MULTISCALE GAUSSIAN-MIXTURE MODELING FOR HMM POST-PROCESSING IN SELECTIVE AUDITORY ATTENTION DECODING

Tianyi Li[1], *Simon Geirnaert*[1,2], *Bert De Smedt*[3], *and Alexander Bertrand*[1]

[1]KU Leuven, Department of Electrical Engineering (ESAT), Leuven, Belgium
[2]KU Leuven, Department of Neurosciences, Research Group ExpORL, Leuven, Belgium
[3]KU Leuven, Faculty of Psychology and Educational Sciences, Leuven, Belgium.

ABSTRACT

Selective auditory attention decoding (sAAD) infers from electroencephalography (EEG) which speaker a listener attends to in multi-speaker scenarios. A widely used approach reconstructs the attended speech envelope from EEG, correlates it with candidate envelopes, and selects the speaker with the highest Pearson correlation. However, correlations in each decision window have high intrinsic variance, making per-window decisions unreliable, especially for short windows. Recent work introduced a hidden Markov model (HMM) that integrates correlation evidence over time and requires state-dependent emission distributions. Since labeled attention states are typically unavailable in practice, these distributions can be inferred from unlabeled observations by fitting a Gaussian mixture model (GMM). However, when the underlying Gaussians are close, unsupervised fitting becomes less reliable with limited samples. We therefore propose a multiscale GMM whose parameters are jointly estimated across multiple window lengths. The model exploits the window-length dependence of Fisher-transformed correlation estimates, whose means are approximately stable while their variances decrease with the number of samples per window. Experiments show that the multiscale GMM estimates emission parameters more accurately than single-scale fitting, particularly at short windows. Its HMM-post-processed accuracy advantage is largest for short recordings and narrows as more data become available, motivating multiscale fitting in data-constrained settings.

***Index Terms*—** auditory attention decoding, hidden Markov model, Gaussian mixture model, multiscale estimation

1. INTRODUCTION

Real-world listening often involves multiple concurrent speakers, where listeners need to focus on a speaker of interest while ignoring competing speakers. Neural tracking studies have shown that attended speech is represented more strongly than unattended speech in cortical responses [1, 2]. Identifying the attended speaker from electroencephalography (EEG), known as selective auditory attention decoding (sAAD), has been widely studied in recent years, with potential applications in neuro-steered hearing devices [3].

A common sAAD strategy is stimulus reconstruction [4, 5, 6, 7], where a neural decoder is trained to reconstruct the attended speech envelope from EEG, and the candidate speaker whose envelope yields the highest Pearson correlation with this reconstruction is identified to be the attended speaker. sAAD has also been studied under more realistic acoustic conditions and with more wearable EEG configurations [3, 8]. However, per-window correlations remain noisy, especially for short decision windows of one or a few seconds [9, 10].

To improve sAAD performance, recent work has proposed a post-processing algorithm based on a hidden Markov model (HMM), which integrates correlation evidence over time [9]. In this framework, the attention state, indicating which speaker is attended, evolves according to a transition probability that reflects the assumption that listeners switch attention relatively infrequently between adjacent windows. The observed correlation values are then modeled generatively through state-dependent emission distributions, which describe how likely different correlation patterns are under each attention state. The resulting predictions have been shown to be more accurate than raw decoder decisions without compromising switch detection time [9].

While the transition probability can be set as a hyperparameter, the state-dependent emission distributions must be estimated from data [9]. A supervised estimate would require windows with known attention states, but such labels are typically unavailable in practice. A natural unsupervised alternative, proposed by Heintz et al. [9], is a Gaussian mixture model (GMM): for each participant, the Fisher-transformed correlation is modeled by a two-component GMM, with the higher-mean component corresponding to the distribution of the correlation with the attended speaker and the lower-mean component corresponding to the distribution of the correlation with the unattended speaker. This assignment follows from the observation that EEG responses generally track attended speech more strongly than unattended speech [1, 2, 4].

