Dynamic Estimation of Auditory Response Function with Confidence Intervals

Sahar Akram[1,2], Jonathan Z. Simon[1,2,3], Behtash Babadi[1,2]

[1]Department of Electrical and Computer Engineering, [2] Institute for Systems Research, [3] Department of Biology

University of Maryland, College Park, MD

Neuroscience 2015

Overview

Cocktail Party Effect

The ability to identify and track a target speaker amid a cacophony of acoustic interference[1].

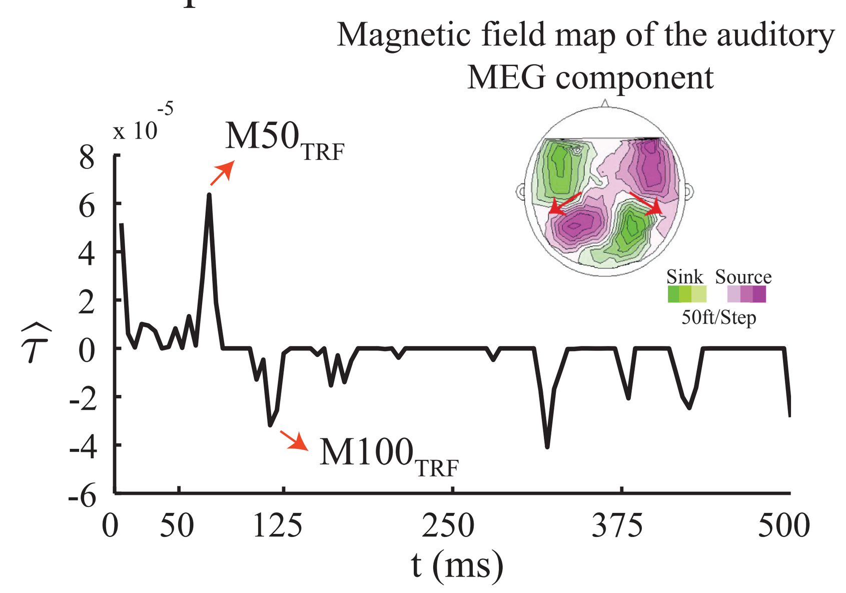
Temporal Response Function (TRF)

A sparse kernel relating the auditory neural response to the envelope of the speech stream.

