Is DTW resilient to noise and effective for EEG functional connectivity assessment?

Maxime BEDOIN[1], Nesma HOUMANI[1], Rita YABOURI[1],
Jérôme BOUDY[1], Kiyoka KINUGAWA[2]

[1] Telecom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France
[2] Sorbonne Universite, CNRS, UMR Biological Adaptation and Aging
AP-HP, Charles Foix Hospital, F-94200 Ivry-sur-Seine, France

[maxime_bedoin, nesma.houmani]@telecom-sudparis.eu

Résumé

Les signaux EEG sont souvent bruités, c'est pourquoi, pour étudier des caractéristiques telles que la connectivité fonctionnelle, nous devons nous assurer que nos mesures de connectivité sont résilientes au bruit. Cette étude propose d'exploiter la distance élastique (DTW, Dynamic Time Warping) comme mesure de connectivité fonctionnelle. Nous montrons que la DTW est plus robuste au bruit que la mesure basée sur l'estimation de phase (PLI, Phase Lag Index), cette dernière étant très utilisée dans la littérature. Nous montrons également que la DTW avec contrainte de Sakoe-Chiba permet une caractérisation fine de la connectivité comparée à la DTW sans contrainte.

Mots-clés

Électroencéphalographie, Bruit, Déformation Temporelle Dynamique, Connectivité fonctionnelle, Maladies neurodégénératives.

Abstract

EEG recordings are often noisy, therefore to study features such as functional connectivity, we need to ensure that our connectivity measures are resilient to noise. This study proposes to exploit elastic distance (DTW, Dynamic Time Warping) as a measure of functional connectivity. We show that DTW is more robust to noise than the measure based on phase estimation (PLI, Phase Lag Index). The latter is widely used in the literature. We also show that DTW with Sakoe-Chiba constraint leads to a finer characterization of connectivity than DTW without constraint.

Keywords

Electroencephalography, Noise, Dynamic Time Warping, Functional connectivity, Neurodegenerative diseases.

1 Introduction

Electroencephalography (EEG) is increasingly recognized as a useful non-invasive neuroimaging technique to measure cortical neurophysiological activity. Indeed, EEG provides an excellent time resolution that is crucial for analyzing fast brain dynamics. Nevertheless, EEG signals are non stationary and often subject to environmental noise, possible acquisition problems in the EEG settings, as well as physiological artifacts such as involuntary contractions of the eyes and the heartbeats of the subjects.

The analysis of EEG signals in the context of neurodegenerative diseases often relies on the computation of functional connectivity. Connectivity is a way to track the dynamics between the signals recorded with different electrodes, thus reflecting the functional interactions among cortical processes. It is generally computed pairwise between EEG signals. A good estimation of connectivity is highly dependent on the level of noise in EEG signals.

It is well admitted that connectivity decreases in patients suffering from dementia, particularly Alzheimer's Disease, comparatively to age-matched healthy subjects [Briels et al., 2020, Abazid, 2022]. Several connectivity measures have been proposed in the literature, such as Phase Lag Index (PLI) [Stam et al., 2007], Amplitude Envelope Correlation [Briels et al., 2020], and Mutual Information [Jeong et al., 2001]. Our research team currently investigates the potential use of Dynamic Time Warping [Senin, 2008] distance (DTW) as a connectivity measure to distinguish between several cognitive disorders.

DTW is an elastic distance that allows to dynamically match two signals in order to follow the temporal fluctuations in such nonlinear signals. It has already been used in EEG analysis, but mainly for artefact detection [Shaw et al., 2017]. To assess the effectiveness of DTW as a connectivity measure in the context of neurodegenerative diseases, we need to study its resilience to noise, since EEG signals are prone to a low signal-to-noise ratio (SNR).

In this work, we intend to study the robustness of DTW to variations in the levels of noise, comparatively to the largely used PLI in the context of neurodegenerative diseases. Also, we propose to investigate the impact of adding a Sakoe-Chiba [Sakoe and Chiba, 1978] constraint on the connectivity characterization. We conduct such analyses by considering three populations : patients with Subjective Cognitive Impairment (SCI), patients with Mild Alzheimer's Disease (AD) and patients with Vascular Dementia (VaD).

TABLE 1 – Characteristics of patients in the cohort. MMSE : Mini-Mental State Examination ; M : Mean, SD : Standard Deviation, BZD : Benzodiazepin use, ADP : Antidepressant use, NL : Neuroleptic use, HN : Hypnotic use.

