KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Geirnaert S., Francart T., Bertrand A. (2021), **Riemannian Geometry-Based Decoding of the Directional Focus of Auditory Attention Using EEG** *Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021)*
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	doi:10.1109/ICASSP39728.2021.9413404
Journal homepage	https://2021.ieeeicassp.org/
Author contact	simon.geirnaert@esat.kuleuven.be *+ 32 (0)16 37 35 36*
Abstract	
IR	

(article begins on next page)

RIEMANNIAN GEOMETRY-BASED DECODING OF THE DIRECTIONAL FOCUS OF AUDITORY ATTENTION USING EEG

Simon Geirnaert[⋆†] *Tom Francart*[†] *Alexander Bertrand*[⋆]

[⋆]KU Leuven, Department of Electrical Engineering (ESAT),
STADIUS Center for Dynamical Systems, Signal Processing, and Data Analytics, Belgium
[†]KU Leuven, Department of Neurosciences, ExpORL, Belgium

ABSTRACT

Auditory attention decoding (AAD) algorithms decode the auditory attention from electroencephalography (EEG) signals that capture the listener's neural activity. Such AAD methods are believed to be an important ingredient towards so-called neuro-steered assistive hearing devices. For example, traditional AAD decoders allow detecting to which of multiple speakers a listener is attending to by reconstructing the amplitude envelope of the attended speech signal from the EEG signals. Recently, an alternative paradigm to this stimulus reconstruction approach was proposed, in which the directional focus of auditory attention is determined instead, solely based on the EEG, using common spatial pattern filters (CSP). Here, we propose Riemannian geometry-based classification (RGC) as an alternative for this CSP approach, in which the covariance matrix of a new EEG segment is directly classified while taking its Riemannian structure into account. While the proposed RGC method performs similarly to the CSP method for short decision lengths (i.e., the amount of EEG samples used to make a decision), we show that it significantly outperforms it for longer decision window lengths.

***Index Terms*—** neuro-steered hearing device, auditory attention decoding, directional focus of attention, brain-computer interface, Riemannian geometry, electroencephalography

1. INTRODUCTION

Previous research has shown that it is possible to decode the auditory attention from brain activity measured by electroencephalography (EEG) sensors [1, 2, 3]. These *auditory attention decoding* (AAD) algorithms fill an important gap in the design of assistive hearing devices (e.g., cochlear implants or hearing aids), as they inform classical speaker separation and noise reduction algorithms about the speaker a user wants to attend to in a multi-speaker scenario. As such, AAD algorithms constitute a fundamental building block of neuro-steered hearing devices.

AAD algorithms traditionally use a stimulus reconstruction approach, in which a spatio-temporal decoder is applied to the EEG to reconstruct the amplitude envelope of the attended speaker [1, 3]. The decoded speech envelope then traditionally shows a higher correlation coefficient with the attended speech envelope than with the unattended speech envelope(s). This approach, however, suffers from low decoding accuracy at high speed, i.e., when using few data to decode the auditory attention [3, 4]. As these short decision windows (i.e., the amount of data used to decode the attention) $< 10\,\mathrm{s}$ are paramount for the practical applicability of AAD algorithms, for example, when the attention is switched between two speakers [4], the stimulus reconstruction approach might be too slow for practical neuro-steered hearing devices or for conducting research experiments that require tracking of attention. Furthermore, this approach requires an error-prone speech separation step, in order to retrieve the individual speech envelopes from the recorded mixture of speech sources [3, 5].

As an alternative paradigm, decoding the directional focus of auditory attention, solely based on the EEG, was proposed in [2]. In this approach, the common spatial pattern (CSP) filtering method is used to discriminate between different angular positions of the attended and unattended speaker(s). This CSP approach significantly outperforms the stimulus reconstruction approach on short decision windows. Furthermore, this paradigm does not require a preceding speech separation step. As such, this alternative paradigm improves the practical applicability of neuro-steered hearing devices.

We propose a new AAD algorithm, capitalizing this new paradigm of decoding the directional focus of auditory attention, but replacing the traditional CSP filter method with a so-called *Riemannian geometry classifier* (RGC). This technique has become very popular in the brain-computer inter-

This research is funded by an Aspirant Grant from the Research Foundation - Flanders (FWO) (for S. Geirnaert), the KU Leuven Special Research Fund C14/16/057, FWO project nr. G0A4918N, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 802895 and grant agreement No 637424), and the Flemish Government (AI Research Program). The scientific responsibility is assumed by its authors.

