A comparison of two measures of subcortical responses to ongoing speech: Preliminary results

FLORINE L. BACHMANN[1]*, EWEN N. MACDONALD[1], AND JENS HJORTKJÆR[1,2]

[1] *Hearing Systems Section, Department of Health Technology, Technical University of Denmark (DTU), 2800 Kgs. Lyngby, Denmark*

[2] *Danish Research Centre for Magnetic Resonance, Centre for Functional Diagnostic Imaging and Research, Copenhagen University Hospital, 2650 Hvidovre, Denmark*

Neural responses in the auditory brainstem and midbrain are traditionally obtained with repetitions of basic stimuli such as clicks and tones. However, two different methods to measure subcortical responses to ongoing speech with non-invasive electroencephalography (EEG) have recently been published: one based on regularised linear regression (Maddox and Lee, 2018), and the other based on cross-correlation (Etard *et al.*, 2009; Forte *et al.*, 2017). Here, we compare these two methods using the same EEG data set. For both measures, we found prominent peaks in the response functions at latencies consistent with wave V of the auditory brainstem response (ABR; mean latency: 8.19 and 5.97 ms, respectively). The peak response latencies in individual participants were correlated between the regression approach and conventional click-evoked auditory brainstem responses (click-ABRs), suggesting a common underlying neural source. However, similar correlations were not found between the two speech-based methods, nor between the correlation approach and click-ABRs. This could arise from either differences in the methodologies or from variability in the measures.

BACKGROUND

Comparing neural processing at different stages of the auditory pathway provides a deeper understanding of the auditory system. Generally, this interplay has been investigated using different stimuli. Brainstem responses are traditionally investigated with basic stimuli such as tones or clicks, but cortical activity has also been assessed with ongoing speech. Using complex stimuli such as ongoing speech to measure responses at subcortical processing stages could shed light on different facets of early speech processing. Furthermore, it would offer the possibility of simultaneously observing neural responses at the subcortical and cortical level to different speech features using the same ongoing speech stimulus. Recent research suggests that this could be possible. Two independent research groups published two different

*Corresponding author: flbach@dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 7: Auditory Learning in Biological and Artificial Systems, August 2019, Nyborg, Denmark. Edited by A. Kressner, J. Regev, J. C.-Dalsgaard, L. Tranebjærg, S. Santurette, and T. Dau. The Danavox Jubilee Foundation, 2019. ISSN: 2596-5522.

approaches for measuring subcortical responses to ongoing speech. Maddox and Lee (2018) used a regularised linear regression approach, and Forte *et al.* (2017) used cross-correlation to assess the association between features of the continuous speech stimulus and the recorded EEG. Both groups reported a peak in their response functions, with a latency similar to that of wave V of the auditory brainstem response (ABR; 6.17 $\pm$ 0.31 ms and 9.3 $\pm$ 0.7 ms, respectively). Maddox and Lee (2018) further compared their response derived from ongoing speech with a classical click-evoked auditory brainstem response (click-ABR) and found high correlations for both peak latencies and amplitudes. Although differences in the two methods exist, the similar morphology of the estimated response functions could indicate that they measure equivalent aspects of the brainstem response to speech. The present study compares these two methods to one another based on the same data set, and to a classical ABR measurement.

METHODS

Data acquisition

Participants listened to an audio book while their neural activity was recorded with an electroencephalogram (EEG) system. Fourteen (7 female) young (M_{age} = 23.12 $\pm$ 2.411) native Danish speakers participated in the study. All participants had pure-tone thresholds better than 20 dB hearing level in both ears (measured at standard audiometric frequencies: 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 6 kHz, and 8 kHz). Each participant provided written informed consent, and all experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

Measurements were conducted in a soundproof, electrically shielded listening booth. Participants were seated in a comfortable chair in front of a computer screen. Experiment presentation and data acquisition were controlled from outside the booth. The audio book was presented at 65 dB SPL through ER-2 insert earphones (Etymotic Research), with a sampling frequency of 44.1 kHz. The audio book consisted of the beginning of the Danish version of Lord of the Flies by William Golding, read by a male narrator. Longer pauses in the audio book were restricted to 450 ms, and the audio book was cut into trial segments of 50 s duration. To ensure that participants attended the story, three multiple-choice questions were asked after every trial. For each segment, one of the three comprehension questions was presented to the participant prior to listening to the segment. The experiment consisted of 36 trials, and answer accuracy was above 80% for all participants ($M_{correct} = 90.74\% \pm 4.49\%$). To get used to the experimental procedure, participants completed a short training session consisting of two trials before starting the experiment. Data from the training session were not included in the analysis.