	SCI (n=32)	AD (n=46)	VaD (n=17)
Age (M ± SD)	68.2 ± 10.4	82.0 ± 8.6	80.0 ± 7.6
Female (%)	81.8%	67.4%	35.3%
MMSE (M ± SD)	28.3 ± 1.6	19.0 ± 5.6	20.0 ± 5.1
BZD use (%)	4 (12.5%)	6 (13.0%)	3 (17.6%)
ADP use (%)	5 (15.6%)	14 (30.4%)	3 (17.6%)
NL use (%)	0 (0%)	1 (2.0%)	1 (5.9%)
HN use (%)	5 (15.6%)	4 (8.7%)	0 (0%)

2 Material and Methods

2.1 Database Description

The cohort used to conduct this retrospective study is composed of resting-state EEG data of 32 SCI patients, 46 Mild AD patients and 17 VaD patients, acquired in a clinical setting at Charles-Foix Hospital (Ivry-sur-Seine, France). Table 1 presents their demographic and clinical characteristics. The study was approved by the institutional review board of the local Ethics Committee Paris 6 and the Ethics Committee of Sorbonne University (N°CER-2021-064).

EEG signals were recorded at rest with eyes closed at a frequency sampling of 256Hz during 20 minutes at least, by taking care that the patients were not falling asleep. Thirty electrodes were used, placed on the scalp according to the 10-20 international system : Fp1, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FC7, FC4, FT8, T3, C3, Cz, C4, T4, TP7, CP3, CPz, CP4, TP8, T5, P3, Pz, P4, T6, O1, Oz, and O2.

The EEG signals were visually inspected to discard the parts of the signals presenting artifacts. Thereby, continuous signals of 20 seconds free from artifacts were then kept for the study. The obtained 20s EEG signals were then band-pass filtered with a third-order Butterworth filter in the frequency range [1-30] Hz, as well in the usual frequency bands of interest : delta [1-4] Hz, theta 1 [4-6] Hz, theta 2 [6-8] Hz, alpha 1 [8-10], alpha 2 [10-12] Hz and beta [12-30] Hz.

2.2 Methodology

For each patient, we compute his/her functional connectivity between all pairs of EEG signals with DTW and PLI that we present in the following sections.

2.2.1 Phase-Lag Index

PLI is generally computed between a pair of signals according to the following formula [Stam et al., 2007] :

$$PLI = | < sign(\Delta\Phi(t_k)) > |$$

where $\langle . \rangle$ represents the mean (over index k), "sign" denotes the signum function that discards phase difference of 0 mod π, $|.|$ is the absolute value and $\Delta\Phi(t_k)$ indicates the phase difference between two time series at time t_k.

2.2.2 Dynamic Time Warping Distance

DTW distance [Senin, 2008] is an elastic matching metric obtained by a dynamic programming algorithm that quantifies the similarity between two time series showing a potential temporal drift or shift.

The computation of DTW distance between two EEG signals S_1 and S_2 of length *N*, consists in a recursive construction of the cost matrix. By design, the last computed value, which has the coordinates (*N*, *N*) contains the value of DTW between the two signals.

In order to improve the above procedure for DTW calculation, it is possible to apply a warping window to limit the shifting that is tolerated when matching observations in the two EEG signals. This method is called the Sakoe-Chiba band constraint [Sakoe and Chiba, 1978].

2.2.3 Study of the Resilience to Noise

To compare DTW and PLI in terms of their resilience to noise, we propose to study the impact of adding Gaussian white noise to original EEG signals on the computed connectivity values for the 32 SCI patients.

This way, we calculate all the pairwise connectivity values for each patient on the original signals, as well as on the six generated signals with different SNR values (0dB, 5dB, 10dB, 20dB, 40dB, 60dB) obtained as follows :

$$SNR_{dB} = 10log_{10}(\frac{P_{signal}}{P_{noise}})$$

Then, we compute connectivity with PLI and DTW, considering for the latter two cases : unconstrained DTW (referred as *DTW*), and DTW with a Sakoe-Chiba constraint (denoted as *DTW_6*) that integrates a warping window size fixed to six.

To assess the effect of added noise on the connectivity values, we measure the deviation between the (30, 30) connectivity matrix obtained on the noisy signals and that obtained from the original signal, by computing the average Euclidean Distance between such matrices for each patient.