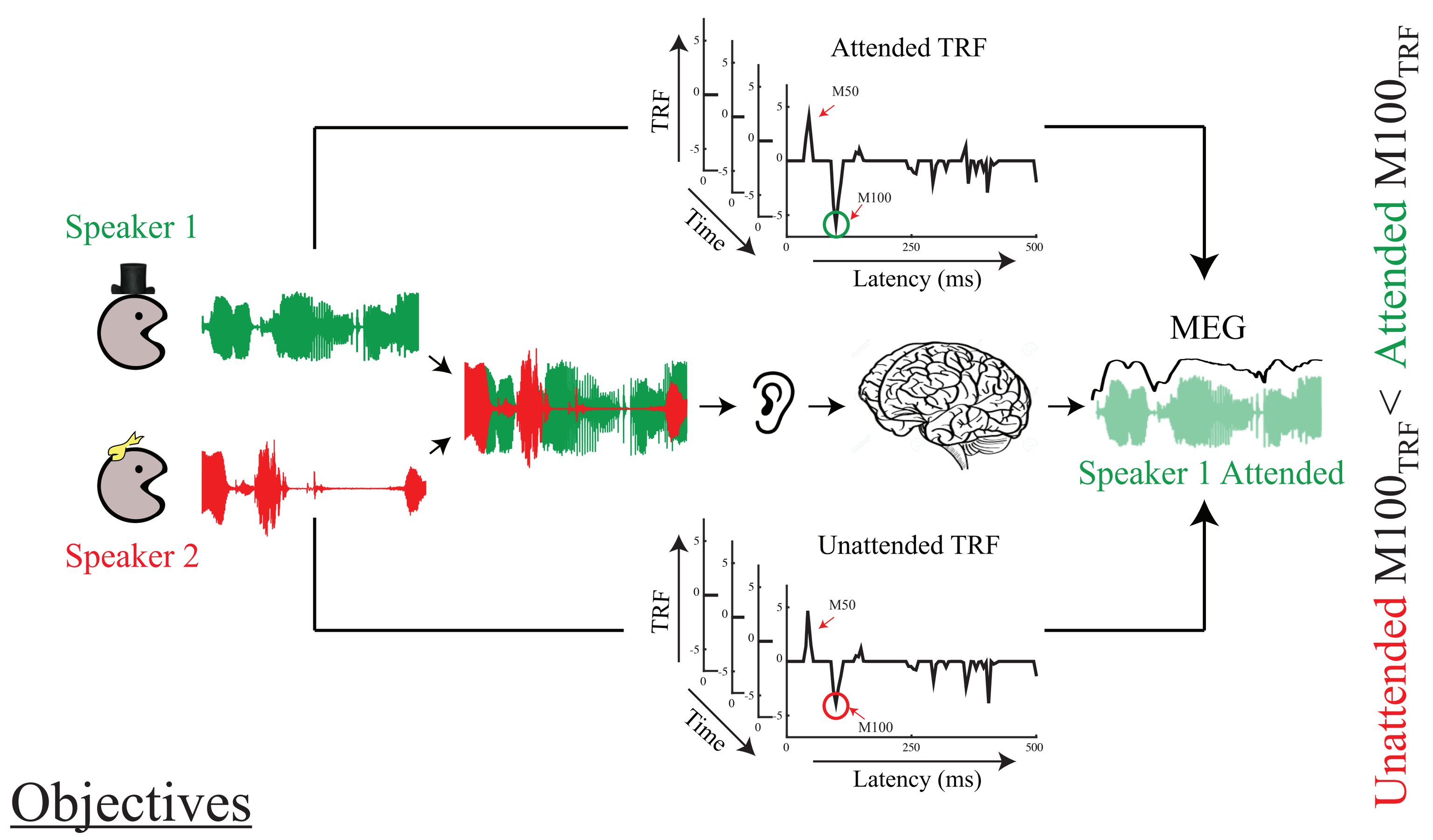
Auditory MEG   Speech Envelope   Noise

$$y(t) = \tau(t) * e(t) + \nu(t)$$

Temporal Response Function



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 $M100_{TRF}$ amplitudes[2,3].

References

[1] Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. The Journal of the acoustical society of America, 25(5), 975-979.

[2] Chait, Maria, Jonathan Z. Simon, and David Poeppel. "Auditory M50 and M100 responses to broadband noise: functional implications." Neuroreport 15.16 (2004): 2455-2458.

[3] Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. Proceedings of the National Academy of Sciences, 109(29), 11854-11859.

[4] Babadi, Behtash, Nicholas Kalouptsidis, and Vahid Tarokh. "SPARLS: The sparse RLS algorithm." Signal Processing, IEEE Transactions on 58.8 (2010): 4013-4025.

[5] Van de Geer, Sara, et al. "On asymptotically optimal confidence regions and tests for high-dimensional models." The Annals of Statistics 42.3 (2014): 1166-1202.

The Proposed Algorithm

Dynamic TRF Estimation

The instantaneous filtering error at time i defined as: $err(i) := y(i) - \boldsymbol{\tau}^T(i)\boldsymbol{e}(i)$, where $\boldsymbol{e}(i) = (e(i), e(i-1), \cdots, e(i-M+1))^T$ and $\boldsymbol{\tau}(i) = (\tau_0(i), \tau_1(i), \cdots, \tau_{M-1}(i))^T$.

