IOP Publishing Journal of Neural Engineering
J. Neural Eng. 12 (2015) 056019 (23pp) doi:10.1088/1741-2560/12/5/056019

EEG biometric identification: a thorough exploration of the time-frequency domain

Marcos DelPozo-Banos[1,2], Carlos M Travieso[1], Christoph T Weidemann[2] and Jesús B Alonso[1]

[1] University Institute for Technical Development and Innovation in Communication (IDeTIC), University of Las Palmas de Gran Canaria (ULPGC), Spain
[2] Department of Psychology, College of Human Health and Science, Swansea University, Wales, UK

Received 6 March 2015, revised 30 July 2015
Accepted for publication 10 August 2015
Published 23 September 2015

Abstract

Objective. Although interest in using electroencephalogram (EEG) activity for subject identification has grown in recent years, the state of the art still lacks a comprehensive exploration of the discriminant information within it. This work aims to fill this gap, and in particular, it focuses on the time-frequency representation of the EEG. *Approach.* We executed qualitative and quantitative analyses of six publicly available data sets following a sequential experimentation approach. This approach was divided in three blocks analysing the configuration of the power spectrum density, the representation of the data and the properties of the discriminant information. A total of ten experiments were applied. *Main results.* Results show that EEG information below 40 Hz is unique enough to discriminate across subjects (a maximum of 100 subjects were evaluated here), regardless of the recorded cognitive task or the sensor location. Moreover, the discriminative power of rhythms follows a W-like shape between 1 and 40 Hz, with the central peak located at the posterior rhythm (around 10 Hz). This information is maximized with segments of around 2 s, and it proved to be moderately constant across montages and time. *Significance*. Therefore, we characterize how EEG activity differs across individuals and detail the optimal conditions to detect subject-specific information. This work helps to clarify the results of previous studies and to solve some unanswered questions. Ultimately, it will serve as guide for the design of future biometric systems.

S Online supplementary data available from stacks.iop.org/jne/12/056019/mmedia

Keywords: brain computer interface (BCI), electroencephalogram (EEG), biometry, subject identification/verification

(Some figures may appear in colour only in the online journal)

Acronyms

ADJUST	automatic EEG artifact detection based on the joint use of spatial and temporal features
AEP	auditory evoked potentials
ANN	artificial neural network
ANOVA	analysis of variance
AR	auto-regressive
AvgMnt	common global average reference montage
BCI	brain–computer interface
BHFDR	Benjamini–Hochberg false discovery rate
BIHMnt	bipolar inter-hemispheric reference montage
CI	confidence interval
CV	cross-validation
CzMnt	common Cz reference montage
EEG	electroencephalogram
ERP	event related potential
FFT	fast Fourier transform

FIR	finite impulse response
IC	independent component
ICA	independent component analysis
MC	Monte Carlo
PRE	percentage reduction of error
PSD	power spectrum density
REC	resting with eyes closed
REO	resting with eyes open
std	standard deviation
STFT	short time Fourier transform
VEP	visual evoked potential

1. Introduction

The search of genetic traits within the EEG began with twin studies by Davis and Davis [11] just 12 years after the first human EEG recordings by Hans Berger [7, 10]. These initial studies inspired follow-up work exploring individual characteristics in artificially elicited brain responses as well as similarities in EEG activity between family members, including twins reared apart in a bid to isolate exogenous factors [55]. The first direct connection between the morphology of the EEG and the genotype of individuals was made by Vogel [54]. Overall, certain EEG traits were concluded to be highly inheritable, especially the alpha power and the peak frequency over occipital regions [53, 55, 62].

These results laid the foundation for the first attempts at automatic EEG-based biometric identification [44, 48]. The field has progressed substantially since then.

Some recent works include the use of functionality connectivity between brain regions [15], spectral coherence [30], wavelet package decomposition [18], similarity-based approaches [17] and features of the N400 [5]. A detailed review of the state of the art can be found in [13]. This review classified EEG identification/verification systems as follows:

- *Resting states*: systems based on resting states, either resting with eyes open (REO), or, especially, resting with eyes closed (REC) are particularly prominent owing to the fact that the alpha rhythm (enhanced in posterior areas when subjects have their eyes closed) has been shown to be highly distinctive in physiological studies. These systems rely mainly on spectral features such as the coefficients of auto-regressive (AR) models, which are then classified by artificial neural networks (ANNs) [9, 28, 42, 43].
- *Event related potential* (*ERP*): attempts to use ERPs to identify individuals have, to date, focused exclusively on visual evoked potentials (VEPs). Although spectral features are used in these systems [40], they mainly rely on topographical characteristic of the VEP signal [45, 47].
- *Multiple-tasks*: studies using neural activity recorded under different cognitive tasks tend to focus on identifying the state which maximizes the performance of a biometric system [38, 56].
- *Indirect*: this category encompasses systems that, although based on EEG signals, do not rely on subject specific characteristics, but on a brain–computer interface (BCI) to, for example, allow the user to enter a PIN code [19, 39].

For further technical details about these and related studies, the reader is referred to [3, 13]. For the purpose of the current study, it is enough to recall the main conclusion: the identification of a lack of a comprehensive study of the spectral discriminant information within the EEG activity. All the biometric studies published to date focused their efforts mainly on improving the accuracy rates through the application of new algorithms. Meanwhile, the exact nature and properties of the processed information (discriminant traits within the EEG) has yet to be fully described from a biometric point of view. Here we attempt to fill this gap with a series of carefully designed experiments. In particular:

(1) we provide a description of the optimal conditions to maximize the quality of the discriminant spectral information;
(2) we present visual evidence of the existence of such information and associated properties by means of a stacked representation of the power spectrum density (PSD);
(3) we propose the use of spectral normalization methods based on measurements robust against outliers, as a booster of the quality of the discriminant data,
(4) we use six publicly available databases (ultimately divided into eight data sets), which allows us to distinguish general properties of the EEG signal from idiosyncrasies in individual data sets; and
(5) the current study is the first attempt to use an auditory evoked potentials (AEPs) database for subject identification.

Before we describe the properties of the subject discriminant information within the EEG spectral domain, we will introduce the databases used for our analyses and the associated preprocessing techniques. We will then elaborate on our experimental approach detailing how latter tests were informed by results of earlier experiments.

2. Materials

Previously published works have been mainly based on analyses of isolated databases, in most cases not publicly available [13]. Unfortunately, this hinders the interpretation and reproduction of their findings. To overcome this, we used a set of heterogeneous databases in an effort to identify commonalities in the properties of the discriminant features.

2.1. Databases

We analysed data from six publicly available databases. The first three; Keirn's [22, 23], Yeom's [57, 58] and Zhang's

[59–61] databases, have been extensively used within the EEG biometrics literature and they were described in detail in [13]. Additionally, we added three new-within-the-context databases: BCI2000 [6, 16], DEAP [12, 25] and Ullsperger's [52] databases.

In some cases it was necessary to exclude data or to reformat events. For example, we only included subjects identified as healthy/normal to avoid classification performance being driven by abnormalities. In addition, we rejected subjects without a minimal number of trials for all tasks to ensure reasonably balanced data sets. In other cases the original EEG events were divided into smaller ones, maximizing the inter-segment time.

The BCI2000 consists of 109 subjects performing different motor/imagery tasks. A total of 64 EEG electrodes, as per the international 10–10 system, were recorded with one of the mastoid channels as reference. Recordings were taken with a sampling rate of 160 Hz. Each subject performed 14 runs in a single session: two 1-min baseline runs (one REO and one REC), and three 2 min runs of four tasks (two motor and two motor-imagery). We extracted two separated sets: baseline and tasks.

The DEAP database (data set for emotion analysis using EEG, physiological and video signals) was originally collected to study emotional responses. A single session was recorded from 32 subjects (50% males, aged between 19 and 37, mean age 26.9) while they visualized 40 60 s music videos which elicited different emotions. Before each session, 2 min baseline activity was recorded with subjects relaxed. EEG and peripheral physiological signals were collected by a Biosemi ActiveTwo system with 32 scalp electrodes at a sampling rate of 512 Hz. We extracted two separate sets: baseline and playback. For the playback set, four conditions corresponding to the quadrants of the arousal-score representation of emotions were identified.

Keirn recorded EEG from seven subjects (six males and one female between the ages of 21 and 48) while they were either resting or performing one of four tasks for 10 s. Each condition was repeated five times over the span of two weeks, two sessions were recorded from each subject. EEG was captured from six electrodes (occipital, parietal and central) at a sampling frequency of 205 Hz.

Ullsperger recorded EEG activity in response to auditory stimuli (words) which had to be classified as either matching or not matching the meaning of a target word. The EEG was recorded from five subjects who each were fitted with 61 electrodes referenced to PCz and CP1 with a sampling rate of 200 Hz. A notch filter between 47 and 53 Hz was applied to remove line noise. From each trial, segments were extracted from 2 s before presentation of the second (test) word to 2 s after.

Yeom recorded EEG activity from 11 male subjects including one pair of monozygotic twins, with ages between 20 and 29 years old (mean 26.67), while they viewed pictures of either themselves or others. All but one participated in two sessions on different days (we used only data from the ten subjects participating in both sessions for our analyses). A Neuroscan SynAmps2 system was used to record the EEG signal from 18 electrodes (International 1020 System). The sampling rate was set to 300 Hz and a band-pass filtered between 0.1 and 100 Hz. Additionally a 60 Hz notch filter was applied. EEG was extracted from 200 ms pre- to 800 ms post-stimulus.

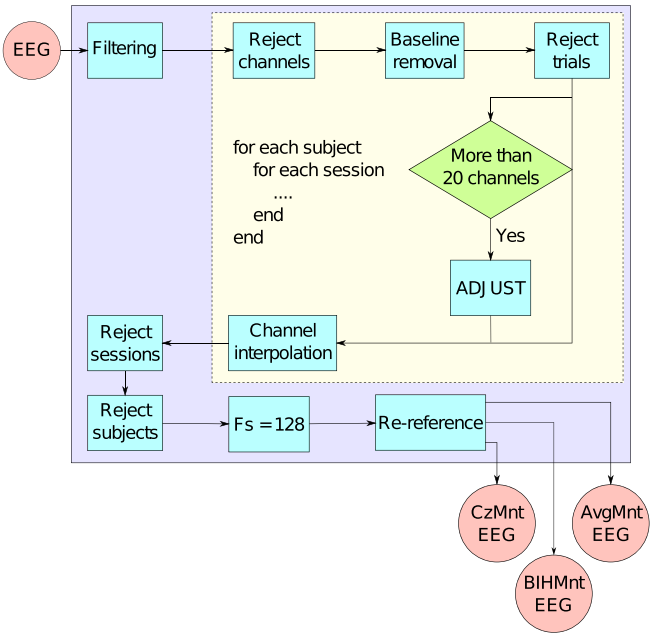

Figure 1. The preprocessing and normalization steps applied to all databases, with ADJUST referring to the automatic artefact rejection system presented in [33] and F_s referring to the EEG sampling frequency.

Zhang's recorded EEG activity from 125 subjects performing a visual matching task. We only used data from the subset of 48 healthy males 25.81 ± 3.38 years old (the remaining 77 subjects were diagnosed with alcoholism). Forty trials were recorded from each subject with an inter-trial-interval of 3.2 s. EEG activity was recorded from 61 channels referred to Cz. The sampling rate was 256 Hz and data were hardware filtered between 0.02 and 50 Hz.

2.2. Preprocessing

We applied a common normalization and preprocessing step to all databases to remove idiosyncrasies that are specific to the particular recording set-up, such as sampling frequency or recording reference (figure 1). This pre-processing procedure, including the applied thresholds, is mainly based on [36].

All databases were filtered using a high-pass finite impulse response (FIR) filter with a low edge frequency of 1 Hz. For data sets that were not already filtered at 50 or 60 Hz to reduce line noise, we applied a FIR notch filter at either 50 Hz (DEAP) or 60 Hz (BCI2000, Keirn's database) as appropriate. Additionally, we low-pass filtered EEG activity from Zhang's database at 40 Hz to homogenize channels.

