PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Decoding the neural processing of selective attention to speech

Tobias REICHENBACH[1]

[1] Imperial College London, U.K.

ABSTRACT

Understanding speech in noisy backgrounds requires selective attention to a particular speaker. Humans excel at this challenging task, while current speech recognition technology still struggles when background noise is loud. The neural mechanisms by which we attend selectively to a particular speech signal remain, however, poorly understood, not least due to the complexity of natural speech. Here we describe recent progress obtained through applying machine-learning to neuroimaging data of humans listening to speech in background noise. In particular, we develop statistical models to relate two characteristic features of speech, pitch and amplitude fluctuations, to neural measurements. We find neural correlates of speech processing both at the subcortical level, related to the pitch, as well as at the cortical level, related to amplitude fluctuations. Our findings may be applied in smart hearing aids that automatically adjust speech processing to assist a user, as well as inform future speech-recognition algorithms.

Keywords: Speech, Noise, Neural processing

1. INTRODUCTION

Hearing impairments affect more than 16% of the adult population worldwide, and more than 5% in children (1-3). Moreover, hearing impairments progress with age and are hence a particular problem in our aging societies: 40% of people above age 50 and 70% of those above age 70 have hearing impairment.

A major difficulty for people with hearing impairment is to understand speech in noisy environments. Background noise, such as other people talking in a loud restaurant or pub, poses indeed a highly difficult problem for speech comprehension at which healthy humans perform remarkably well.

The neural mechanisms through which we can attend to a particular speaker in background noise and understand him or her still remain largely unclear. A better understanding of these neural processes could lead to a better assessment as well as treatment of hearing impairments, as well as inspire the next generation of automatic speech recognition technology.

A main difficulty with investigating the neural machinery that allows us to understand speech in noise lies in the complexity of speech signals, which unfold rapidly over time and over many different frequency bands. However, recent progress has been made through applying tools from statistical modeling to identify neural correlates of different aspects of speech.

2. Auditory brainstem response to continuous speech

Many parts of speech are voiced, that is, they possess a temporal fine structure characterized by a fundamental frequency and many higher harmonics. The fundamental frequency is typically in the range of 100 - 300 Hz. Neurons in the auditory brainstem can phase lock their responses to a pure tone within this frequency range, resulting in the frequency-following response (FFR) that can be recorded from scalp electrodes.

The frequency content of voiced speech is more complex than that of a pure tone. In particular, the fundamental frequency is not constant, and many higher harmonics are present. Nonetheless, work on brainstem responses to single syllables has shown that the brainstem responds at the fundamental frequency, as well as, to a much lesser degree, at higher harmonics (4-6). However, because the

[1] gf@aaa.com

brainstem response is tiny, its measurement required hundred- to thousandfold repetition of the short speech token.

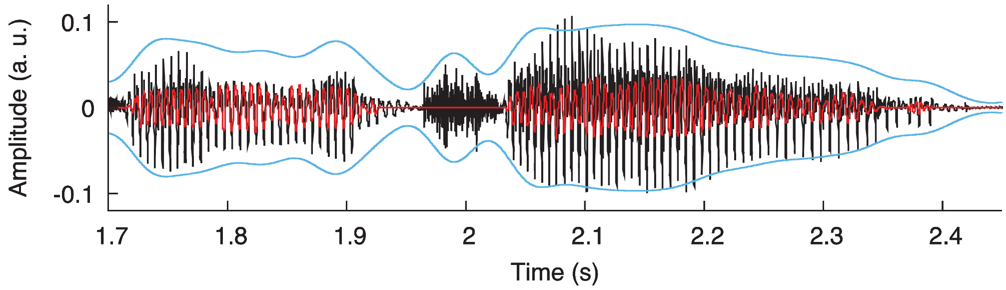

Figure 1 – Speech (black) is a complex signal. Two important quantities are its envelope (blue) as well as the fundamental frequency of the voiced parts of speech that give rise to the fundamental waveform (red).

We wondered if we could measure the brainstem response to continuous speech, without repeating the sound. The average over many repeated stimuli would therefore need to be replaced by a statistical analysis of the brainstem response to the temporally-varying speech signal. We therefore extracted a waveform that, at each time instance, oscillated at the fundamental frequency and at a corresponding amplitude (7). We termed it the 'fundamental waveform' (Figure 1). It was obtained through a Hilbert-Huang transform of the speech signal that decomposed the speech into a series of empirical modes; the fundamental waveform could be viewed as such a mode.

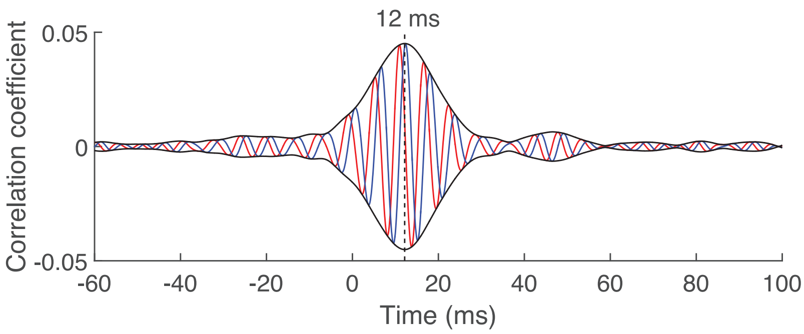

Figure 2 – The cross-correlation of the fundamental waveform of continuous speech with the scalp-recorded brainstem response (blue) peaks at around 12 ms. We show also its Hilbert transform (red) and the envelope (black).

