *Spectrograms for the three listening conditions*

signals were resampled to 100 Hz for computing the decoding models.

2.4. Analysis

We trained participant-specific linear decoding models for each listening condition, using regularised ridge regression as implemented in the multivariate Temporal Response Function Toolbox (mTRF) [2]. In this case, the models are applied "backwards", in that they reconstruct the speech envelope from the EEG responses. Formally, the decoding model $g(\tau, n)$ returns the stimulus $\hat{s}(t)$ as a linear convolution of the neural response $r(t, n)$ over a range of time lags τ, or

$$\hat{s}(t) = \sum_n \sum_\tau r(t + \tau, n)\, g(\tau, n)\,.$$

The accuracy of this reconstruction, estimated as a Pearson's correlation *r* to the true stimulus, provides a measure of speech encoding by the cortex. We trained the decoder by minimising the mean squared error using 4-fold, leave-one-out cross-validation, as described in detail in [2, 28]. The model incorporated time lags $0 - 450$ ms post-stimulus. The optimal ridge parameter λ was obtained from logarithmically spaced values ranging from $10^{-6} - 10^{6}$. For each listening condition, 60 s of data were set aside to be reconstructed, with the remainder ($\sim$10min per condition) forming the training set. To test model significance, we generated a null distribution for each listening condition by segmenting the continuous stimulus into short segments (Mean duration 2.03 s, SD 1.30 s), shuffling the segments, and re-concatenating them to form a new envelope. Segmentation was applied at naturally occurring transient pauses and segments were smoothly joined using spline interpolation, thereby avoiding the introduction of artefacts to the permuted signal. We then trained and tested new models using the shuffled stimulus and original neural response, for each listening condition and participant, 1000 times to obtain a critical value, $p = 0.05$. To measure our effects of interest, we performed linear mixed effect modelling in `R` 3.5.0 [29] with functions from `lme4` [30] and `emmeans` [31]. "Participant" was included as a random intercept for each model. Model residuals were inspected visually and by using diagnostic tests of normality. The final model was chosen based on the Akaike information criterion (AIC). Bonferroni correction was employed for multiple comparisons and we report adjusted *P*-values where relevant with Kenward-Roger degrees of freedom.

3. Results

3.1. Subjective ratings and auditory detection task

We first examined the effect of listening condition on the two self-reported measures, which were how well the participant felt they could follow the previous listening block, and how engaging they had found that part of the story to be, each on a scale of $1 - 7$. We averaged responses over the two blocks per experimental condition. In comparison to acoustic speech (Mean 6.53, SD 0.87), listeners reported that vocoded speech with current spread (Mean 5.00, SD 0.93) was significantly more difficult to follow (Estimate = 1.52, SE = 0.17, *t*(1,42.10) = 9.02, p_{bonf} < 0.001). Vocoding with current spread was also more difficult to follow than vocoded speech (Mean 6.33, SD 0.69; Estimate = 1.32, SE = 0.17, *t*(1,42.10) = 7.84, $p_{bonf} < 0.001$). Acoustic and vocoded speech did not differ significantly from one another ($p_{bonf} = 0.73$). For the engagement participants felt when listening to the story, there was again a negative effect of vocoding with current spread (Mean 5.15, SD 1.19), which was significantly less engaging than acoustic speech (Mean 6.28, SD 0.70; Estimate = 1.13, SE = 0.23, *t*(1,42.10) = 4.92, p_{bonf} < 0.001) and vocoded speech (Mean 5.92, SD 0.96; Estimate = 0.78, SE = 0.23, *t*(1,42.10) = 3.39, $p_{bonf} = 0.005$). Again, we found no statistical difference in engagement between acoustic and vocoded speech ($p_{bonf} = 0.40$).

Turning to the repeated phrase detection task, performance was generally high across listening conditions. The mean Hit Rate ranged from 0.99 for acoustic speech, to 0.97 for vocoding with current spread. Similarly, the mean False Alarm rate was low, from 0.02 for acoustic speech, to 0.04 for vocoding with current spread. Hence, despite degraded spectral quality, participants achieved close to ceiling performance throughout the task. When modelling their reaction times to the auditory targets, we found that the best fitting model included main effects of, and an interaction term between, listening condition and self-reported levels of engagement. Holding engagement levels constant, acoustic speech was associated with faster median reaction times (Mean 1.02 s, SD 0.21 s) in comparison to vocoding with current spread (Mean 1.19 s, SD 0.31 s; Estimate = -0.13, SE = 0.05, *t*(1,49.50) = -2.76, $p_{bonf} = 0.03$). Vocoded speech (Mean 1.08 s, SD 0.24 s) did not differ from either of the other two listening conditions ($p_{bonf} > 0.14$; Fig. 2). The linear trend of engagement predicted slower reaction times for vocoded speech with current spread (Estimate = -0.09, SE = 0.03, $p_{bonf} = 0.002$), but did not meet significance for either acoustic or vocoded speech ($p_{bonf} > 0.66$).