2.2.4 Study of the impact of Sakoe-Chiba Constraint

We propose to analyze the quality of characterization of the connectivity in SCI, AD and VaD populations on original signals, using DTW with and without Sakoe-Chiba constraint.

Thereby, after computing the (30, 30) DTW matrices, we take the inverse of the obtained distances to have similarity values (connectivity values), and normalize them between '0' and '1' per electrode with min-max procedure.

Then, we define 8 brain regions for connectivity analysis : prefrontal/frontal (Fp1, Fp2, Fz), frontal left (F7, F3, FT7, FC3), central (FCz, C3, CZ, C4), frontal right (F4, F8, FC4, FT8), temporal left (T3, TP7, CP3, T5), parietal (P3, Pz, P4), temporal right (T4, CP4, TP8, T6), and occipital (O1, Oz, O2) region. We estimate the intra-region connectivity by averaging the connectivity values on all pairs of EEG signals associated to the considered region. We also estimate inter-regions connectivity by averaging the connectivity values on all pairs of EEG signals associated to such regions. Hence, for each patient, we obtain a (8, 8) symmetric region-based connectivity matrix, where the diagonal contents the 8 intra-region connectivity values, the upper or lower triangular matrix contains the 28 inter-regions connectivity values.

We also define three types of connectivity ranges : (i) "short-range" connectivity corresponding to the intra connectivity values ; (ii) "mid-range" connectivity defined as the connectivity between two regions that are 1-hop neighbors ; and (iii) "long-range" connectivity that gathers all the connectivities between two regions that are 2-hops neighbors or more.

3 Experimental Results

3.1 Assessment of Resilience to Noise

Figure 1 displays the deviation values computed with Euclidean distance between the connectivity matrix obtained on noisy signals and that obtained on the original signal, using PLI, *DTW* and *DTW_6*. We report the average deviation values on the 32 SCI patients, as well as the 95% confidence intervals. For clarity of visualization, we display the values on the Y axis with a square root scale.

We first observe that as the amount of noise in data increases (from SNR=60dB to SNR=0dB), the deviation values increase for all connectivity measures, albeit more markedly for PLI than the two DTW configurations until SNR=20dB. For very low SNR (10dB and 0dB), the behavior of the three measures become similar.

Besides, we observe that PLI is more affected than *DTW_6* and *DTW* for all frequency bands even for low noise (SNR=60dB). Such effect increases for increased noise (SNR=40dB), while both DTW configurations show more stable values for SNR=60dB and SNR=40dB. Therefore, DTW-based connectivity measure is more resilient to noise than PLI ; however, we do not observe a difference between DTW computed with and without Sakoe-Chiba constraint in terms of robustness to noise.

3.2 Impact of Sakoe-Chiba Constraint on the Characterization of Connectivity

We report in this section only the results obtained in theta 2 but similar conclusions can be drawn from the other bands. Figure 2 shows the (30, 30) connectivity matrices of three SCI patients obtained on the one hand with *DTW_6* (on the left of Figure 2), and on the other hand with *DTW* (on the right of Figure 2).

We note that matrices obtained with *DTW* are less contrasted than those obtained with *DTW_6*, presenting more uniform connectivity values with high values concentrated in the top right corner. This first result may indicate that *DTW_6* gives rise to a fine characterization of functional connectivity.

To go deeper in our analysis, we generate the connectivity matrices with *DTW* and *DTW_6* for all SCI, AD and VaD patients. Then, we compute 8 intra- and 28 inter-region connectivity values per patient, as explained in Section 2.2.4.

To investigate the impact of using *DTW* and *DTW_6* on the overall population structure and characterization, we apply Principal Component Analysis (PCA) [Maćkiewicz and Ratajczak, 1993] on the obtained feature vectors of SCI, AD and VaD patients, each contains 36 connectivity values.

Figures 3.a and 3.b show SCI, AD and VaD patients projected onto the PCA space spanned by the first two principal components, based on connectivity values computed with *DTW_6* and *DTW*, respectively. We observe that the distinction between the three classes seems to be clearer in the case of *DTW_6* than in the case of *DTW*. Indeed, the samples on the top right corner that mainly correspond to VaD patients (in blue) are better separated from the rest of patients with *DTW_6*. Also, SCI and AD patients that in majority are in the bottom left corner seem to be more well separated with *DTW_6* than with *DTW*.