HMM-based sAAD is particularly attractive at short operating windows because they preserve the temporal resolution needed to detect attention switches [9]. Yet short-window correlations are noisy, making their emission distributions difficult to estimate reliably with a single-scale GMM. We address this problem by using longer-window correlations to assist the estimation of short-window HMM emissions. Correlations at different window lengths reflect the same underlying attended and unattended neural responses, but with different estimation noise variances. After Fisher transformation, their means are approximately invariant across window lengths, whereas their variances scale approximately inversely with the number of samples [10, 11]. Based on these properties, the proposed *multiscale GMM* jointly estimates the emission parameters from observations at all window lengths under this scaling law.

The proposed method is particularly useful in two settings.

This work was supported by Internal Funds KU Leuven (projects IDN/23/006, C14/25/108, and C3/25/107), FWO projects G081722N and G026026N, and the FWO Junior Postdoctoral Fellowship for Fundamental Research awarded to S. Geirnaert (No. 1242524N). The authors thank Davina Van den Broek and Elien Bellon for their collaboration in the broader research project.

First, for short-window operation, the multiscale GMM incorporates longer-window correlation observations into the estimation of short-window emissions. Second, with limited available data, its cross-scale constraints regularize GMM estimation. We evaluate these settings by measuring emission-parameter recovery across window lengths and HMM-post-processed classification accuracy at different available recording durations.

2. MULTISCALE GMM

This section reviews HMM post-processing and introduces the multiscale GMM used to estimate its state-dependent emission distributions. We start from a standard stimulus-reconstruction sAAD pipeline with 2 speakers. A backward decoder reconstructs the attended speech envelope from EEG and can be trained in either a supervised [4, 5] or unsupervised [12] fashion. Within each decision window of length L, we compute two Fisher-transformed Pearson correlations $\mathbf{z} = (z_1, z_2)$ (hereafter Fisher-z correlations),

$$z_i = \operatorname{atanh}\big(\operatorname{corr}(\hat{y}, y_i)\big), \quad i \in \{1, 2\}, \tag{1}$$

where $\hat{y}$ is the reconstructed speech envelope (decoded from EEG recordings) and y_i is the envelope of candidate speaker i. If speaker 1 is the attended speaker, then we expect $\operatorname{corr}(\hat{y}, y_1) > \operatorname{corr}(\hat{y}, y_2)$ (hence $z_1 > z_2$ due to monotonicity of the Fisher transform) and vice versa. The Fisher transformation makes the correlation distribution approximately Gaussian [11].

For a sequence of T decision windows, let $\mathbf{z}_{1:T} = \{\mathbf{z}_t\}_{t=1}^{T}$ denote the Fisher-z correlation observations, and let $s_t \in \{1, 2\}$ denote the attended speaker at window t. The HMM postprocessor defined in [9] combines a transition model $p(s_t \mid s_{t-1})$ with an emission model $p(\mathbf{z}_t \mid s_t)$. We use forward-backward inference to compute $p(s_t \mid \mathbf{z}_{1:T})$ and select the state with the largest posterior probability.

Following [9], we assume conditional independence between z_1 and z_2 given the attention state s. The two correlation streams follow the attended density $p(z \mid \text{att})$ and unattended density $p(z \mid \text{unatt})$ according to the attention state. The HMM emission density then factorizes as

$$p(z_1, z_2 \mid s) = \begin{cases} p(z_1 \mid \text{att})\, p(z_2 \mid \text{unatt}), & \text{if } s = 1, \\ p(z_2 \mid \text{att})\, p(z_1 \mid \text{unatt}), & \text{if } s = 2. \end{cases} \tag{2}$$

The attended and unattended densities are modeled as Gaussian components with distinct means μ_{att} and μ_{unatt} and a common[1] scale-dependent variance σ_L^2, where L is the decision window length over which the correlation is computed.

$$\begin{aligned} p(z|\text{att}) &= \mathcal{N}(z; \mu_{\text{att}}, \sigma_L^2), \\ p(z|\text{unatt}) &= \mathcal{N}(z; \mu_{\text{unatt}}, \sigma_L^2). \end{aligned} \tag{3}$$

The attended mean is assumed to be larger than the unattended mean, because the EEG reconstruction is expected to correlate more strongly with attended speech [4].