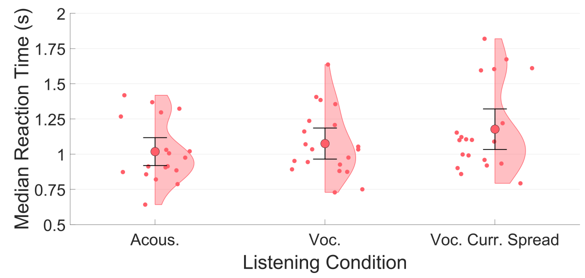

Figure 2: *Median reaction times in the repeated phrase auditory detection task, by listening condition. Bold markers indicate the mean and 95% confidence intervals of the mean.*

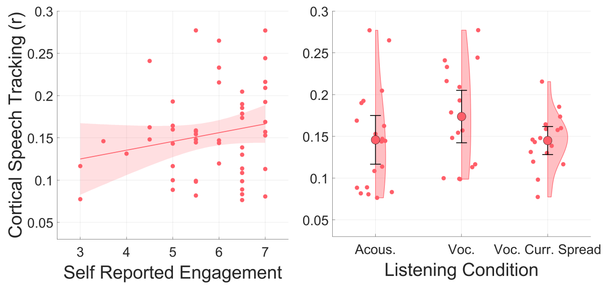

Figure 3: *Cortical tracking of the speech signal, left: as a function of self-reported engagement by the story, with shaded region representing 95% confidence intervals of the linear fit; and right: by listening condition. Bold markers indicate the mean and 95% confidence intervals of the mean.*

3.2. Cortical speech tracking

We first compared observed r values to those generated from models trained on a null distribution (i.e., with permuted stimulus envelopes). A model was deemed significant if its stimulus reconstruction $r \geq$ the 95th percentile of the null distribution, or $p = 0.05$. Based on this criteria, we retained 52 of 60 r values. We compared linear mixed models that included the following terms: listening condition; the two self-reported measures of ability to follow and engagement; and median reaction time. According to the AIC, the model that best fit the data contained a single predictor, which was the subjective rating of engagement. This indicates that cortical tracking did not statistically differ across our speech stimuli–acoustic, vocoded, or vocoded with current spread (Fig. 3). Self-reported engagement with the story, however, failed to meet significance in the final model (Estimate 0.01, 95% CI $[-0.00, 0.03]$, $t(1{,}46.59) = 1.49$, $p = 0.14$, $R^2_{sp} = 0.04$). Given that reported engagement varied by listening condition, and most (7/8) of the non-significant decoders were trained on spectrally degraded speech, we ran an exploratory model that included the non-significant r values. In this case, engagement does show a tenuous statistical association with cortical tracking, but this predictor would nonetheless fail to meet corrected significance (Estimate 0.02, 95% CI $[0.00, 0.03]$, $t(1{,}57.77) = 2.09$, $p = 0.04$, $R^2_{sp} = 0.07$).

4. Discussion and Conclusion

In this EEG experiment with typically hearing participants, we simulated cochlear implant listening using vocoded speech with and without current spread, a common complication affecting CI listeners that is thought to disrupt speech processing. Despite participants' subjective reports of difficulty comprehending and decreasing engagement, we found no evidence that spectral degradation negatively affects cortical tracking to the speech envelope. There are, however, limitations to the current study that should be considered for future work.

