Multi-task Feature Learning for EEG-based Emotion Recognition Using Group Nonnegative Matrix Factorization

Ayoub Hajlaoui*†, Mohamed Chetouani* and Slim Essid†

*Institut des Systèmes Intelligents et de Robotique, Université Pierre et Marie Curie

† LTCI, Télécom ParisTech, Université Paris–Saclay

Abstract—**Electroencephalographic sensors have proven to be promising for emotion recognition. Our study focuses on the recognition of valence and arousal levels using such sensors. Usually, ad hoc features are extracted for such recognition tasks. In this paper, we rely on automatic feature learning techniques instead. Our main contribution is the use of Group Nonnegative Matrix Factorization in a multi-task fashion, where we exploit both valence and arousal labels to control valence-related and arousal-related feature learning. Applying this method on HCI MAHNOB and EMOEEG, two databases where emotions are elicited by means of audiovisual stimuli and performing binary inter-session classification of valence labels, we obtain significant improvement of valence classification F1 scores in comparison to baseline frequency-band power features computed on predefined frequency bands. The valence classification F1 score is improved from 0.56 to 0.69 in the case of HCI MAHNOB, and from 0.56 to 0.59 in the case of EMOEEG.**

Index Terms—**Electroencephalography, Valence, Arousal, Nonnegative Matrix Factorization, Group NMF, Common Spectral Patterns**

1. Introduction

Electroencephalographic (EEG) recording has attracted the attention of researchers in the field of affective computing as part of the effort to perform human-behaviour analysis tasks, especially automatic emotion recognition. Indeed, EEG signals have proven to be a precious clue in emotion classification [1]. Emotions are often represented in a two-dimensional valence-arousal space [2], which respectively describe the pleasure or displeasure felt by a person and her degree of “excitement”. Though the classification of valence and arousal levels can be performed independently one from another, it has been shown that multi-task learning, that is learning to classify valence and arousal labels jointly, can improve emotion classification performance [3], [4]. The interdependence between valence and arousal [5] can explain such an improvement.

In this work, we take this interdependence into account as we devise a novel feature learning strategy, instantiating the multi-task learning paradigm, that proves to be effective for automatic valence/arousal classification from EEG recordings. Our approach is clearly different from the ones adopted in previous instantiations of the multi-task learning paradigm in the context of physiological signal processing, where the focus was rather on coping with subject variability, using the data of other subjects to learn the model of a target subject [6].

Our method relies on nonnegative matrix factorization (NMF) that is applied to a time-frequency representation of the EEG data, motivated by several studies showing the importance of the brain activity in predefined frequency bands, such as the β or γ bands, in emotional and cognitive processes [7], [8]. NMF presents the advantage of not relying on such ad hoc features.

The novelty of this paper mainly lies in the exploitation of arousal labels to control valence-related feature learning using Group Nonnegative Matrix Factorization (GNMF), motivated by previous works on valence/arousal interdependencies [9].

Our NMF-based models were thus applied on the HCI MAHNOB and EMOEEG [10], [11]. They present synchronized physiological recordings, among which are EEG recordings of subjects while short video stimuli were presented to them. We chose to perform single-channel based emotion classification since it opens the way to easier applicability in real-world scenarios with more lightweight devices than full headsets.

The remainder of the paper is organised as follows. Section 2 offers a brief review of the state-of-the-art in EEG data recording and feature extraction for emotion classification. Section 3 presents our NMF-related architecture for multi-task feature learning. Section 4 details the experimental setup and system hyper-parameter tuning, whereas Section 5 presents the classification results we obtained and discusses them.

2. State of the art

Data recording. To address the challenge of assessing the generalization abilities of EEG-based classification systems across subjects, some existing databases such as HCI MAHNOB and DEAP [10], [12] included a relatively

ISBN 978-90-827970-1-5 © EURASIP 2018

high number of participants (respectively 27 and 32 each). Others such as eNTERFACE'06 and EMOEEG [11], [13] (with respectively 5 and 8 participants) chose to sacrifice the number of subjects for the benefit of another consideration. Because of the subject-dependent nature of EEG responses [14], the number of trials or recording sessions per participant is an important factor. For instance, in the eNTERFACE'06 database, there are 30 trials per subject. As for the HCI MAHNOB and the EMOEEG datasets, on which we have performed our study, they respectively contain 50 and 20 trials per session. A usual order of magnitude for the duration of one trial is 15-20 seconds.