Next, extremely noisy EEG channels were automatically detected and rejected to avoid any interference in subsequent

computations. In particular, the time kurtosis

$$K_c = \mathrm{rn}\Big(\big|\mathrm{kurt}(X_c) - 3\big|\Big) \text{ for } c = 1, 2, \ldots, C, \quad (1)$$

where X_c is the EEG of the cth channel, was used as a noise indicator feature, together with the weighted correlation between channels

$$R_c = \mathrm{rn}\Big(\mathrm{median}\big(\Phi_c, \forall\, j \neq c\big)\Big) \text{ for } i = 1, 2, \ldots, C \quad (2)$$

$$\text{and } \Phi_c = \Big\{1 - \big|\mathrm{corr}\big(X_{ci}, X_{cj}\big)\big|e^{-d_{ci,cj}}, \forall\, j \neq c\Big\}, \quad (3)$$

where $d_{ci,cj}$ is the arc distance between channels ci and cj in a sphere of radius one. In both cases, rn is the robust normalization function

$$\mathrm{rn}(v) = \frac{20(v - \mathrm{median}(v))}{27 \cdot \mathrm{iqr}(v)}, \quad (4)$$

where v is a feature vector and median and iqr are the median and interquartile functions respectively. Note that $(27/20) \cdot \mathrm{iqr}(v_\mathrm{n}) \simeq \theta_\mathrm{n}$ for a normal distribution v_n with standard deviation (std) θ_n.

We considered channels for which either of the features exceeded three for rejection. If the proportion of such channels fell below 10%, they were excluded, otherwise we rejected the 10% of the channels with the largest scores $\sqrt{K_c^2 + R_c^2}$.

EEG voltage ranges were then normalized by equation (4). This was preferred over the z-score normalization due to the level of artefacts in some of the databases.

When the number of available trials allowed it, the noisiest were discarded. Again, the criterion applied to define the number of maintained trials relied on the desire to have a balanced database across subjects, conditions and sessions.

Hence, noise in events was characterized by the voltage range

$$A_e = \left|rn\Big(\Big\langle p_{(T)}^{(95)}(X_e) - p_{(T)}^{(5)}(X_e)\Big\rangle_{(C)}\Big)\right| \text{ for } e = 1, 2, \ldots, E \quad (5)$$

and the voltage variance

$$V_e = \left|rn\Big(\Big\langle \mathrm{var}(T)(X_e)\Big\rangle_{(C)}\Big)\right| \text{ for } e = 1, 2, \ldots, E, \quad (6)$$

where X_e is the EEG of the eth event, $p_{(T)}^{(n)}$ and $\mathrm{var}_{(T)}(\cdot)$ are the n% percentile and variance across time (T) and $\langle \cdot \rangle_{(C)}$ represents the average across channels. Events were then sorted based on their score $\sqrt{A_e^2 + V_e^2}$ and those with the highest values were rejected until the desired amount was reached.

Once artefactual events had been removed, rejected channels were interpolated back in the data set. A spherical interpolation method was applied for this purpose.

When the number of subjects available allowed it, those with the worst signal quality were discarded. The rejection was based on the score

$$\frac{\Big\langle \sqrt{A^2 + V^2}\Big\rangle_{(E)}}{2} + 30\frac{\hat{C} + 1}{C}, \quad (7)$$

where A and V are the features defined in (5) and (6) and $\langle \cdot \rangle_{(E)}$ is the average along the events dimension.

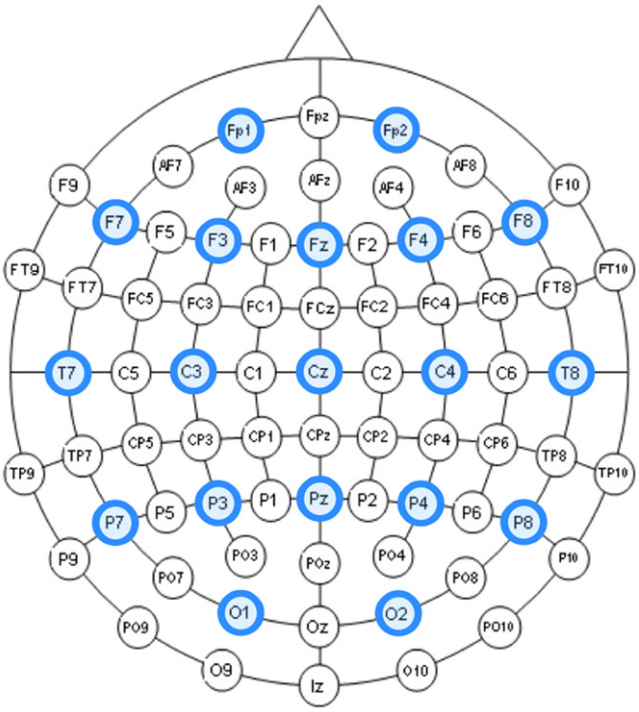

Figure 2. Diagram of the 10–20 international system for EEG channel locations. Highlighted channels are kept for experimentation.

Finally, in order to reduce the volume of data, only a set of 16 channels distributed around the scalp were kept (figure 2), except for Keirn's and Yeom's databases, from which all available channels were kept. In addition, databases were down-sampled to 128 Hz.

A second version of databases with more than 20 channels (i.e. all but Keirn's and Yeom's) was computed by applying the automatic EEG artifact detection based on the joint use of spatial and temporal features (ADJUST) algorithm [33] between the stages of event rejection and channel interpolation. This tool automatically identifies artefactual independent components from time and topological features by means of an unsupervised classification method. In particular, a slightly modified version of the algorithm was used, which accounted for missing features by simply ignoring them from the computation. This artefact-free version was used to assess the effect of noise on the discriminant information. Note, however, that automatic artefact rejection methods are still a hot topic of research. Results obtained with the ADJUST processed data sets should therefore be interpreted accordingly (see appendix A for a short discussion on the effects of ADJUST on the classification results).

2.3. Diversity of databases

Table 1 summarizes the properties of the extracted databases. A total of eight data sets, originated from six different publicly available databases, were used in the forthcoming experiments. Having said so, their diversity is as important as their number. In particular,

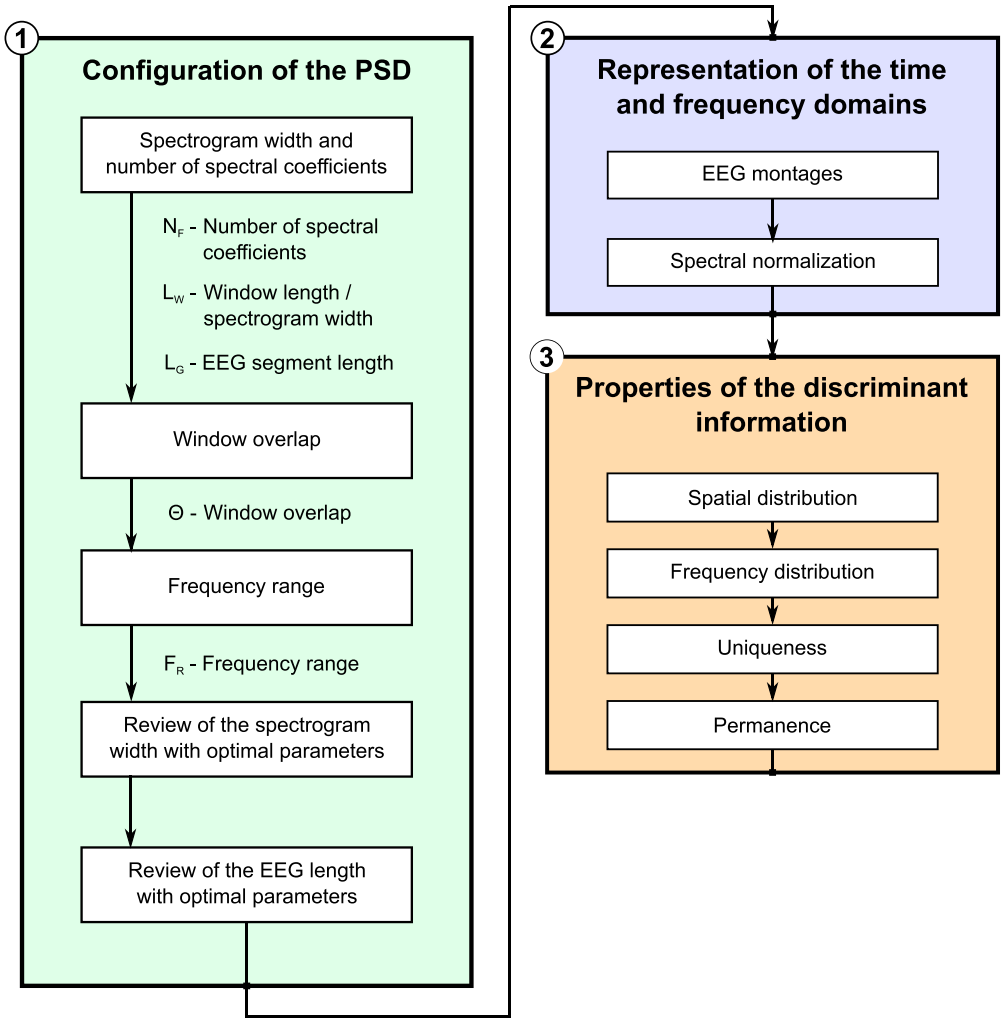

Figure 3. Diagram of the experimentation methodology. (1) We identified the configuration of the PSD that maximizes the discriminant information. (2) We find the representation of the time and frequency domains that maximizes the discriminant information. (3) We characterized the properties of the EEG discriminant information. This figure complements the information presented in table 2.

Table 1. Normalized databases used for experimentation. Columns show the name of the database (Database), the number of subjects (Subj.), the number of conditions (Cond.), the number of sessions (Sess.), the number of trials (Trials), the length of the EEG segments in seconds (Len.) and the descriptive keywords (Keywords).

Database	Subj.	Cond.	Sess.	Trials	Len.	Keywords
BCI2000-Baseline	100	2	1	5	10	REO; REC
BCI2000-Tasks	100	12	1	[9, 10]	4	Multiple-tasks; motor real/imagery
DEAP-Baseline	20	1	1	5	20	REO
DEAP-Playback	20	4	1	[5, 10]	20	Elicited emotions
Keirn's	5	5	2	[8, 10]	4	Multiple-tasks; intellectual
Ullsperger's	5	2	1	180	4.1	Auditory evoked potentials (AEP); synonyms versus non-synonyms
Yeom's	10	2	2	900	1	VEP; self-representation
Zhang's	30	3	1	[15, 20]	1	VEP

Table 2. Phases of the experimentation methodology. This table complements the information presented in figure 3.

1. *Configuration of the PSD*	
Spectrogram width and number of spectral coefficients	Analysis of number of spectral coefficients (N_F), spectrogram window length (L_W) and EEG length (L_G)
Window overlap	Analysis of spectrogram window overlap (Θ)
Frequency range	Analysis of maximum (F_{max}) and minimum (F_{min}) cut-off frequencies
Review of spectrogram width with optimal parameters	Analysis of spectrogram window length (L_W), with specific number of coefficients (N_F), window overlap (Θ) and EEG length (L_G)
Review of EEG length with optimal parameters	Analysis of the EEG length (L_G), with specific number of coefficients (N_F), spectrogram width (L_W) and window overlap (Θ)
2. *Representation of the time and frequency domains*	
EEG montages	Analysis of multiple EEG montages
Spectral normalization	Analysis of different normalization methods
3. *Properties of the discriminant information*	
Spatial distribution	Analysis of the spatial distribution of the EEG discriminant information
Frequency distribution	Analysis of the frequency distribution of the EEG discriminant information
Uniqueness	Analysis of the EEG discrimination power with increasing number of subjects
Permanence	Analysis of time stability of the EEG discrimination information

(1) BCI2000-Baseline and DEAP-Baseline databases represent resting conditions,
(2) BCI2000-Tasks accounts for real and imagery motor conditions,
(3) Keirn's data set provides multiple intellectual tasks,
(4) DEAP-Playback assesses the effects of emotions,
(5) Yeom's and Zhang's databases contain two extensively studied visual evoked potentials (VEPs), and
(6) Ullsperger's data introduces the use of AEP.