Cost function: $f(err(1), err(2), ..., err(n)) := \sum_{i=1}^{n} \lambda^{n-i} \mid err(i) \mid^2, \ 0 < \lambda < 1$

$$\min_{\boldsymbol{\tau}(n)} \ f(err(1), err(2), ..., err(n))$$

$\mathcal{L}_1$ -Regularized Maximum Likelihood Estimate:

$$\hat{\boldsymbol{\tau}}(n) = \underset{\boldsymbol{\tau}(n)}{\operatorname{argmin}} \frac{1}{2\sigma^2} \left\| \mathbf{D}^{1/2}(n)\mathbf{y}(n) - \mathbf{D}^{1/2}(n)\mathbf{E}(n)\boldsymbol{\tau}(n) \right\|_2^2 + \eta \|\boldsymbol{\tau}(n)\|_1$$

Where $D(n) := diag(\lambda^{n-1}, \lambda^{n-2}, ..., 1)$, $\boldsymbol{y}(n) = (y(1), y(2), \cdots, y(n))$
$\boldsymbol{E}(n) = (\boldsymbol{e}(1), \boldsymbol{e}(2), \cdots, \boldsymbol{e}(n))^T$, $\nu_t \sim \mathcal{N}(0, \sigma^2)$, and η : regularization parameter.

An Efficient Recursive Expectation-Maximization Approach (SPARLS[4]):

E-step: $\boldsymbol{Q}(\boldsymbol{\tau}|\hat{\boldsymbol{\tau}}(n)) := -\frac{1}{2\alpha^2} \left\| \boldsymbol{r}^{(\ell)} - \boldsymbol{\tau} \right\|_2^2 - \eta \|\boldsymbol{\tau}\|_1$,
where $\boldsymbol{r}^{(\ell)}(n) := (\boldsymbol{I} - \frac{\alpha^2}{\sigma^2}\boldsymbol{E}^T(n)\boldsymbol{D}(n)\boldsymbol{E}(n))\hat{\tau}^{(\ell)}(n) + \frac{\alpha^2}{\sigma^2}\boldsymbol{E}^T(n)\boldsymbol{D}(n)\boldsymbol{y}(n)$

M-step: $\hat{\boldsymbol{\tau}}^{(\ell+1)}(n) := \underset{\boldsymbol{\tau}}{\operatorname{argmax}} \, \boldsymbol{Q}(\boldsymbol{\tau}|\hat{\boldsymbol{\tau}}(n)) = \mathcal{S}(\boldsymbol{r}^{(\ell)})$
where $\mathcal{S}(.) : \mathbb{R}^M \to \mathbb{R}^M$ is an element-wise soft thresholding function:
$(\mathcal{S}(w))_i := sgn(\omega_i)(max(|\omega_i| - \eta\alpha^2), 0)$.

Dynamic Computation of Confidence Intervals

> Removing the bias of the ℓ_1-regularized estimate[5]:

$$\hat{\boldsymbol{\tau}}_{biased} = \hat{\boldsymbol{\tau}}_{unbiased} - \hat{\boldsymbol{\Theta}}(n)\boldsymbol{E}^T(n)\boldsymbol{D}(n)\boldsymbol{\nu}(n)$$

where $\hat{\boldsymbol{\Theta}}(n) :=$ relaxed inverse approx. of $\hat{\boldsymbol{\Sigma}}(n) = \boldsymbol{E}^T(n)\boldsymbol{D}^2(n)\boldsymbol{E}$, computed via node-wise regression:

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

$$\hat{\boldsymbol{\gamma}}_j(n) := \underset{\boldsymbol{\gamma} \in \boldsymbol{R}^{M-1}}{\operatorname{argmin}} (\|\boldsymbol{D}^{1/2}(n)\boldsymbol{E}_j(n) - \boldsymbol{E}_{-j}(n)\boldsymbol{D}^{1/2}(n)\boldsymbol{\gamma}\|_2^2 + \eta\|\boldsymbol{\gamma}\|_1)$$

for $j = 1, \ldots, M$, where $\boldsymbol{E}_j$: the j^{th} column of $\boldsymbol{E}$, and $\boldsymbol{E}_{-j}$: the submatrix of $\boldsymbol{E}$, without the j^{th} column.