S. Geirnaert and A. Bertrand are also affiliated with Leuven.AI - KU Leuven Institute for AI, Belgium.

face (BCI) community [6] and outperforms the classical CSP approach in various BCI applications, in particular in motor imagery paradigms [6, 7, 8]. In Section 2, we explain how this RGC can be used to classify the directional focus of auditory attention. In Section 3, we compare the proposed RGC classifier with the state-of-the-art CSP method and stimulus reconstruction approach. Conclusions are drawn in Section 4.

2. RIEMANNIAN GEOMETRY-BASED CLASSIFICATION

In recent years, a new class of RGCs has gained a lot of attention in the BCI community [6]. Instead of pre-filtering the EEG using data-driven filters based on the EEG covariance structure (as is the case in CSP filtering [9]), the EEG covariance matrices are classified directly, as it is assumed that all spatial (and potentially temporal) information concerning different conditions is encoded in these covariance matrices [7, 8]. However, covariance matrices are symmetric positive definite (SPD), such that they live on a differentiable Riemannian manifold rather than in a Euclidean space. RGCs take this specific structure into account to improve classification performance. More details about RGCs and their use in BCIs can be found in [6, 7, 8].

As covariance matrices live on a *differentiable* Riemannian manifold, a tangent space at each point (i.e., covariance matrix) can be computed. Such a tangent space, containing symmetric matrices, is Euclidean, where Euclidean distances between tangent vectors approximate Riemannian distances (i.e., distances between covariance matrices on the Riemannian manifold) of the (projected) covariance matrices. As traditional classifiers rely on Euclidean metrics, which conflict with the Riemannian structure of the manifold on which covariance matrices live, it is preferred to first project all covariance matrices onto the tangent space of a reference matrix. This is the crucial difference with a straightforward direct classification of covariance matrices, which assumes a Euclidean structure of the covariance matrices. In the RGC, the intermediate *tangent space mapping* (TSM) assures that Euclidean metrics are applicable. For the tangent space to be a good local approximation of the Riemannian manifold, where Euclidean distances between tangent vectors closely approximate Riemannian distances between the covariance matrices, a good choice of the reference point of the TSM is the geometric or Riemannian mean.

Let $\{\mathbf{X}_k, y_k\}_{k=1}^K$ be a training set containing K segments of bandpass filtered EEG data $\mathbf{X}_k \in \mathbb{R}^{C\times T}$, with C channels and T time samples, and with known labels $y_k \in \{-1, 1\}$ (e.g., attending to the left or right speaker). The corresponding covariance matrices are defined as

$$\mathbf{R}_k = \frac{1}{T-1}\mathbf{X}_k\mathbf{X}_k^{\mathsf{T}} \in \mathbb{R}^{C\times C}. \tag{1}$$

As in [2], we estimate the covariance matrices using ridge regression, where the regularization hyperparameter is determined automatically using the method proposed in [10]. This hyperparameter estimation method is considered to be the state-of-the-art in BCI research [6].

The geometric or Riemannian mean of these K covariance matrices is then given by the SPD matrix $\mathbf{R}_{\mathfrak{G}}$ that minimizes the mean squared Riemannian distance from each $\mathbf{R}_k$ to $\mathbf{R}_{\mathfrak{G}}$ [7]:

$$\mathbf{R}_{\mathfrak{G}} = \mathfrak{G}(\mathbf{R}_1, \ldots, \mathbf{R}_K) = \underset{\mathbf{R}\text{ is SPD}}{\operatorname{argmin}} \sum_{k=1}^{K} \delta_R^2(\mathbf{R}_k, \mathbf{R}), \tag{2}$$

where $\delta_R(\mathbf{R}, \mathbf{S})$ denotes the Riemannian distance between two SPD matrices $\mathbf{R}$ and $\mathbf{S}$, which can be computed as [7]:

$$\delta_R(\mathbf{R}, \mathbf{S}) = \left|\left|\log\left(\mathbf{R}^{-1}\mathbf{S}\right)\right|\right|_F, \tag{3}$$

with $\log(\cdot)$ the matrix-logarithm. Given a diagonalizable matrix $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, the matrix-logarithm of $\mathbf{A}$ is defined as:

$$\log(\mathbf{A}) = \mathbf{V}\log(\mathbf{\Lambda})\mathbf{V}^{-1}, \tag{4}$$

with $\log(\mathbf{\Lambda})$ a diagonal matrix with diagonal elements $\log(\lambda_i)$. The Riemannian mean in (2) can only be computed in an iterative way, by iteratively computing the Euclidean mean in the tangent space mapping, or can be approximated using log-euclidean metrics [11]:

$$\mathbf{R}_{\mathfrak{G}} \approx \exp\left(\frac{1}{K}\sum_{k=1}^{K}\log(\mathbf{R}_k)\right), \tag{5}$$

where the matrix-exponential $\exp(\cdot)$ is defined similarly as the matrix-logarithm in (4). We here use the latter estimation method in (5) to efficiently compute the Riemannian mean covariance matrix.

