
Decoding EEG responses to the speech envelope using deep neural networks

Michael Thornton^{*†1}, Tobias Reichenbach^{‡2}, and Danilo Mandic³

¹Imperial College London, Department of Bioengineering – United Kingdom

²Friedrich-Alexander Universität Erlangen-Nürnberg – Germany

³Imperial College London, Department of Electrical and Electronic Engineering – United Kingdom

Abstract

Background: Unimpaired human listeners are remarkably good at attending to a target speaker whilst filtering out background sounds. Distinct representations of attended and unattended speakers can be decoded from electrophysiological recordings, and it is anticipated that advances in so-called auditory attention decoding (AAD) methodologies will one day lead to a neuro-steered hearing aid for hearing-impaired listeners. For this application, auditory attention decoding from EEG recordings is required to run with a high accuracy, a low latency, and it needs to work in various listening conditions. We present an investigation into how the use of deep neural networks (DNNs) might address these challenges.

Methods: We compared the standard linear technique of regularised least-squares (ridge regression) against two distinct neural networks for reconstructing the envelope of the attended speech stream from a listener’s EEG recordings. Our dataset included several listening conditions: clean speech in native English and foreign Dutch; native speech in background babble noise; and a competing-speaker scenario. An additional clean-speech dataset was used to train listener-independent decoders.

Results: For reconstructing the envelope of clean speech, listener-specific DNNs were shown to offer a considerable improvement over listener-specific linear methods. Even when listener-independent methods were used, the DNNs performed significantly better than ridge regression. Furthermore, whilst the listener-independent methods were trained using EEG recorded under native clean speech conditions, they generalised well to new listeners and different listening conditions. The pre-trained DNNs achieved a significantly greater reconstruction score than the pre-trained ridge regressor across all listening conditions.

Conclusions: We showed that linear methods and deep neural networks which were trained to reconstruct the envelope of native clean speech from EEG recordings can be applied effectively across a variety of listening conditions. DNNs are known to suffer from overfitting issues and can generalise poorly, so it is significant that the deep neural networks were capable of generalising between listeners and listening conditions. These results therefore suggest that the use of DNNs offer good prospects for real-world auditory attention decoding.

Funding: This work was supported by the UKRI CDT in AI for Healthcare <http://ai4health.io> (Grant No. EP/S023283/1).

*Speaker

†Corresponding author: mdt20@ic.ac.uk

‡Corresponding author: tobias.j.reichenbach@fau.de