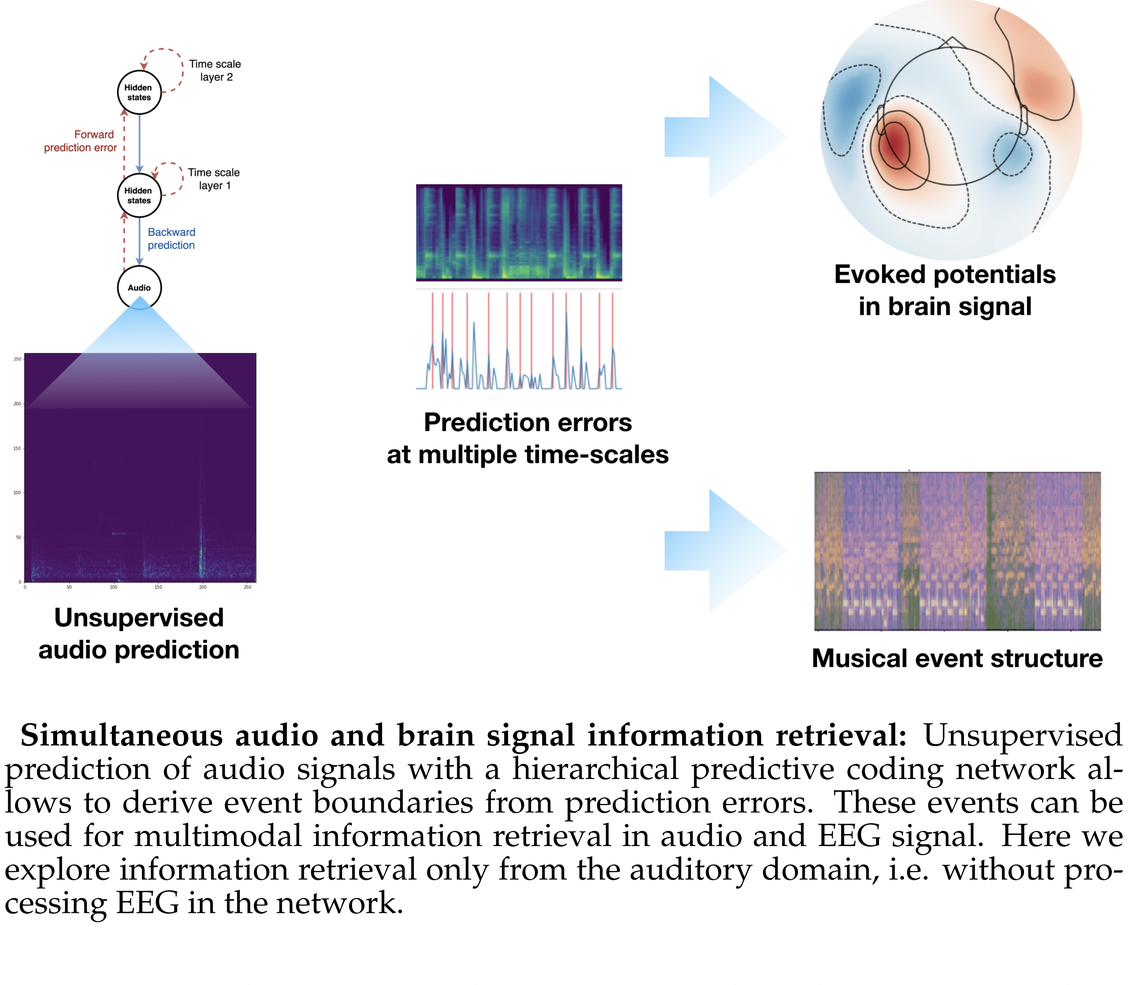
Auditory Segmentation with Deep Predictive Coding Locates Candidate ERPs in EEG

André Ofner Sebastian Stober
Otto-von-Guericke University, Magdeburg, Germany

(ofner,stober)@ovgu.de

Joint audio and brain signal information retrieval with bio-plausible neural networks?