Elicited emotion classification tasks are known to be complex. Most works focus on the classification of emotions elicited by blocks of images [13]. A more complex task is the classification of emotional states elicited by audiovisual stimuli, as done in [10], where the recognition of auto-assessed emotional states elicited by means of short video excerpts was performed.

Finally, valence and arousal are usually discretized. Most discretizations decompose each axis into two or three labels, respectively low/high and low/average/high.

Feature extraction and feature learning. Commonly extracted features are the spectral power for each considered electrode in specific frequency bands (theta, slow alpha, alpha, beta, gamma) that are well known for their role in emotional and cognitive processes [8], [15]. Spectral moments of different orders and heuristic spectral shape descriptors have also been used [10], [16]. In the multi-channel case, the spectral power asymmetry between specific pairs of electrodes can be computed in the frequency bands mentioned earlier [17]. Other approaches such as Common Spatial Patterns (CSP) [18]–[20] rather focus on the spatial aspect of the activity on the skull.

What all these representations have in common is the fact they rely on expert knowledge and a feature engineering effort. The new trend in machine learning is to learn representations adapted to the subsequent classification stage. Along this line, Nonnegative Matrix Factorization [21], which is an an unsupervised feature extraction method, has been used for EEG-based motor imagery classification tasks [22].

Multi-task feature learning has been used in a subject-to-subject transfer fashion, where priors for feature dictionaries are shared across different subjects. Kang et al. [6] used multi-task feature learning in such a way, improving classification accuracy obtained from CSP filters on a motor imagery task. They obtained an average accuracy of 0.54 across all subjects, whereas the average accuracy reached almost 0.57 in the multi-task feature learning case.

3. NMF-based architecture for multi-task feature learning

3.1. Non-negative Matrix Factorization (NMF)

Let us recall the NMF model [21] that we use for feature learning. The idea is to approximate a given non-negative

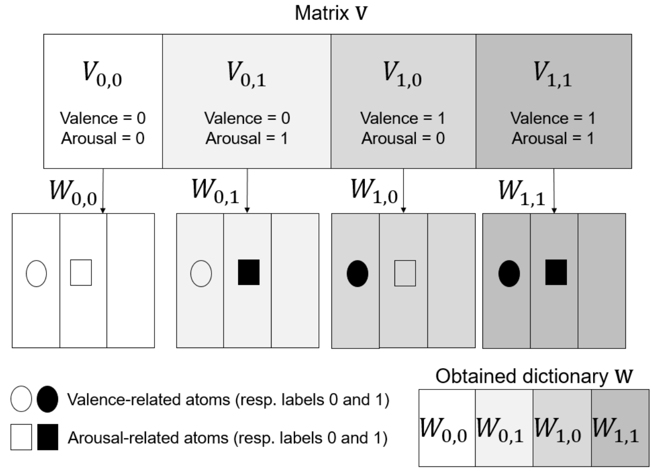

Figure 1. Learning a dictionary matrix **W** with GNMF

matrix $\mathbf{V} \in \mathbb{R}_+^{F \times N}$ by a product of non-negative matrices $\tilde{\mathbf{V}} = \mathbf{WH}$ with $\mathbf{W} \in \mathbb{R}_+^{F \times K}$ and $\mathbf{H} \in \mathbb{R}_+^{K \times N}$. Assuming **V** represents observations (the activity of F frequency bands across T time frames), **W** is a dictionary of K *atoms* (or latent variables) whose activation in time is indicated by the rows of the activation matrix **H**. In the EEG-based classification problem, **V** can be the Power spectrogram related to the activity at one particular electrode or a linear combination of electrodes.