To compare speech EEG recordings with standard ABRs, basic click-ABR responses were obtained after the speech experiment. A 10 Hz click train with alternating

polarities was presented at 93 dB peak-to-peak equivalent SPL (to a 1 kHz sinusoid) for five minutes, resulting in 3000 click repetitions. No jitter was applied to the click train.

The EEG was recorded using the Active Two system (BioSemi) with a sampling rate of 16384 Hz. Electrical potentials were measured from 32 scalp electrodes placed according to the 10-20 system, and 4 external electrodes placed on the left and right mastoid bones, as well as over and below the right eye to measure the electrooculogram (EOG).

Analysis

The EEG data was pre-processed using Matlab (MathWorks) and the Fieldtrip toolbox (Oostenveld *et al.*, 2011). Pre-processing of the EEG data was identical for both speech-EEG methods and the click-ABR. It entailed high-pass filtering at 1 Hz to exclude slow electrode drifts and re-referencing to the average of the two mastoid electrodes, after which the mastoid channels were discarded. Noisy EEG channels were further identified through visual inspection and discarded (on average, 0.36 channels per participant were discarded). All analyses reported here focused on electrode Pz. Both speech-based approaches were computed twice, once for the audio segment that was heard during the respective EEG recording (corresponding audio), and once for all other audio segments, which had been presented at another time during the study (random audio). After the analyses described below, the responses at Pz for each participant were averaged over trials. From this calculated individual response, the largest local maximum between 1 and 11 ms was identified as the response peak, and the respective time point as peak latency or delay. For one participant, no clear local maximum between 1 and 11 ms could be identified with the regularized linear regression approach. This participant was therefore excluded from average latency estimations for the regression approach, and all correlations entailing the regression approach. All filters applied to the audio and the EEG in the common pre-processing such as at later processing stages were applied both in forward and reverse direction, compensating for filter delays.

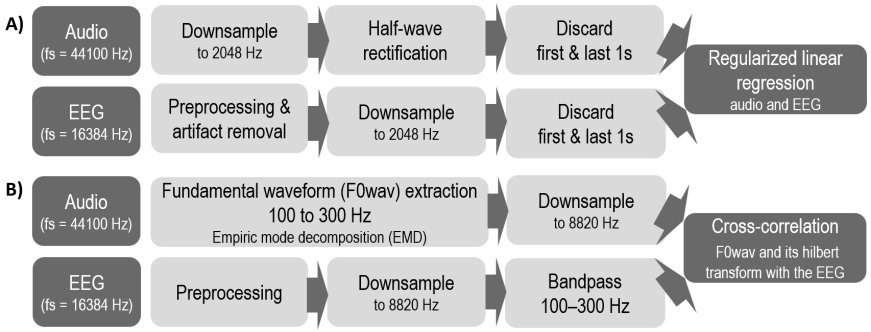

Fig. 1: Schematic description of the analysis pipelines of the (A) regression approach, and (B) correlation approach.

Regularized linear regression approach

The analysis pipeline for the subcortical regularised linear regression approach is depicted in Fig. 1 A, and is based on Maddox and Lee (2018). This analysis is similar to techniques used for measuring speech entrainment at the cortical level (Hjortkjær *et al.*, 2018; Lalor *et al.*, 2009).

In a first step, the audio was down-sampled to 2048 Hz. To account for cochlear processing, and for better comparison with the EEG data, half-wave rectification was applied to the audio. The EEG signal was first pre-processed and artefacts removed, after which signals were down-sampled to 2048 Hz. Muscle and eye movement artefacts were identified as extreme values of the z-scored EEG and the EOG channels, respectively, using individual cut-offs (average cut-off: z-value of 5.07 for muscle artefacts, and 0.57 for eye artefacts). The EOG channels were discarded thereafter. Both EEG recordings and audio from the affected time points were not considered in the analysis.

