European Commission | Horizon 2020 European Union funding for Research & Innovation

H2020 ICT-22-2014
COCOHA
Project No.: 644732

Cognitive Control of a Hearing Aid

edited by Alain de Cheveigné, 13 September 2017

COCOHA report on Audio preprocessing - v2

This document provides an overview of preprocessing techniques for audio signals to be matched with EEG signals for the purpose of hearing aid control. It is part of Deliverable D2.2 of the COCOHA project, funded by the H2020 ICT programme of the European Union under grant number 644732. The COCOHA project aims to help hearing impaired persons so that they can deal with challenging noisy environments, by providing them with the means to steer a sophisticated hearing aid with signals derived directly from the brain. This document is available at https://cocoha.org/cocoha-reports/.

Executive summary

1. COCOHA aims allow the user of a hearing aid to control acoustic processing (e.g. microphone arrays) so as to isolate one target source among many for amplification. This is done using of signals recorded from the brain using EEG.
2. A prominent control strategy is to correlate each audio stream with EEG signals measured from the brain, choosing for amplification that stream which yields the highest correlation.
3. This strategy requires that the audio signal be *preprocessed* into a representation that can be usefully correlated with the EEG, as the raw audio and EEG signals occupy different frequency regions and cannot be directly compared. The aim of preprocessing is to derive from the audio a quantity predictive of the EEG response.
4. Two variants are used: the forward model in which EEG signals are *predicted* from the stimulus representation (so-called temporal response function, TRF, models), and the backward model in which the stimulus representation is *inferred* from the EEG (so-called stimulus reconstruction models). A third approach, deployed in COCOHA, uses *canonical correlation analysis* to transform both audio and EEG so as to maximize correlation.
5. A wide range of preprocessing schemes have been proposed, from low level features (such as *waveform envelope* and *spectrogram*) to higher level features (cortical model, modulation spectrum) and symbolic representations.
6. Pioneering studies used the simplest representations (waveform envelope and spectrogram) which are still remarkably effective.
7. Recent studies investigate more sophisticated representations (cochlear and cortical models, MFCC, or symbolic representations). These provide a benefit that can be significant, but usually small.
8. The most promising direction involves *combining* features to obtain a more reliable predictor of EEG signals (and thus more reliable classification), and *switching* between models according to their validity.
9. The COCOHA project is investigating these directions.

1. The COCOHA context

The COCOHA project (http://cocoha.org/) funded by the EU H2020 initiative aims at developing a "smart" hearing aid that puts acoustic processing under direct control of the user's brain, using signals recorded by electroencephalography (EEG) or other means. Such a device should enable a hearing-impaired listener to attend to one particular sound source (for example a person speaking) and ignore competing voices and sounds. Our intact auditory is adept at performing such a task, usually without our noticing, but this ability is greatly reduced by impairment. The COCOHA project aims to restore this ability by artificial means.

Processing techniques to analyze complex acoustic scenes, separate sources, and improve intelligibility are reviewed in the COCOHA Report on Acoustic Signal Processing for Hearing Aids (https://cocohablog.files.wordpress.com/2017/01/acoustical-signal-processing-v4.pdf). The present document deals with a different task: preprocessing acoustic signals to allow comparison with brain signals, so as to allow the hearing aid to select which particular sound stream the user wishes to attend to.

2. Cognitive control strategies

Within COCOHA we consider several strategies to control a hearing aid based on brain signals. These are:

- Correlation of brain signals and acoustic signals. Each available audio stream is correlated with EEG signals and the stream that yields the best correlation is selected for amplification.
- Decoding cues to spatial attention attention and inattention. The spatial focus of attention is decoded from the EEG signal and audio streams that match that position are selected for amplification.
- Exploiting hybrid control cues (head and eye position, haptic control, etc. together with EEG). EEG signals are used to complement other control modalities.

The first strategy is mainly investigated within the project, but we remain open to every possibility to combine it with other strategies to improve the applicability, ergonomy, and performance of cognitive control.

3. Relating audio to EEG

Mainly two approaches have been used in the past: temporal response function (TRF) or *forward model*, and stimulus reconstruction or *backward model*. The first approach attempts to predict EEG responses based on the audio, using an audio-to-EEG forward model fit to the data. The second approach aims to infer the audio from the EEG based on a backward model. Forward and backward approaches are usually based on the same linear model that is "learned" from the data.

In addition to these two classic approaches, a new approach based on Canonical Correlation Analysis (CCA) have more recently been introduced. In this approach, audio and EEG features are processed together, projecting both into a common space where they can be compared.