The optimal **W** and **H** minimize a divergence between **V** and **WH**, which is the sum of scalar divergences between the coefficients of **V** and the coefficients of **WH**. The Itakura-Saito (IS) divergence [23] has the desirable property of scale invariance, meaning that in the minimization of the divergence between **V** and **WH**, no particular advantage will be given to high-value coefficients of **V** at the expense of the low-value ones. The optimization problem can be solved using so-called multiplicative update algorithms [24].

3.2. Multi-task GNMF-based feature learning

We exploit the Group-NMF (GNMF) model, which allows one to account for similarity between sub-groups of atoms [25]. A sub-group of **V** is a subset of columns of **V** that were selected according to specific conditions. For instance, if **V** stores the EEG signal recorded from a subject while he/she watches a series of video stimuli, the chunk of signal corresponding to a specific stimulus is a sub-group of **V**.

Given a definition of sub-groups of **V**, GNMF adds to NMF some constraints that force the atoms extracted from the same sub-group of **V** to be similar, by adding to the objective function to be minimized a term controlling the distances between the atoms of the same sub-group. Within this framework, Serizel et al. proposed a model which can tackle two types of dependencies [26], that is two kinds of groups at the same time, with a focus on modeling session dependency. Their model inspired our derivations of GNMF.

ISBN 978-90-827970-1-5 © EURASIP 2018

We call session the recording of a given subject at a given time of the day. A subject can take more than one session. In our multi-task approach, the considered sub-groups are signal chunks corresponding to different pairs of valence and arousal labels. GNMF uses valence and arousal information to compute the atoms of the session dictionary **W** across sessions. This dictionary is the concatenation of sub-dictionaries related to each pair of valence/arousal labels.

Let $\mathcal{V}\mathrm{al} = \{0, 1\}$ and $\mathcal{A}\mathrm{rsl} = \{0, 1\}$ be respectively the sets of valence and arousal labels. Let $V_{v,a}$ be the sub-matrix of V_{train} corresponding to valence label v and arousal label a (that is to say, the chunk of the data corresponding to the valence annotation v and the arousal annotation a). Let $W_{v,a}$ be the sub-dictionary corresponding to valence label v and arousal label a. In such sub-dictionary:

- $W^{\mathrm{val}}_{v,a}$ is composed of K_{val} atoms that must be similar to the other W^{val}_{v,a_2} $(v_2 \neq v)$
- $W^{\mathrm{aro}}_{v,a}$ is composed of K_{sess} atoms that must be similar to other $W^{\mathrm{sess}}_{v_2,s}$ $(v_2 \neq v)$
- $W^{\mathrm{res}}_{v,a}$ is composed of K_{res} atoms upon which no additional constraints are added

In Figure 1, valence-dependent atoms are constrained to show some similarity between $W_{0,0}$ and $W_{0,1}$ on the one hand, and between $W_{1,0}$ and $W_{1,1}$ on the other hand (same valence, different arousals). Likewise, another constraint lies between arousal-dependent atoms.

Let $D(.|.)$ denote the matrix divergence to minimize between **V** and its approximation **V'**, and $D_2(.|.)$ another divergence. Then the objective function to minimize in the case of this specific GNMF model is:

$$\mathcal{F}_{GNMF} = \sum_{v=0}^{1}\sum_{a=0}^{1} D_1(V_{v,a}|W_{v,a}H_{v,a})$$

$$+ \quad \lambda_{\mathrm{val}} \sum_{v=0}^{1} D_2(W^{\mathrm{val}}_{v,0}|W^{\mathrm{val}}_{v,1}) + \lambda_{\mathrm{aro}} \sum_{a=0}^{1} D_2(W^{\mathrm{aro}}_{0,a}|W^{\mathrm{aro}}_{1,a})$$

The term $\lambda_{\mathrm{valence}}$ controls the similarity between sub-dictionaries corresponding to the same valence labels, whereas $\lambda_{\mathrm{arousal}}$ controls the similarity between sub-dictionaries corresponding to the same arousal labels. We chose $\lambda_{\mathrm{valence}} > \lambda_{\mathrm{arousal}}$ for the valence classification task, and vice versa. Following the framework described in [26], similarities between valence and arousal-related atoms are expressed in terms of Euclidean distance. The other atoms, or residual atoms, allow a degree of freedom in the GNMF minimization, so that the additional sub-group distances constraint does not hamper the divergence minimization between **V** and **WH**.