Firstly, the most challenging listening condition, vocoding with current spread, was still associated with relatively good intelligibility (Mean 5.00 on a scale of $1-7$). The beneficial effects of prior knowledge for speech intelligibility are well known and have been specifically studied using vocoded speech; hence, it is possible that the current results are driven by so-called top-down influences that repair the degraded speech signal [32, 33]. Next steps should seek to disentangle the cognitive, linguistic, and phonological factors that may shape listeners' expectations, thereby helping guide acoustic processing and enabling degraded speech to "pop out". The current results should also be extended in future work with more severely degraded speech, as well as other, CI-relevant factors; for example, neural deterioration of the auditory system is not directly captured with vocoders, but accounts for a large heterogeneity among CI listeners [34]. Another drawback is our using of only decoding or backwards models, which are non-informative regarding the time series and topography of neural activations. It is likely that more subtle differences on the basis of listening condition may emerge by comparing the spatio-temporal profile of the brain response, rather than the absolute magnitude of cortical tracking. Although cortical tracking is more amenable to a clinical context as an objective measure simple to interpret, studying encoding or forward models may point to important distinctions between acoustic and spectrally degraded speech processing that our methods were not sensitive to. Finally, there was considerable inter-subject variability in the brain response to the three listening conditions, and our sample size may have been a limitation in detecting the effect of self-reported engagement, which was associated with a non-significant trend in the current study. Although the null finding for listening condition should be replicated in a larger sample, it does not support a clear deleterious effect of spectral degradation when speech intelligibility is good and auditory attention is sustained, suggesting that the EEG detection of cortical tracking is feasible even when some degree of spectral degradation, including CI-specific current spread, is present. Together, the results contribute to our understanding of cortical tracking as a potentially clinically valuable tool and a candidate objective measure for speech transmission.

5. References

[1] N. Ding and J. Z. Simon, "Neural coding of continuous speech in auditory cortex during monaural and dichotic listening," *Journal of neurophysiology*, vol. 107, no. 1, pp. 78–89, 2012.

[2] M. J. Crosse, G. M. Di 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.

[3] J. Vanthornhout, L. Decruy, J. Wouters, J. Z. Simon, and T. Francart, "Speech intelligibility predicted from neural entrainment of the speech envelope," *Journal of the Association for Research in Otolaryngology*, vol. 19, pp. 181–191, 2018.

[4] E. B. Petersen, M. Wöstmann, J. Obleser, and T. Lunner, "Neural tracking of attended versus ignored speech is differentially affected by hearing loss," *Journal of neurophysiology*, vol. 117, no. 1, pp. 18–27, 2017.

[5] R. Schmitt, M. Meyer, and N. Giroud, "Better speech-in-noise

comprehension is associated with enhanced neural speech tracking in older adults with hearing impairment," *cortex*, vol. 151, pp. 133–146, 2022.

[6] I. Boisvert, M. Reis, A. Au, R. Cowan, and R. C. Dowell, "Cochlear implantation outcomes in adults: A scoping review," *PLoS One*, vol. 15, no. 5, p. e0232421, 2020.

[7] J. Badajoz-Davila, J. M. Buchholz, and R. Van-Hoesel, "Effect of noise and reverberation on speech intelligibility for cochlear implant recipients in realistic sound environments," *The Journal of the Acoustical Society of America*, vol. 147, no. 5, pp. 3538–3549, 2020.

[8] A. J. Oxenham and H. A. Kreft, "Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing," *Trends in Hearing*, vol. 18, p. 2331216514553783, 2014.

[9] J. A. Grange, J. F. Culling, N. S. Harris, and S. Bergfeld, "Cochlear implant simulator with independent representation of the full spiral ganglion," *The Journal of the Acoustical Society of America*, vol. 142, no. 5, pp. EL484–EL489, 2017.

[10] T. Goehring, A. W. Archer-Boyd, J. G. Arenberg, and R. P. Carlyon, "The effect of increased channel interaction on speech perception with cochlear implants," *Scientific Reports*, vol. 11, no. 1, pp. 1–9, 2021.

[11] R. P. Carlyon and T. Goehring, "Cochlear implant research and development in the twenty-first century: a critical update," *Journal of the Association for Research in Otolaryngology*, vol. 22, no. 5, pp. 481–508, 2021.

[12] A. W. Archer-Boyd, R. V. Southwell, J. M. Deeks, R. E. Turner, and R. P. Carlyon, "Development and validation of a spectrotemporal processing test for cochlear-implant listeners," *The Journal of the Acoustical Society of America*, vol. 144, no. 5, pp. 2983–2997, 2018.

[13] A. W. Archer-Boyd, A. Harland, T. Goehring, and R. P. Carlyon, "An online implementation of a measure of spectro-temporal processing by cochlear-implant listeners," *JASA Express Letters*, vol. 3, no. 1, p. 014402, 2023.

[14] L. M. Friesen, R. V. Shannon, D. Baskent, and X. Wang, "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *The Journal of the Acoustical Society of America*, vol. 110, no. 2, pp. 1150–1163, 2001.

