AN ALTERNATIVE APPROACH FOR AUDITORY ATTENTION TRACKING USING SINGLE-TRIAL EEG

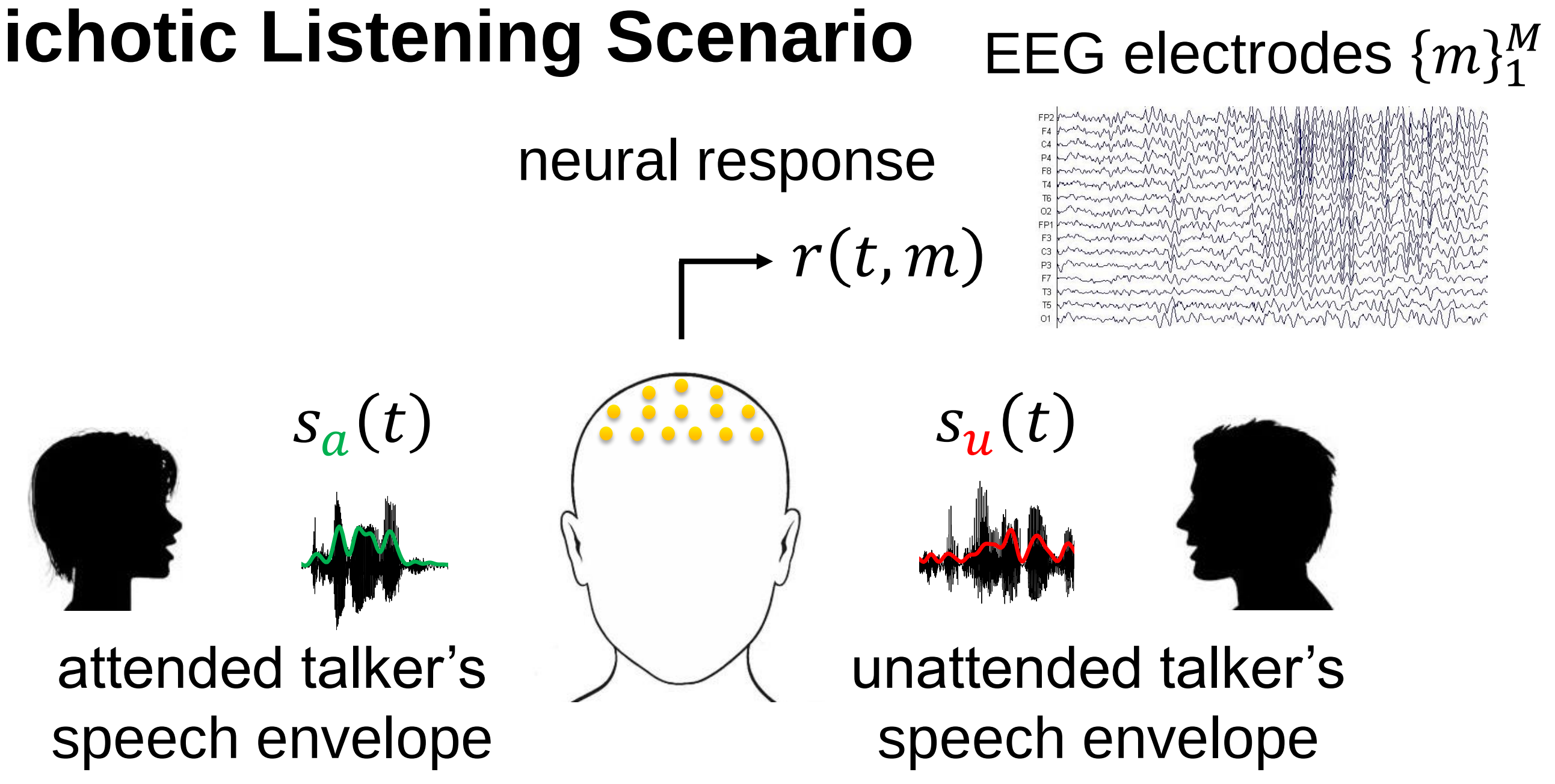
Bradley Ekin and Les Atlas, Department of Electrical Engineering
Majid Mirbagheri and Adrian KC Lee, Institute for Learning & Brain Sciences
UNIVERSITY of WASHINGTON

Overview

- **Goal**: Given two simultaneous speech sources, detect which is being “attended” to and which is being “unattended” to, using the listener’s electroencephalography (EEG) data.
- **Innovation**: Extensions to conventional methods for auditory attention detection are proposed:
  (1) selective channel deconvolution,
  (2) maximally correlated multimodal projections,
  (3) balanced correlation decoders.
- **Importance:** These are tools aimed to improve the understanding of how humans solve “the cocktail party problem.” Applications include, for example, attention driven acoustic beamforming for hearing prostheses.