The first and the last second were discarded from both the audio and the EEG signal, and the two pre-processed signals were then fed into a ridge regression analysis. Using the Telluride Decoding Toolbox (Akram *et al.*, 2017), a forward model was computed for a ridge parameter of $\lambda = 2^{12}$. Time lags between -10 and 23 ms were considered for this analysis. The resulting regression weights or temporal response function (TRF) map from the time-lagged audio stimulus linearily to the EEG response, and characterises the stimulus-evoked neural response (Fuglsang *et al.*, 2017; Ding and Simon, 2012b; Ding and Simon, 2012a; Lalor *et al.*, 2009). The TRF recorded at electrode Pz was up-sampled to the original recording sampling rate of 16384 Hz, and interpreted as the response.

Cross-correlation approach

Figure 1 B shows the analysis pipeline for the cross-correlation approach, which was conducted similarly to how Forte *et al.* (2017) applied it. In contrast to the regression approach, the fundamental waveform of the audio signal was extracted prior to the analysis. The fundamental waveform was extracted according to Kegler (2019), using empirical mode decomposition. No half-wave rectification was applied in the process. The fundamental waveform was restricted between 100 and 300 Hz and down-sampled to 8820 Hz. The EEG recording was pre-processed, but unlike the regularized linear regression approach, no further artefact rejection was applied for the cross-correlation approach. The EEG signal was also down-sampled to 8820 Hz, and then band-pass filtered between 100 and 300 Hz to offer a fair comparison between audio and EEG recording.

The cross-correlation of the EEG recorded at Pz and the fundamental waveform, and the imaginary part of its Hilbert transform, were then computed and interpreted as the real and imaginary part of a complex correlation function, respectively. Time lags between -60 to 60 ms were considered. The computed cross-correlation functions

were up-sampled to the original recording sampling rate of 16384 Hz, and the magnitude peak of the complex correlation function was identified.

Click-ABRs

Classical ABRs at electrode Pz were computed from EEG recordings to the click stimuli. After pre-processing, the EEG data recorded for every click was defined as one trial, and aligned. Trials with voltages exceeding 50 μV were interpreted as including artefacts and excluded from further analysis (3.56 $\pm$ 7.19 trials excluded on average). The ABR data were analysed without down-sampling.

RESULTS

Results from all three compared methods showed response peaks with latencies below 10 ms, consistent with a brainstem or midbrain origin of the response. For the click-ABR, wave V occurred at a latency of 6.69 $\pm$ 0.28 ms. On average, the peak responses were observed at earlier times with the cross-correlation approach (5.97 $\pm$ 1.45 ms), than with the regression approach (8.19 $\pm$ 0.46 ms; Fig. 2).

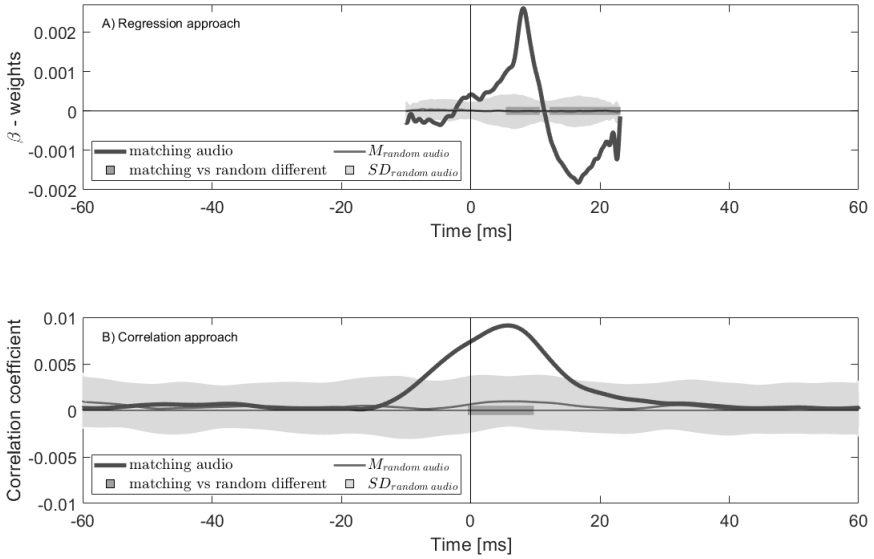

Fig. 2: Comparison of subcortical responses to ongoing speech computed with the two methods: (A) temporal response function (TRF; $\lambda = 2^{12}$), (B) correlation function. Both responses were calculated at electrode Pz. Due to computational limitations, only time lags between -10 and 23 ms were considered for (A).