2.1. Distributional properties of Fisher-z correlations

A conventional single-scale GMM estimates the emission model independently at the operating window length L. In this work, we instead exploit the approximate distribution of Fisher-z correlations and its dependence on L. For a window of length L, let $N_L = Lf_s$ denote the number of samples used to compute the correlation, where f_s is the sampling rate. For a Pearson correlation coefficient r with underlying true correlation ρ, the standard Fisher-Hotelling approximation models the Fisher-z statistic $z = \operatorname{atanh}(r)$ as approximately Gaussian with mean and variance given by:

$$\mathbb{E}[z \mid \rho, L] \approx \operatorname{atanh}(\rho) + \frac{\rho}{2(N_L - 1)}, \quad \operatorname{Var}(z \mid \rho, L) \approx \frac{1}{N_L - 1}. \tag{4}$$

This type of correlation-distribution modeling has been used to characterize AAD performance across window lengths [10]. Related scale-dependent behavior of correlation statistics was examined by Lopez-Gordo et al. [13] for unsupervised AAD accuracy estimation. Here, we apply this idea to the Fisher-z correlations used as HMM observations.

Approximate mean stability. The dominant term in (4) is $\operatorname{atanh}(\rho)$, which depends on the underlying attended or unattended correlation regardless of the window length. The first-order Hotelling correction introduces only a weak scale dependence, $\operatorname{atanh}(\rho) + \rho/[2(N_L - 1)]$. For the sampling rate and window lengths considered here, this term is small compared with the dominant Fisher-z mean. A corrected-mean sensitivity analysis also had negligible effect. We therefore use the scale-invariant mean model:

$$\forall\, L : \mu_{\text{att}}(L) \approx \mu_{\text{att}}, \quad \mu_{\text{unatt}}(L) \approx \mu_{\text{unatt}}. \tag{5}$$

Variance scaling. Following the inverse-sample-size dependence in (4), the GMM component variance is modeled as

$$\sigma_L^2 = \frac{C}{N_L - 1}. \tag{6}$$

Under the ideal Hotelling approximation, $C = 1$. Here, estimating C as an unknown parameter allows the model to capture the empirical variance level while preserving the expected dependence on window length.

Together, (5) and (6) imply that correlations measured at different window lengths follow a parametric family governed by a small set of parameters shared across window lengths. These two properties are empirically examined in Section 4.

2.2. Proposed multiscale GMM algorithm

Let $\mathcal{L} = \{L_1, \ldots, L_K\}$ be the set of window lengths used in the multiscale GMM fitting. For each window length $L \in \mathcal{L}$, let $\{z_t^{(L)}\}_{t=1}^{T_L}$ contain the Fisher-z observations pooled from both candidate speakers. Here, T_L counts the pooled scalar observations, which depends on the window length L since windows are not overlapping. The multiscale GMM parameters are $\theta = \{\mu_{\text{att}}, \mu_{\text{unatt}}, C, \pi\}$, where μ_{att} and μ_{unatt} are the scale-invariant component means, C is the variance scale parameter, and π is the mixture weight of the GMM. The mixture density at scale L is

$$\begin{aligned} p(z \mid \theta, L) = \pi\, &\mathcal{N}(z \mid \mu_{\text{att}}, \sigma_L^2) \\ &+ (1 - \pi)\, \mathcal{N}(z \mid \mu_{\text{unatt}}, \sigma_L^2), \end{aligned} \tag{7}$$

with $\sigma_L^2 = C/(N_L - 1)$. The parameters are estimated by maximizing the composite log-likelihood pooled across scales,

$$\hat{\theta} = \arg\max_{\theta} \sum_{L \in \mathcal{L}} \sum_{t=1}^{T_L} \log p(z_t^{(L)} \mid \theta, L). \tag{8}$$

[1] For simplicity, and as empirically shown in [13], both components share the same σ_L, which will also be theoretically and empirically motivated further on (see equation (4) and Fig. 1, respectively).