Paradigm

Dichotic Listening Scenario



Experiment: An EEG subject is presented with 36 one-minute duration segments of competing spoken stories, one in each ear, and attends to only one.

Auditory Attention Decoding Goal

- Using training data (35 of 36 segments in a leave-one-out cross validation approach), learn linear filters to reconstruct $s_a(t)$ or $s_u(t)$ from $r(t,m)$.
- Using test data (the remaining segment), predict $\hat{s}_a(t)$ and $\hat{s}_u(t)$ from $r(t,m)$ and compare to ground truth $s_a(t)$ and $s_u(t)$.

Assumed Linear System

neural response
additive noise

$$r(t,m) = \int_{\tau=0}^{\tau_{max}} \left(a(\tau,m)s_a(t-\tau) + u(\tau,m)s_u(t-\tau)\right) + \eta(t,m)$$

attended and unattended channels

Discrete Time Definitions

$$\boldsymbol{r} = \begin{bmatrix} r(t+\tau_{max},1) \\ \vdots \\ r(t,1) \\ r(t+\tau_{max},2) \\ \vdots \\ r(t,M) \end{bmatrix} \qquad \boldsymbol{s}_a = \begin{bmatrix} s_a(t+2\tau_{max}) \\ \vdots \\ s_a(t+1) \\ s_a(t) \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} a(0,1) & \cdots & a(\tau_{max},1) & 0 & \cdots & 0 \\ & \ddots & & & \ddots & \\ 0 & \cdots & 0 & a(0,1) & \cdots & a(\tau_{max},1) \\ a(0,2) & \cdots & a(\tau_{max},2) & 0 & \cdots & 0 \\ & \ddots & & & \ddots & \\ 0 & \cdots & 0 & a(0,M) & \cdots & a(\tau_{max},M) \end{bmatrix}$$

Discrete time is assumed for both t and τ. $\boldsymbol{\eta}$, $\boldsymbol{s}_u$, and $\mathbf{U}$ are analogous to the equations above, respectively.

Compact model: $\boldsymbol{r} = \mathbf{A}\boldsymbol{s}_a + \mathbf{U}\boldsymbol{s}_u + \boldsymbol{\eta}$

Baseline Method: Stimulus Reconstruction

MMSE: *Minimum-Mean Square Error*
Learn a reconstruction filter, $g(\tau,m)$, to linearly combine spatiotemporal EEG observations using MMSE criteria.

$$\boldsymbol{g}_{a-MMSE} = \underset{\boldsymbol{g}}{\operatorname{argmin}}\, \mathbb{E}\{|\boldsymbol{g}^T\boldsymbol{r} - s_a(t)|^2\}$$

(1) Selective Channel Deconvolution

MVDR: *Minimum Variance Distortionless Response*
Reconstruct “attended” stimulus while minimizing any presence of deconvolved “unattended” stimulus & noise.

$$\boldsymbol{g}_{a-MVDR} = \underset{\boldsymbol{g}}{\operatorname{argmin}}\, \mathbb{E}\{|\boldsymbol{g}^T(\mathbf{U}\boldsymbol{s}_u + \boldsymbol{\eta})|^2\}$$
$$\text{s.t.:} \quad \boldsymbol{g}^T\mathbf{A} = [0,\dots,0,1]$$

(2) Multimodal Projections

Multimodal (EEG & Speech) Data Structure

$$\boldsymbol{r}_c = \begin{bmatrix} r(t,1) \\ \vdots \\ r(t,M) \end{bmatrix} \quad \boldsymbol{s}_{a,c} = \begin{bmatrix} s_a(t+\tau_{max}) \\ \vdots \\ s_a(t) \end{bmatrix} \quad \boldsymbol{s}_{u,c} = \begin{bmatrix} s_u(t+\tau_{max}) \\ \vdots \\ s_u(t) \end{bmatrix}$$