4. The need for audio preprocessing

The audible frequency range extends from roughly 20 Hz to 20 kHz, whereas EEG signals occupy a lower range, from 0.1 Hz or lower for very low frequency activity (Vanhatalo et al 2005) to a few tens of Hz for gamma activity. At higher frequencies brain signals measured by EEG are often weak due to low-pass characteristics of brain source-to-electrode propagation and mixing, and dominated by muscle artifact or other sources of noise (Yuval-Greenberg et al 2008; Whitham et al 2007). According to the Wiener-Khinchine theorem the cross-correlation function and cross-spectrum are Fourier pairs:

$$C_{xy}(\tau) \Leftrightarrow \overline{X(f)}Y(f)$$

which implies that the cross-correlation between signals is zero unless the spectra overlap. The audio signal has a different spectral content from the EEG, and thus must be *preprocessed*.

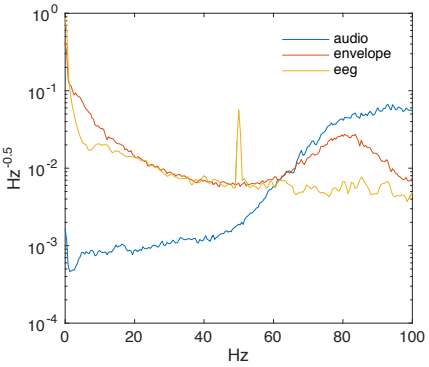
A simple form of preprocessing consists of simple demodulation of the acoustic signal to derive a *waveform envelope*. More sophisticated preprocessing can involve decomposition by a filterbank, or cepstral processing, auditory processing models (e.g. cochlear or cortical models), higher-order signal processing models (e.g. modulation filterbank or scattering transform), symbolic representations (phonemes, words, etc.) or neural network activations. These approaches are reviewed in this document. The aim in every case is the same: derive from the audio signal some quantity that is reliably *predictive* of some aspect of measurable brain activity. Different parts of the brain may care for different aspects of the signal: primary auditory cortex may care for low level features (waveform envelope or spectrogram), whereas secondary regions may care for linguistic information, or higher-order statistical structure of the acoustic waveform.

Progress is measured in terms of correlation values between speech and EEG, and ultimately the score of classification algorithms that rely on these data. The need for real-time processing within the device may impose additional constraints.

5. Audio preprocessing for EEG-audio decoding

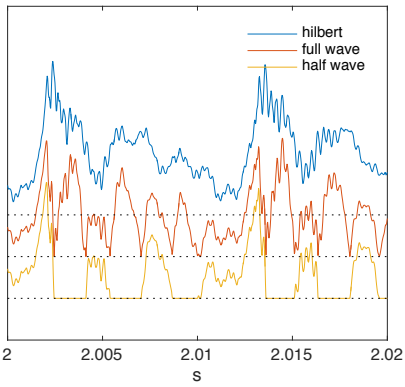
5.1 Waveform envelope

The acoustic waveform spectrum content is usually too high to be commensurable with EEG, but the envelope *modulation spectrum* of audio signals such as speech occupies a roughly similar spectral range to EEG (Obleser et al 2012).



Power spectra of audio, audio envelope and EEG in the low frequency region. The raw audio signal has little power in the range where EEG power is concentrated, implying poor correlation. The *envelope* of that signal has power in the same range as EEG.

The modulation spectrum, distinct from the acoustic spectrum, is an important feature characteristic of speech and other sounds (Dau et al. 1997; Hermansky 2010; Jepsen et al. 2008; Lorenzi et al. 2001; Singh and Theunissen 2016; Xiang et al 2013). The envelope may be derived from the waveform by applying the Hilbert transform, or merely by half-wave or full-wave rectification followed by low-pass filtering.



Waveform envelope calculated using various methods. Spectral properties are roughly equivalent in the low-frequency range where EEG is concentrated.

The resulting envelope waveform may be submitted to a non-linear transform, typically compressive power function (e.g. exponent 2/3) or logarithm. These details usually have a minor effect on performance (Biesmans et al 2017).

Additional schemes have been proposed, for example the half-wave-rectified *derivative* of the envelope signal (Sturm et al 2015). The rationale here is that brain responses tend to be triggered by the onset of acoustic events. While attractive, this transform has not proved superior to the envelope itself.