Figure 4 displays the absolute contributions of the 36 features on the top three components retrieved with both connectivity measures. The first component for *DTW_6* shows high connectivity values on the matrix diagonal, corresponding to the short-range connectivity (intra-region). This observation is confirmed by the average contributions (Cr) of short-, mid-, and long-range connectivity values reported in Table 2. Indeed, with *DTW_6*, the first component is mostly defined by short-range connectivity (Cr=0.458) and mid-range connectivity (Cr=0.377). Long-range values are less correlated to the first component (Cr=0.165).

The second component of *DTW_6* visually presents homogeneous connectivity values, with slightly higher values in the long-range. Notably, in Table 2, the second component is mostly defined by long-range connectivity values (Cr=0.477) and mid-range values (Cr=0.321), and less by short-range values (Cr=0.211). The contribution of the short-range is even lower (Cr=0.166) on the third component that is mainly correlated to mid- and long-range connectivity.

By contrast, with *DTW*, the three components are characterized similarly in terms of long-, mid- and short-range connectivity, with high contribution of long- and mid-range, and low contribution of short-range connectivity.

These results show that the principal components convey relevant information in terms of connectivity when considering *DTW_6*, bringing a fine characterization of individuals.

TABLE 2 – Contributions of short-, mid- and long-range on the top three components of the PCA space for *DTW_6* and *DTW*. PC : principal component ; SR : Short-range, MR : Mid-range, LR : Long-range.

Connectivity	PC	SR	MR	LR
DTW_6	1	0.458	0.377	0.165
	2	0.211	0.321	0.477
	3	0.166	0.413	0.421
DTW	1	0.263	0.306	0.432
	2	0.276	0.395	0.330
	3	0.210	0.370	0.420

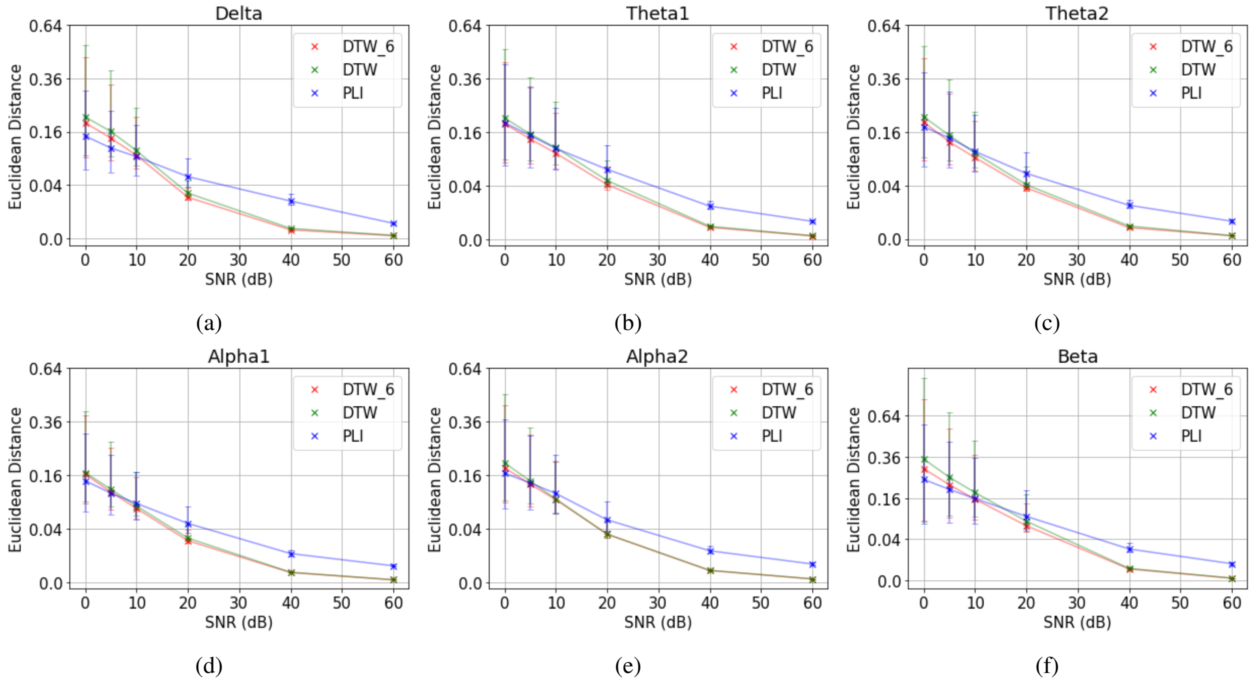


FIGURE 1 – Square Root of average Euclidean Distance computed at different SNR values for *DTW_6* (red), *DTW* (green) and PLI (blue) in all frequency bands.