As a result, findings that are consistent across data sets can be safely generalized as inherent characteristics of the EEG signal and not due to idiosyncrasies of the set-up.

3. Methodology

The experimentation process was divided into three blocks, each composed of several tests focussed on a common goal (figure 3 and table 2). Configuration details missing in this section will be given within the corresponding results section, as their rationale depends on the results of previously evaluated steps. Complementary qualitative and quantitative studies were carried out during each experiment. In both cases, a short time Fourier transform (STFT) spectrogram, computed with a Hamming window, was used to extract the time-frequency representation of the EEG signal. All the experiments were ran on the preprocessed databases listed in table 1. From each data set, except Keirn's and Yeom's, two versions were evaluated: one containing the raw (preprocessed) EEG, and a second one containing the artefact free EEG after applying ADJUST (section 2.2).

It is imperative to keep in mind that, contrary to the state of the art, the current study at no point aimed to obtain high accuracy rates during classification. Instead, it intended to describe in some detail the subject discriminant information within the time–frequency representation of EEG activity, and to characterize the effects of each of the parameters in the classification problem. Indeed, some of the decisions taken on the design and analysis of experiments are founded on this characteristic goal.

3.1. Configuration of the PSD

During the first experimentation stage, we defined the configuration of the STFT that maximizes the discriminant information. Specifically, we explored the number of fast Fourier transform (FFT) coefficients (N_F), the window length or spectral width (L_W), the window overlap percentage (Θ), the EEG segment length (L_G) and the frequency range (F_R). With this aim, we ran the following experiments:

(1) *Spectrogram width and number of spectral coefficients*: Hamming window lengths $L_W \in [0.1, 60]$ seconds and the number of FFT spectral coefficients $N_F \in [32, 1024]$ were evaluated on a grid-like set of experiments where the window overlap Θ was fixed to 0.
(2) *Window overlap*: overlap percentage values $\Theta = 0$, 25, 50 and 75 were tested on the best performing configurations (L_W and N_F) of the previous step.

Table 3. PSD normalization functions. P is a PSD matrix with dimensions N_FxN_T, with N_T the number of time points, and $p_{(F)}^{n\%}$ is the n-% percentile applied along the frequency dimension.

Name	Equation	Name	Equation
powNorm	$\frac{P}{\sum_{\forall f} P}$	prcNorm	$\frac{P - p_{(F)}^{5\%}(P)}{p_{(F)}^{95\%}(P) - p_{(F)}^{5\%}(P)}$
normNorm	$\frac{P}{\sqrt{\sum_{\forall f} P^2}}$	iqrNorm	$\frac{P - p_{(F)}^{25\%}(P)}{p_{(F)}^{75\%}(P) - p_{(F)}^{25\%}(P)}$
zNorm	$\frac{P - \langle P \rangle_{(F)}}{\mathrm{std}_{(F)}(P)}$	rNorm	$\frac{P - \mathrm{median}_{(F)}(P)}{\mathrm{iqr}_{(F)}(P)}$

(3) *Frequency range*: the width of the spectral band used was modified by varying the maximum $F_{\max}$ and minimum $F_{\min}$ frequencies between 10 and 60 Hz on individual experiments.
(4) *Review of the spectrogram width with optimal parameters*: to clarify the effects of the STFT window length on the system's performance, optimal configurations were represented against L_W. That is, $L_W \in [0.1, 60]$ was tested with the remaining parameters (N_F, L_G and Θ) set based on conclusions from previous experiments.
(5) *Review of the EEG length with optimal parameters*: to clarify the effects of the EEG length on the system's performance, optimal configurations were represented against L_G. That is, $L_G \in [L_W, L]$, with L the available length of signal, was tested with the remaining parameters (N_F, L_W and Θ) set based on conclusions from previous experiments.

3.2. Representation of the time and frequency domains

Once the STFT was fully configured, we evaluated different time and frequency representations of the data. In particular, we ran the following experiments:

(1) *Montages*: transformations of the signal in the time domain were assessed by EEG montages common global average reference montage (AvgMnt), bipolar inter-hemispheric reference montage (BIHMnt) and common cz reference montage (CzMnt).
(2) *Spectral normalization*: in a bid to find the optimal representation of the spectral information for the problem, the normalization functions described in table 3 were applied to the PSD prior to the classification.

3.3. Properties of the discriminant information

This experimentation block ran a number of tests to describe some of the properties of the discriminant PSD information. In particular:

(1) *Spatial distribution*: each EEG channel was evaluated individually to assess the performance of neural activity from different sensor locations (which vary in sensitivity to activity from different brain areas).
(2) *Frequency distribution*: each frequency was evaluated individually to asses the performance of different EEG rhythms.
(3) *Uniqueness of patterns*: to evaluate the uniqueness of individual spectral patterns, experiments varying the number of users N_S in the system were run, so that the stability of the performance with increased N_S could be assessed.
(4) *Permanence of patterns*: this step aimed to judge the independence of the subject characteristic patterns with respect to time. To do so, Keirn's and Yeom's data sets were cross-validated considering each recording session as an indivisible unit. That is, for each subject, training and testing sets contained samples from different sessions. For simplicity, this cross-validation (CV) methodology will be referred to as *session*-CV.

3.4. Qualitative study

To simplify inspection of the EEG PSD, the spectrograms from selected trials, sessions, conditions and channels were stacked along their time axis, resulting in a single, piecewise-continuous (in time) spectrogram (figure 4). This representation of the data helped to visually confirm the existence of different spectral neural signatures across individuals. It also helped to understand the effects that the evaluated parameters have on this signature. Hence, we performed these qualitative analyses on all experiments, to complement quantitative tests.

3.5. Quantitative study

To quantify the observations made during visual inspection of the spectrogram, we performed a series of classification experiments. Because the aim of such experiments was not to obtain high accuracy rates, we applied a simple design based on a Bayes classifier.

To this end, we z-normalized PSD coefficients across samples and used them as input to the classifier. The normalizing factors were computed only with the training set, and uninteresting filtered frequencies were removed from the analysis.

Different final responses of the Bayes classifier were computed from the sum-score fusion of N_G neighbouring windows, with $N_G = 1, 2, \ldots, N_W$, where N_W is the number of windows extracted from a single trial. Hence, for each N_G value, a total of $N_W - N_G + 1$ responses where computed by shifting the fusing scope a single PSD window to the right each time. This process was done to differentiate between the effects of the length of the Hamming window L_W and the length of the EEG segment

$$L_G = L_W\left[N_G - \left(1 - N_G\right)\frac{\Theta}{100}\right], \tag{8}$$

with Θ the window overlap percentage, used to compute a response. When representing results against L_G, intermediate values (not multiples of L_W) were interpolated whenever possible for ease of interpretation.

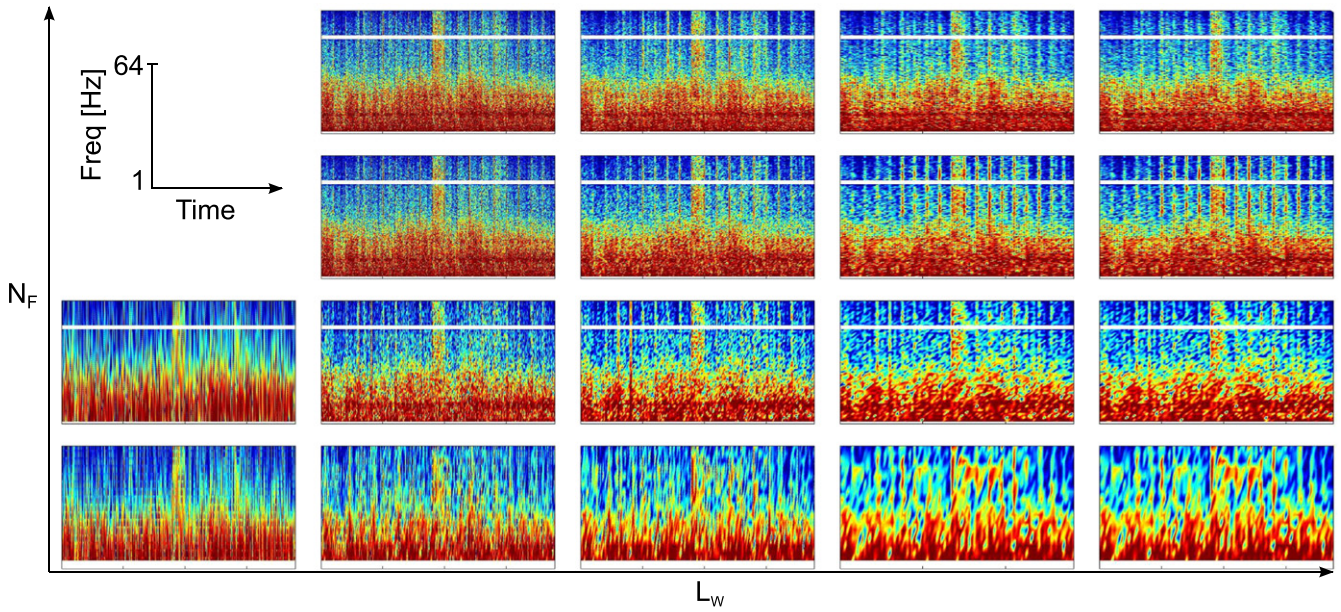

Figure 4. PSD with different STFT window length (L_W) and number of FFT coefficients (N_F). The PSD, in dB scale, corresponds to one of the subjects of BCI2000-Baseline database. Within each PSD, the missing frequency band around 50 Hz (vertical axis) was filtered in the pre-processing and manually removed from the analysis. The horizontal axis (time) is piecewise-continuous and contains all the samples from the subject.

Table 4. Fusion modes used during experimentation.

Mode	Description
No-fusion	The system is evaluated using a single PSD coefficient, individually for each EEG channel.
Freq-fusion	The system is evaluated using all PSD coefficients simultaneously (from the selected frequency range), but individually for each EEG channel.
Ch-fusion	The system is evaluated using a single PSD coefficient, but combining all EEG channels.
Full-fusion	The system is evaluated using all PSD coefficients (from the selected frequency range) from all EEG channels simultaneously.

In addition, four fusion modes were differentiated based on the level of feature fusion (table 4):

- *Full-fusion*: coefficients from all frequencies and channels are fused into a single vector.
- *Freq-fusion*: coefficients from all frequencies are fused into a single vector and each channel is evaluated individually.
- *Ch-fusion*: coefficients from all channels are fused into a single vector, and each frequency is evaluated individually.
- *No-fusion*: the coefficient from each channel and frequency is evaluated individually.

For CV, we combined stratified K-folds and Monte Carlo (MC) techniques in order to benefit from the stability (lack of bias) of the former and the low-variance of the latter [26]. To this end, we repeated a K-fold process with five folds ten times.

For databases where the number of available subjects exceeded the number used for experimentation (N_S), we used a different subset of subjects on each of the MC repetitions. Such subsets were built in a balanced way, i.e. by trying to use each subject the same number of times across the whole experimentation process. Unless otherwise specified, we used a maximum of $N_S = 20$ subjects as a compromise between the number of subjects and the number of available databases able to accommodate N_S. The latter was important to avoid any unexpected interaction between the tested factor and N_S.

In addition, to ensure the validity of the results, during the segmentation of the databases all windows extracted from the same trial were kept in the same fold. This prevented the inclusion of EEG segments on both training and testing sets when the window overlap parameter was greater than 0%.