We solve (8) using expectation-maximization (EM). The component means are initialized at the 0.33 and 0.67 empirical quantiles of the pooled observations, C is initialized such that the average of (6) across scales corresponds to the empirical average , and $\pi^{(0)} = 1/2$. Although the present data are balanced, we keep π learnable for additional flexibility in the unsupervised fit. The lower- and higher-mean components are assigned to the unattended and attended states, respectively.

The E-step computes the responsibility of the attended component,

$$\gamma_t^{(L)} = \frac{\pi\,\mathcal{N}(z_t^{(L)} \mid \mu_{\mathrm{att}}, \sigma_L^2)}{\pi\,\mathcal{N}(z_t^{(L)} \mid \mu_{\mathrm{att}}, \sigma_L^2) + (1-\pi)\,\mathcal{N}(z_t^{(L)} \mid \mu_{\mathrm{unatt}}, \sigma_L^2)}. \tag{9}$$

Since $N_L - 1$ is proportional to the precision at scale L, the M-step updates for the component means and variance scale are

$$\mu_{\mathrm{att}} = \frac{\sum_{L,t}(N_L-1)\,\gamma_t^{(L)}\,z_t^{(L)}}{\sum_{L,t}(N_L-1)\,\gamma_t^{(L)}}, \tag{10}$$

$$\mu_{\mathrm{unatt}} = \frac{\sum_{L,t}(N_L-1)\,(1-\gamma_t^{(L)})\,z_t^{(L)}}{\sum_{L,t}(N_L-1)\,(1-\gamma_t^{(L)})}, \tag{11}$$

$$C = \frac{1}{T_{\mathrm{tot}}}\sum_{L,t}(N_L-1)\Big[\gamma_t^{(L)}(z_t^{(L)}-\mu_{\mathrm{att}})^2 + (1-\gamma_t^{(L)})(z_t^{(L)}-\mu_{\mathrm{unatt}})^2\Big], \tag{12}$$

$$\sigma_L^2 = \frac{C}{N_L - 1}. \tag{13}$$

The mixture weight is updated as $\pi = T_{\mathrm{tot}}^{-1}\sum_{L,t}\gamma_t^{(L)}$. After each M-step, the components are reordered if necessary to maintain $\mu_{\mathrm{att}} > \mu_{\mathrm{unatt}}$, and the iterations are stopped when the relative log-likelihood increase falls below a preset tolerance or the maximum number of iterations is reached. The multiscale GMM estimates four parameters from data pooled across all K scales, instead of $4K$ parameters for independent single-scale GMMs. At inference time, the fitted densities are inserted into (2) to construct the two HMM-state emissions.

3. EXPERIMENTAL SETUP

All experiments use a supervised leave-one-participant-out (LOSO) decoder trained on the other 15 participants. GMM fitting and HMM post-processing are performed separately for each held-out participant, with no parameter sharing across participants. Held-out labels are used only for evaluation, not for GMM fitting.

3.1. Dataset

We evaluate the method on the publicly available KU Leuven sAAD dataset [5, 14], containing 64-channel EEG from 16 normal-hearing participants listening to two competing Dutch speakers. Each participant completed 20 trials with one attended speaker. After truncation to whole minutes, eight trials are 6 min and twelve are 2 min, yielding 72 min. In each LOSO fold, all 20 trials from each of the 15 training participants are used to train the decoder. All 20 trials from the held-out participant are used for GMM fitting and HMM post-processing.

3.2. Preprocessing

We follow the public KUL preprocessing pipeline [5]. Speech envelopes are extracted from the original speech signals using a gammatone filterbank and power-law compression with exponent 0.6. Both EEG and envelopes are bandpass filtered to 1-9 Hz, downsampled to $f_s = 32$ Hz, and the EEG is re-referenced to Cz.