[15] T. Goehring, J. G. Arenberg, and R. P. Carlyon, "Using spectral blurring to assess effects of channel interaction on speech-in-noise perception with cochlear implants," *Journal of the Association for Research in Otolaryngology*, vol. 21, pp. 353–371, 2020.

[16] B. Somers, E. Verschueren, and T. Francart, "Neural tracking of the speech envelope in cochlear implant users," *Journal of neural engineering*, vol. 16, no. 1, p. 016003, 2018.

[17] E. Verschueren, B. Somers, and T. Francart, "Neural envelope tracking as a measure of speech understanding in cochlear implant users," *Hearing research*, vol. 373, pp. 23–31, 2019.

[18] W. Nogueira, H. Dolhopiatenko, I. Schierholz, A. Büchner, B. Mirkovic, M. G. Bleichner, and S. Debener, "Decoding selective attention in normal hearing listeners and bilateral cochlear implant users with concealed ear eeg," *Frontiers in neuroscience*, vol. 13, p. 720, 2019.

[19] R. P. Carlyon, F. Guérit, J. M. Deeks, A. Harland, R. Gransier, J. Wouters, S. R. de Rijk, and M. Bance, "Using interleaved stimulation to measure the size and selectivity of the sustained phase-locked neural response to cochlear implant stimulation," *Journal of the Association for Research in Otolaryngology*, vol. 22, pp. 141–159, 2021.

[20] J. E. Peelle, J. Gross, and M. H. Davis, "Phase-locked responses to speech in human auditory cortex are enhanced during comprehension," *Cerebral cortex*, vol. 23, no. 6, pp. 1378–1387, 2013.

[21] E. Sohoglu and M. H. Davis, "Rapid computations of spectrotemporal prediction error support perception of degraded speech," *elife*, vol. 9, p. e58077, 2020.

[22] D. Byrne, H. Dillon, K. Tran, S. Arlinger, K. Wilbraham, R. Cox, B. Hagerman, R. Hetu, J. Kei, C. Lui *et al.*, "An international comparison of long-term average speech spectra," *The journal of the acoustical society of America*, vol. 96, no. 4, pp. 2108–2120, 1994.

[23] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen, "Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data," *Computational intelligence and neuroscience*, vol. 2011, pp. 1–9, 2011.

[24] A. de Cheveigné and D. Arzounian, "Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data," *NeuroImage*, vol. 172, pp. 903–912, 2018.

[25] M. Klug and N. A. Kloosterman, "Zapline-plus: A zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in m/eeg," *Human Brain Mapping*, vol. 43, no. 9, pp. 2743–2758, 2022.

[26] Y. Oganian and E. F. Chang, "A speech envelope landmark for syllable encoding in human superior temporal gyrus," *Science advances*, vol. 5, no. 11, p. eaay6279, 2019.

[27] A. D. MacIntyre, C. Q. Cai, and S. K. Scott, "Pushing the envelope: Evaluating speech rhythm with different envelope extraction techniques," *The Journal of the Acoustical Society of America*, vol. 151, no. 3, pp. 2002–2026, 2022.

[28] M. J. Crosse, N. J. Zuk, G. M. Di Liberto, A. R. Nidiffer, S. Molholm, and E. C. Lalor, "Linear modeling of neurophysiological responses to speech and other continuous stimuli: methodological considerations for applied research," *Frontiers in Neuroscience*, p. 1350, 2021.

[29] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/

[30] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *arXiv preprint arXiv:1406.5823*, 2014.

[31] R. V. Lenth, "Least-squares means: the r package lsmeans," *Journal of statistical software*, vol. 69, pp. 1–33, 2016.

[32] M. H. Davis, I. S. Johnsrude, A. Hervais-Adelman, K. Taylor, and C. McGettigan, "Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences." *Journal of Experimental Psychology: General*, vol. 134, no. 2, p. 222, 2005.

[33] R. E. Millman, S. R. Johnson, and G. Prendergast, "The role of phase-locking to the temporal envelope of speech in auditory perception and speech intelligibility," *Journal of cognitive neuroscience*, vol. 27, no. 3, pp. 533–545, 2015.

[34] C. Garcia, T. Goehring, S. Cosentino, R. E. Turner, J. M. Deeks, T. Brochier, T. Rughooputh, M. Bance, and R. P. Carlyon, "The panoramic ecap method: estimating patient-specific patterns of current spread and neural health in cochlear implant users," *Journal of the Association for Research in Otolaryngology*, vol. 22, pp. 567–589, 2021.