The normalized TSM of the covariance matrix $\mathbf{R}_k$ onto the tangent space at reference point $\mathbf{R}_{\mathfrak{G}}$ (2) is then equal to [7]:

$$\mathbf{T}_k = \log\left(\mathbf{R}_{\mathfrak{G}}^{-\frac{1}{2}}\mathbf{R}_k\mathbf{R}_{\mathfrak{G}}^{-\frac{1}{2}}\right). \tag{6}$$

The TSM $\mathbf{T}_k$ is then half-vectorized (i.e., a vectorization over the lower-triangular part only, as it is a symmetric matrix), which leads to the feature vector $\mathbf{f}_k \in \mathbb{R}^{\frac{C(C+1)}{2}\times 1}$, representing EEG segment $\mathbf{X}_k$ of the training set. Similarly, for a new test segment $\mathbf{X}^{(\text{test})}$, the test feature vector can be found by computing the TSM of its covariance matrix using the Riemannian mean $\mathbf{R}_{\mathfrak{G}}$ over the training set.

The generated feature vectors with the aforementioned method can then be classified using any (Euclidean) classifier, trained with the training set $\{\mathbf{f}_k, y_k\}_{k=1}^K$. We here choose a support vector machine (SVM) classifier with a linear kernel. Such a classifier works well in high-dimensional feature spaces, which we are dealing with here. Note that combining the TSM with a linear SVM can be interpreted as applying an SVM with a *Riemannian kernel* on the half-vectorized original covariance matrix [8]. The classification algorithm is summarized in Algorithm 1.

Algorithm 1 Riemannian geometry-based classification

Input: Test EEG segment $\mathbf{X}^{(\text{test})} \in \mathbb{R}^{C \times T}$ and given Riemannian mean $\mathbf{R}_{\mathfrak{G}}$ over a training set and (linear) SVM classifier $D(\cdot)$

Output: Class label $y^{(\text{test})}$ (e.g., left or right attended)

1: Bandpass filter $\mathbf{X}^{(\text{test})}$ between 12–30 Hz

2: Compute a regularized covariance matrix:

$$\mathbf{R}^{(\text{test})} = \frac{1}{T-1}\mathbf{X}^{(\text{test})}\mathbf{X}^{(\text{test})^{\mathsf{T}}} + \delta\mathbf{I},$$

with regularization constant δ

3: Compute the tangent space mapping onto Riemannian mean $\mathbf{R}_{\mathfrak{G}}$:

$$\mathbf{T}^{(\text{test})} = \log\left(\mathbf{R}_{\mathfrak{G}}^{-\frac{1}{2}}\mathbf{R}^{(\text{test})}\mathbf{R}_{\mathfrak{G}}^{-\frac{1}{2}}\right)$$

4: Compute the feature vector as the half-vectorization $\mathbf{f}^{(\text{test})} = \text{vech}\left(\mathbf{T}^{(\text{test})}\right)$ of the TSM

5: Classify: $y^{(\text{test})} = \text{sign}(D(\mathbf{f}))$

3. EXPERIMENTS AND RESULTS

We compare the proposed RGC method with the CSP method [2], which is the state-of-the-art method for decoding the directional focus of auditory attention. In the CSP method used in [2], features are generated by applying six spatial filters that maximize discriminability [9] and are classified with a linear discriminant analysis (LDA) classifier. The state-of-the-art stimulus reconstruction method (canonical correlation analysis (CCA) + LDA), as shown in [12, 3], is also added as a reference. For the CCA method, the same preprocessing steps and design choices as in [2] are used.

3.1. AAD dataset

The comparison is performed on a publicly available dataset, which is recorded for the purpose of AAD [13, 14]. This dataset contains the EEG of 16 subjects, attending to one of two simultaneously active competing speakers, located at $\pm 90^\circ$ along the azimuth direction. Per subject, 72 minutes of data is available. The EEG is recorded using a $C = 64$-channel BioSemi ActiveTwo system. For more details, we refer to [13, 14].