Micro-accuracies were computed for each of the MC repetitions as follows. Let $M_{(i,j)}$ be the $N_S \times N_S$ confusion matrix of the jth K-fold iteration of the ith MC repetition, so that $M_i = \sum_{\forall j} M_{(i,j)}$ is the aggregated confusion matrix for the ith MC repetition. Let A_i be the corresponding mean accuracy rate. We converted accuracy rates to percentage reduction of error (PRE) values for the evaluation of the

Table 5. Focus modes used during experimentation.

Mode	Description
Detailed-focus	The system is evaluated using data from a single cognitive task/condition and a single recorded session.
Task-focus	The system is evaluated using data from a single cognitive task/condition but all recorded sessions.
Sess-focus	The system is evaluated using data from all cognitive tasks/conditions but a single recorded session.
No-focus	The system is evaluated using data from all cognitive tasks/conditions and all recorded sessions.

results [21]. Specifically, we compared A_i with a random process, such that

$$\mathrm{PRE}_i = \frac{A_i - \frac{1}{N_S}}{1 - \frac{1}{N_S}}. \quad (9)$$

The mean (μ_{PRE}) and 95% confidence interval (CI) of the above PRE values were then computed and reported. To avoid cluttering the text with figures and tables, we only inserted the most representative examples in each case. Supplementary results and statistical tests can be found in appendix C of the supplementary data (stacks.iop.org/jne/12/056019/mmedia).

We use PREs instead of the more common absolute accuracy values for practical reasons. First, we aimed to shed light on the properties of the discriminant information within the EEG data, relegating absolute performance to a secondary role (as long as it is above chance level). This is especially true given that results were analysed by comparison rather than inspecting isolated values. Second, PREs as measured here will be 0 if the system performs at chance levels and >0 otherwise, regardless of the number of subjects included in the analysis. In other words, the range of values of the results have a common meaning across databases. This allows for a better representation and more direct interpretation and comparison of the results. Having said that, if required, performance values can be easily extracted from the given PREs by inverting equation (9) (see appendix B in the supplementary data (stacks.iop.org/jne/12/056019/mmedia) for step-by-step examples).

Finally, we differentiated between the following four experimental modes based on the level of focus (table 5):

- *Detailed-focus*: experiments were run individually for each combination of session and task.
- *Task-focus*: experiments were run individually for each task, ignoring session labels during the CV partitioning. Mixing sessions on training and testing sets is not the standard procedure from a biometric point of view. However, in the current scenario, there are random factors linked with the set-up process, such as the exact position of sensors or their contact quality, which affect the recorded EEG. One way to alleviate the influence of such factors is to average across sessions. Note that we also executed experiments cross-validating sessions to keep the study relevant within the biometric field (section 3.3).
- *Sess-focus*: experiments were run individually for each session, ignoring task labels during the CV partitioning.
- *No-focus*: experiments were run over all the available data, ignoring task and session labels during the CV partitioning.

4. Configuration of the PSD: results, discussion and conclusions

Within the first study block, we analysed the effects of some basic parameters on the quality of the extracted EEG discriminant information. Specifically, we considered the number of FFT coefficients (N_{F}), the length of the spectral window (L_{W}) and its overlap (Θ), and the length of the EEG signal used to compute the final identification response (L_G). We executed *full-fusion* experiments with all focus modes (see tables 4 and 5).

4.1. Spectrogram width and number of FFT coefficients

First, we ran a number of experiments varying the number of FFT coefficients (N_{F}), the length of the spectral window (L_{W}) and the length of the EEG signal used to compute the final identification response (L_G) (Θ was set to 0%). As we suspected a triple interaction between these parameters, we executed all possible combinations in a cube like methodology with the following ranges $N_{\mathrm{F}} = [16, 2048]$, $L_W = [0.1, \max(20, L_A)]$ and $L_G = [L_W, L_A]$; with L_{A} the length of the available signal.

Results. Overall, a common behaviour was observed across databases. Visually, the PSD became more stable with longer windows (above 1 s). At the same time, some EEG traits and details are not apparent until N_{F} is higher than 64 (figure 4). In quantitative terms, representing the μ_{PRE} surface of L_{W} against N_{F} corroborated the qualitative observations (figures 5 and C.1). Overall, the system reached quasi-optimal performance at the diagonal $N_{\mathrm{F}} \approx L_W^* F_{\mathrm{s}}$, with F_{s} the sampling frequency (table).

Although less prominent and less homogeneous across databases, an increase in L_G was generally followed by a gain in μ_{PRE} values (figures 5 and C.2). In some instances, such as DEAP and Keirn's databases, this leap was as high as 10 percentage points. This could be expected, as increasing L_G means more information being fed to the system to produce the

Table 6. Quantitative results for different configurations. Mean PRE and 95% CI obtained with different configurations. Data correspond to raw databases tested during *full-fusion no-focus* experiments. Within data sets, performances statistically different than the maximum are pointed by $*$ (single tail *t*-tests with BHFDR adjusted $p > 0.05$). A maximum of 20 subjects was used in each experimental iteration. Refer to table 1 for details on databases' code names, and to table 13 for further related results.

Conf	BB	BT	DB	DP
Conf-FullLen	68.79$*$ [64.52, 73.06]	88.88 [87.67, 90.10]	64.11$*$ [62.58, 65.63]	**95.82** [**95.61, 96.04**]
Conf-1s	92.74 [89.85, 95.62]	87.80 [85.43, 90.17]	92.63$*$ [91.69, 93.57]	81.99$*$ [81.70, 82.27]
Conf-2s	**93.16** [**91.60, 94.72**]	**89.32** [**87.25, 91.40**]	**93.68** [**93.33, 94.04**]	88.51$*$ [88.26, 88.75]
Conf	*K*	*P*	*Y*	*Z*
Conf-HalfLen	— —	— —	**52.89**$*$ [52.86, 52.92]	**78.82** [**75.26, 82.37**]
Conf-FullLen	74.42 [73.68, 75.17]	**94.22** [**94.15, 94.30**]	**57.47** [**57.45, 57.50**]	77.88 [76.05, 79.71]
Conf-1s	73.74$*$ [73.57, 73.92]	92.83$*$ [92.71, 92.94]	— —	— —
Conf-2s	**74.70** [**73.81, 75.59**]	93.72$*$ [93.61, 93.84]	— —	— —

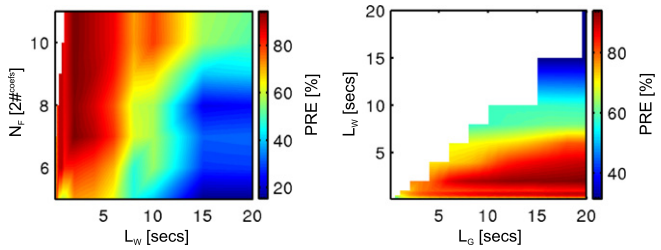

Figure 5. Quantitative analysis of the number of spectral coefficients (N_F), the SFTF window length (L_W) and the length of the EEG signal (L_G). Mean PRE results obtained with DEAP-Baseline database on an L_W versus N_F grid with $L_G = 10$ s (left) and an L_G versus L_W grid with $N_F = 128$ coefficients (right). In both cases, Θ was fixed to 0% and the results were obtained during the *full-fusion no-focus* experiments. A maximum of 20 subjects was used in each experimental iteration. Refer to figures C.1 and C.2 for further related results.

final response. Overall, we found the performance peak at $L_W < L_G$, i.e. when the available signal was divided into individually processed segments instead of in a unique long chunk. The only exceptions to this were DEAP-Playback results of only *no-focus* experiments, and Yeom's and Zhang's databases. In the latter pair, the effect might still be there, but we are unable to see it due to the limited EEG available (only 1 s).

In terms of the stability of the results, databases with fewer than 20 subjects (i.e., all subjects were used in each iteration of MC) showed, as expected, lower levels of θ_{PRE}. The interaction between the std and the parameters considered here were not that clear. In general, minimum values of θ_{PRE} were located within the optimal area, i.e. above the N_F-L_W diagonal, while L_G had little or no effect in this regard.

When comparing results from each of the experimentation modes, the use of multiple tasks or sessions had no effect other than the expected averaging of μ_{PRE} and the increase of θ_{PRE}. The application of ADJUST for artefact rejection increased the amount of discriminant information, which was reflected in both mean and std PRE values.

Discussion. In terms of the parameters N_F and L_W, the optimal diagonal $N_F \approx L_W^* F_s$ suggests that it is desirable to retain the maximum frequency resolution. This conflicts with architectures such as [43, 44], which focused on the power of relatively wide spectral bands. At the same time, the relationship between performance and frequency resolution is in line with the findings of [42] regarding the need for higher orders of the AR model; and hence better frequency resolutions, to counteract the rise in the number of users. According to this, wide spectral bands should only be used on neurophysiological studies—where brain rhythms are well described—or as supplementary features into a more complex feature vector.

With no obvious relationship between L_G and the system's performance, a 'maximum' approach could be recommended, i.e. use as much EEG data that is available. Perhaps, the most interesting effect of L_G is its ability to dilate the 'optimal configuration area', i.e. the set of parameters that yield results similar to the observed best performance. It is this property which allows to reduce L_W and subsequently N_F, maintaining the level of discrimination. A more detailed study of L_W and L_G will be presented in sections 4.4 and 4.5.

4.2. Window overlap

Subsequently, the overlap percentage parameter Θ was optimized by testing configurations of Θ, L_W and N_F, which accounted for the interaction between these three parameters. In this case, based on the previous experiment, the range of N_F was shrunk to [64, 256] in a bid to reduce the volume of experimentation.

Results. Results were less uniform across data sets and system's configurations than in the previous experiment. In general, higher values of overlap yielded similar or better results (figures 6 and C.3). The most obvious effect of Θ was its ability to widen the optimal configuration area (as described with L_G). The effects of Θ remained relatively stable across experiments (*detailed-focus*, *task-focus*, *sess-focus* and *no-focus*) and after the application of the artefact rejection methods. The most prominent exception was, once again, obtained with the DEAP-Playback data set during *no-focus* experiments—the behaviour just explained was observed in *task-focus* experiments.

Discussion. *A priori*, it seems reasonable to use some degree of window overlap, as higher Θ translates into more

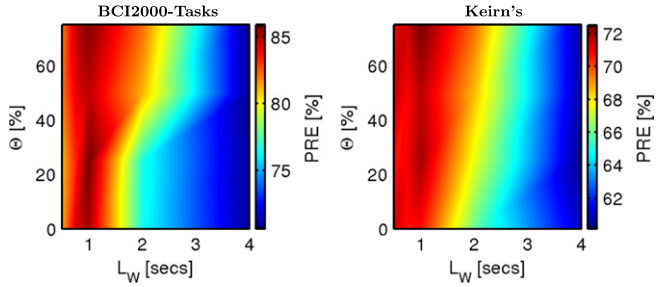

Figure 6. Quantitative analysis of the SFTF window overlap (Θ). Mean PRE results on a L_W versus Θ grid with $N_F = 128$ and L_G equal to the maximum EEG available. Results correspond to BCI2000-Tasks (left) and Keirn's (right) data sets, during *full-fusion no-focus* experiments. A maximum of 20 subjects was used in each experimental iteration. Refer to figure C.3 for further related results.

information being fed to the system to produce the final response—similar to L_G. However, although results supported this to some degree, the observed effects are not strong enough to make a final statement in this regard. With this in mind, we decided to use an overlap of 75% in subsequent experiments. This is, however, a somehow arbitrary decision, as we have no strong evidence that this overlap value works better than any other. Therefore, lower overlaps could be used if, for example, processing time or data volume are a concern in the design of the system.

4.3. Frequency range

The visual inspection of EEG PSD undertaken on previous steps suggests that spectral neural signatures stabilize for frequencies above 30–40 Hz (figure 4). To explore this further we systematically limited the maximum $F_{\max}$ and minimum $F_{\min}$ frequencies in separate experiments. Once more, to reduce the volume of experimentation, only two configurations were tested, named *Conf-HalfLen* and *Conf-FullLen*. The latter uses L_W equal to the available EEG signal and hence $\Theta = 0$, while *Conf-HalfLen* utilizes L_W equal to half the available signal and $\Theta = 75$. In both cases, N_F was set to the power of two closest from the right to $L_W^* F_s$.