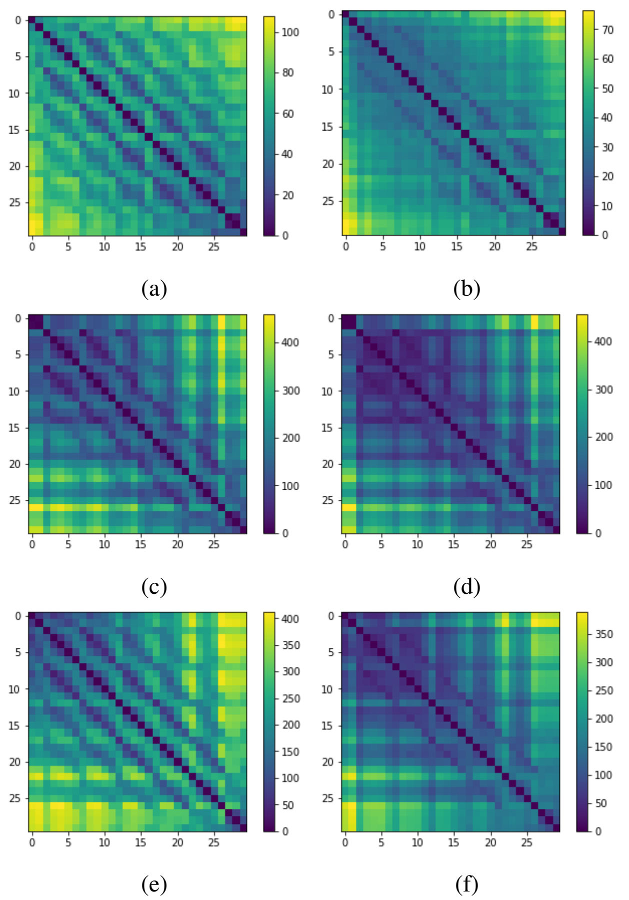
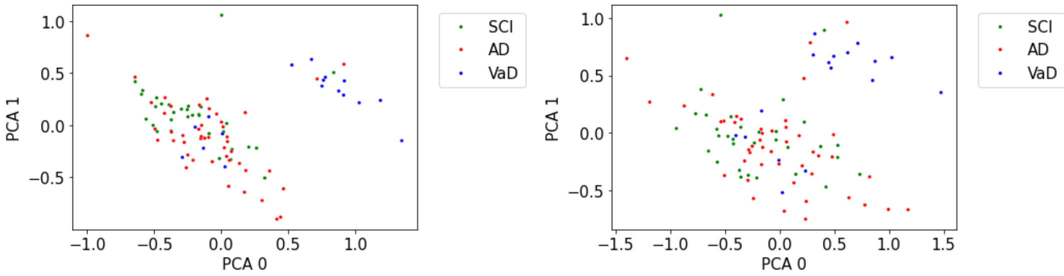


FIGURE 2 – Connectivity matrices of three SCI patients (one patient per row) with *DTW_6* (on the left) and *DTW* (on the right) in Theta 2.

4 Discussion and Conclusion

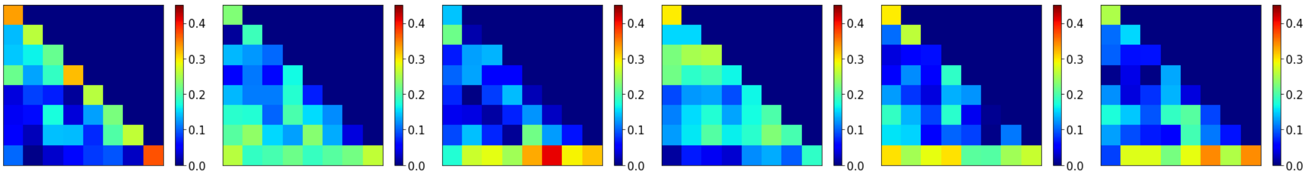
The intrinsic noisy nature of EEG signals and their low SNR make the extraction of efficient features a challenge for EEG data analysis and interpretation. Comparatively to PLI that is widely used in the literature, both DTW measures, with and without the warping window constraint, show better robustness to added Gaussian white noise. Nevertheless, no difference has been observed between the two DTW measures in terms of noise resilience. DTW consists in a nonlinear time-warping alignment between two time series. More precisely, it is an optimization algorithm that relies on finding the optimal temporal matching between two signals, before computing the similarity. This infers to DTW a temporal elasticity that makes it more robust against noise.