Simultaneous audio and brain signal information retrieval: Unsupervised prediction of audio signals with a hierarchical predictive coding network allows to derive event boundaries from prediction errors. These events can be used for multimodal information retrieval in audio and EEG signal. Here we explore information retrieval only from the auditory domain, i.e. without processing EEG in the network.

Deep predictive coding on multiple time scales

Inspired by the hierarchical organisation of cortical areas in the brain we suggest a deep hierarchical generative model that predicts audio spectrograms from stochastic hidden states. Layers higher in the hierarchy encode expectations about the hidden state parameters of lower layers. The network propagates prediction errors between layers internally, allowing to efficiently merge local predictions with top-down information from larger temporal context.

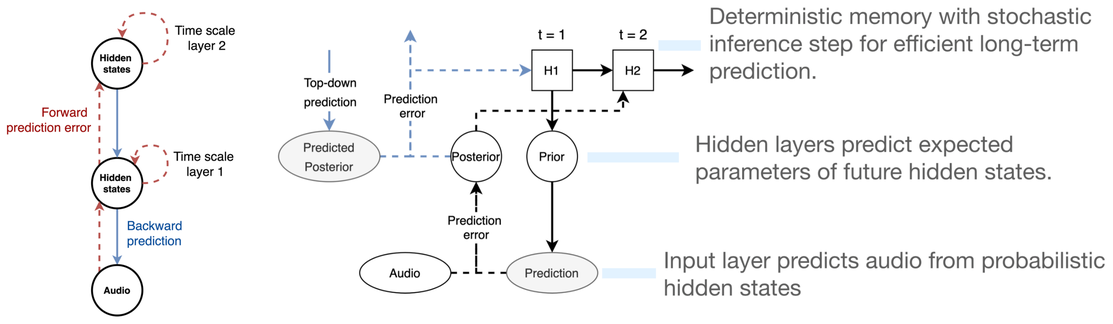


Figure 1: Predictive coding network for audio representation learning.

Forward connections in the model (red) propagate prediction errors while backward connects (blue) carry predictions about lower layer activity. Hidden layers operate a slower pace than layers closer to the data.

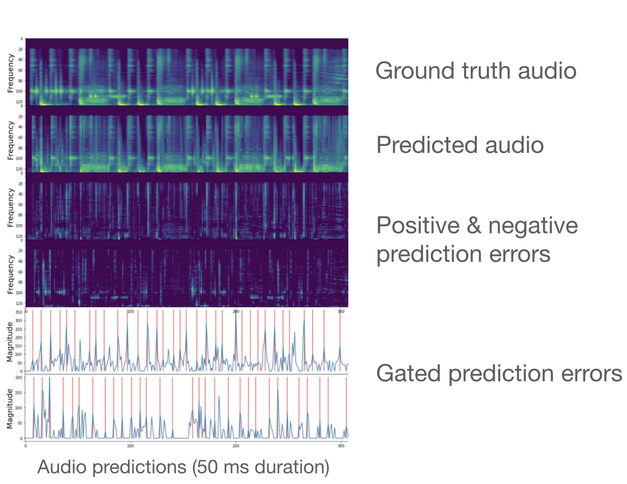


Figure 2: Deriving event onsets from gated prediction errors.

Using the model for short-term autoregressive prediction of audio spectrograms allows to visualise positive and negative prediction errors, corresponding to over and underestimated loudness at a specific frequency. These fluctuations in prediction errors were then gated with a fixed threshold, providing temporal onset markers for downstream tasks.

Dataset and experiments

The network was trained on the ”small” partition of the Free Music Archive (FMA) dataset, featuring 8000 songs with 30 seconds duration. Testing the networks was done on audio signals from a EEG dataset, without retraining. The Naturalistic Music EEG Dataset—Tempo (NMED-T) features EEG recordings from 10 commercially available music pieces and 20 healthy subjects, spanning 55 to 150 BPM in various genres.