The majority of studies of audio/EEG decoding assume a waveform envelope representation (e.g. Powers et al 2012;

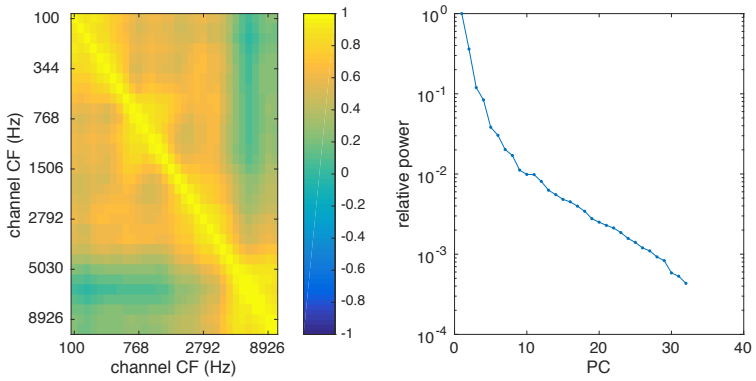
5.2. Filter-bank

While the waveform envelope has been successful in early studies, it is a very crude representation of the ongoing acoustic signal. Processing within the auditory system begins with filtering within the cochlea, and thus a filter-bank representation (for example based on a model of cochlear filtering) would seem more promising than the waveform envelope. The filter-bank splits the acoustic signal into multiple channels, each of which can be processed individually (by demodulation as above), to obtain a set of envelope signals with slow variations commensurate with EEG. Envelopes from multiple channels then form a multidimensional signal representative of the sound.

Nonetheless, the small number of demodulated filter channels constitute an impoverished representation of the speech signal, although multichannel representations of this sort have been shown to support a high level of intelligibility (Shannon et al 1995). Regardless, this representation has some hope to better predict cortical responses than the waveform envelope.

The filterbank representation has been investigated in several decoding studies (Di Liberto 2015; Biesmans et al 2017; Martin et al 2014; Mesgarani and Chang 2012; Pasley et al 2012), for both forward (TRF) and backward (reconstruction models). Reconstruction from invasive measurements (ECoG) has met a surprising degree of success (e.g. Mesgarani and Chang 2012; Pasley et al 2012), but reconstruction from EEG or MEG is less successful. In contrast,

forward models benefit from the greater detail of spectrogram vs waveform envelope representations ((Di Liberto 2015; Biesmans et al 2017).



Left: cross-correlation coefficient between channels of an auditory spectrogram. The filterbank was simulated in the frequency domain (FFT) with channel frequency spacing and bandwidths calculated to simulate human peripheral selectivity, and temporal resolution calculated to approximate human temporal resolution while maintaining uniform alignment and resolution across frequency bands (~20ms) so that the spectrogram is separable. Note the large values of correlation between most channels. Right: relative power of each individual principal component of the spectrogram.

Filterbank representations are contingent on type and parameters of the filterbank. These include the bandwidth and shape of each filter, the density and distribution of centre frequencies (CF), and the nature of the demodulation and eventual nonlinear scaling of the outputs. Convolution by the filter impulse response entails *temporal smoothing* that is greater for narrow filters. Many filterbanks are logarithmic (e.g. wavelet) or approximately logarithmic (e.g. cochlear), with narrower filters at low CFs. The power of audio signals such as speech is often correlated across frequency bands, leading to a high degree of correlation between channels, such that the envelope of any band is not too different from the envelope of the raw waveform.

Interestingly, filterbank representations are often summarized by averaging over frequency dimension (Ding and Simon 2013), effectively producing a "broadband" envelope similar to that discussed earlier.

5.3. Cortical model

Cochlear filtering is but the first stage of auditory processing. It captures well our sensitivity to spectral features, but not the effects of temporal and spectral modulation, that are presumably calculated at higher auditory stages such as the cortex. Such higher-order processing stages have also been modeled, in particular cortical processing of spectral and temporal modulation (Chi et al 2005), and some success has been reported in using them to relate to brain responses, in particular invasive recordings (Pasley et al 2012).

5.4. Higher-order signal-processing models

Many schemes have been proposed to describe sound (e.g. Alías 2016). To the extent that they are relevant to predict meaning or aspects of sound that determine the response of a listener, they are all of potential use to predict the listener's brain response.

The Mel frequency cepstral coefficient (MFCC) is commonly used in Automatic Speech Recognition (ASR). The MFCC captures well the overal shape of the spectrum, with macroscopic features (e.g. spectral tilt) coded in the lowest coefficients and details (e.g.

harmonic structure) in the highest coefficients. MFCCs have been used as an alternative to a spectrogram representation (Chakrabarti et al 2013; Chan et al 2014).