3.3. Decoder training

Following [5], we use a spatio-temporal least-squares backward decoder to reconstruct the attended speech envelope from EEG. The decoder uses $N_c = 64$ channels and $J = 17$ post-stimulus lags, corresponding to 0-500 ms at $f_s = 32$ Hz [15]. For each LOSO fold, the normal equations are pooled over data from the 15 training participants and solved to obtain one subject-independent decoder. It is applied to the held-out participant to obtain the reconstructed envelope, which is correlated with each candidate envelope and Fisher-transformed as in (1).

3.4. Multiscale GMM and HMM inference

On the evaluation trials, we compute Fisher-z correlations at $K = 8$ window lengths, $L \in \mathcal{L} = \{1, 2, 3, 5, 6, 10, 15, 30\}$ s. These window lengths cover short, intermediate, and long windows while allowing each retained trial to be divided into complete windows at every scale. The multiscale GMM is fitted jointly over all K window lengths.

We compare the proposed multiscale GMM with the single-scale GMM used in [9] that fits the emission model at window length L, using the same EM initialization and update rules. For each target L, the evaluation trials are concatenated trial by trial to form one sequence per participant. Forward-backward inference uses a symmetric transition probability $p_{\mathrm{switch}}(L) = 0.01L$.

Emission-parameter recovery is reported at all eight scales using the complete 72-min fitting set. For HMM post-processing, we focus on short operating windows at $L^* = 1$ and 2 s as suggested by [9], while longer-window observations serve as auxiliary data in the multiscale fit. To assess performance as a function of available data, we vary the recording duration from 6 to 72 min in 6-min steps. At each duration, accuracy is averaged within participant over up to ten data-subset realizations. Trials within each realization are ordered so that the attended speaker alternates at every trial boundary. Each realization is used for both GMM fitting and HMM post-processing and is identical for both methods.

3.5. Evaluation metrics

Emission-parameter estimation is evaluated against label-based empirical references. For each participant and window length, observations from both candidate streams are pooled across both speakers and split according to whether the corresponding speaker was attended or unattended. Their sample means and pooled standard deviation are used as supervised "ground-truth" references for the Gaussian parameters.

Specifically, we evaluate $a \in \{\mu_{\mathrm{att}}, \mu_{\mathrm{unatt}}, \sigma^*, \Delta\mu\}$, where $\Delta\mu = \mu_{\mathrm{att}} - \mu_{\mathrm{unatt}}$ and $\sigma^* = \sqrt{C/(N_{L^*} - 1)}$, with L^* denoting the target correlation window length used by the HMM. Let a^{fit} and a^{ref} denote the fitted and empirical values, respectively. Parameter recovery is quantified by the normalized relative error $E_a = |a^{\mathrm{fit}} - a^{\mathrm{ref}}|/|a^{\mathrm{ref}}|$.

HMM-post-processed classification accuracy is computed within each trial as the percentage of decision windows for which the

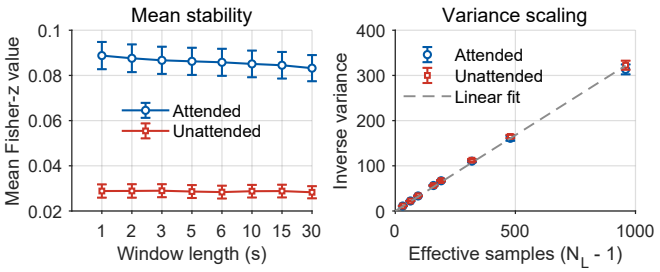

Fig. 1. Empirical attended and unattended Fisher-z means (left) and inverse variance versus $N_L - 1$ (right), averaged across participants. Error bars denote mean $\pm$ SEM; the dashed line is a linear fit through the origin.