A two-sided two-sample t-test between the average responses to the matching and the random audio was conducted at every time point and significant differences were observed for both approaches (from 5.49 to 10.62 ms such as from 12.27 ms on for the regression approach and from -0.37 to 9.70 ms for the correlation approach; $\alpha = 0.05$,

no correction applied). The regression approach produced a sharper response peak compared to the cross-correlation approach.

The range of the response peaks was similar across the three methods (Fig. 3 A). Paired-sample t-tests with Bonferroni correction were conducted pairwise between all methods. Latencies obtained with the regression approach were significantly different from both those measured with the correlation approach, and click-ABR wave V (both $p < 0.001$). Latencies from the correlation approach and the click-ABR did not differ significantly ($p = 0.071$).

The average peak obtained with the three different methods for each individual were compared (Fig. 3 B and C). Pearson's correlation coefficient was computed for all three comparisons. The correlations between the regression and the correlation approach and the other two approaches did not yield significance ($p = 0.363$ and $p = 0.136$, for regression and click-ABR respectively). However, latencies of the regression approach and the click-ABRs were highly correlated ($\rho = 0.844$; $p < 0.001$ after Bonferroni correction).

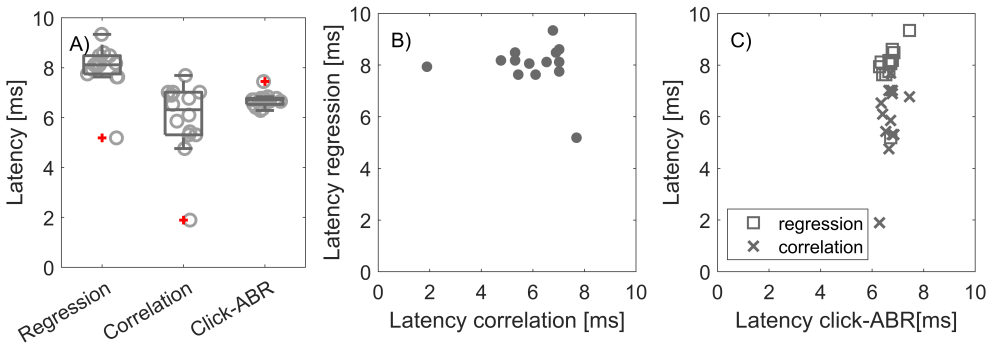

Fig. 3: Comparison of the time lags between the different methods. (A) Box plot of latencies obtained with the different methods, (B) comparison of latencies between regression and correlation approach, (C) comparison of latencies from both regression and correlation approach with the click-ABR.

DISCUSSION

The purpose of the study was to compare two novel methods for deriving a subcortical response to ongoing speech using the same EEG data set. The response latencies obtained for both the regression and the correlation approach lay consistently within a similar range to that reported by the reference studies (Forte *et al.*, 2017; Etard *et al.*, 2009; Maddox and Lee, 2018), and are similar to the latencies of wave V in traditional click-ABRs, both observed here and in previous studies (Garrett and Verhulst, 2019; Maddox and Lee, 2018). Latencies obtained with the regression approach and the click-ABR were correlated. Taken together, our results confirm that both approaches measure aspects related to brainstem processing. Participants'

latencies were not significantly correlated between the two speech-based methods, nor between the correlation approach and the click-ABRs. This may be because the individual differences across the young normal-hearing listeners tested here were small relative to the variance in the latency estimation inherent in each method. If the three methods measure equivalent aspects of the subcortical response, then significant correlations might be observed in studies that include more participants with a broader age range and/or listeners with hearing loss.

