Real-Time Decoding of Auditory Attention from EEG via Bayesian Filtering

Sina Miran, Sahar Akram, Alireza Sheikhattar,
Jonathan Z. Simon, Tao Zhang, Behtash Babadi

UNIVERSITY OF MARYLAND

Starkey Hearing Technologies

Research Supported by

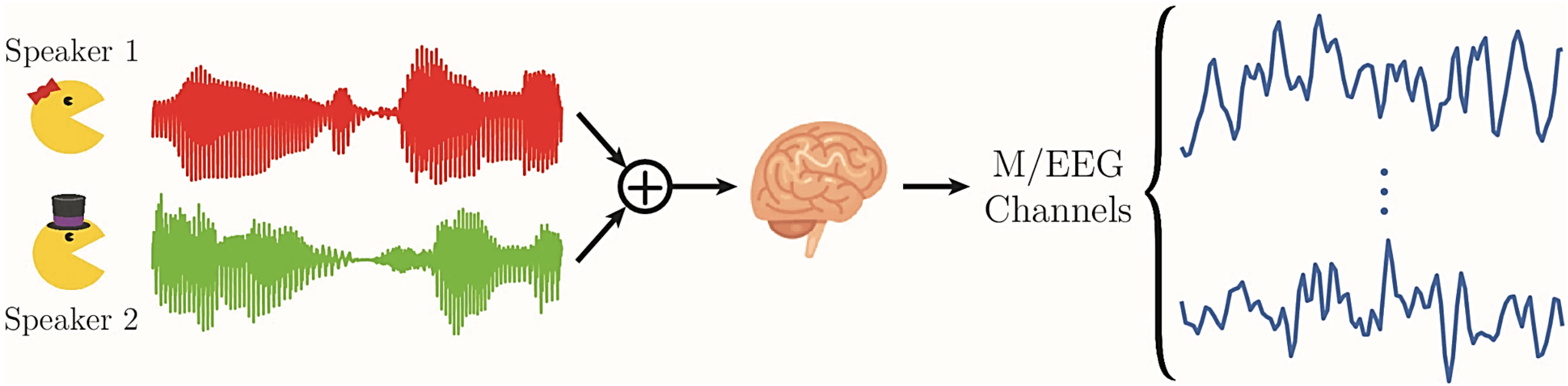
NSF

Starkey

- Introduction
- Existing Methods
- Proposed Method
- Results
  - Simulation
  - EEG Recoding
- Conclusion & Future Work

Cocktail Party Effect: The ability to select a single speaker in an auditory scene, consisting of multiple competing speakers, and maintain attention to that speaker

Simple Attention Decoding Experiment: Subject instructed to attend to speaker 1 or 2



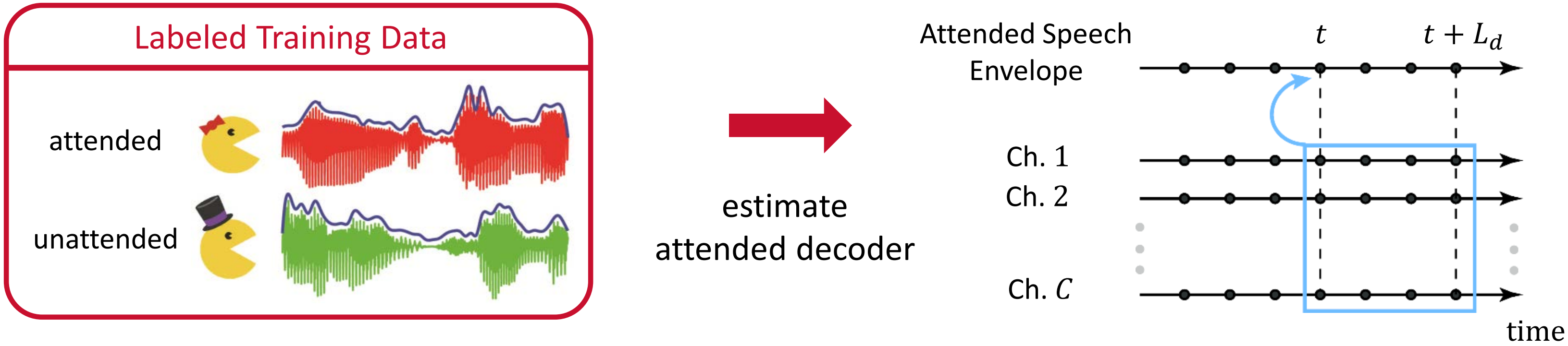
Attention Decoding Algorithm:

- **Input:** clean speech data (speech envelopes), EEG channel recordings
- **Output:** the attended speaker at each time

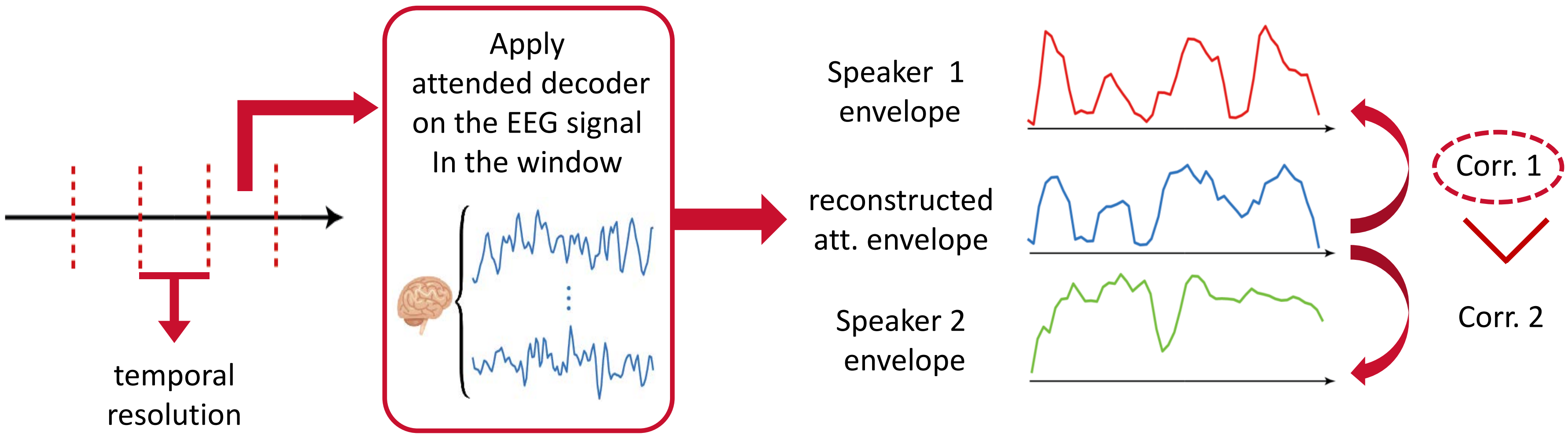
Applications: Brain-Computer Interface (BCI) systems, smart hearing aids

Overview:

1. **Attended Decoder Estimation:** use extensive labeled training data to learn a linear mapping from EEG recordings to the speech envelope of the attended speaker



2. **Classification:** use the attended decoder estimate for classification in test trials

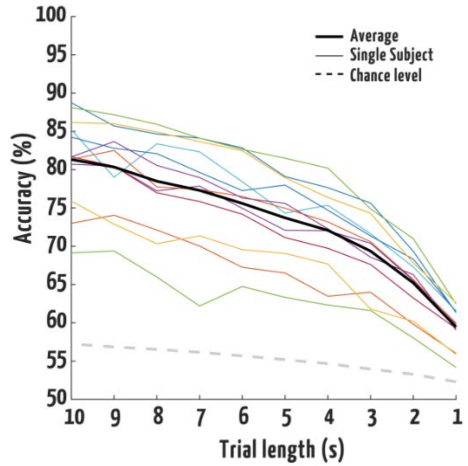


Problems with Existing Methods in Application to *Real-Time* Attention Decoding:

1. **requiring large labeled training datasets to pre-estimate an attended decoder:**
   - might not be available in real-time applications
   - costly recalibration