4. Experiments

4.1. Database and hyper-parameters

Table 1 summarizes the HCI MAHNOB and EMOEEG databases characteristics, on which our study was performed. After watching a video stimulus, the subject assesses his/her global valence and arousal levels during the stimulation (static annotation). The values assessed are between 1 (very negative) and 9 (very positive). We binarize such values considering two classes for each modality: low and average/high. To minimize class-imbalance in the training data, the low-value class corresponds to values from 1 to 4 in the case of HCI MAHNOB, but only from 1 to 3 in the case of EMOEEG. The main reason for such a difference is that EMOEEG annotations are more biased towards negative values, due to a focus on negative emotions in the choice of the stimuli [11].

TABLE 1. HCI MAHNOB AND EMOEEG CHARACTERISTICS

Database	HCI	EMOEEG
Nb of sessions (used for classification)	24	8
Nb of video stimuli per session	20	50
Duration of a video stimulus	$\tilde{2}$5s	15s
Nb of electrodes	32	20

TABLE 2. SPECTROGRAM AND UNSUPERVISED NMF PARAMETERS

Downsampling factor	2 (256 Hz)
Considered frequencies	4 to 45 Hz
Tested electrode combinations	[Cz Pz Cz-Pz Fz-Pz]
Divergence	Itakura Saito
Number of initializations	5

Table 2 reports the parameters used for the computation of the power spectrograms and the NMF. The chosen electrode was the central electrode Cz. The number of NMF random initializations reported in Table 3 improves the performance of this non-convex optimization. The number of atoms K was empirically set to 100 in the unsupervised NMF case.

An equivalent number of atoms K_{total} was used for GNMF, as shown in Table 3. $\lambda_{\mathrm{valence}}$ and $\lambda_{\mathrm{arousal}}$ are the weights respectively given to the conditions of similarity of atoms which share the same valence label and atoms which share the same arousal label. Serizel et al.'s code was used for GNMF [26].

4.2. Feature learning

The matrix **V** is a concatenation of power spectrogram matrices over the time dimension, corresponding to a concatenation on trials and sessions. One part, V_{train}, of **V** is used as the training feature matrix, whereas the remaining part V_{test} is used as the test feature matrix. We consider

TABLE 3. GNMF PARAMETERS IN THE CASE OF VALENCE CLASSIFICATION (CONTROLLED BY BOTH VALENCE AND AROUSAL LABELS)

Divergence	Itakura-Saito
$K_{\mathrm{valence}}, K_{\mathrm{arousal}}, K_{\mathrm{residual}}$	15, 5, 5
K_{total}	(15+5+5)×4=100
$\lambda_{\mathrm{valence}}, \lambda_{\mathrm{arousal}}$ (HCI)	10^{-4}, 10^{-5}
$\lambda_{\mathrm{valence}}, \lambda_{\mathrm{arousal}}$ (EMOEEG)	0.5, 0.05

ISBN 978-90-827970-1-5

TABLE 4. F1-SCORES FOR EMOTION CLASSIFICATION OBTAINED FOR DIFFERENT METHODS ON HCI MAHNOB

Dimension	Valence	Arousal
Band power baseline	0.56	0.55
NMF	0.68	0.56
GNMF	0.69	0.59

TABLE 5. F1-SCORES FOR EMOTION CLASSIFICATION OBTAINED FOR DIFFERENT METHODS ON EMOEEG

Dimension	Valence	Arousal
Band power baseline	0.56	0.51
NMF	0.57	0.53
GNMF	0.59	0.51

an inter-session scheme, where $\mathbf{V}$ is the concatenation of power-spectrogram matrices over the trials of all sessions. V_{train} is composed of the recordings of all sessions but one, whereas V_{test} is composed of the recording of the remaining session (with a view to performing leave-one-session-out classification).