Then:

$$\hat{\boldsymbol{C}}(n) := \begin{bmatrix} 1 & -\hat{\gamma}_{1,2}(n) & \ldots & -\hat{\gamma}_{1,M}(n) \\ -\hat{\gamma}_{2,1}(n) & 1 & \ldots & -\hat{\gamma}_{2,M}(n) \\ \vdots & \vdots & \ddots & \vdots \\ -\hat{\gamma}_{M,1}(n) & -\hat{\gamma}_{M,2}(n) & \ldots & 1 \end{bmatrix}, \text{ where } \hat{\boldsymbol{\gamma}}_j(n) = \{\hat{\gamma}_{j,k}(n); k = 1, \cdots, M, k \neq j\},$$

and

$$\hat{\omega}_j^2(n) := \left\| \boldsymbol{D}^{1/2}(n)\boldsymbol{E}_j(n) - \boldsymbol{D}^{1/2}(n)\boldsymbol{E}_{-j}(n)\hat{\boldsymbol{\gamma}}_j(n) \right\|_2^2 + \frac{\eta}{2}\|\hat{\boldsymbol{\gamma}}_j(n)\|_1, \text{ for } j = 1, \ldots, M.$$

Set: $\hat{\boldsymbol{T}}^2(n) := diag(\hat{\omega}_1^2(n), ..., \hat{\omega}_M^2(n))$

Compute: $\hat{\boldsymbol{\Theta}}(n) := \hat{\boldsymbol{T}}^{-2}(n)\hat{\boldsymbol{C}}(n)$

> Confidence intervals at significance level of ϵ:

$$\pm\Phi^{-1}(1 - \epsilon/2)\sigma\sqrt{\left(\hat{\boldsymbol{\Theta}}(n)\hat{\boldsymbol{\Sigma}}(n)\hat{\boldsymbol{\Theta}}^T(n)\right)_{(j,j)}}$$

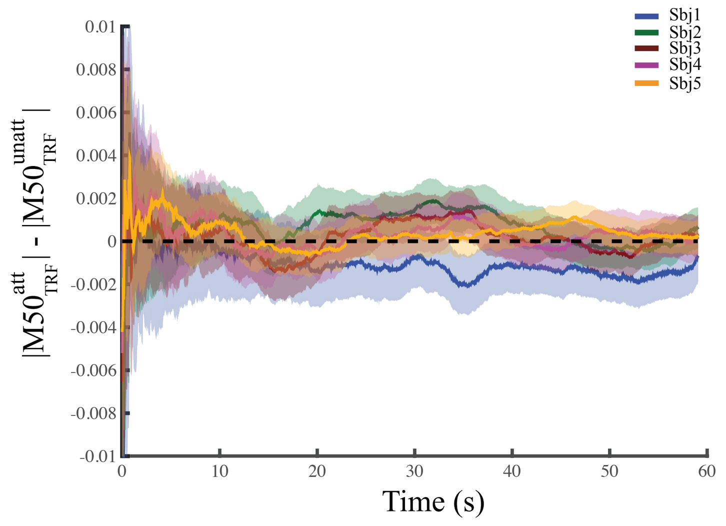
where $\Phi(.) :=$ c.d.f of $\mathcal{N}(0, 1)$.