3.2. Design choices

3.2.1. Bandpass filtering

According to the analysis of the filterband importance in the state-of-the-art CSP approach [2], the β-band (12–30 Hz) is the most useful EEG frequency band to decode the directional focus of attention. As such, both for the baseline CSP algorithm, as for the proposed RGC method, the EEG is prefiltered in the β-band using an 8th-order Butterworth filter and downsampled to 64 Hz.

3.3. Performance evaluation

The proposed RGC method is tested in a subject-specific way using ten-fold cross-validation. Therefore, the 72 minutes of EEG data per subject are split into 60 s segments, which are randomly distributed across ten folds. Note that these 60 s segments are normalized by setting the mean per channel to zero, as well as setting the Frobenius norm across all channels to one. The decision window length is defined as the length of the EEG window over which a single AAD decision is made (this usually results in a trade-off between AAD accuracy and decision latency [4]). In the case of our RGC framework, the decision window length is defined by the number of samples T over which the covariance matrices are estimated. To evaluate the AAD accuracy for various decision window lengths, all 60 s segments are split into shorter decision windows. The Riemannian mean in (2) and linear SVM are retrained for every decision window length. The significance level for above-chance AAD accuracy is computed based on the inverse binomial distribution [1]. Note that shorter decision window lengths result in more decisions over the test fold, resulting in a lower significance level. A similar ten-fold cross-validation procedure is used for the CSP and CCA method.

Evaluating the AAD accuracy across different decision window lengths is important, for example, in the context of detecting switches in auditory attention. To resolve the traditional trade-off between accuracy and decision window length, the minimal expected switch duration (MESD) metric [s] is used, as proposed in [4]. This single-number AAD performance metric quantifies the minimal expected time it takes to switch the gain from one speaker to another, following a switch in attention, based on an optimized stochastic model of a robust (i.e., assuring stable operation above a pre-defined comfort level) attention-steered gain control system.

3.4. Results

Figure 1 shows the AAD accuracies as a function of decision window length for the RGC, CSP, and CCA method. Below 1 s decision window lengths (i.e., using $T = 64$ samples at $f_s = 64$ Hz), the RGC and CSP methods have very similar accuracies. Between 1 s and 5 s, there is a much faster increase in performance for the RGC method than for the CSP method. This is mostly due to the quickly improving covariance matrix estimation (required for the RGC method) at these shorter decision window lengths. Indeed, as more data become available for increasing decision window lengths to estimate the covariance matrix, less regularization is required, introducing a smaller bias on the estimated covariance matrix. There is no similar effect for the CSP method, as there is no

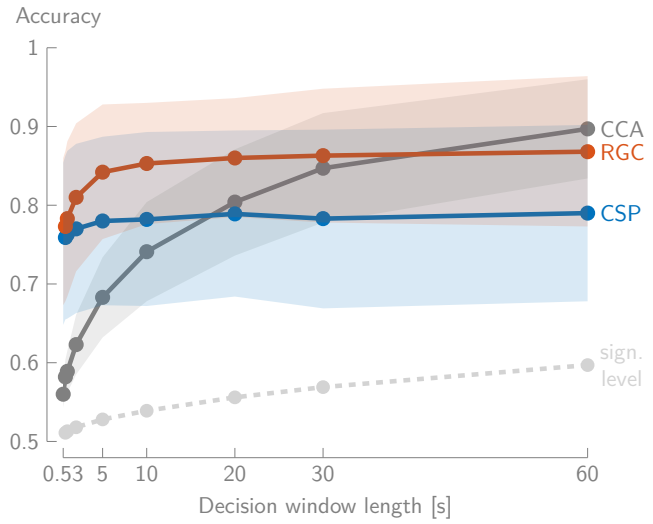

Fig. 1: The mean AAD accuracy across subjects ($\pm$ standard deviation across subjects) shows that the RGC outperforms the CSP approach on almost all decision window lengths, but exhibits a faster decrease in performance on very short decision window lengths, resulting in very similar performances on 531 ms decision windows.

direct covariance matrix estimation involved. Note that, potentially, the RGC method could be improved on these very short decision windows by applying an intelligent dimensionality reduction or feature selection method, which is beyond the scope of this paper. Beyond 5 s, the performance levels off in both cases and the RGC method outperforms the CSP method with $\approx 6\%$.

