Dynamic Estimation of Auditory Response Function with Confidence Intervals

Sahar Akram1,2, Jonathan Z. Simon1,2,3, Behtash Babadi1,2

1Department of Electrical and Computer Engineering, 2 Institute for Systems Research, 3 Department of Biology
University of Maryland, College Park, MD

**Overview**

**Cocktail Party Effect**
The ability to identify and track a target speaker amid a cacophony of acoustic interference.

**Temporal Response Function (TRF)**
A sparse kernel relating the auditory neural response to the envelope of the speech stream.

**Attention Decoding via Dynamic TRF Estimation**

**Objectives**
1) Dynamic TRF estimation with high temporal resolution.
2) Dynamic estimation of confidence intervals.
3) Tracking listeners’ auditory attention, using M100\textsubscript{TRF} amplitudes.

**References**


**Dynamic TRF Estimation**
The instantaneous filtering error at time i defined as: $\epsilon_t = y_t - \phi(X_t)^T \hat{a}_t$, where $y_t = (y_t, \ldots, y_t)$, and $X_t = \mathcal{X}(t, t - M + 1, \ldots, t - 2, t - 1, t)^T$. The cost function: $f(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-M}) = \sum_{j=0}^{M-1} \epsilon_t^2(1 - \lambda)^j$, where $0 < \lambda < 1$.

$$\min_{\hat{a}_t} f(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-M})$$

**The Proposed Algorithm**

**Dynamic Computation of Confidence Intervals**

> Removing the bias of the $\hat{a}$-regularized estimate:

$$\hat{a}_t^{\text{biased}} = \hat{a}_t^{\text{unbiased}} - \hat{a}_t R^{-1} \hat{a}_t^T \hat{a}_t$$

where $\hat{a}_t^{\text{unbiased}}$ is relaxed inverse approx. of $E(\hat{a}_t) = R^{-1} E(\hat{a}_t) R$, computed via node-wise regression:

A 2nd SPARLS implemented in parallel with the TRF estimation to solve:

$$\hat{a}_t = \arg\min_{\hat{a}_t} \| D^{1/2}(\hat{a}_t) E(\hat{a}_t) - E(\hat{a}_t) E(\hat{a}_t)^T \hat{a}_t + \nu \|$$

where $E(j)$ is the $j$th column of $E$, and $E_{(j)}$ the submatrix of $E$ without the $j$th column.

Then: $\hat{a}_t = \{ \hat{a}_t(1), \ldots, \hat{a}_t(K) \}_{K \neq j}$,

and $\hat{a}_t^{\text{biased}} = \{ \hat{a}_t^{\text{unbiased}}(1), \ldots, \hat{a}_t^{\text{unbiased}}(K) \}_{K \neq j}$;

Set: $\hat{a}_t = \text{diag} \{ \hat{a}_t(1), \ldots, \hat{a}_t(K) \}$

> Confidence intervals at significance level of $\epsilon$:

$$\pm \epsilon^{-1} (1 - \epsilon/2) \text{c.d.f.} \left( \Phi \left( \frac{\mathcal{X}(X_{(i)}) \cdot \hat{a}_t}{\hat{a}_t^T \hat{a}_t} \right) \right)$$

where $\Phi(.)$ is c.d.f. of $\chi^2(0.5)$.

**Amplitude Modulations**

M50\textsubscript{TRF}: Not significantly modulated by attention.

M100\textsubscript{TRF}: Strongly modulated by attention.

**Attention Decoding**

Constant-attention: Listeners attended to a target speaker during 60 s trials.

Attention-switch: Listeners switched their attention from one speaker to the other, after 30 s.

**Acknowledgement**

The authors thank Alessandro Prossaco for providing part of the data for this study.