Results. As expected from the observation of the PSD representation, the system reached maximum performance at $F_{\max}$ equal 30 or 40 Hz in all databases with the exception of Yeom's data set, which reaches it at 50 Hz (figure 7 left). On DEAP databases, there was a decrease of less than 10 percentage points in PRE when raising $F_{\max}$ from 30 to 40 Hz (tables C.2 and C.3). Experiments varying $F_{\min}$ resulted in a μ_{PRE} curve with two more pronounced increases, one when the first frequencies are added (between 50 and 60 Hz) and the second one with the last frequencies—frequencies below 20 or 30 Hz (figure 7 right and tables C.2 and C.3). This behaviour was also observed after rejecting artefacts with ADJUST (figure C.4 and table C.4).

Discussion. In line with [9], a high frequency cut-off of 30 or 40 Hz can be established based on the results. Indeed, most of the systems in the literature have used a maximum frequency below 40 Hz [13]. The described results highlight that architectures using only high frequencies [41, 46] are not necessarily the optimal approach.

The outlier behaviour in Yeom's database may be a result of the classifier capitalizing on differences between users' artefacts—muscle artefacts have been identified to overlap with frequencies above 20 Hz [35]. A post-hoc examination of the spectra from Yeom's database corroborated the existence of noise at high frequencies.

Indeed, EEG artefacts have been shown to be subject discriminant and used successfully for identification [4]. Hence, the performance of systems using frequencies above 20 Hz cannot be considered to be solely based on the users' neural signature without a proper artefact analysis. In our case, the exact same behaviour was obtained with artefact free databases, as processed by ADJUST. In fact, techniques based on independent component analysis have been described as the 'more promising approaches that have been used for attenuating muscle artefact' [35]. Hence, although it is not possible at the moment to be certain about the source of the performance above 20 Hz, it seems unlikely that it comes from muscle artefacts. Unfortunately, due to an insufficient number of EEG sensors, we lacked an artefact free version of Yeom's database (which we noted had high frequency artefacts) to further test this.

4.4. Review of spectrogram width with optimal parameters

Although previous experiments have shed light onto the optimal value of the spectral window length (L_W), they did not address the issue in a definitive way due to the relationship between L_W and the other parameters. To circumvent this, we re-tested a range of window lengths, setting the remaining parameters to their defined optimal values. That is, N_F was set to the maximum power of two closest to $L_W^* F_s$, L_G was set to the available signal length, Θ was set to 75% whenever possible (i.e. when L_W is small enough to allow multiple windows with 75% overlap on the available signal length) and 0% otherwise, and $F_{\max}$ was set to 40 Hz. Results obtained with Yeom's and Zhang's will be presented but not considered in the analysis, as they only contain 1 s of EEG.

Results. Across databases, maximum performance was reached at L_W between 1 and 2 s (figure 8 and table C.6). After this point, performances either stabilised within the 95% CI or decreased. The main exception encountered was on DEAP-Playback data set, for which the maximum was reached between 6 and 8 s.

The rejection of artefacts reduced the differences between window lengths below 2 s, boosting the performance of the shortest windows. On DEAP-Playback data set, cleaning the EEG shifted the optimal point to earlier in L_W, reaching it between 4 and 6 s on ADJUST processed data.

Discussion. Similar behaviour was observed across databases. Overall, we obtained optimal performances with window lengths 1 or 2 s, with the exception of DEAP-Playback. In addition, the flattening observed on the

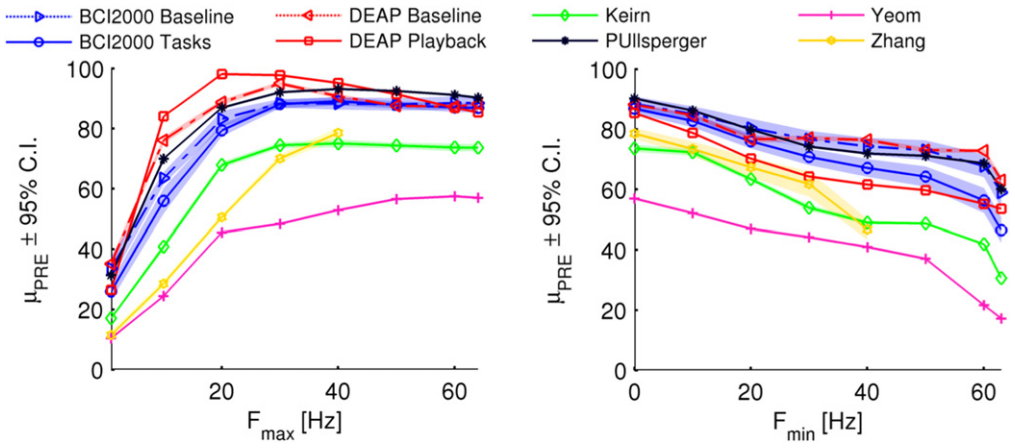

Figure 7. Quantitative analysis of the maximum (F_{max}) and minimum (F_{min}) cut-off frequencies. Mean and 95% CI of PRE results obtained in *full-fusion no-focus* experiments with different F_{max} (left) and F_{min} (right) values, corresponding to *Conf-HalfLen* system. A maximum of 20 subjects was used in each experimental iteration. Refer to figure C.4 for further related results.

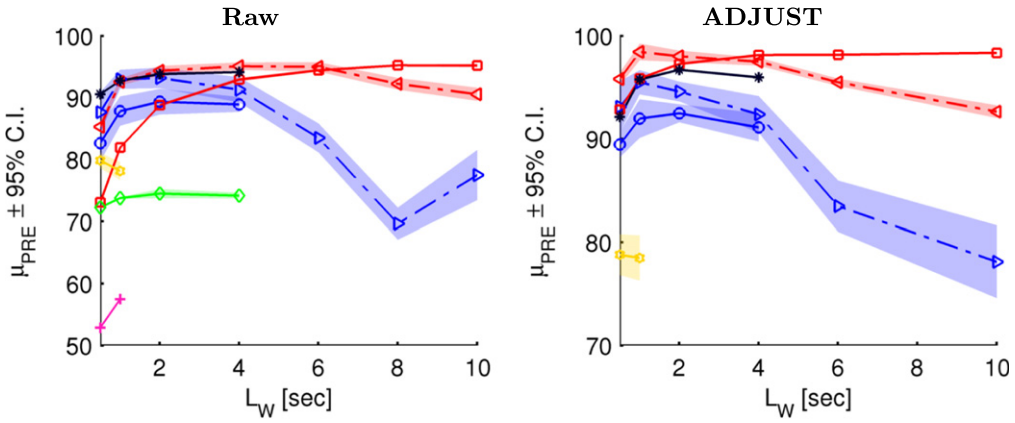

Figure 8. Quantitative analysis of the STFT window length (L_W). Mean PRE and 95% CI (shaded area) obtained with different L_W. The remaining parameters were set to their optimal values according to previous experiments. Data correspond to *full-fusion* set-up in *no-focus* experiments with raw databases (left) and after the application of ADJUST processing (right). A maximum of 20 subjects was used in each experimental iteration. Refer to figure 7 for details on the legend.

performance curves for windows sorter than 2 s suggests that these configurations are more sensitive to noise.

4.5. Review of EEG segment length with optimal parameters

A review similar to the previous was run for the length of the EEG signal used to compute the final response (L_G). We defined here configurations *Conf*-1*s* and *Conf*-2*s*. These had $\Theta = 75\%$ and a L_W of 1 and 2 s respectively, with N_F and L_G set to their optimal values (section 4.7). Note that, since the duration of events in Yeom's and Zhang's databases was 1 s, only *Conf-HalfLen* and *Conf-FullLen* were possible (a 75% overlap as used in *Conf*-1*s* in not achievable in this case). Additionally, in BCI2000-Tasks and Keirn's databases, configurations *Conf-HalfLen* and *Conf*-2*s* are equivalent.

Results. In this case, results are remarkably similar across databases, with PRE curves showing an asymptotic behaviour with increasing L_G (figures 9 and C.5). On databases with EEG signals too sort to show the asymptote itself, the PRE curves still behave in a way similar to databases with longer signals, clearly suggesting the presence of such a limit. Overall, the maximum performance is obtained using all the available EEG.

To study the described behaviour, we fitted the rational model

$$\mathrm{PRE} = \frac{a + bL_G}{c + dL_G}, \tag{10}$$

with *a*–*d* fitted factors, and used $LW = 60$ s as the maximum PRE point. Such model can be seen to accommodate well the dynamics of the data. Exceptions were found on the curves of Ullsperger's database after rejection of artefacts, and on DEAP-Baseline with *Conf-HalfLen* system after the application of ADJUST (figure C.5). For all data sets (raw and artefact free) and all systems, we found that [92.5%–95%] of the maximum performance (asymptote) was obtained at L_G between 4 and 6 s.

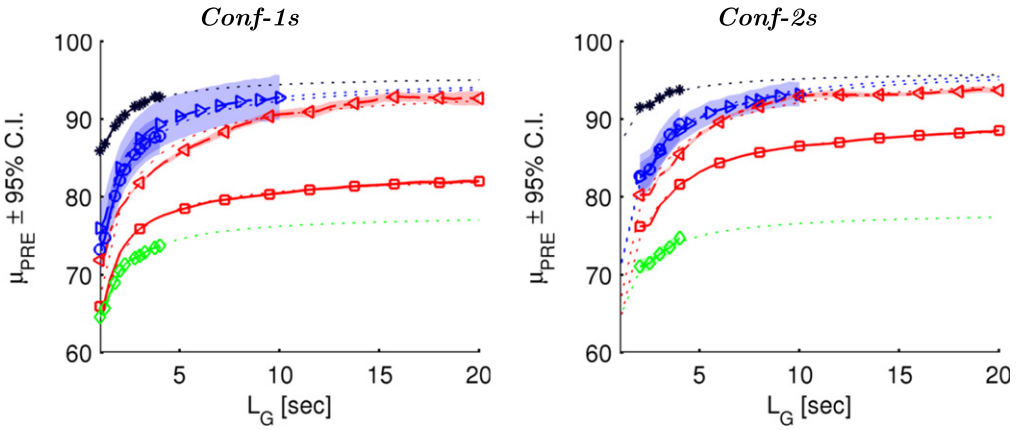

Figure 9. Quantitative analysis of the EEG segment length (L_G). Mean PRE and 95% CI (shaded area) obtained with different L_G. Data correspond to *full-fusion* set-up in *no-focus* experiments using *Conf*-1*s* (left) and *Conf*-2*s* (right) systems. A maximum of 20 subjects was used in each experimental iteration. Refer to figure 7 for details on the legend, and to figure C.5 for further related results.

Discussion. An increase in performance with longer EEG recordings is to be expected, as this translates in more information being fed into the system. Once the system has enough information to overcome the distortion of noisy segments, the performance levels out. This asymptotic behaviour has been studied here by means of a first-degree/first-degree rational model. For the shorter databases (less than 4 s), the amount of data available to fit the model was relatively small, and results should be considered accordingly. Having said that, the homogeneity across all (sort and long) data sets is obvious from the represented curves, which suggest that, although more experimentation is needed to make a final statement, the described behaviour may be a good approximation of the true interaction between performance and L_G.

4.6. A comparison across systems

To finalize the analysis of the results obtained in all the previous experiments, we compare here the performance of the systems defined in sections 4.3 and 4.5, i.e. *Conf-HalfLen*, *Conf-FullLen*, *Conf*-1*s* and *Conf*-2*s* (tables 6 and C.7). Overall, the fragmentation of the EEG signal into shorter overlapping windows (between 1 and 2 s) outperformed the use of all the EEG signal at once. DEAP-Playback data set was the main exception, but only in *no-focus* experiments. Raw processed Ullsperger's database also performed better with *Conf-FullLen*, but by less than 1 percentage point. Yeom's and Zhang's databases should be considered aside, as they only allow configurations *Conf-HalfLen* and *Conf-FullLen* due to the limited amount of EEG signal available (1 s).