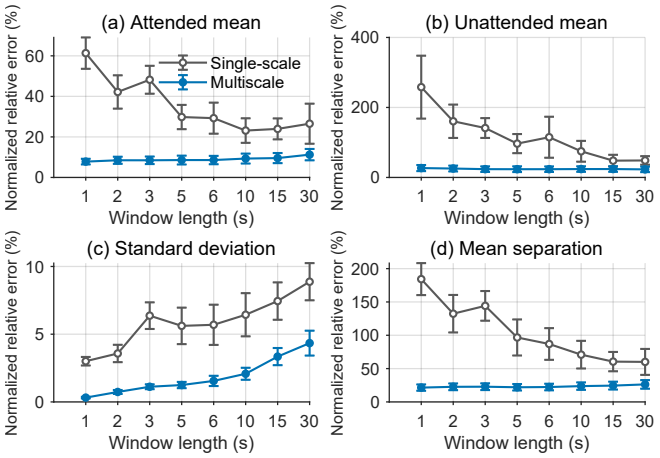

Fig. 2. Normalized relative error across window lengths for the attended mean, unattended mean, standard deviation, and mean separation $\Delta\mu = \mu_{\text{att}} - \mu_{\text{unatt}}$. Markers and error bars denote mean $\pm$ SEM across participants.

HMM's maximum-posterior state matches the true attended speaker, and then averaged across trials for each participant [9]. Paired comparisons at the 6- and 72-min recording-duration endpoints use one-sided Wilcoxon signed-rank tests for higher multiscale accuracy, with Holm correction across the four comparisons arising from the two target window lengths ($L^* = 1$ and 2 s) and the two recording-duration endpoints. Emission-parameter recovery and intermediate data-amount trends are otherwise summarized using the mean and standard error of the mean (SEM) across participants.

4. RESULTS AND DISCUSSION

Figure 1 examines the two multiscale assumptions. For this diagnostic only, the true labels are used to pool observations into attended and unattended groups. Their means remain separated and vary only weakly with window length. Meanwhile, inverse variance is approximately linear in $N_L - 1$. Long-window observations can therefore inform the component means at short operating windows, while the variance law maps their precision across scales.

As shown in Fig. 2, multiscale fitting yields lower mean errors than single-scale fitting for all four evaluated quantities at both $L^* = 1$ and $L^* = 2$ s, with larger reductions at 1 s. At longer windows, lower correlation variance separates attended and unattended samples more clearly, making the single-scale GMM easier to fit. Thus, long windows are most valuable as auxiliary data for stabilizing noisy short-window HMM emissions.

We next evaluate HMM post-processing under the default full-data setting. The mean classification accuracies for single-scale and multiscale fitting were 76.17% and 77.98% at 1 s, respectively, corresponding to a difference of 1.82 percentage points (pp) (Holm-corrected $p = 0.0077$). At 2 s, the corresponding accuracies were 77.24% and 78.32%, with a difference of 1.08 pp (Holm-corrected $p = 0.0232$).

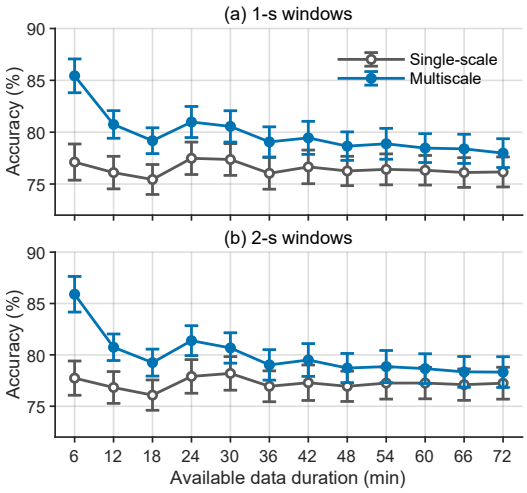

Fig. 3. HMM-post-processed classification accuracy at 1- and 2-s windows versus available recording duration (mean $\pm$ SEM across participants, averaged over data-subset realizations).

Figure 3 shows how classification accuracy changes with available recording duration. With only 6 min of data, the mean multiscale-minus-single-scale accuracy differences are 8.32 pp at 1 s and 8.16 pp at 2 s (both Holm-corrected $p < 0.001$). As the recording duration increases, the two curves gradually approach each other. Shorter subsets contain fewer trial boundaries and therefore fewer attention-state changes, which affects the absolute accuracy levels. The larger paired differences at short durations highlight the value of multiscale fitting when only a short recording is available.