CCA: *Canonical Correlation Analysis*
Bypass direct estimation of $\hat{s}(t)$ to produce maximally correlated projections for both modalities, reducing overall number of feature dimensions by using spatial structure of EEG and temporal structure of stimuli.

$$\boldsymbol{g}_{r_a-CCA}, \boldsymbol{g}_{s_a-CCA} = \underset{\boldsymbol{g}_r,\boldsymbol{g}_s}{\operatorname{argmin}}\, \mathbb{E}\left\{\left|\boldsymbol{g}_r^T\boldsymbol{r}_c - \boldsymbol{g}_s^T\boldsymbol{s}_{a,c}\right|^2\right\}$$
$$\text{s.t.:} \quad \boldsymbol{g}_r^T\mathbb{E}\{\boldsymbol{r}_c\boldsymbol{r}_c^T\}\boldsymbol{g}_r = \boldsymbol{g}_s^T\mathbb{E}\{\boldsymbol{s}_{a,c}\boldsymbol{s}_{a,c}^T\}\boldsymbol{g}_s = 1$$

(3) Balanced Correlation Decoders

BMMSE: *Balanced MMSE*
Optimize the reconstruction filter to the detection statistic by jointly maximizing the correlation & anticorrelation of the reconstruction with the “attended” & “unattended” stimuli, respectively.

$$\boldsymbol{g}_{a-BMMSE} = \underset{\boldsymbol{g}}{\operatorname{argmin}}\, \mathbb{E}\left\{\left|\boldsymbol{g}^T\boldsymbol{r} - \left(s_a(t) - s_u(t)\right)\right|^2\right\}$$

BCCA: *Balanced CCA*
Similar to BMMSE, but maximizing canonical correlation and canonical anticorrelation of multimodal projections.

$$\boldsymbol{g}_{r_a-BCCA}, \boldsymbol{g}_{s_a-BCCA} = \underset{\boldsymbol{g}_r,\boldsymbol{g}_s}{\operatorname{argmin}}\, \mathbb{E}\left\{\left|\boldsymbol{g}_r^T\boldsymbol{r}_c - \boldsymbol{g}_s^T\left(\boldsymbol{s}_{a,c} - \boldsymbol{s}_{u,c}\right)\right|^2\right\}$$
$$\text{s.t.:}\ \boldsymbol{g}_s^T\mathbb{E}\left\{\left(\boldsymbol{s}_{a,c} - \boldsymbol{s}_{u,c}\right)\left(\boldsymbol{s}_{a,c} - \boldsymbol{s}_{u,c}\right)^T\right\}\boldsymbol{g}_s = 1$$
$$\boldsymbol{g}_r^T\mathbb{E}\{\boldsymbol{r}_c\boldsymbol{r}_c^T\}\boldsymbol{g}_r = 1$$