As is also shown in [2], the traditional stimulus reconstruction method (CCA) outperforms the CSP method for the - less practical - long decision windows > 20 s. As the RGC method outperforms the CSP method on almost all decision window lengths, the region in which the CCA method is the best has decreased to the range > 40 s. If one would construct an AAD algorithm combining both approaches (RGC + CCA), the envelope would largely, and in the most important regions, be dominated by the RGC method.

The per-subject MESD values are all < 5 s (except for two outliers due to poorer performing subjects with MESDs > 5 s, but < 24 s), with median MESD $= 2.26$ s and $[25, 75]\%$-quantiles $= [2.13, 2.62]$s. Note that the MESD values of the CCA method are all above 5 s (due to poor performance at short decision windows, median MESD $= 16.07$ s). The median MESD of the CSP method is $= 2.34$ s, with $[25, 75]\%$-quantiles $= [2.12, 2.61]$s. For the CSP and RGC method, as there are still relatively high accuracies on the very short decision windows, the optimal trade-off point between AAD accuracy and decision window length is very often located at the shortest decision window lengths. As both methods have very similar accuracies there (see Figure 1), the MESD values are also very similar across both methods, with similar median values. Furthermore, a paired Wilcoxon signed-rank test ($n = 16, p = 0.0627$) shows no significant difference between both methods.

4. DISCUSSION AND CONCLUSION

We have shown that the proposed RGC is capable of outperforming the state-of-the-art CSP method to decode the directional focus of auditory attention by $\approx 6\%$ on most decision window lengths. However, two limitations are to be noted. Firstly, the RGC method performs similarly to the CSP method on very short decision windows (see Figure 1), due to the worse covariance matrix estimation on small sample sizes. As the MESD values indicate that these very short decision windows are most relevant in the context of attention switching, the RGC method achieves a similar overall MESD as the CSP method. Furthermore, this RGC method has due to the TSM in (6) a higher computational load than applying a simple spatial filter. Both limitations need to be considered for the real-time AAD application in neuro-steered hearing devices.

To conclude, the large increase in AAD accuracy compared to the state-of-the-art CSP method makes the proposed method a good candidate to decode the auditory attention, given that it also outperforms the stimulus reconstruction (i.e., CCA) approach for decision window lengths below 40 s. This makes the RGC-based decoding of the directional focus of auditory attention one of the best AAD methods to date.

5. REFERENCES

[1] J. A. O'Sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. Shamma, and E. 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, 2014.

[2] S. Geirnaert, T. Francart, and A. Bertrand, "Fast EEG-based decoding of the directional focus of auditory attention using common spatial patterns," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 5, pp. 1557–1568, 2020.

[3] S. Geirnaert, S. Vandecappelle, E. Alickovic, A. de Cheveigné, E. C. Lalor, B. T. Meyer, S. Miran, T. Francart, and A. Bertrand, "EEG-based Auditory Attention Decoding: Towards Neuro-Steered Hearing Devices," *IEEE Signal Processing Magazine*, 2021.

[4] S. Geirnaert, T. Francart, and A. Bertrand, "An Interpretable Performance Metric for Auditory Attention Decoding Algorithms in a Context of Neuro-Steered Gain

Control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 307–317, 2020.

[5] N. Das, J. Zegers, H. Van hamme, T. Francart, and A. Bertrand, "Linear versus deep learning methods for noisy speech separation for EEG-informed attention decoding," *Journal of Neural Engineering*, vol. 17, no. 4, p. 046039, 2020.

[6] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.

[7] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass Brain-Computer Interface Classification By Riemannian Geometry," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2012.

[8] ——, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *Neurocomputing*, vol. 12, pp. 172–178, 2013.

[9] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing Spatial Filters for Robust EEG Single-Trial Analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2007.

[10] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.

[11] M. Congedo, B. Afsari, A. Barachant, and M. Moakher, "Approximate Joint Diagonalization and Geometric Mean of Symmetric Positive Definite Matrices," *PLoS ONE*, vol. 10, no. 4, p. 121423, 2015.

[12] A. de Cheveigné, D. D. E. Wong, G. M. Di Liberto, J. Hjortkjær, M. Slaney, and E. C. Lalor, "Decoding the auditory brain with canonical component analysis," *NeuroImage*, vol. 172, pp. 206–216, 2018.

[13] N. Das *et al.*, "Auditory Attention Detection Dataset KULeuven," Zenodo, 2019. [Online]. Available: https://zenodo.org/record/3997352

[14] W. Biesmans, N. Das, T. Francart, and A. 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, 2017.