Several studies have postulated an intimate connection between the rhythmic structure of speech (as captured by modulation spectral representations) and cortical rhythms (Ghitza 2011; Zoefel and Van Rullen 2016; Peelle et al 2012; Giraud and Poppel 2012), although the well-groundedness of this hypothesis has been questioned (Cummins 2012; Obleser 2012). Rhythmic structure on the (slow) scale of syllables or articulatory gestures can be characterized directly from temporal variation of acoustic features (envelope, spectrogram) or quantified more abstractly by higher-order modulation features, such as offered by certain auditory models (e.g. Dau 1997; Jepsen et al 2008). Modulation features are captured by the cortical model of Chi et al (2005) that has been applied to EEG/audio decoding (Pasley et al 2012).

Audio signals have structure on multiple scales, ranging from the short scale captured by spectral representations to the scales characteristic of textures, or speech or musical structure that may be characterized by methods such as the Scattering Transform (Andèn et al 2015). These models are of potential interest to the extent that cortical events may be triggered by events within a complex hierarchical structure, in addition to the low-level patterns that have been characterized in past studies. The potential of such models in the context of the COCOHA application is that the richer representation that they offer they may lead to better discrimination between attended and unattended streams.

5.5. Symbolic representations

In addition to low-level acoustic features characteristic of the waveform, speech may be indexed with higher-level labels of phonemes (or phonemic traits), words, and so-on. Several studies have shown that EEG responses can be related to these labels, leading to better decoding performance than with acoustic features only (Di Liberto et al 2015; Di Liberto et al 2017; Tankus et al 2012; Mesgarani et al 2015; Khalighinejad et al 2017). This result is drawn from studies where speech data with phonemic labels was available, but the same information could in principle be extracted from the speech using an automatic speech recognition system.

5.6. Multimodal representations

Most speech decoding studies assume that only the audio signal is available, but in practial scenarios it is not unreasonable to assume that visual cues are available, at least part of the time. They may even be crucially needed in noisy situations, if the listener is impaired, or if the SNR of the desired source is initially too low to allow audio-based decoding (the so-called "bootstrap problem" of a decoding-based control system). Cortical responses to speech are indeed enhanced in the presence of visual cues (Zion Golumbic et al 2013; Crosse et al 2015; Crosse et al 2016; Peelle et al 2015). In order to benefit from these cues, an actual system would need to analyze the scene using machine vision techniques for cues to visible events (e.g. speaker articulator movements).

A useful feature of such higher-order, symbolic and multimodal representations is that they may involve cortical sources (most likely from secondary areas) distinct from low level features (most likely from primary areas). Distinct sources are likely to have distinct *spatial signatures* that can be used to derive distinct discriminative dimensions for decoding and control.

5.7. Deep Neural Networks

The most remarkable recent trend in information processing is the development of deep learning techniques. These have been applied to EEG signal analysis (e.g. Stober et al 2016; Suckling et al 2015; Kwak et al 2017; Zheng & Lu 2015). DNNs do not yet seem to have been applied to the task of extracting an EEG-predictive representation from audio streams.

6. Combining representations

Different representations may tap different levels of processing. To the extent that they come from spatially distinct sources (e.g. from different levels of the processing hiearchy) they provide distinct and complementary discriminatory dimensions. A general tool for combining distinct representations, extensively used within the COCOHA project, is *canonical correlation analysis* (Hotelling 1936).

Summary

The COCOHA project aims to develop a hearing aid with sophisticated acoustic processing under cognitive control. One strategy for control calls for the EEG signal from the user to be correlated with each of the acoustic streams that can be isolated by the acoustic processing module, so as to select one for amplification. For this to be successful the audio must be preprocessed so as to isolate features most predictive of the EEG evoked by an attended stream. A wide range of features is available, from low-level envelope or spectrotemporal features, to high-level structural or symbolic representations. Low-level features are implemented in the COCOHA toolbox, others are under investigation.

References

Alías, F., Socoró, J., & Sevillano, X. (2016). A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds. *Applied Sciences*, *6*(5), 143. https://doi.org/10.3390/app6050143

Andén, J., Lostanlen, V., & Mallat, S. (2015). Joint Time-Frequency Scattering for Audio Classification. https://doi.org/10.1109/MLSP.2015.7324385

Biesmans, W., Das, N., Francart, T., & Bertrand, A. (2017). Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*, 402–412.

