Applications of Non-linear Component Extraction to Spectrogram Representations of Auditory Data

Jörg Bornschein and Jörg Lücke
Frankfurt Institute for Advanced Studies, Goethe-University, Germany

Bernstein Focus: Neurotechnology Frankfurt

FIAS Frankfurt Institute for Advanced Studies

Abstract

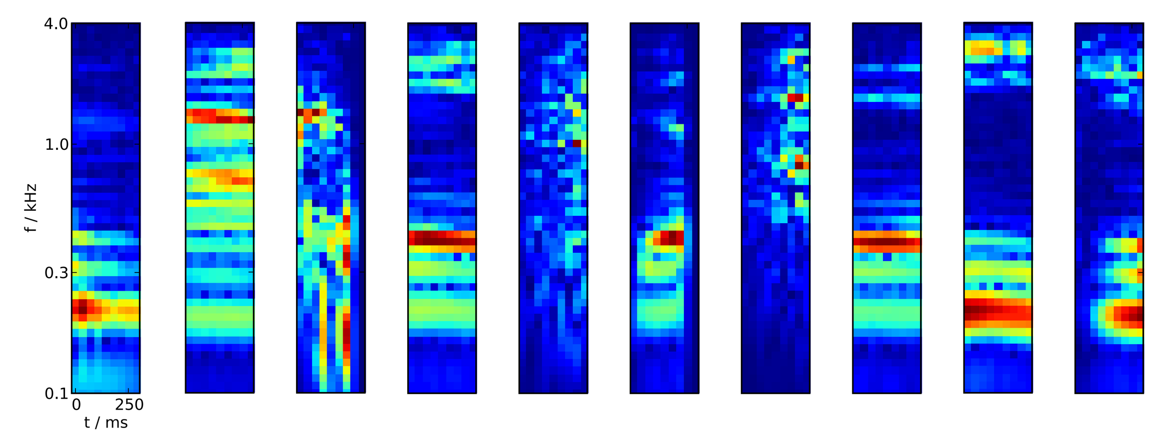
We study the extraction of non-linear components from spectrogram data of natural sounds. Such data is processed by the primary auditory cortex after preprocessing by the cochlea. The elementary components of log-spectrograms combine according to a point-wise maximum [1]. We therefore apply with Maximal Causes Analysis (MCA) a component extraction algorithm that assumes a maximum non-linearity in the place where most other algorithms assume linear superposition.

Conclusion

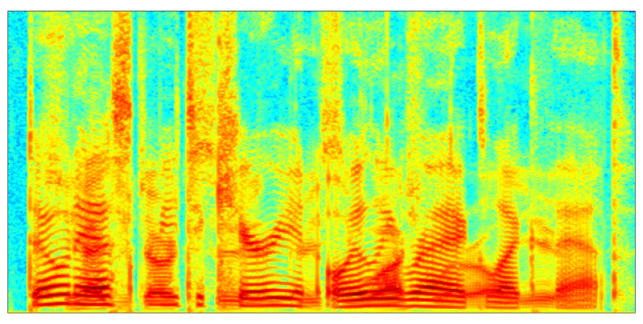
We find that Maximal Causes Analysis successfully extracts the elementary components of sound spectrograms. This suggests that it might be better suited to model component extraction in auditory cortex than linear approaches such as ICA or sparse coding.

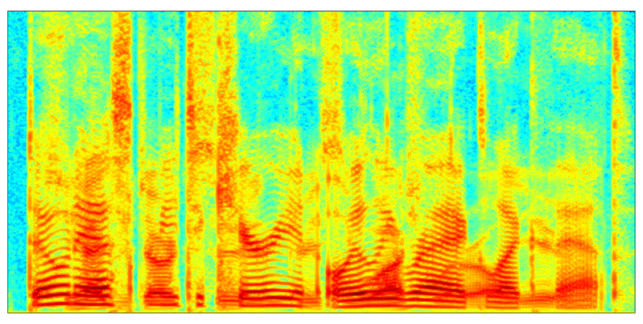
Log-Spectrogram Representation of Audio Data

The data processed by the human auditory system just after the cochlea is closely aligned with the log-spectrogram representation of audio signals:



The linear combination of waveforms does not translate into a linear superposition of their respective log-spectrogram representations. Therefore, when operating on log-spectrogram data, a non-linear component extraction algorithm should to be applied.

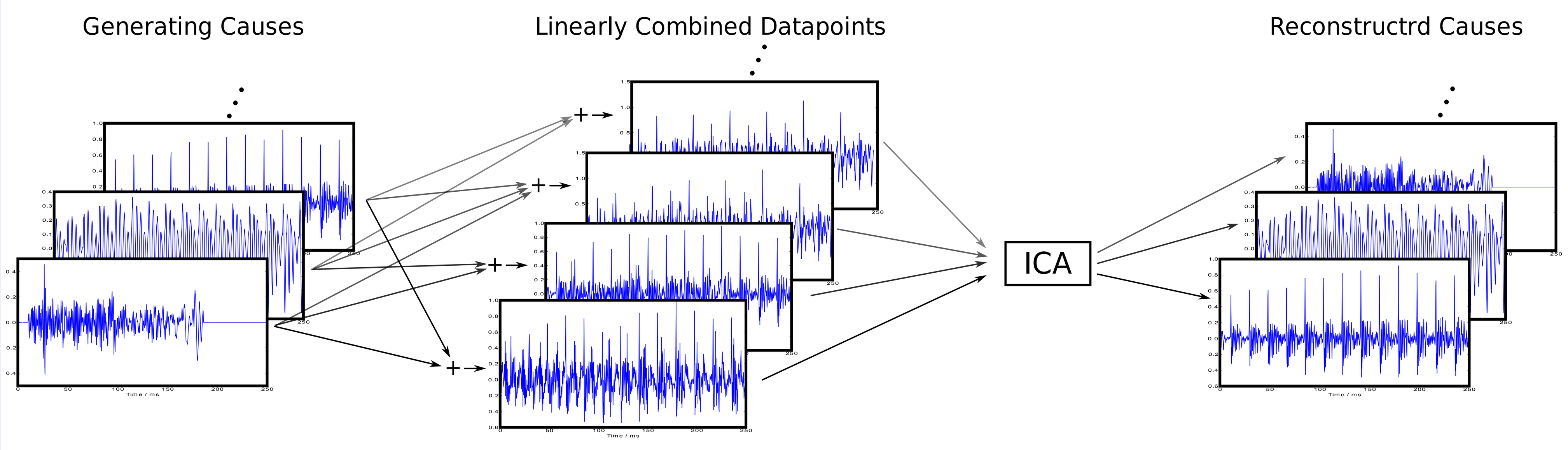

Log-spectrogram of linear mixture


Max-combination of individual log-spectrograms

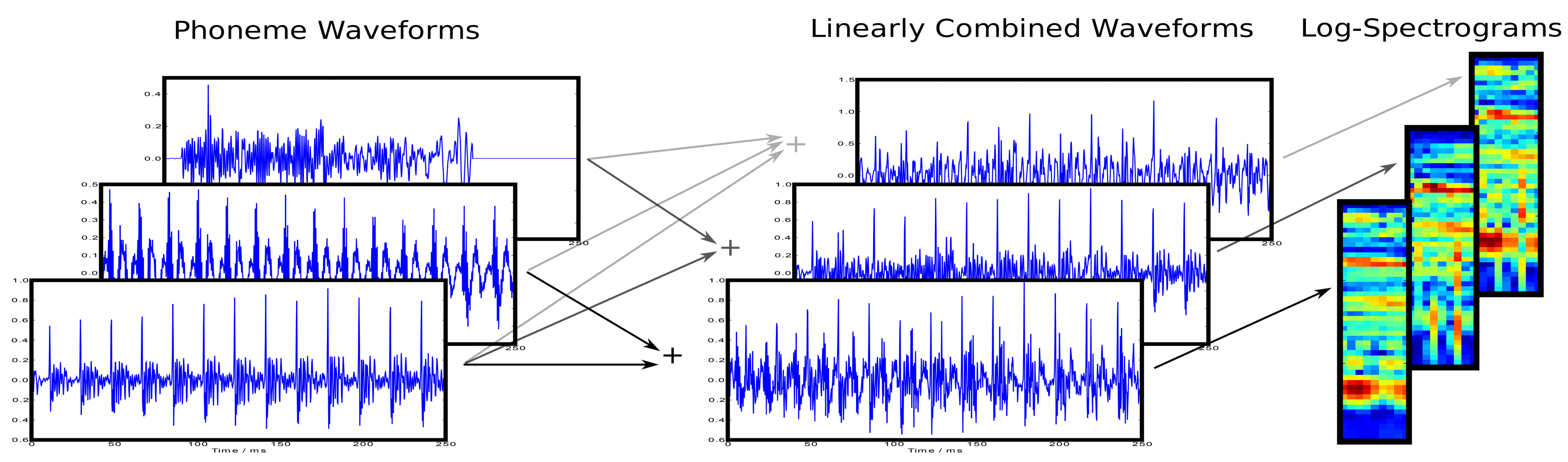
(Log-spectrograms taken from [1])

ICA: Component Extraction for Linearly Combined Data

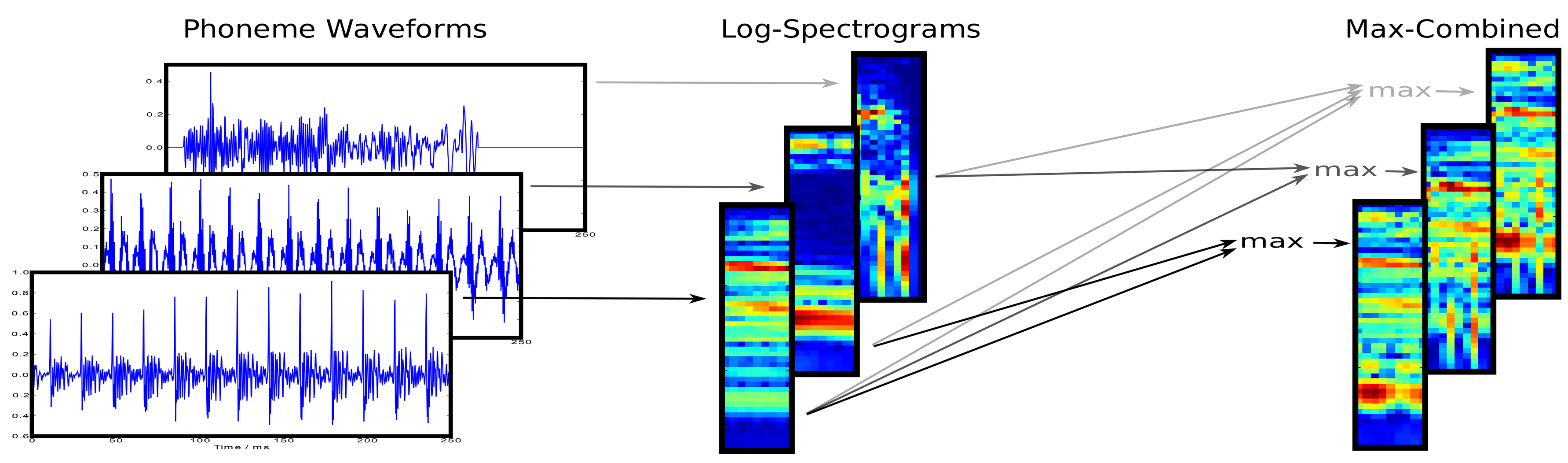
Independent Component Analysis (ICA) is a state-of-the-art technique to decompose data-points into generating causes. ICA assumes that the observed data can be explained by linearly combining generating fields [2].



Log-Max Approximation



The linear superposition of waveform data can be approximated by a point-wise maximum in the log-spectrogram domain [1]:



MCA: Maximal Causes Analysis

Maximal Causes Analysis (MCA; [3]) is a learning algorithm based on generative models. It assumes that data-points can be explained by maximum-combining generative fields. Using deterministic annealing, the algorithm searches for the maximum likelihood solution. Training such models is computationally expensive but can be made feasible by an approximation scheme. Here we use an improved training algorithm based on the preselection of the most probable causes per data-point (see [3] for a similar approach).

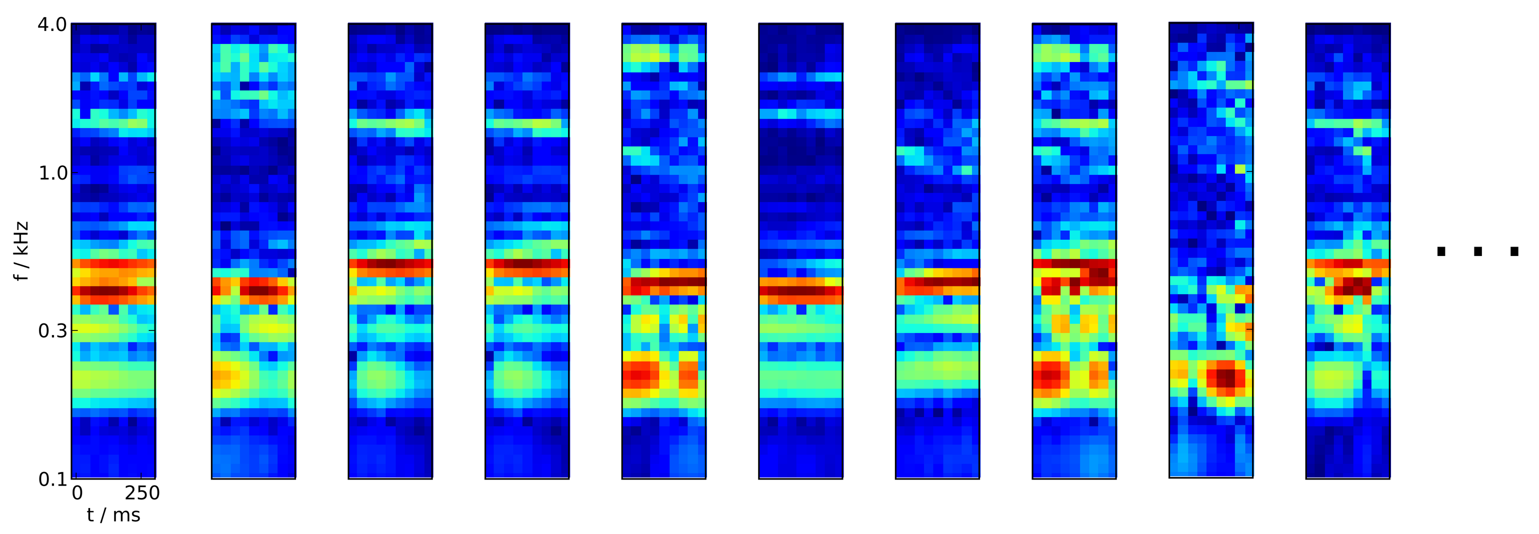
$$y_d^{(n)} = \max_h\{s_h^{(n)} W_{dh}\} + \eta_d$$