We followed the usual procedure for NMF-based classification: we jointly learned the dictionary $\mathbf{W}$ (as shown in Figure 1) and computed the training feature matrix H_{train} from the training matrix V_{train}, after which we computed H_{test} by performing unsupervised NMF with $\mathbf{W}$ fixed on H_{test}.

4.3. Classifier training and evaluation

While the feature learning stage was multi-tasked with GNMF, single-task classifiers were used, that is classifiers for valence and arousal were learned separately. For each session, NMF and GNMF were performed on the spectrogram matrix V_{train} corresponding to all sessions but one. This process was repeated for each session, so as to perform leave-one-session-out cross-validation. The training outputs were used to train a linear SVM classifier, for which the best values of the regularization parameter C were found by performing a grid search on the following grid : $\{2^{-5}, 2^{-4.5}, ..., 2^{4.5}, 2^{5}\}$.

5. Results

Tables 4 and 5 present the obtained F1 scores for the valence and arousal classification tasks on HCI MAHNOB and EMOEEG. In addition to a frequency-band power feature baseline used in [10], that is to say θ, slow α, α, β, γ Power Spectral Density (PSD) and PSD asymmetry, and to an unsupervised NMF-based classification, valence classification was performed using GNMF explicitly accounting for both valence and arousal variability. In other words, the variability related to the arousal label is used to improve the performance of the valence classification task. Therefore, the task of classifying valence labels is controlled by both valence and arousal labels. For arousal classification, GNMF also accounted for valence and arousal variability.

In the case of HCI MAHNOB, NMF performs better than the baseline in terms of valence classification, with an F1 score of 0.68 against 0.56. GNMF yields a score of 0.69, making a slight improvement compared to NMF. As for arousal classification, Table 4 shows that GNMF (0.59) yields a more significant improvement of the baseline (0.55) than NMF (0.56).

In the case of EMOEEG, GNMF slightly improves the valence classification score, as shown in Table 5. However, the weak arousal F1 score obtained with the baseline is not improved.

An interesting comparison point between the two databases is the fact that the baseline performed better valence classification for EMOEEG than for HCI, whereas GNMF performed much better for HCI. The reason why EMOEEG did not benefit from GNMF as HCI could be that the arousal annotations are less reliable in EMOEEG, as suggested by the weaker baseline arousal classification. Such annotations are used not only for GNMF-based arousal classification, but also for GNMF-based valence classification, which would explain why the valence classification score stagnates for EMOEEG.

Even if some encouraging results were obtained using arousal labels to control valence-related feature learning, the way to further improvement is clear, especially for the EMOEEG dataset.

6. Conclusion and future work

Two pre-existing multimodal databases, namely the HCI MAHNOB and EMOEEG datasets, have been exploited to test a novel EEG-based emotion recognition system. We have used GNMF for feature learning, using valence labels to control the learning of an arousal classifier, and vice-versa. This is realized in a multi-task fashion. F1-scores obtained prove that the use of arousal labels to control the valence classification task improves its performance. However, the arousal classification task is more challenging.

Future work will seek the extension of such classification tasks to multi-label cases, with a particular attention to ensure that the constraint of similarity is bigger for GNMF sub-dictionaries corresponding to closer labels than for sub-dictionaries corresponding to distant ones. More specifically, the valence classification task is more challenging when the arousal is low, which has to be tackled with a particular focus on sub-dictionaries corresponding to low arousal labels.

ISBN 978-90-827970-1-5 © EURASIP 2018

Acknowledgments

This research has been supported by the Laboratory of Excellence SMART (ANR-11-LABX-65) supported by French State funds managed by the ANR within the Investissements d'Avenir programme (ANR-11-IDEX-0004-02).

References

[1] V. Bajaj and R. Pachori, "Human emotion classification from eeg signals using multiwavelet transform," in *Medical Biometrics, 2014 International Conference on*. IEEE, 2014, pp. 125–130.

[2] A. Mehrabian and J. Russell, *An approach to environmental psychology*. the MIT Press, 1974.

[3] M. Kandemir, A. Vetek, M. Gönen, A. Klami, and S. Kaski, "Multi-task and multi-view learning of user state," *Neurocomputing*, vol. 139, pp. 97–106, 2014.