Such behaviour may be explained by the way *Conf*-1*s* and *Conf*-2*s* process the data. By breaking the available data into segments, we isolate localized noise into individual segments, or more precisely, into a set of them as we allow 75% overlap between windows. At the same time, segments with good quality signal are also obtained. While noisy segments may yield random outputs from the system, clean segments will presumably result in accurate responses with high confidences (scores). When averaging outputs across segments, we expected the response from clean segments to be selected.

4.7. Conclusions

Throughout these experiments, we have defined the optimal configuration of the STFT to maximize subject-discriminant information. In particular, the fragmentation of the EEG signal into shorter overlapping windows (between 1 and 2 s) has been identified as the best overall approach. The spectral domain of each segment should then be computed without compromising in frequency resolution ($N_F \approx L_W^* F_s$), while also keeping in mind that discriminative information was found primarily under 40 Hz.

With regard to the amount of EEG needed, it is a matter of maximizing the information fed to the system. Having said that, from our analysis we conclude that $L_G = [4, 6]$ s results in performances between 92.5% and 95% of the maximum achievable by the data. However, we can expect this to vary depending on the quality of the data.

5. Representation of the time and frequency domains: results, discussion and conclusions

Once the optimal configuration of N_F, L_W, Θ and L_G was found, we assessed the effects of different time and frequency representations. We compared the discriminant information of AvgMnt, BIHMnt and CzMnt montages, as well as the effects of several PSD normalization methods. In the following experiments, we executed the four focus modes (table 5) in *full-fusion* with the four defined systems (*Conf-HalfLen*, *Conf-FullLen*, *Conf*-1*s* and *Conf*-2*s*).

5.1. Montages

The above experiments have all been executed with common global average reference montage (AvgMnt) reference. Here, we evaluate the effects of bipolar inter-hemispheric reference

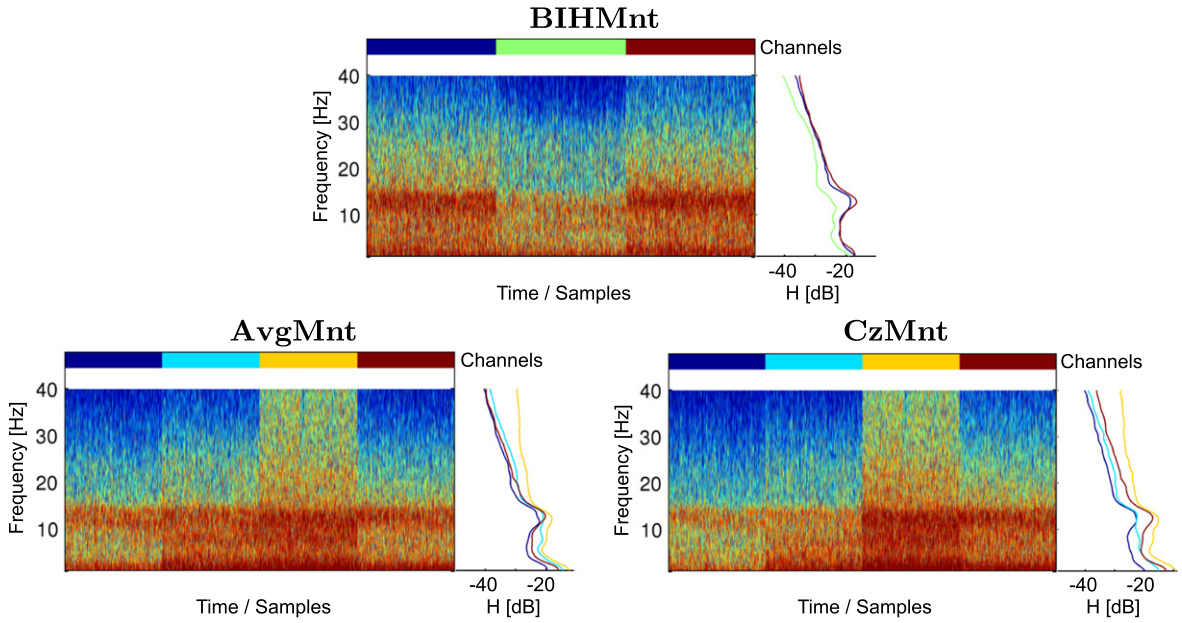

Figure 10. Qualitative analysis of EEG montages. PSD (in dB) of a subject from Yeom's database computed with BIHMnt (top), AvgMnt (bottom-left) and CzMnt (bottom-right). The colour bar-code above each PSD represents changes in channels, in the piecewise-continuous time axis. Each channel's spectrum (H) is added to the right of the spectrogram using the same colour scheme as the channel bar-code.

montage (BIHMnt) and common Cz reference montage (CzMnt) montages on the EEG discriminant information compared to AvgMnt.

Results. Qualitatively, the differences across montages were quite subtle (figure 10). Overall, AvgMnt seemed to have slightly more stable spectral patterns. Occasionally, BIHMnt showed a pronounced loss of information, probably due to highly correlated inter-hemispheric channels[3].

Quantitative results between and within databases were quite heterogeneous. In general, mean PRE values between montages were within 5 percentage points in *Conf-1s* and *Conf-2s* systems, and within 10 percentage points in *Conf-HalfLen* and *Conf-Full-Len* (figure 11 and tables C.8–C.10). The application of artefact rejection had virtually no effect on the described overall relationship between montages, other than define their differences by reducing the dispersion of the results (figure C.6, and tables C.8–C.11).

Discussion. The inhomogeneity of the results should come as no surprise. Optimal montages have been described before as highly dependent on the recording paradigm and level of noise [31]. Interestingly, BIHMnt, although reducing the volume of data by almost half, maintained a remarkably similar performance (within 5% or 10% points), especially with *Conf-1s* and *Conf-2s* systems. This makes it an attractive option in scenarios where the volume of data is a concern.

[3] In cases where all rejected sensors cover a common area; e.g left frontal lobe, the result of the interpolation of the sensors closer to the saggital midline may be extremely correlated with that of its hemispheric pair, resulting in an almost flat response of the BIHMnt.

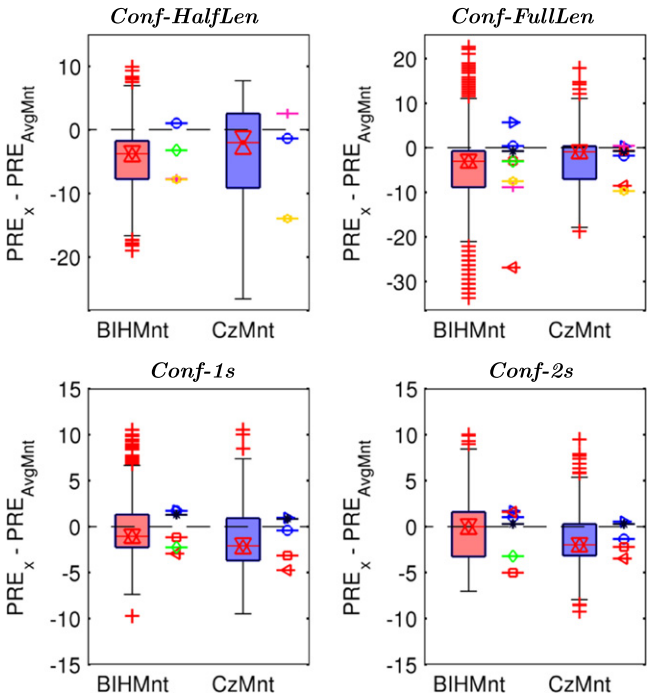

Figure 11. Quantitative analysis of EEG montages with raw data sets. Relative PRE values between *BIHMnt–AvgMnt* and between *CzMnt–AvgMnt*. Boxes show results stacked across databases. Box limits are 25 and 95 percentiles, while black bars shows maximum and minimum values after excluding outliers (red crosses). The red line within each box and triangle markers shows median values and their 95% CI. To the right of each box, corresponding mean PRE values from each database are shown following the legend of figure 7. A maximum of 20 subjects was used in each experimental iteration. Refer to figure C.6 for further related results.

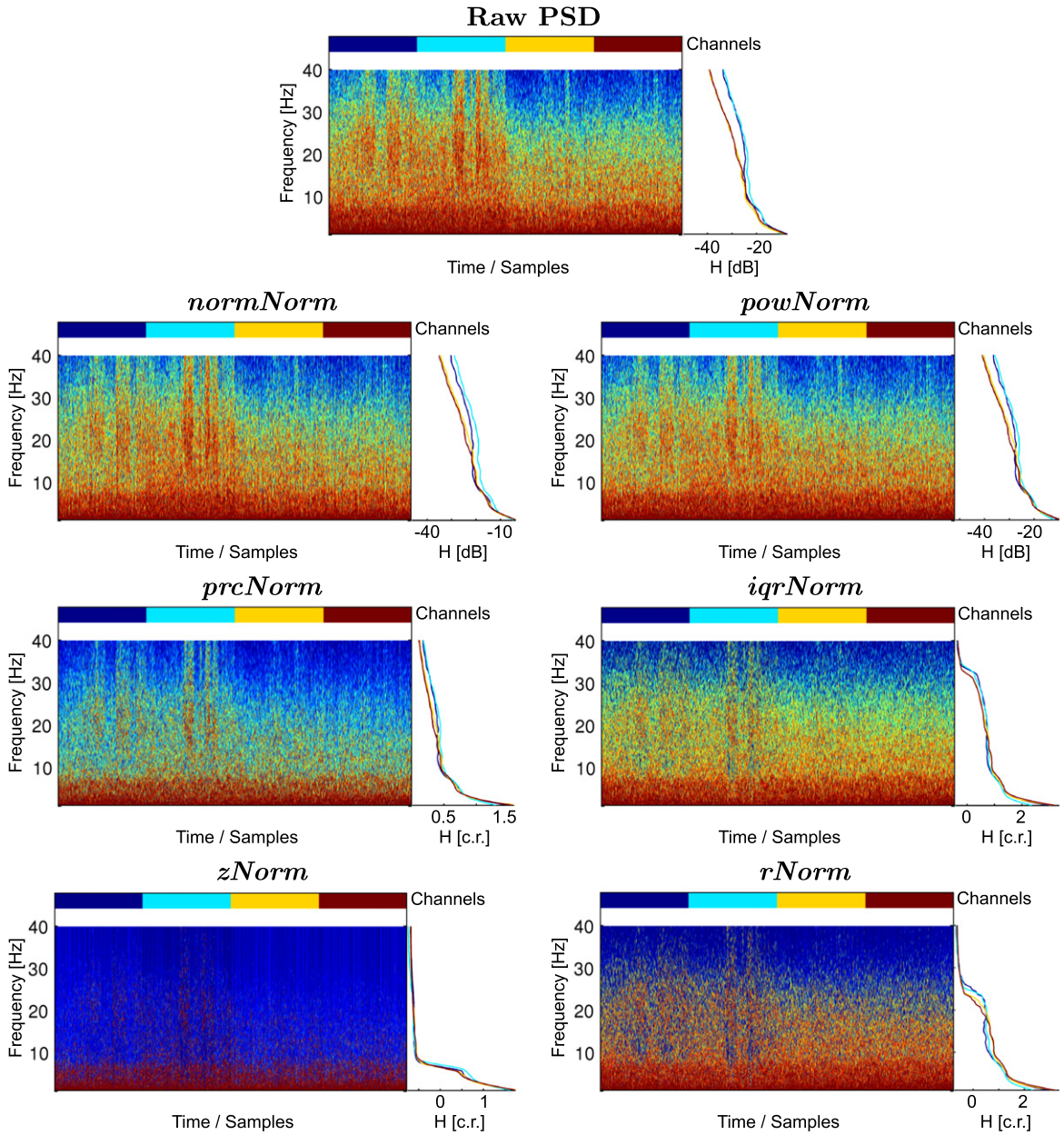

Figure 12. Qualitative analysis of PSD normalization. *Conf*-2*s* PSD, in dB, of a subject from Ullsperger's database. prcNorm, iqrNorm, zNorm and rNorm are cubic root scaled (c.r.) instead, as they contained negative values.