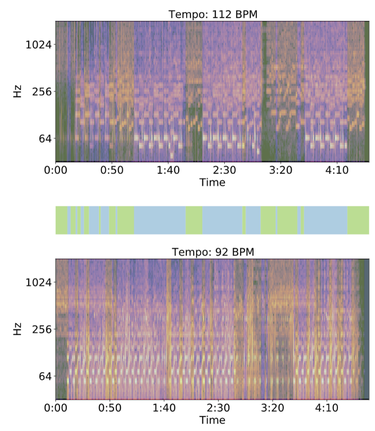
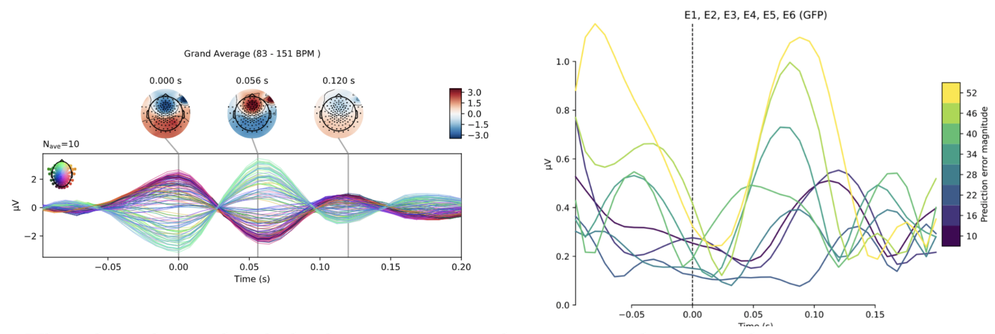
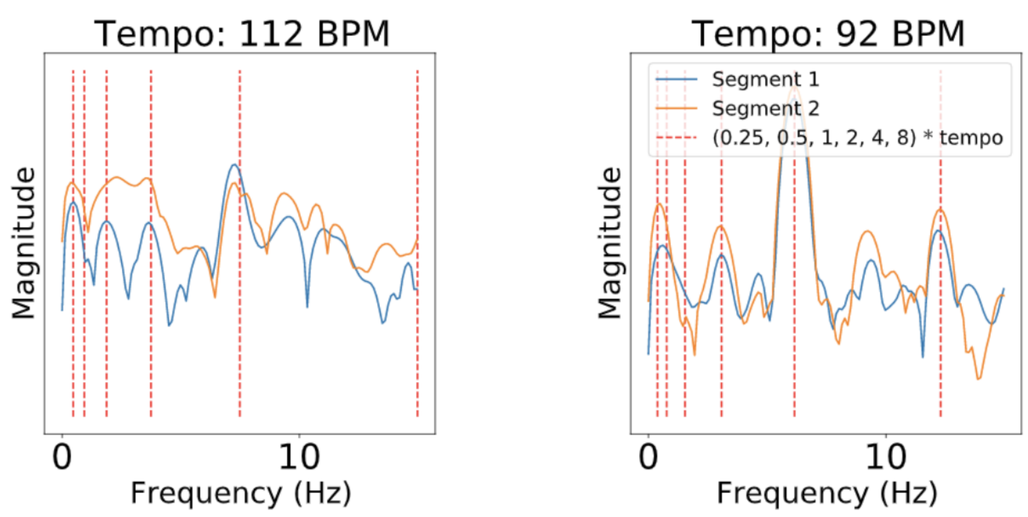


Figure 3: Song-level segmentation with multi-step predictions.

Song-level temporal segmentation generated using multi-step predictions in the hidden layers. The resulting binary segmentation reflect shifts from positive to negative prediction error and vice versa. The segmented parts visually match global changes in song structure.



The local and global event markers can be used to aggregate temporally aligned evoked responses in EEG. The evoked brain response is proportional to the model prediction error for large error values. Evoked responses at smaller error peaks are more dependent on the song tempo.



Low frequency components of the EEG signal show characteristic peaks at multiples of the song tempo. Magnitude shifts of these peaks between the derived segments indicate changes in rhythmic processing in the brain.

Conclusion and future work

These results indicate that hierarchical predictive coding allows to retrieve short and long-term musical event structures while also providing useful markers for simultaneous brain signal information retrieval. However, more elaborate prediction mechanisms and quantitative benchmarks are necessary. Future work could also include multi-modal neural processing, e.g. by processing audio and brain signal in parallel or annotating audio with neural responses captured in EEG.

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG