Decoding Auditory Attention (in Real Time) with EEG

Edmund Lalor[1], Nima Mesgarani[2], Siddharth Rajaram[3], Adam O'Donovan[4], James Wright[5], Inyong Choi[3], Jonathan Brumberg[3,6], Nai Ding[4], Adrian KC Lee[7], Nils Peters[8], Sudarshan Ramenahalli[9], Jeffrey Pompe[9], Barbara Shinn-Cunningham[3], Malcolm Slaney[10,11,7], Shihab Shamma[4]

[1]Trinity College Dublin, [2]University of California, San Francisco, [3]Boston University, [4]University of Maryland, College Park, [5]University of Western Sydney, [6]University of Kansas, [7]University of Washington, [8]University of California, Berkeley, [9]Johns Hopkins University, [10]Microsoft Research, [11]Stanford CCRMA

2012 Neuromorphic Cognition Engineering Workshop

Background

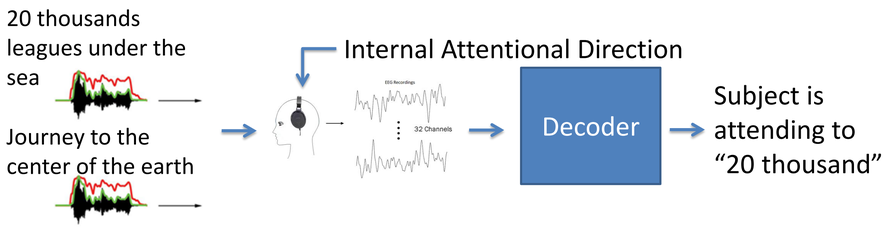
Both magnetoencephalography and electrocorticography recordings have been used to decode which of two competing sources a listener is attending. However, it is not clear whether these techniques might work with Electroencephalography (EEG), particularly in a real-time system. We therefore set out to decode a listener's attentional focus from EEG signals in real time, knowledge that could be incorporated into next-generation assistive listening devices.

Methods

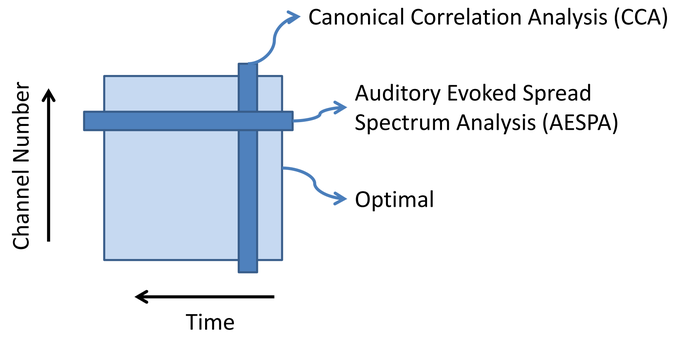
Offline, we acquired EEG data when a subject listened to a single speech source, from which we estimated a mapping from the EEG data to the perceived speech. The subject then attended to one of two simultaneous speech streams, presented dichotically. The previously estimated system transfer function from the single-source presentation was used to estimate the attended stream in real time. Whichever input speech stream more closely resembled the estimated input was deemed to be the attended stream. Three decoding methods were tested. The first approach, canonical correlation analysis (CCA), is based on measuring the correlation between audio streams and the EEG signals. The second two approaches estimate the mapping from the EEG to the input stimulus. This can be done using a single channel at a time and summing the result, or by finding a single multi-channel filter that represents all the signals.

Single Source (Typical BCI)

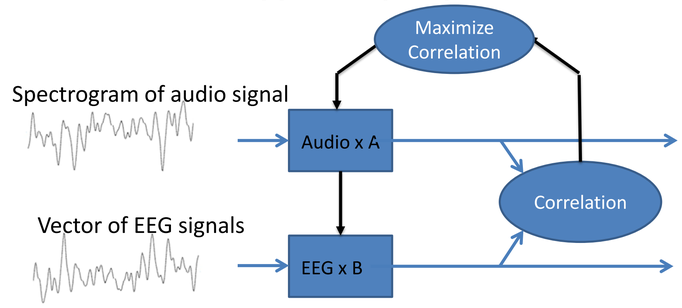
Two Sources with Attention



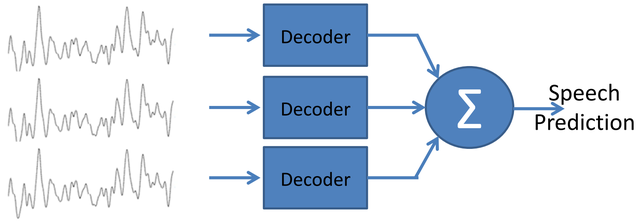
Three Decoding Approaches over Time and Channel



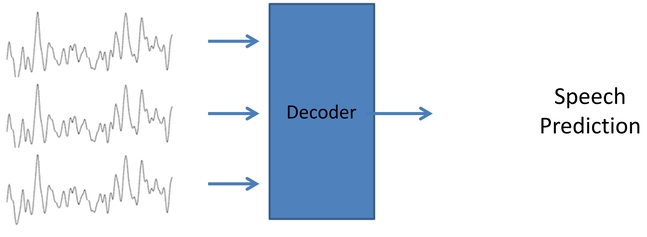
Canonical Correlation Approach (CCA)



Single Channel Inversion (AESPA)



All Channel Inversion (Optimal)



Results

We found the best results by estimating a multivariate linear filter that incorporates the channel covariance structure in the least-squares estimation of the impulse response, similar to the approach described by Mesagarani & Chang (2012). Using this approach we could estimate single-speaker data with high accuracy. Notably this approach yielded estimates of the speech envelope that were better correlated with the original speech (r ~= 0.08) than the other two methods. Applying this to the attention paradigm (after training on single-speaker data), we could predict the focus of attention with 95% accuracy for a one-minute-long sample of dichotic speech. As we shortened the amount of data used to decode, our accuracy fell almost linearly to about 65% for 10 seconds. Other presentation conditions (i.e., diotic and HRTF) were decoded with lower accuracy than dichotic.

Conclusion

EEG signals can be decoded in real time to determine what natural speech stream a listener is attending with relatively high accuracy.

Speech Decoding Accuracy, 20ms Frame

Performance	Accuracy	Correlation
CCA	>90%	0.02-0.04
Single Channel	>90%	0.02-0.04
Optimal	>90%	0.04-0.08

Attention Decoding Accuracy, 60s Look

	Accuracy
CCA	65-80%
Single Channel	65-80%
Optimal	75-95%

Attention Decoding Accuracy (%) vs. Look Time (s)