Across databases, AvgMnt is the safest option among the montages tested.

5.2. Spectral normalization

Next, we normalized the PSD of each window segment with the methods described in table 3. Following previous results, we again used AvgMnt.

Results. In this case, the effect of each normalization varied greatly across system configurations. Overall, each had a homogenizing effect on the spectral pattern of the EEG, especially for frequencies above ∼20 Hz (figure 12). This effect was particularly strong for prcNorm, iqrNorm and rNorm methods.

Quantitatively, results where again quite heterogeneous (figure 13 and tables C.12–C.15). While Keirn's database always benefited substantially from normalization—with 20 percentage points of μ_{PPRE} improvement on average—the others showed great variability across systems. Overall, normNorm, prcNorm and zNorm were the worst performing normalizations. rNorm and, to a lesser extent, iqrNorm gave equal or better performance than the raw PSD in almost all cases, especially in *Conf*-1*s* and *Conf*-2*s* systems.

The rejection of artefacts by ADJUST (figure C.7 and tables C.16–C.19) had a great impact on the relationship between normalization methods and the raw PSD. In particular, the latter gained the most from the cleaning,

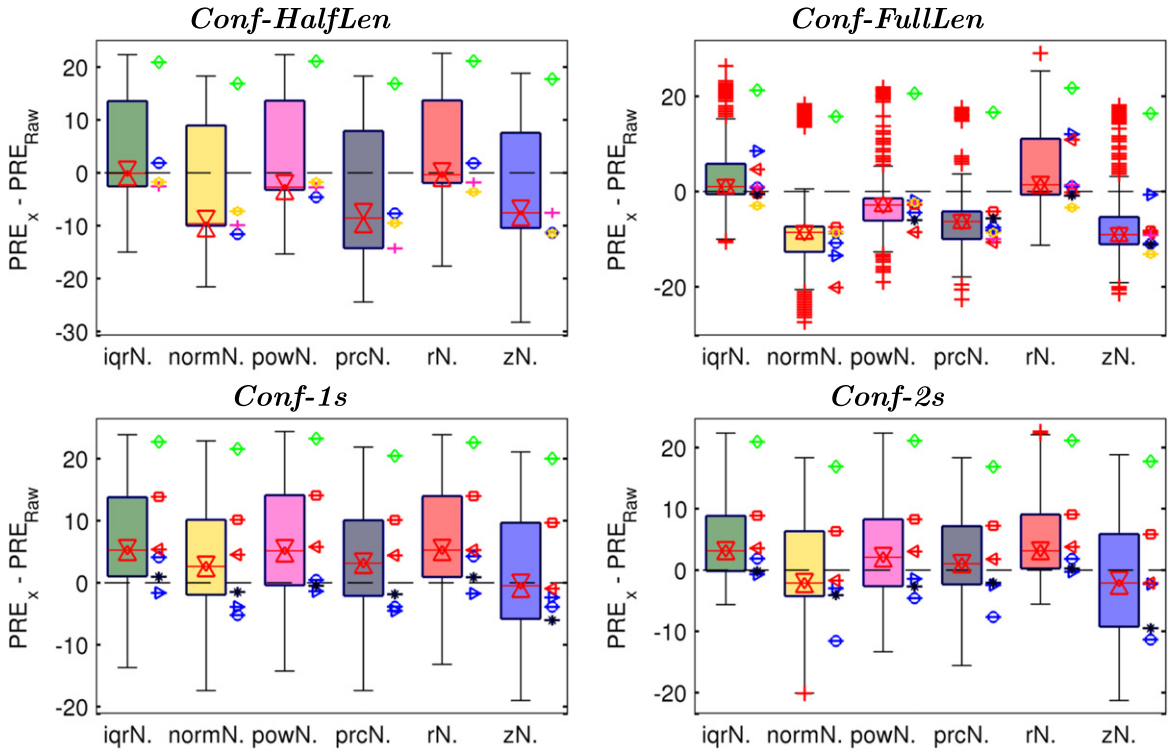

Figure 13. Quantitative analysis of PSD normalization. Relative PRE values between the PSD normalized by each of the methods in table 3 and the raw PSD. Results are stacked across databases. A maximum of 20 subjects was used in each experimental iteration. Refer to caption of figure 11 for details about the meaning of symbols within the image, and to figure C.7 for further related results.

which translated in normalization methods being less beneficial, or even disadvantageous.

Discussion. In line with [37, 41] results, normalization approaches based on magnitudes sensitive to outliers had a negative effect under some configurations. On the other hand, those based on more robust measurements, such as rNorm and iqrNorm, boosted the performance in almost all cases, especially with optimal configurations (*Conf*-1*s* and *Conf*-2*s*). This, together with the modulating behaviour of the artefact rejection methods, suggests that:

(1) the discriminant information is coded within the relationship between spectral coefficients (i.e. the global spectral shape), rather than in the instant absolute power of each frequency;
(2) and the enhancement observed via robust normalizations is, at least in part, due to the counteraction of effects from artefacts.

5.3. Conclusions

In this experimentation block, we have assessed for the first time the effects of EEG montages (time representation) on the EEG subject discriminant information. Although AvgMnt turned out to be the best performing montage on average across system configurations, the performance of BIHMnt was strikingly similar, especially when considering the highly reduced data volume of BIHMnt. In addition, normalizing the spectral coefficients using measurements robust against outliers reduced the effect of noise and boosted the quality of the data in almost all cases. Hence, it may be applied as an alternative to artefact rejection methods.

6. Properties of the discriminant information: results, discussion and conclusions

Next, we described some of the properties of the discriminant information identified in the previous experimentation blocks. Specifically, we studied the spatial and frequency distribution of the discriminant information, as well as its uniqueness across individuals and permanence along time.

We evaluated the systems *Conf-HalfLen*, *Conf-FullLen*, *Conf*-1*s* and *Conf*-2*s* (sections 4.3 and 5) with the raw PSD coefficients and with their rNorm version.

6.1. Spatial distribution

The discriminant information of different sensor locations was evaluated through *no-focus* and *sess-focus* experiments in *freq-fusion* architectures, i.e. running isolated experiments within sensors.

Results. Looking at the PSD representations, although there were instances of patterns with large variability across sensors, there were also numerous cases where the pattern remained relatively equal across channels (figure 10). Quantitative results showed no clear pattern in the spatial distribution of the performance across data sets, tasks or

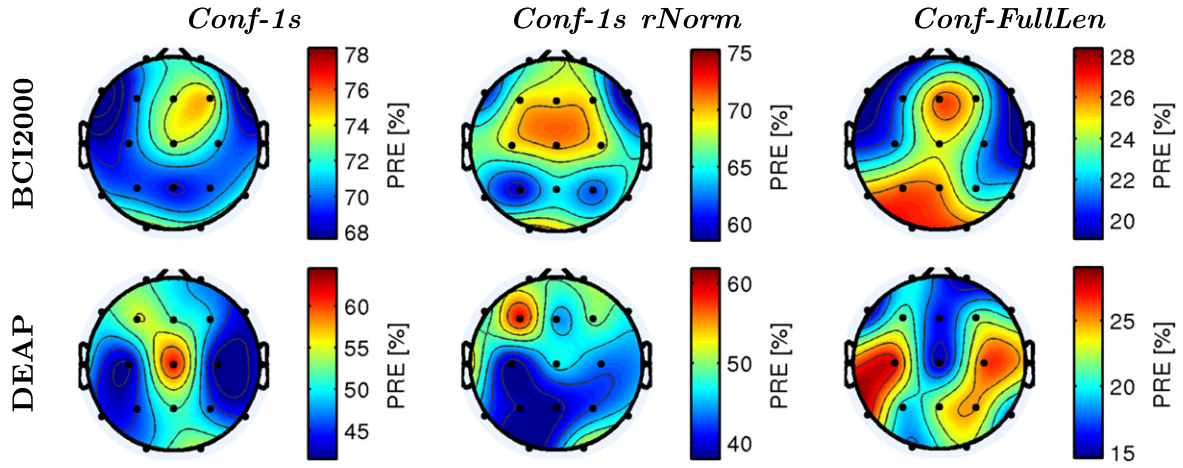

Figure 14. Quantitative analysis of the spatial distribution of the discriminant information. Mean PRE values obtained at each location with the REO condition of BCI2000 (top) and DEAP (bottom) data sets when applying the systems *Conf*-1*s* (left), *Conf*-1*s* rNorm (centre) and *Conf-FullLen* (right). Results correspond to *freq-fusion ch-focus* experiments, with a maximum of 20 subjects used in each experimental iteration. Refer to figure C.8 for further related results.

systems. In fact, this distribution varied even within conditions. For example, focusing on the REO condition of BCI2000 and DEAP databases, we observed a great variability across system configurations as well as between data sets (figures 14 and C.8). The removal of artefacts brought the performance between channels closer by boosting those with the worst μ_{PRE} values.

Discussion. The lack of uniformity across and within databases, tasks and systems suggests that there is no obvious 'most-discriminative' region. Rather, the homogenizing effect of removing the artefacts hints that the performance depends more on the strengths of the applied system and on the idiosyncrasies of each session's set-up, which affects the quality of the signal individually on each channel. This would explain the lack of consensus in the literature regarding the performance of sensor locations during subject identification. Looking at research with similar results, such as in [9, 28, 34] or [1, 2, 50, 51], they all relied on EEG signals recorded from a single task and from the same database[4].

On the other hand, one has to consider that the systems applied here rely on the general spectral shape of the EEG. Systems that focus on specific characteristics of the EEG during a particular task, such as the power of the alpha rhythm during REC or the P300 amplitude during VEP, can be expected to have a more defined spatial distribution of the discrimination power.

6.2. Frequency distribution

To study the effect of each frequency component, we ran *no-focus* and *sess-focus* experiments with *ch-fusion* and *no-fusion* architectures.

[4] Because databases are not generally labelled, it is difficult to be certain whether the same database was used for different analyses. However, the descriptions of the databases suggest that they used the same database.

Results. In general, quantitative experiments showed noisy μ_{PRE} curves idiosyncratic to each database (figure 15). However, within this variability, we observed some common characteristics across data sets, the most distinctive being a peak within the alpha rhythm. After this peak, the PRE raised again passed 15 or 20 Hz, until it reached the global maximum—in some cases, curves of Keirn's database decreased after reaching this maximum. With *Conf-FullLen* systems, the PRE also increased to the left of the alpha peak (towards lower frequencies) and reached a local maximum at 1 Hz. This behaviour was also reproduced in *no-fusion* experiments within each channel.

The rejection of artefacts had no major effects on the described behaviour (figure C.9). The performance of frequencies above the alpha rhythm remained virtually the same. With *Conf*-1*s* and *Conf*-2*s* systems, the discrimination power of frequencies below the alpha peak raised, especially with ADJUST processed databases where, on occasion, the gain was as large as 30 percentage points. In some cases, this increase created another local peak between 1 and 8 Hz.

Discussion. The observed behaviour is in line with those presented in section 4.3. They also correspond with other genetic [14, 53] and biometric [28, 43] studies at least with regard to the amount of information within the delta and alpha rhythms. In addition to this, our results suggest that frequencies corresponding to the beta rhythm and up to 40 Hz carry as much discriminant information as the delta and alpha rhythms. Furthermore, as reported in [24, 45], the high-beta and gamma band (up to 40 Hz) reached performance levels, on occasion, above those of lower bands.