Auditory Attention Detection Statistic

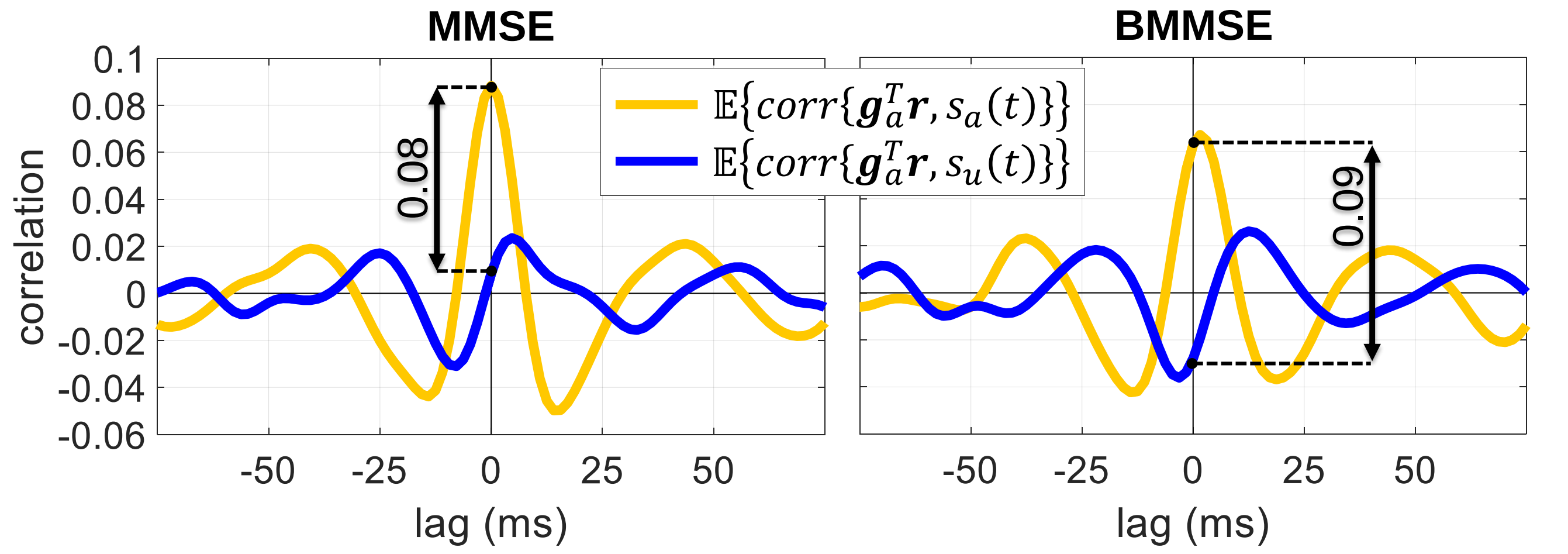
MMSE, MVDR, and BMMSE Decoders

Attended detected if: $corr\{\boldsymbol{g}_a^T\boldsymbol{r}, s_a(t)\} > corr\{\boldsymbol{g}_a^T\boldsymbol{r}, s_u(t)\}$
Unattended detected if: $corr\{\boldsymbol{g}_u^T\boldsymbol{r}, s_u(t)\} > corr\{\boldsymbol{g}_u^T\boldsymbol{r}, s_a(t)\}$

CCA and BCCA Decoders

Attended detected if: $corr\{\boldsymbol{g}_{r_a}^T\boldsymbol{r}_c, \boldsymbol{g}_{s_a}^T\boldsymbol{s}_{a,c}\} > corr\{\boldsymbol{g}_{r_a}^T\boldsymbol{r}_c, \boldsymbol{g}_{s_a}^T\boldsymbol{s}_{u,c}\}$
Unattended detected if: $corr\{\boldsymbol{g}_{r_u}^T\boldsymbol{r}_c, \boldsymbol{g}_{s_u}^T\boldsymbol{s}_{u,c}\} > corr\{\boldsymbol{g}_{r_u}^T\boldsymbol{r}_c, \boldsymbol{g}_{s_u}^T\boldsymbol{s}_{a,c}\}$