[4] M. Abadi, A. Abad, R. Subramanian, N. Rostamzadeh, E. Ricci, J. Varadarajan, and N. Sebe, "A multi-task learning framework for time-continuous emotion estimation from crowd annotations," in *Proceedings of the 2014 International ACM Workshop on Crowdsourcing for Multimedia*. ACM, 2014, pp. 17–23.

[5] P. Kuppens, F. Tuerlinckx, M. Yik, P. Koval, J. Coosemans, K. J. Zeng, and J. A. Russell, "The relation between valence and arousal in subjective experience varies with personality and culture," *Journal of personality*, 2016.

[6] H. Kang and S. Choi, "Bayesian multi-task learning for common spatial patterns," in *Pattern Recognition in NeuroImaging (PRNI), 2011 International Workshop on*. IEEE, 2011, pp. 61–64.

[7] N. Rowland, M. Meile, S. Nicolaidis *et al.*, "Eeg alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes," 1985.

[8] M. Li and B.-L. Lu, "Emotion classification based on gamma-band eeg," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2009, pp. 1223–1226.

[9] F. Schweitzer and D. Garcia, "An agent-based model of collective emotions in online communities," *The European Physical Journal B*, vol. 77, no. 4, pp. 533–545, 2010.

[10] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.

[11] A.-C. Conneau, A. Hajlaoui, M. Chetouani, and S. Essid, "Emoeeg: A new multimodal dataset for dynamic eeg-based emotion recognition with audiovisual elicitation," in *Signal Processing Conference (EUSIPCO), 2017 25th European*. IEEE, 2017, pp. 738–742.

[12] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[13] A. Savran, K. Ciftci, G. Chanel, J. Mota, L. Hong Viet, B. Sankur, L. Akarun, A. Caplier, and M. Rombaut, "Emotion detection in the loop from brain signals and facial images," 2006.

[14] M. Thulasidas, C. Guan, and J. Wu, "Robust classification of eeg signal for brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 1, p. 24, 2006.

[15] R. Yuvaraj, M. Murugappan, N. Ibrahim, M. Omar, K. Sundaraj, K. Mohamad, R. Palaniappan, and M. Satiyan, "Emotion classification in parkinson's disease by higher-order spectra and power spectrum features using eeg signals: A comparative study," *Journal of integrative neuroscience*, vol. 13, no. 01, pp. 89–120, 2014.

[16] A.-C. Conneau and S. Essid, "Assessment of new spectral features for eeg-based emotion recognition," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 4698–4702.

[17] Y.-J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, and Y. Shi, "Real-time movie-induced discrete emotion recognition from eeg signals," *IEEE Transactions on Affective Computing*, 2017.

[18] S. Koelstra, A. Yazdani, M. Soleymani, C. Mühl, J.-S. Lee, A. Nijholt, T. Pun, T. Ebrahimi, and I. Patras, "Single trial classification of eeg and peripheral physiological signals for recognition of emotions induced by music videos," in *International Conference on Brain Informatics*. Springer, 2010, pp. 89–100.

[19] W. Samek, M. Kawanabe, and K.-R. Muller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 50–72, 2014.

[20] A. Dupres, F. Cabestaing, and J. Rouillard, "Sélection par un expert humain des intervalles temps-fréquence dans le signal eeg pour les interfaces cerveau-ordinateur," in *9e Conférence Handicap (Handicap 2016)*, 2016, pp. 45–50.

[21] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[22] H. Lee, Y.-D. Kim, A. Cichocki, and S. Choi, "Nonnegative tensor factorization for continuous eeg classification," *International journal of neural systems*, vol. 17, no. 04, pp. 305–317, 2007.

[23] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.

[24] D. Lee and H. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.

[25] H. Lee and S. Choi, "Group nonnegative matrix factorization for eeg classification." in *AISTATS*, 2009, pp. 320–327.

[26] R. Serizel, S. Essid *et al.*, "Group nonnegative matrix factorisation with speaker and session variability compensation for speaker identification," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 5470–5474.