5. CONCLUSION

We have proposed a multiscale GMM that jointly estimates shared emission parameters from correlation observations at multiple window lengths. This allows the more reliable statistics at longer windows to improve the estimation of the HMM emission probabilities based on noisy short-window correlations. Experiments showed more accurate short-window emission-parameter recovery than single-scale fitting. The data-amount analysis further showed larger mean classification-accuracy gains for shorter available recordings, highlighting the value of multiscale estimation in data-constrained settings.

6. REFERENCES

[1] Nima Mesgarani and Edward F. Chang, "Selective cortical representation of attended speaker in multi-talker speech perception," *Nature*, vol. 485, no. 7397, pp. 233–236, 2012.

[2] Alan J. Power, John J. Foxe, Emma-Jane Forde, Richard B. Reilly, and Edmund C. Lalor, "At what time is the cocktail party? a late locus of selective attention to natural speech," *European Journal of Neuroscience*, vol. 35, no. 9, pp. 1497–1503, 2012.

[3] Bojana Mirkovic, Martin G. Bleichner, Maarten De Vos, and Stefan Debener, "Target speaker detection with concealed EEG around the ear," *Frontiers in Neuroscience*, vol. 10, pp. 349, 2016.

[4] James A O'Sullivan, Alan J Power, Nima Mesgarani, Siddharth Rajaram, John J Foxe, Barbara G Shinn-Cunningham, Malcolm Slaney, Shihab A Shamma, and Edmund C Lalor, "Attentional selection in a cocktail party environment can be decoded from single-trial EEG," *Cerebral cortex*, vol. 25, no. 7, pp. 1697–1706, 2015.

[5] Wouter Biesmans, Neetha Das, Tom Francart, and Alexander Bertrand, "Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 25, no. 5, pp. 402–412, 2016.

[6] Alain De Cheveigné, Daniel DE Wong, Giovanni M Di Liberto, Jens Hjortkjær, Malcolm Slaney, and Edmund Lalor, "Decoding the auditory brain with canonical component analysis," *NeuroImage*, vol. 172, pp. 206–216, 2018.

[7] Michael J. Crosse, Giovanni M. Di Liberto, Adam Bednar, and Edmund C. Lalor, "The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli," *Frontiers in Human Neuroscience*, vol. 10, pp. 604, 2016.

[8] Søren Asp Fuglsang, Torsten Dau, and Jens Hjortkjær, "Noise-robust cortical tracking of attended speech in real-world acoustic scenes," *NeuroImage*, vol. 156, pp. 435–444, 2017.

[9] Nicolas Heintz, Tom Francart, and Alexander Bertrand, "Post-processing of EEG-based auditory attention decoding decisions via hidden markov models," *arXiv preprint arXiv:2506.24024*, 2025.

[10] Simon Geirnaert, Jonas Vanthornhout, Tom Francart, and Alexander Bertrand, "Performance modeling for correlation-based neural decoding of auditory attention to speech," in *2025 33rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2025, pp. 1467–1471.

[11] Ronald A. Fisher, "On the probable error of a coefficient of correlation deduced from a small sample," *Metron*, vol. 1, pp. 3–32, 1921.

[12] Simon Geirnaert, Tom Francart, and Alexander Bertrand, "Unsupervised self-adaptive auditory attention decoding," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3955–3966, 2021.

[13] Miguel A Lopez-Gordo, Simon Geirnaert, and Alexander Bertrand, "Unsupervised accuracy estimation for brain-computer interfaces based on selective auditory attention decoding," *IEEE Transactions on Biomedical Engineering*, 2025.

[14] Neetha Das, Tom Francart, and Alexander Bertrand, "Auditory attention detection dataset kuleuven," *Zenodo*, 2019.

[15] Jonas Vanthornhout, Lieselot Decruy, Jan Wouters, Jonathan Z. Simon, and Tom Francart, "Effect of task and attention on neural tracking of speech," *Frontiers in Neuroscience*, vol. 13, pp. 977, 2019.