Chakrabarti, S., Krusienski, D. J., Schalk, G., & Brumberg, J. S. (2013). Predicting mel-frequency cepstral coefficients from electrocorticographic signals during continuous speech production. *6th International IEEE EMBS Conference on Neural Engineering*, *7*(5), 1064912.

Chan, A. M., Dykstra, A. R., Jayaram, V., Leonard, M. K., Travis, K. E., Gygi, B., … Cash, S. S. (2014). Speech-specific tuning of neurons in human superior temporal gyrus. *Cerebral Cortex*, *24*(10), 2679–2693. https://doi.org/10.1093/cercor/bht127

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *35*(42), 14195–204. https://doi.org/10.1523/JNEUROSCI.1829-15.2015

Crosse, M. J., Di Liberto, G. M., & Lalor, E. C. (2016). Eye Can Hear Clearly Now: Inverse Effectiveness in Natural Audiovisual Speech Processing Relies on Long-Term Crossmodal Temporal Integration. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *36*(38), 9888–95. https://doi.org/10.1523/JNEUROSCI.1396-16.2016

Di Liberto, G. M., & Lalor, E. C. (2017). Indexing cortical entrainment to natural speech at the phonemic level: Methodological considerations for applied research. *Hearing Research*, *348*, 70–77. https://doi.org/10.1016/j.heares.2017.02.015

Cummins, F. (2012). Oscillators and syllables: A cautionary note. *Frontiers in Psychology*, *3*(OCT), 1–2. https://doi.org/10.3389/fpsyg.2012.00364

Dau T., Kollmeier B., and Kohlrausch A. (1997a). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. 102, 2892–2905

Ding, N., & Simon, J. Z. (2013). Adaptive Temporal Encoding Leads to a Background Insensitive Cortical Representation of Speech. *J. Neurosci*, *33*, 5728–5735.

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. *Current Biology : CB*, *25*(19), 2457–2465. Retrieved from http://dx.doi.org/10.1016/j.cub.2015.08.030

Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, *2*(JUN), 1–13. https://doi.org/10.3389/fpsyg.2011.00130

Giraud, A., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063.Cortical

Herff, C., Heger, D., de Pesters, A., Telaar, D., Brunner, P., Schalk, G., & Schultz, T. (2015). Brain-to-text: Decoding spoken phrases from phone representations in the brain. *Frontiers in Neuroscience*, *9*(JUN), 1–11. https://doi.org/10.3389/fnins.2015.00217

Hermansky, H. (2010). History of modulation spectrum in ASR. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 5458–5461). IEEE. Retrieved from http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5494907

Hotelling, Harold, 1936. Relations between two sets of variates. Biometrika 28 (3/4), 321–377.

Huigen, E.; Peper, A.; Grimbergen, C.A. Investigation into the origin of the noise of surface electrodes. Med. Biol. Eng. Comput. 2002, 40, 332–338.

Jepsen, M. L., Ewert, S. D., & Dau, T. (2008). A computational model of human auditory signal processing and perception. *The Journal of the Acoustical Society of America*, *124*(1), 422.