We then recorded the brainstem response to continuous speech of about 10 minutes in duration from a few scalp electrodes that were positioned at the mastoids as well as at the vertex. We computed the difference between each mastoid electrode and the one at the vertex, and cross-correlated this recording to the fundamental waveform of the speech signal. The result from one exemplary volunteer shows a clear peak at a delay of 12 ms. This delay corresponds to the delay of the brainstem response in the subject, and the spread around the peak comes from the autocorrelation of the fundamental waveform. This shows that the auditory brainstem response to continuous speech can be measured from scalp electrodes.

We used this method to investigate whether this brainstem response is modulated by selective attention to one of two competing speakers. We found indeed that the brainstem response to the fundamental waveform of a particular speaker is indeed higher when that speaker is attended than when it is ignored (7).

3. Cortical responses to amplitude fluctuations

Amplitude fluctuations in speech occur at the rate of 1 - 15 Hz, a frequency range that contains the range of delta oscillations in the brain (1 - 4 Hz), theta oscillations (4 - 8 Hz), and alpha oscillations (8

- 12 Hz). Recent electroencephalographic (EEG) and magnetoencephalographic (MEG) studies have shown that the neural activity in these frequency ranges tracks the amplitude fluctuations that are evident in the speech envelope (8-11). The tracking has been hypothesized to play a role in speech processing in the brain.

We have measured EEG responses from 18 native English speakers to continuous speech. Through linear regression we have then computed an estimate of the speech envelope, at different delays of the speech signal relative to the EEG recordings and in different frequency bands (Figure 3). The results show that both the theta and the delta band at a delay of around 100 ms yield good reconstruction.

Moreover, we have measured EEG responses of the volunteers to speech in their native language English, as well as in a foreign language Dutch that they did not understand. Both languages were presented in different levels of background noise. While the speech comprehension varied with the level of background noise for the English stimuli, it was consistently nil for the Dutch speech stimuli. We reconstructed the speech envelope of both the English and the Dutch stimuli, and then trained machine-learning classifiers to predict both speech comprehension as well as speech clarity. The latter was defined as the signal-to-noise ratio of the speech in the background noise, independently of whether it was English or Dutch. We found that the theta band allowed to predict speech clarity, whereas the delta band allowed to decode speech comprehension.

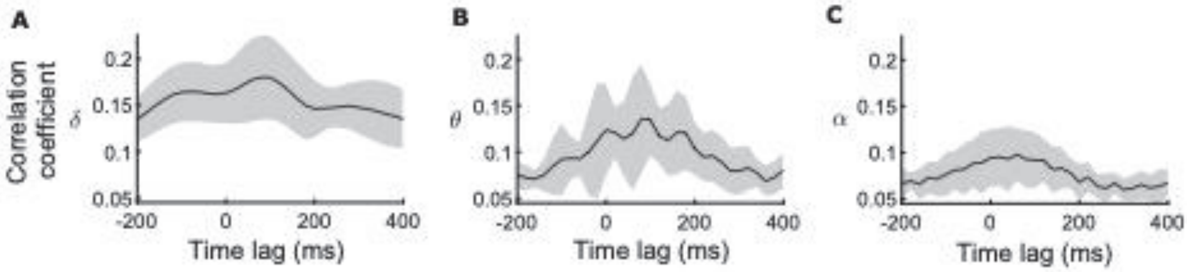

Figure 3 – Reconstruction of the speech envelope from EEG recordings. The correlation coefficient between the reconstructed envelope and the actual one measures the goodness of the reconstruction. We show the correlation coefficient for reconstructions using temporal lags from -200 ms to 400 ms, and for the delta frequency band (A), the theta frequency band (B), and the alpha frequency band (C).

4. CONCLUSIONS

We have shown that neural activity tracks both the temporal fine structure of speech as well as its amplitude fluctuation. The subcortical response at the fundamental frequency is modulated by attention, and the neural response to the amplitude fluctuations allows to decode speech comprehension.

Beyond the relevance of these findings for neurobiology, they may also be employed in technology. In fact, both attention as well as speech comprehension can be decoded efficiently from short EEG recordings of a few seconds in duration. This suggests that the decoding may be employed for near-real-time control of settings in a hearing aid. The hearing aid could, for instance, amplify an attended speaker more, or improve the complex settings if speech comprehension drops.

ACKNOWLEDGEMENTS

This research was supported by the Royal British Legion Centre for Blast Injury Studies, as well as by EPSRC grants EP/M026728/1 and EP/R032602/1.

REFERENCES

1. Davis AC. The prevalence of hearing impairment and reported hearing disability among adults in Great Britain. Int J Epidemiol. 1989;18(4):911-7.

2. Moore DR, Rosen S, Bamiou D-E, Campbell NG, Sirimanna T. Evolving concepts of developmental auditory processing disorder (APD): a British Society of Audiology APD special interest group'white paper'. Int J Audiol. 2013;52(1):3-13.

3. Davis A, Wood S. The epidemiology of childhood hearing impairment: factors relevant to planning of services. Br J Audiol. 1992;26(2):77-90.

4. Skoe E, Kraus N. Auditory brainstem response to complex sounds: a tutorial. Ear Hear. 2010;31(3):302.

5. Chandrasekaran B, Kraus N. The scalp‐recorded brainstem response to speech: Neural origins and plasticity. Psychophysiol. 2010;47(2):236-46.

6. Musacchia G, Strait D, Kraus N. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hear Res. 2008;241(1):34-42.

7. Forte AE, Etard O, Reichenbach T. The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. eLife. 2017;6:e27203.

8. Ding N, Simon JZ. Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum Neurosci. 2014;8:311.

9. Di Liberto GM, O'Sullivan JA, Lalor EC. Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Curr Biol. 2015;25(19):2457-65.

10. Peelle JE, Gross J, Davis MH. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cerebr Cort. 2013;23(6):1378-87.

11. Giraud A-L, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci. 2012;15:511-7.