Significant differences were observed for average latencies between both speech-based methods, such as between the regression approach and the click-ABRs. Given the differences across methods, there are several possible explanations for this result which are still consistent with the three methods measuring similar aspects of the auditory brainstem response. First, the artefact rejection procedure varied between the three methods. For the click-ABR, trials that exceeded 50 μV were excluded, whereas for the regression approach, a statistical analysis of EOG and EEG data was used to identify muscle and eye movement artefacts. For the cross-correlation approach, no artefact rejection was applied. These differences may have contributed to the observed differences in relative latency. In addition, the latency introduced by the analysis window used to extract F0 in the cross-correlation approach may differ from that in the peripheral auditory system, biasing the results from this approach.

In the present study, only the latency of the peak response was considered across the three methods. However, other aspects, such as peak amplitude, may be of interest for investigating group differences in sub-cortical processing. Thus, further work is needed to compare these methods using other metrics and to investigate how robust the two approaches are.

SUMMARY

The regularized linear regression approach of Maddox and Lee (2018) and the cross-correlation approach of Etard *et al.* (2009) were applied to the same EEG data to derive a subcortical response to speech. The response latencies of both measures were similar to each other and to that of a traditional click-ABR approach.

ACKNOWLEDGEMENTS

This work was partially financially supported by the Sonova Holding AG, and J.H. was supported by the Novo Nordisk Foundation, synergy grant NNF17OC0027872 (UHeal). The authors would like to thank Jonatan Märcher-Rørsted for his support with parts of the analysis, and Rikke Skovhøj Sørensen for conducting the audiometric testing of the participants.

REFERENCES

Akram, S., A. de Cheveigné, P. U. Diehl, E. Graber, C. Graversen, J. Hjortkjaer, N. Mesgarani, L. Parra, U. Pomper, S. Shamma, J. Simon, M. Slaney, and D. Wong (**2017**). *Telluride Decoding Toolbox*. https://github.com/neuromorphs-2017-decoding/telluride-decoding-toolbox.

Ding, N. and J. Z. Simon (**2012a**). "Emergence of neural encoding of auditory objects while listening to competing speakers," Proc. Natl. Acad. Sci. U.S.A., **109** (29), 11854–11859, doi: 10.1073/pnas.1205381109.

Ding, N. and J. Z. Simon (**2012b**). "Neural coding of continuous speech in auditory cortex during monaural and dichotic listening," J. Neurophysiol., **107** (1), 78–89, doi: 10.1152/jn.00297.2011.

Etard, O., M. Kegler, C. Braiman, A. E. Forte, and T. Reichenbach (**2009**). "Decoding of selective attention to continuous speech from the human auditory brainstem response," NeuroImage, **200**, 1–11, doi: 10.1016/j.neuroimage.2019.06.029.

Forte, A. E., O. Etard, and T. Reichenbach (**2017**). "The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention," eLife, **6**, e27203, doi: 10.7554/elife.27203.001.

Fuglsang, S. A., T. Dau, and J. Hjortkjær (**2017**). "Noise-robust cortical tracking of attended speech in real-world acoustic scenes," NeuroImage, **156**, 435–444, doi: 10.1016/j.neuroimage.2017.04.026.

Garrett, M. and S. Verhulst (**2019**). "Applicability of subcortical EEG metrics of synaptopathy to older listeners with impaired audiograms," Hear. Res., **380**, 150–165, doi: 10.1016/j.heares.2019.07.001.

Hjortkjær, J., J. Märcher-Rørsted, S. A. Fuglsang, and T. Dau (**2018**). "Cortical oscillations and entrainment in speech processing during working memory load," Eur. J. of Neurosci., 1–11, doi: 10.1111/ejn.13855.

Kegler, M. (**2019**). *Fundamental waveforms extraction*. https://github.com/MKegler/fundamental_waveforms_extraction.

Lalor, E. C., A. J. Power, R. B. Reilly, and J. J. Foxe (**2009**). "Resolving precise temporal processing properties of the auditory system using continuous stimuli," J. Neurophysiol., **102** (1), 349–359, doi: 10.1152/jn.90896.2008.

Maddox, R. K. and A. K. Lee (**2018**). "Auditory brainstem responses to continuous natural speech in human listeners," eNeuro, **5** (1). doi: 10.1523/eneuro.0441-17.2018.

Oostenveld, R., P. Fries, E. Maris, and J.-M. Schoffelen (**2011**). "FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data," Comput. Intell. and Neurosci., **2011**, 1, doi: 10.1155/2011/156869.