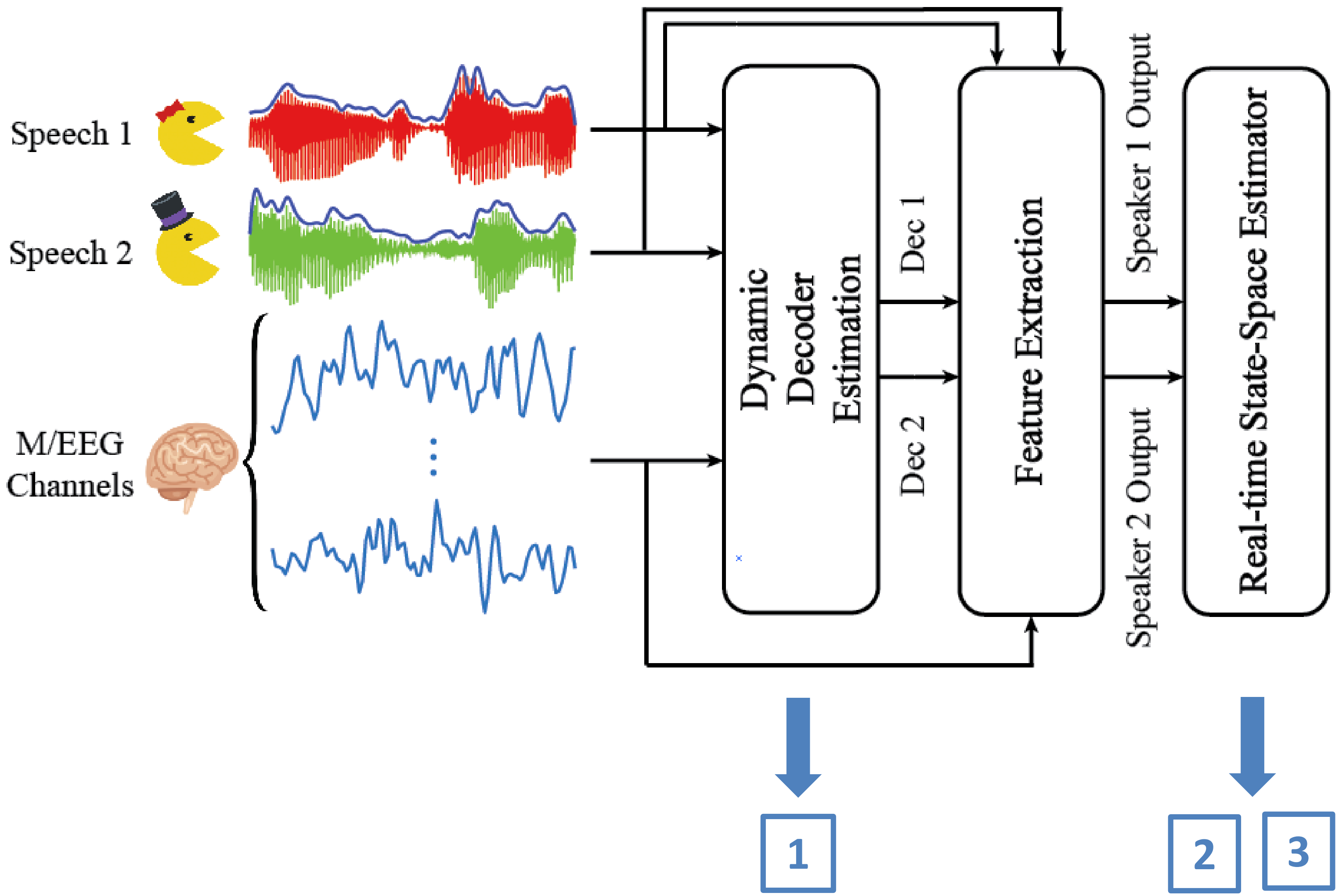
2. **reduced attention decoding accuracy at high temporal resolutions:**
   - reliable attention decoding available for windows in the order of 10s of seconds
   - no way to correct for large stochastic fluctuations at high temporal resolutions

[Zink et al. (2017)]



3. **lack of a *robust probabilistic measure* as the attentional state estimate:**
   - useful for soft-decision making in BCI applications or smart hearing aids

Modular Design of the Proposed Framework:



1. Dynamic Decoder Estimation:

- Break a trial of length T into consecutive non-overlapping intervals of W samples, i.e., $T = KW$
- W determines the temporal resolution in attention decoding, e.g., $W = 0.25 f_s$
- Update the decoder estimate $\widehat{\boldsymbol{\theta}}_k$ for each speaker in each window through the minimization problem below

$$\hat{\boldsymbol{\theta}}_k = \arg\min_{\boldsymbol{\theta}} \sum_{j=1}^{k} \lambda^{k-j} \| \mathbf{y}_j - \mathbf{X}_j \boldsymbol{\theta} \|_2^2 + \gamma \| \boldsymbol{\theta} \|_1 , \quad k = 1, 2, \ldots, K$$

Forgetting Factor:
- Chosen large enough to ensure robustness
- Chosen small enough to capture dynamicity

Target Signal:
Speech envelope of the speaker

Covariate Matrix:
Recorded EEG

ℓ_1 Regularization:
- Chosen by cross-validation
- Results in sparse estimates

2. Feature Extraction:

- Calculate an attention-modulated feature for each speaker in each window as:

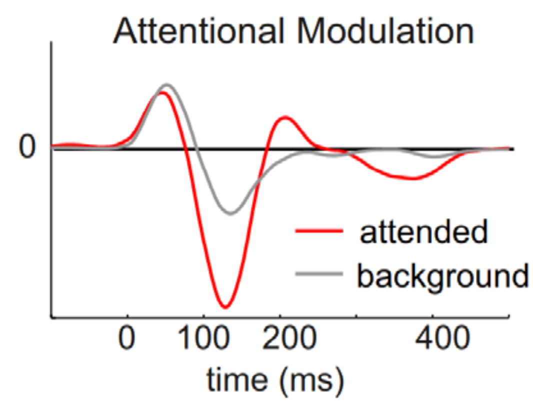
$m_k^{(i)} = f\left(\widehat{\boldsymbol{\theta}}_k^{(i)}, \boldsymbol{y}_k^{(i)}, \boldsymbol{X}_k\right)$, with $m_k^{(i)} > 0$ for $k = 1, \ldots, K$ and $i = 1,2$

- Examples:
  - reconstruction error of the decoder: $m_k^{(i)} = \mathrm{Corr}\left(\boldsymbol{y}_k^{(i)}, \boldsymbol{X}_k \widehat{\boldsymbol{\theta}}_k^{(i)}\right)$

    rationale: attended stimulus is more dominant in neural response
  - ℓ_1 norm of the estimated decoder: $m_k^{(i)} = \left|\widehat{\boldsymbol{\theta}}_k^{(i)}\right|_1$

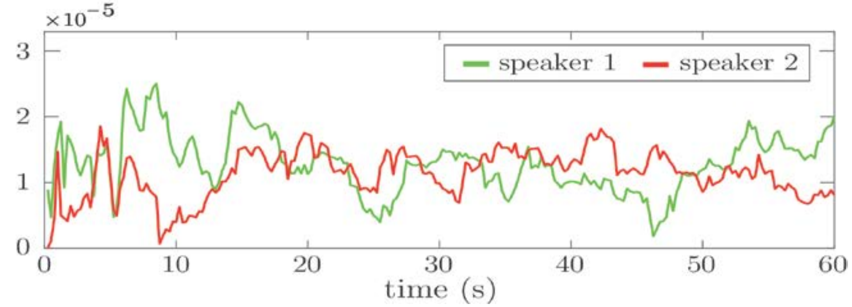
    rationale: detection of strong peaks in the decoder for the attended speaker

[Ding et al. (2012)]



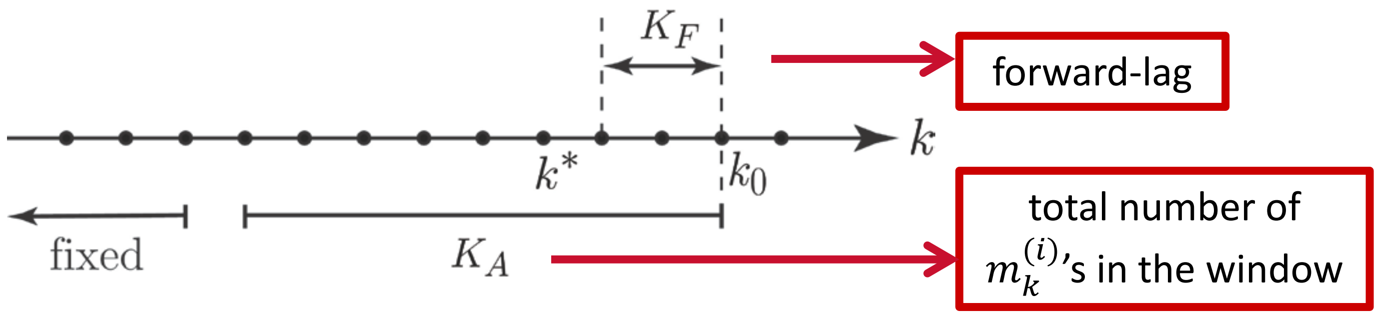
3. Real-Time State-Space Estimator:

- lots of stochastic fluctuations in features due to:
  background neural activity,
  high temporal resolution, …



- **Goal:** transform the extracted $m_k^{(1)}$'s and $m_k^{(2)}$'s for $k = 1, \dots, K$ into *dynamic* and *probabilistic* measures of the attentional state with *confidence intervals*, which are *robust* to the stochastic fluctuations in the extracted features

- fixed-lag sliding window framework:



when at $k = k_0$, estimate the attentional state at $k = k_0 - K_F$

K_F creates a tradeoff between *robustness* and *delay* in the estimates

- Fit the following model on $m_k^{(i)}$ for $i = 1,2$ and $k = 1,2,\dots,K_A$ in the sliding window: n_k is a binary RV determining the attended speaker at *instance* k ($n_k = 1$ or 2)

state-space model

$$\begin{cases} p_k = \mathrm{P}\left(n_k = 1\right) = \frac{1}{1+\exp(-z_k)} \\ \boxed{z_k} = c_0 z_{k-1} + w_k \\ w_k \sim \mathcal{N}(0, \eta_k) \\ \boxed{\eta_k} \sim \text{Inverse-Gamma}\left(a_0, b_0\right) \end{cases}$$

observation model

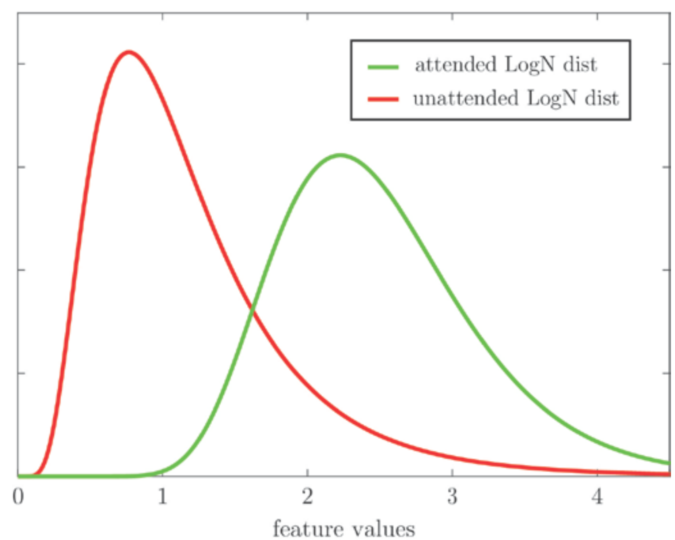
$$\begin{cases} m_k^{(i)} \,\Big|\, n_k = i \sim \text{Log-Normal}\left(\rho^{(a)}, \mu^{(a)}\right), \quad i = 1, 2 \\ m_k^{(i)} \,\Big|\, n_k \neq i \sim \text{Log-Normal}\left(\rho^{(u)}, \mu^{(u)}\right), \quad i = 1, 2 \\ \boxed{\rho^{(a)}} \sim \text{Gamma}\left(\alpha_0^{(a)}, \beta_0^{(a)}\right), \quad \boxed{\mu^{(a)}} \Big| \rho^{(a)} \sim \mathcal{N}\left(\mu_0^{(a)}, \rho^{(a)}\right) \\ \boxed{\rho^{(u)}} \sim \text{Gamma}\left(\alpha_0^{(u)}, \beta_0^{(u)}\right), \quad \boxed{\mu^{(u)}} \Big| \rho^{(u)} \sim \mathcal{N}\left(\mu_0^{(u)}, \rho^{(u)}\right) \end{cases}$$

- Parameters:

$$\boldsymbol{\Omega} = \left\{ z_{1:K_A}, \eta_{1:K_A}, \rho^{(a)}, \mu^{(a)}, \rho^{(u)}, \mu^{(u)} \right\}$$

- Output:

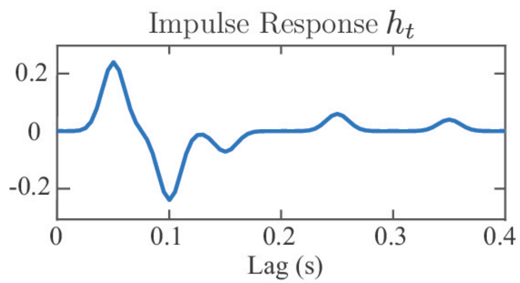
Plot $\hat{p}_k = \frac{1}{1+\exp(-\hat{z}_k)}$ with its confidence intervals as the estimated probability of attending to Sp. 1



Example Simulation Results:

Forward model for neural response:

$$e_t = w_t^{(1)}\left(s_t^{(1)} * h_t\right) + w_t^{(2)}\left(s_t^{(2)} * h_t\right) + \mu + n_t$$

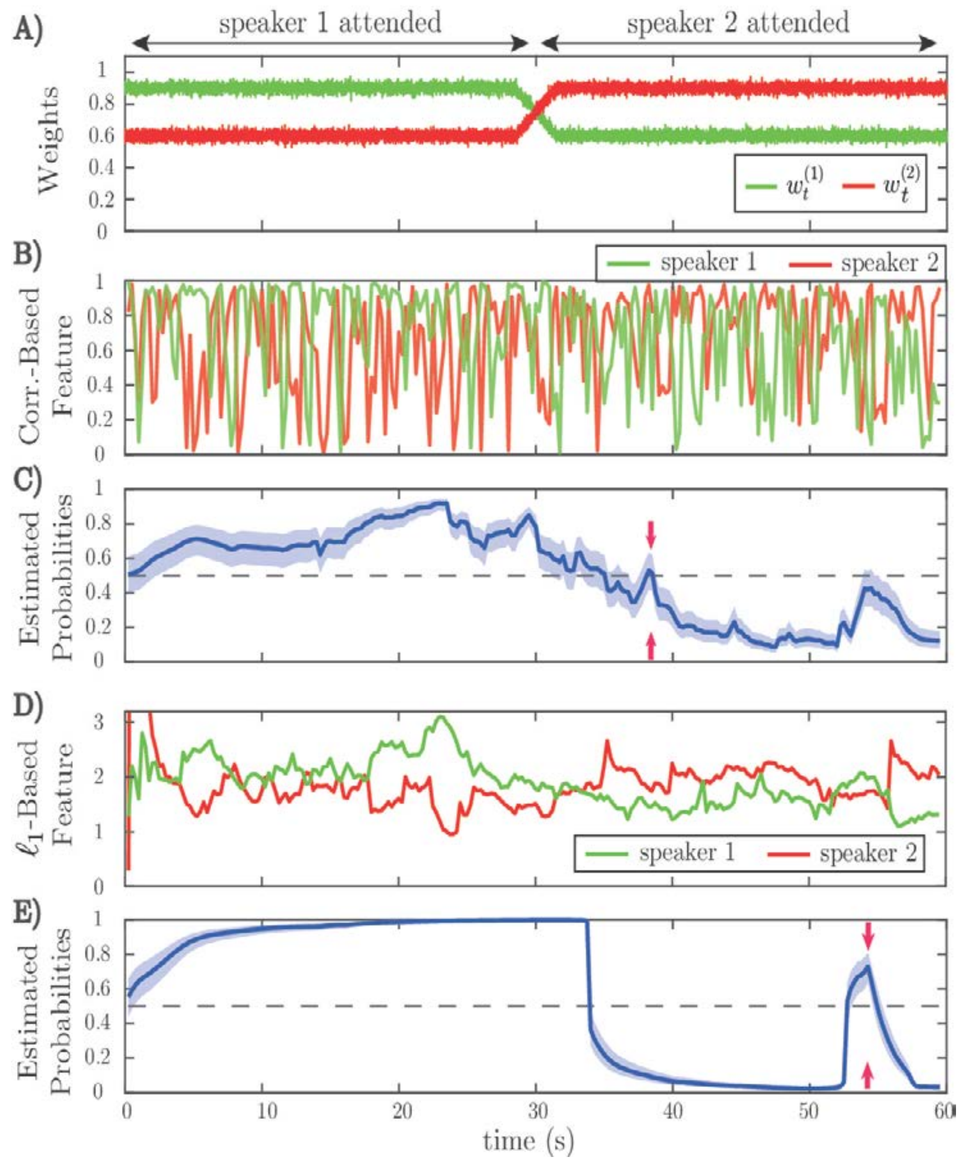


$w_t^{(1)}$ and $w_t^{(2)}$ determine the relative presence of envelopes $s_t^{(1)}$ and $s_t^{(2)}$ in the neural response

Estimation Settings:

$$W = 0.25 f_s,\ L_d = 0.4 f_s,\ K_F = 1.5 f_s / W$$

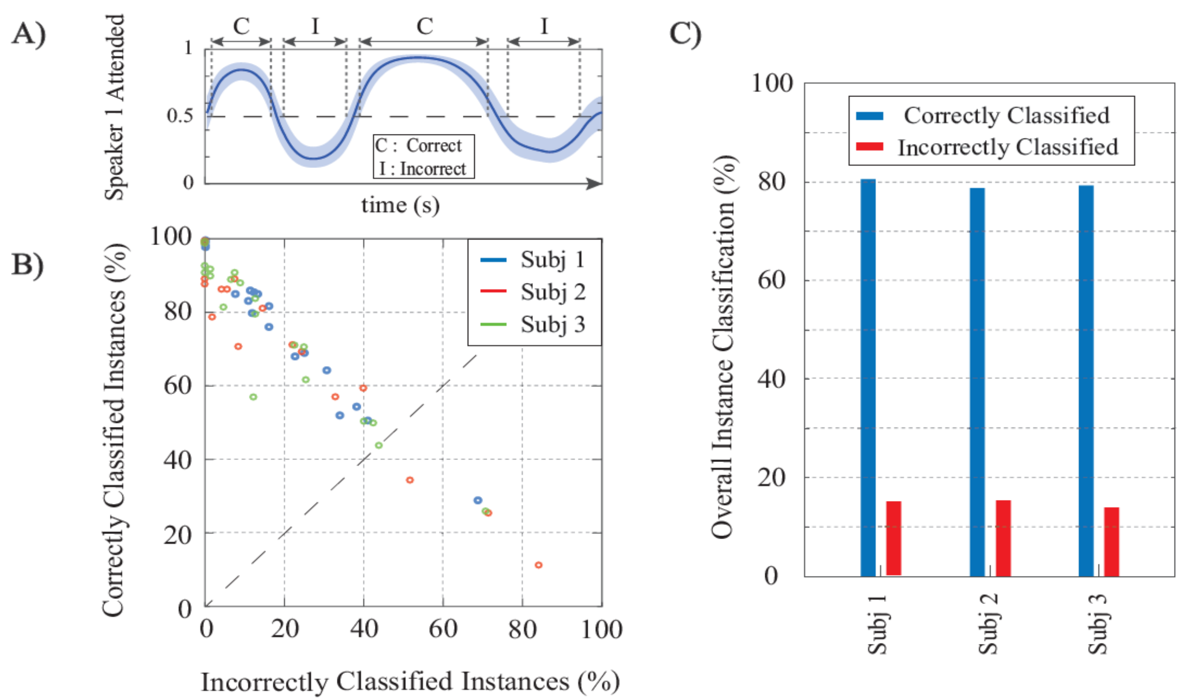
1.9 s built-in delay



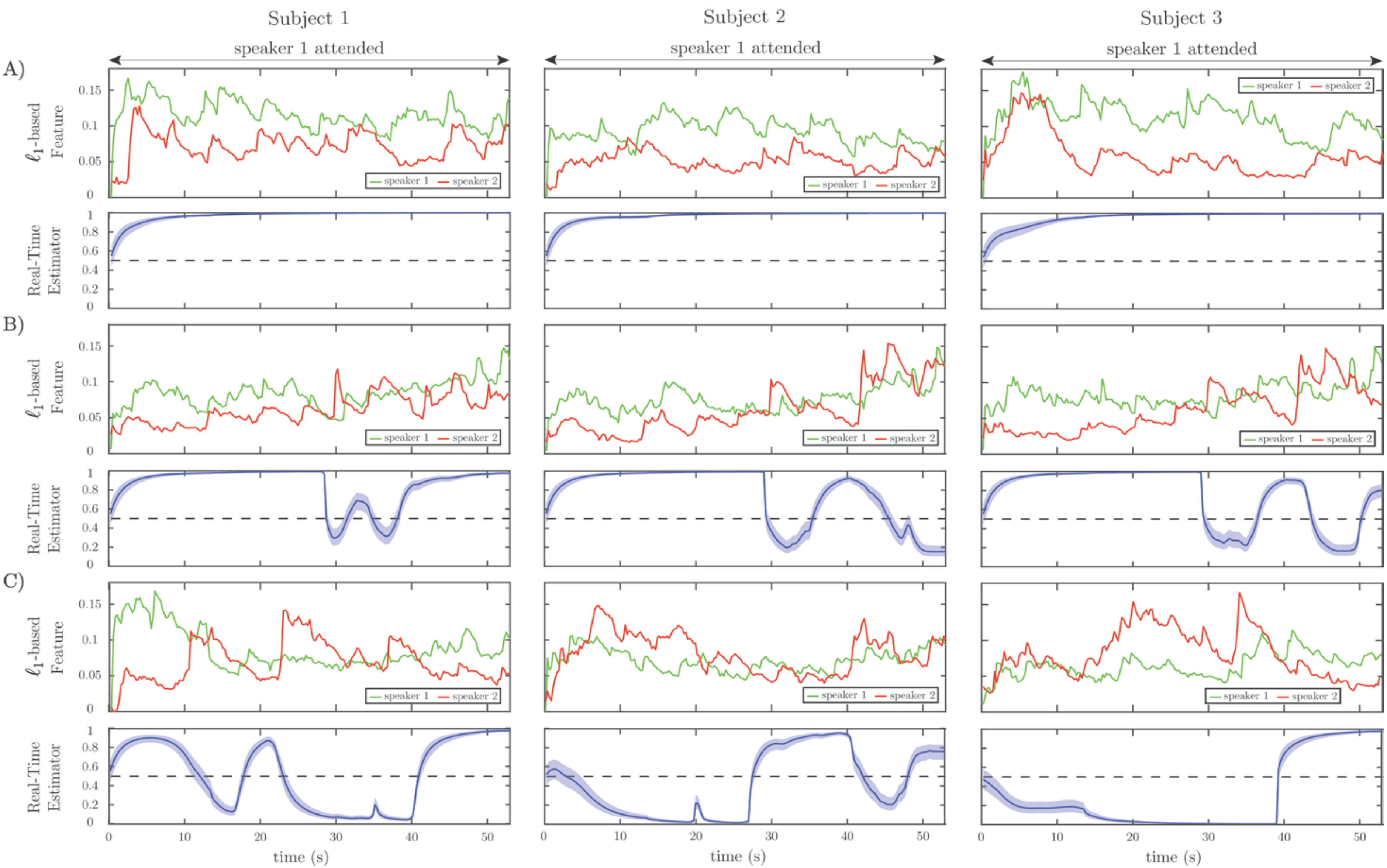
Recorded EEG Results:

- 3 subjects, 24 trials each, each trial 60s, two male speakers
- subjects instructed to maintain constant attention on speaker 1 during trials
- EEG downsampled to 64Hz, 64-channels reduced to 28 frontal channels (comp. cost)
- $W = 0.25 f_s, \lambda = 0.975, \gamma = 0.4$ (cross-validation), $K_F = 1.5 f_s / W, K_A = 15 f_s / W$
- decoder length $L_d = 0.25 f_s$, resulting in a total delay of 1.75 s in attention decoding
- ℓ_1-based feature showed a better attention modulation effect

- Classification Results:



Estimation outputs for sample trials:



Introduced a framework for (near) real-time attention decoding resulting in a robust and statistically interpretable measure of the attentional state. All processing is done in an online fashion and usage of training data has been minimized.

Journal version including encoding models, MEG analysis, and inference algorithms:

S. Miran, S. Akram, A. Sheikhattar, J.Z. Simon, T. Zhang, and B. Babadi, “Real-Time Tracking of Selective Auditory Attention from M/EEG: A Bayesian Filtering Appriach”, *Frontiers in Neuroscience, Vol. 12, pp. 262*, May 2017 .

Future Work:

- New EEG dataset collected at Starkey Hearing Technologies which includes:
  - babbling background noise
  - three speakers
  - reverberation effects
  - attention switching
  - more subjects and trials
- Moving beyond linear decoders

Thank You!

Questions?