Attention Decoding

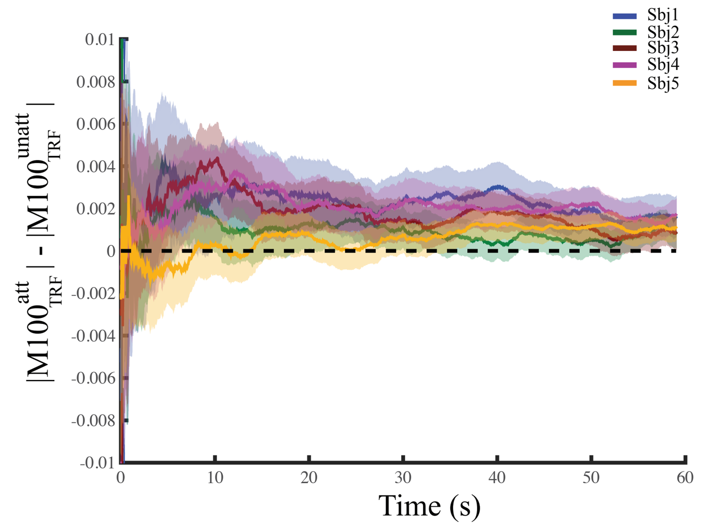
Amplitude Modulations

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

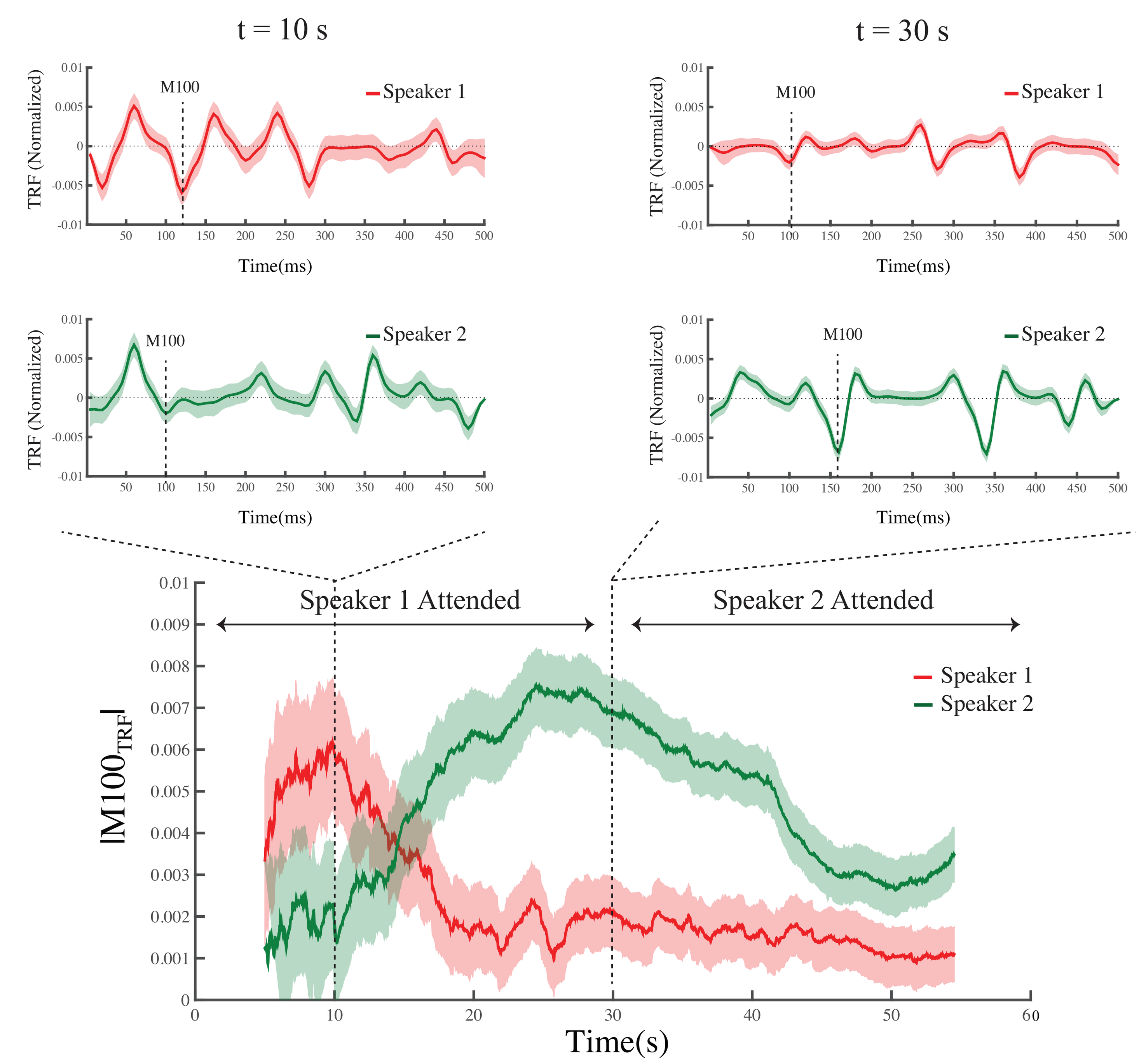
$M50_{TRF}$: Not significantly modulated by attention



$M100_{TRF}$: Strongly modulated by attention



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



Auditory attention tracked via comparing $M100_{TRF}$ amplitudes, estimated from the attended and unattended speech signals.

Acknowledgement

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