Then, we studied the added value of considering Sakoe-Chiba constraint in the computation of connectivity. We found that connectivity matrices obtained with *DTW_6* are visually more contrasted than those obtained with unconstrained DTW (i.e. *DTW*). This points out that integrating Sakoe-Chiba constraint leads to fine and relevant connectivity matrices in terms of information content. This is because when using the Sakoe-Chiba constraint, we force the algorithm to search for temporal alignments that are physically realistic. When no temporal constraint is used, two observations in the two signals can be matched no matter how far apart they are in time. Consequently, alignments that deviate from the true correspondence between the two signals appear, decreasing the DTW score in an unrealistic way. For example, a matching of two observations that are separated of 1 second in time is very unlikely. Setting



(a) (b)

FIGURE 3 – Distribution of patients on the first two components of PCA for (a) *DTW_6* and (b) *DTW*.



(a) (b) (c) (d) (e) (f)

FIGURE 4 – Absolute contributions of the 36 variables on the top three components with *DTW_6* (a, b, c) and *DTW* (d, e, f).

the Sakoe-Chiba constraint at 6 allows a matching between points of two signals that are $6/256 \approx 0.02s = 20ms$ apart.

The multivariate analysis of SCI, AD and VaD patients highlights a better visual distinction between the three classes with *DTW_6* in the PCA space spanned by the first two principal components. Even if we only consider two components which might not be representative of the whole set of selected components, we can say that the two components with maximum variance allow a better separation between classes with *DTW_6* than with *DTW*.

When looking at the contributions of the features and the proportions of short-, mid- and long-range connectivities on the top three components, we find that *DTW_6* brings out principal components that are more indicative of a certain range of connectivity. Indeed, for *DTW_6*, the first component is strongly associated with short- and mid-range connectivity. This result is not observed with any of the first three components obtained with *DTW*. The second component is strongly associated with mid- and long-range for *DTW_6*, and with a mix of the three connectivity ranges for *DTW*. In the third component, we observe a very low contribution of the short-range connectivity for *DTW_6*. It is also the case for *DTW* but with more balanced contributions . These results highlight the effectiveness of using Sakoe-Chiba constraint in finely characterizing the functional connectivity in SCI, AD and VaD patients.

To conclude, DTW with narrow Sakoe-Chiba band is an interesting alternative for connectivity assessment in EEG : it is quite robust to noise and conveys fine and relevant information on functional connectivity. In future research, we will extend the use of DTW with Sakoe-Chiba constraint to study connectivity in patients with heterogeneous profiles.

Acknowledgments

This research was partially funded by IP Paris and Hi ! PARIS Center on Data Analytics and Artificial Intelligence.

Références

[Abazid, 2022] Abazid, M. (2022). *Topological study of the brain functional organization at the early stages of Alzheimer's disease using electroencephalography*. These de doctorat, Institut polytechnique de Paris.

[Briels et al., 2020] Briels, C. T. et al. (2020). Reproducibility of EEG functional connectivity in Alzheimer's disease. *Alzheimer's Research & Therapy*, 12(1) :68.

[Jeong et al., 2001] Jeong, J. et al. (2001). Mutual information analysis of the EEG in patients with Alzheimer's disease. *Clinical Neurophysiology*, 112(5) :827–835.

[Maćkiewicz and Ratajczak, 1993] Maćkiewicz, A. and Ratajczak, W. (1993). Principal components analysis (PCA). *Computers & Geosciences*, 19(3) :303–342.

[Sakoe and Chiba, 1978] Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1) :43–49.

[Senin, 2008] Senin, P. (2008). Dynamic time warping algorithm review.

[Shaw et al., 2017] Shaw, L. et al. (2017). A robust motifs based artifacts removal technique from EEG. *Biomedical Physics & Engineering Express*, 3(3) :035010.

[Stam et al., 2007] Stam, C. J. et al. (2007). Phase lag index : Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Human Brain Mapping*, 28(11) :1178–1193.