Retrieved from http://scitation.aip.org/content/asa/journal/jasa/124/1/10.1121/1.2924135
Khalighinejad, B., Cruzatto da Silva, G., & Mesgarani, N. (2017). Dynamic Encoding of Acoustic Features in Neural Responses to Continuous Speech. *The Journal of Neuroscience*, *37*(8), 2176–2185. https://doi.org/10.1523/JNEUROSCI.2383-16.2017
Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. Nature Neuroscience, 12(5), 535–540. Retrieved from http://www.nature.com/doifinder/10.1038/nn.2303
Kwak, N. S., Müller, K. R., & Lee, S. W. (2017). A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. *PLoS ONE*, *12*(2), 1–20. https://doi.org/10.1371/journal.pone.0172578
Lorenzi, C., Soares, C., & Vonner, T. (2001). Second-order temporal modulation transfer functions. *The Journal of the Acoustical Society of America*, *110*(2), 1030. Retrieved from http://scitation.aip.org/content/asa/journal/jasa/110/2/10.1121/1.1383295
Ma J, Tao P, Bayram S, Svetnik V (2012) Muscle artifacts in multichannel EEG: characteristics and reduction. Clinical Neurophysiology 123:1676–1686.
Martin, S., Brunner, P., Holdgraf, C., Heinze, H., & Crone, N. E. (2014). Decoding spectrotemporal features of overt and covert speech from the human cortex. *Frontiers in Neuroengineering*, *7*(May), 1–15. http://doi.org/10.3389/fneng.2014.00014
Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, *485*, 233–236. http://doi.org/10.1038/nature11020
Mesgarani, N., Cheung, C., Johnson, K., Chang, E. F., & Francisco, S. (2015). HHS Public Access, *343*(6174), 1006–1010. https://doi.org/10.1126/science.1245994.Phonetic
Obleser, J., Herrmann, B., & Henry, M. J. (2012). Neural Oscillations in Speech: Don't be Enslaved by the Envelope. *Frontiers in Human Neuroscience*, *6*, 250. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3431501/
Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., … Chang, E. F. (2012). Reconstructing speech from human auditory cortex. *PLoS Biology*, *10*(1). https://doi.org/10.1371/journal.pbio.1001251
Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*(SEP), 1–17. https://doi.org/10.3389/fpsyg.2012.00320
Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex*, *68*, 169–181. https://doi.org/10.1016/j.cortex.2015.03.006
Power, A. J., Foxe, J. J., Forde, E. J., Reilly, R. B., & Lalor, E. C. (2012). At what time is the cocktail party? A late locus of selective attention to natural speech. *European Journal of Neuroscience*, *35*(9), 1497–1503. https://doi.org/10.1111/j.1460-9568.2012.08060.x
Singh, N. C., & Theunissen, F. E. (2016). Modulation spectra of natural sounds and ethological theories. *Journal of the Acoustical Society of America*, *114*(6), 3394–3411. http://doi.org/10.1121/1.1624067
Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., Ekelid, M. 1995 Speech recognition with primarily temporal cues. *Science* **270,** 303-304.
Stober, S., Sternin, A., Owen, A. M., & Grahn, J. A. (2015). Deep Feature Learning for EEG Recordings. *Arxiv*, (https://arxiv.org/pdf/1511.04306.pdf), 1–24. Retrieved from http://arxiv.org/abs/1511.04306
Sturm, I., Dähne, S., Blankertz, B., & Curio, G. (2015). Multi-variate EEG analysis as a novel tool to examine brain responses to naturalistic music stimuli. *PLoS ONE*, *10*(10), 1–30. https://doi.org/10.1371/journal.pone.0141281
Suckling, J., Balaguer-Ballester, E., Manor, R., & Geva, A. B. (2015). Convolutional Neural Network for Multi-Category Rapid Serial Visual Presentation BCI. *Frontiers in Computational Neuroscience Front. Comput. Neurosci*, *9*(9), 1463389–146. https://doi.org/10.3389/fncom.2015.00146

Tankus, A., Fried, I., & Shoham, S. (2012). Structured neuronal encoding and decoding of human speech features. *Nature Communications*, *3*, 1015. https://doi.org/10.1038/ncomms1995

Whitham EM, Pope KJ, Fitzgibbon SP, Lewis T, Clark CR, Loveless S, Broberg M, Wallace A, DeLosAngeles D, Lillie P, Hardy A, Fronsko R, Pulbrook A, Willoughby JO (2007) Scalp electrical recording during paralysis: quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG. Clinical Neurophysiology 118:1877–1888.

Xiang, J., Poeppel, D., & Simon, J. Z. (2013). Physiological evidence for auditory modulation filterbanks: Cortical responses to concurrent modulations. *The Journal of the Acoustical Society of America*, *133*(1), EL7. Retrieved from http://scitation.aip.org/content/asa/journal/jasa/133/1/10.1121/1.4769400

Yuval-Greenberg S, Tomer O, Keren AS, Nelken I, Deouell LY (2008) Transient Induced Gamma-Band Response in EEG as a Manifestation of Miniature Sac- cades. Neuron 58:429–441.

Zheng, W. L., & Lu, B. L. (2015). Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Transactions on Autonomous Mental Development*, *7*(3), 162–175. https://doi.org/10.1109/TAMD.2015.2431497

Zion Golumbic, E., Cogan, G. B., Schroeder, C. E., & Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party". *The Journal of Neuroscience*, *33*(4), 1417–1426. https://doi.org/10.1523/JNEUROSCI.3675-12.2013

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., … Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." *Neuron*, *77*(5), 980–991. https://doi.org/10.1016/j.neuron.2012.12.037

Zoefel, B., & VanRullen, R. (2016). EEG oscillations entrain their phase to high-level features of speech sound. *NeuroImage*, *124*, 16–23. https://doi.org/10.1016/j.neuroimage.2015.08.054