Observed data-points: $\vec{y}^{(n)} \in \mathbb{R}^D$ Receptive-/generative fields: $W \in \mathbb{R}^{H\times D}$
Binary hidden causes: $\vec{s}^{(n)} \in \{0,1\}^H$ Gaussian noise: η_d

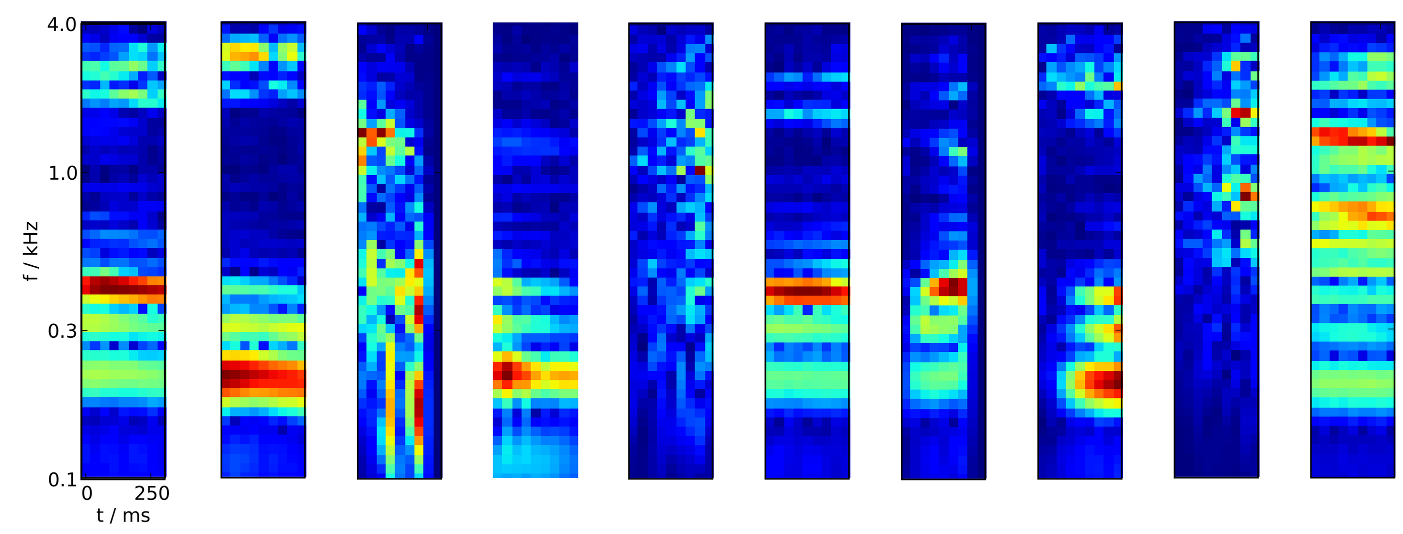
Results

We generated the training data by linearly combining phoneme waveforms and applied MCA to the log-spectrogram representation of these combined signals. The algorithm successfully extracts the log-spectrogram representation of the individual, generating phonemes.

Selection of training data-points:



Generating fields after learning:



References

[1] Automatic Speech Processing by Inference in Generative Models, S. T. Roweis (2004)
(Roweis quotes Moore (1983) as the first pointing out the log-max approximation).
[2] Independent component analysis, a new concept? P. Comon, Signal Processing, 36(3):287–314 (1994).
[3] Maximal Causes for Non-linear Component Extraction, J. Lücke and M. Sahani (2008).

This project was supported by the German Federal Ministry of Education and Research (BMBF) within the "Bernstein Focus: Neurotechnology Frankfurt" through research grant 01GQ0840.