Estimation & Detection Accuracy



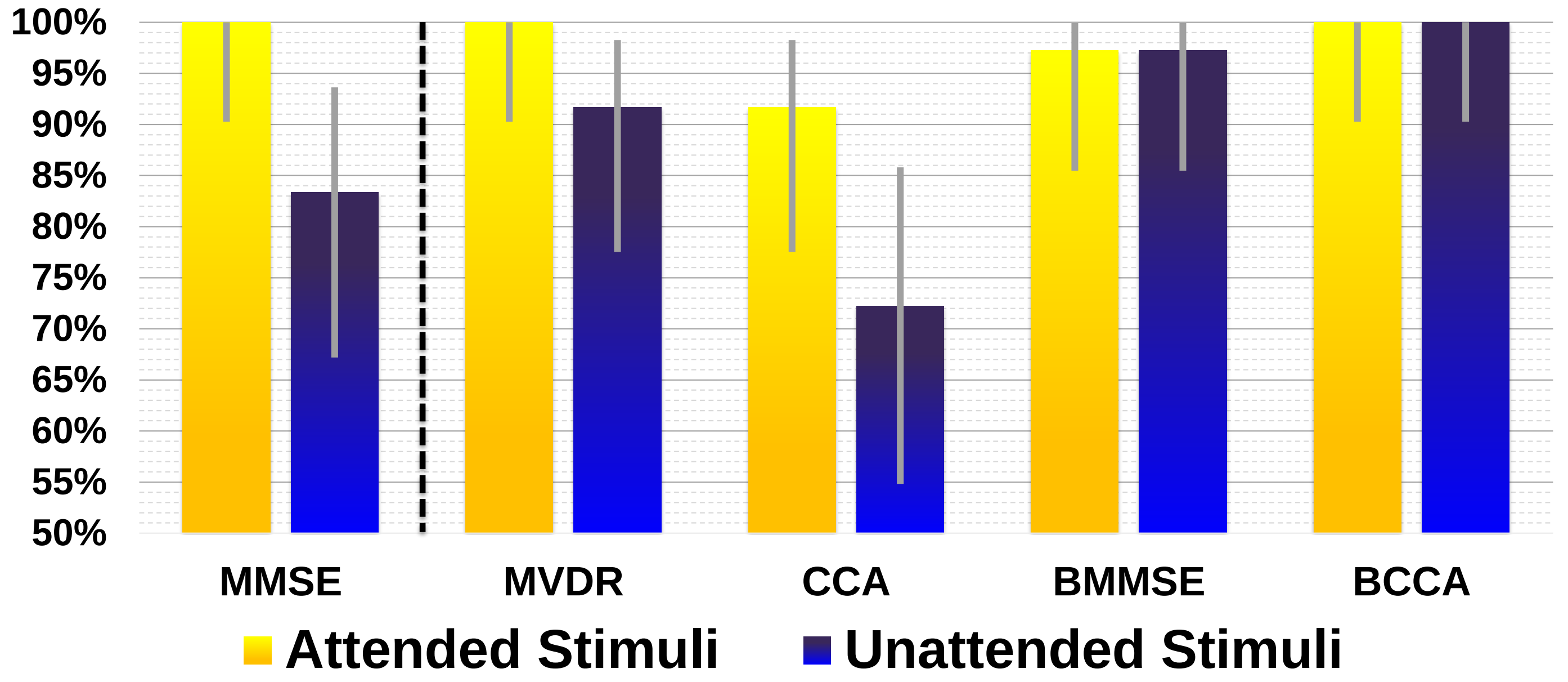
Average cross-correlation of reconstructed stimuli with true stimuli shows how BMMSE takes advantage of selective-attention structures embedded within EEG to maximize correlation separation used in the detection statistic.

	MMSE	MVDR	BMMSE
$\mathbb{E}\{corr\{\boldsymbol{g}_a^T\boldsymbol{r}, s_a(t)\} - corr\{\boldsymbol{g}_a^T\boldsymbol{r}, s_u(t)\}\}$	0.0800	0.0740	0.0908
$\mathbb{E}\{corr\{\boldsymbol{g}_u^T\boldsymbol{r}, s_u(t)\} - corr\{\boldsymbol{g}_u^T\boldsymbol{r}, s_a(t)\}\}$	0.0405	0.0367	0.0908

	CCA	BCCA
$\mathbb{E}\left\{corr\{\boldsymbol{g}_{r_a}^T\boldsymbol{r}_c, \boldsymbol{g}_{s_a}^T\boldsymbol{s}_{a,c}\} - corr\{\boldsymbol{g}_{r_a}^T\boldsymbol{r}_c, \boldsymbol{g}_{s_a}^T\boldsymbol{s}_{u,c}\}\right\}$	0.0643	**0.0967**
$\mathbb{E}\left\{corr\{\boldsymbol{g}_{r_u}^T\boldsymbol{r}_c, \boldsymbol{g}_{s_u}^T\boldsymbol{s}_{u,c}\} - corr\{\boldsymbol{g}_{r_u}^T\boldsymbol{r}_c, \boldsymbol{g}_{s_u}^T\boldsymbol{s}_{a,c}\}\right\}$	0.0336	**0.0967**

Mean difference in detection statistics for decoder models.

Auditory Attention Detection Accuracy



Mean detection accuracies for each decoding method. Error bars based on a binominal distribution ($p = 0.5$, $n = 36, \alpha = 0.05$). Methods evaluated on one subject.

Summary

- We developed 4 auditory attention decoders, each an extension to the traditional MMSE optimization criteria.
- See paper for details on how utilizing channel estimations via **MVDR** improves decoding accuracy.
- Best accuracy for attention detection occurs using **BCCA** by balancing canonical projections using both stimuli.

sites.google.com/a/uw.edu/isdl/publications

{bekin,atlas,mbagheri,akclee}@uw.edu

Funding from Army Research Office grant numbers W911NF-12-1-02770 and W911NF-15-1-0450, and the Office of Naval Research Young Investigator Program award N00014-15-1-2124.