6.3. Uniqueness

The current experimental step aims to asses the uniqueness of the neural signature. If this signature is comprised of a finite number of categorical characteristics of EEG activity, it will only be able to classify subjects into said categories. Hence, individual discrimination beyond these categories would not be

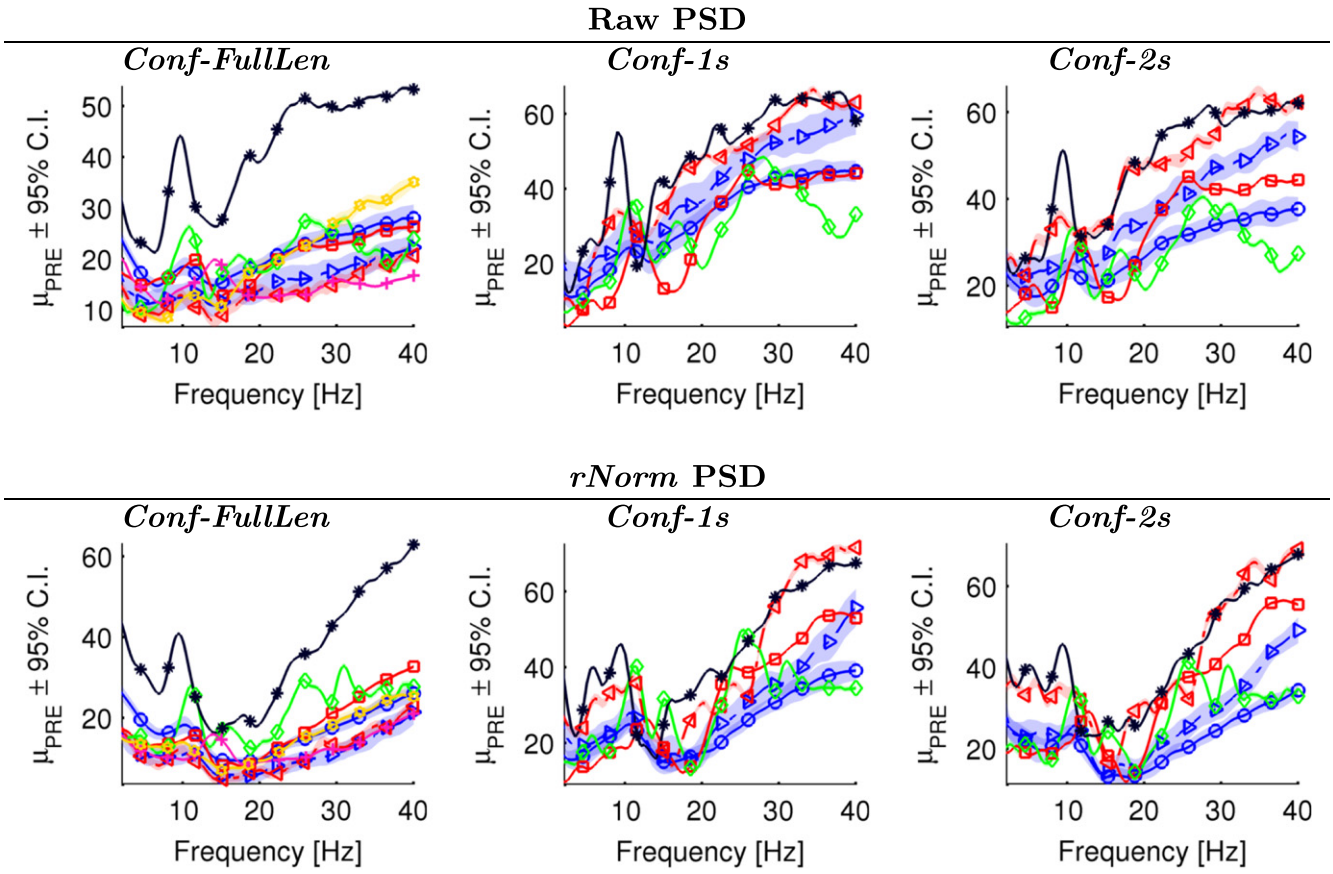

Figure 15. Quantitative analysis of the discrimination power frequency distribution. Mean PRE and 95% CI (shaded area) obtained with each frequency (*ch-fusion* experiments). Curves where smoothed by local regression, using weighted linear least squares and a first degree polynomial model with a 3 Hz span. Graphs correspond to results obtained with the raw PSD (top) and rNorm PSD (bottom). A maximum of 20 subjects was used in each experimental iteration. Refer to figure 7 for details on the legend, and to figure C.9 for further related results.

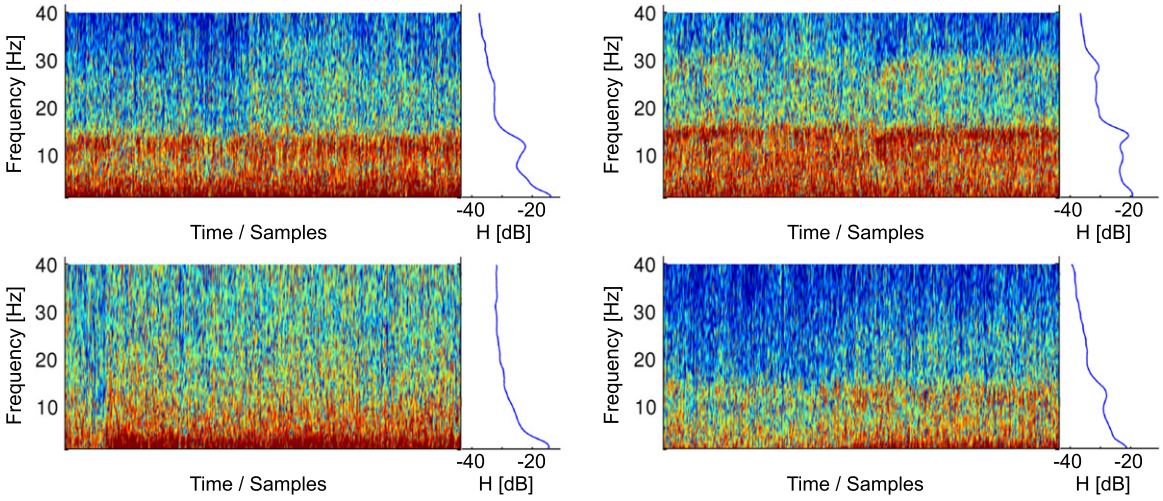

Figure 16. Qualitative analysis of uniqueness. PSD, in dB, of four subjects from Yeom's database, corresponding to the EEG sensor P3 and a single task (self-representation).

possible. This will be reflected as a performance slump when the number of subjects in the system increases. In order for the system to unequivocally classify among a large number of individuals, the neural signature must be a continuous characteristic. Thus, we tested the uniqueness of the neural signature by evaluating the systems with an increased number of subjects and analysing the dynamics of the obtained PRE curve.

Results. Differences across subjects in the shape of their EEG spectrum are obvious (figure 16). From experiments,

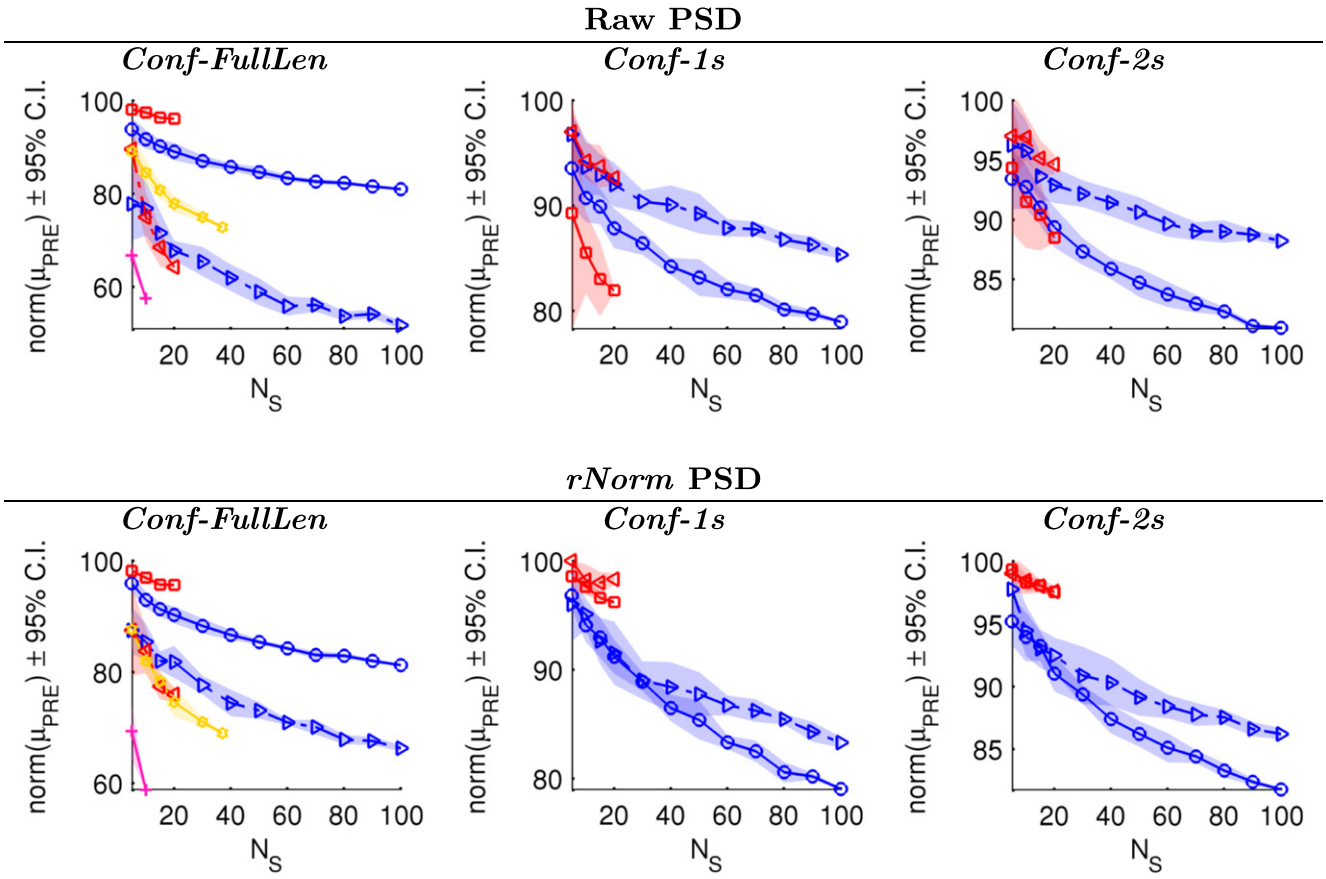

Figure 17. Quantitative analysis of uniqueness. Mean PRE and 95% CI (shaded area) for different numbers of subjects (N_S) in the system and each database. Results correspond to the rNorm system. Refer to figure 7 for details on the legend, and to figure C.10 for further related results.

systems *Conf*-1*s* and *Conf*-2*s* were the most robust against the rise in the number of subjects. In particular, their μ_{PRE} dropped less than 10% when increasing the number of users from 5 to 20 (figure 17). With BCI2000 databases, their performance fell less than 20% between 5 and 100 subjects. The application of ADJUST artefact rejection attenuated such decay by as much as 6 percentage points (figure C.10).

Discussion. Results show that EEG spectral patterns are distinctive enough to discriminate between 100+ subjects. A PRE of 80%, the lowest observed with *Conf*-1*s* and *Conf*-2*s* systems for 100 subjects, represents an accuracy rate of almost 80%. This is an encouraging result, especially considering that the systems used here are fairly simple ones. Overall, in line with [49], our results suggest that the neural signature is comprised of continuous features. In this regard, larger databases are necessary to find the true potential of the EEG discrimination power.

6.4. Permanence

In a bid to evaluate the stability of the EEG subject traits, we ran *session-CV* experiments. To allow a direct comparison of results, we also ran extra tests using normal CV (no *sess-CV*) with a K (from K-folds) equal to the number of sessions.

Results. From the representation of the spectrograms, time seemed to have a moderate to large effect on the neural signature (figure 18). Quantitatively, median PRE values decreased less than 5 percentage points on Keirn's database and between 17 and 22 percentage points on Yeom's data set when comparing normal and *sess-CV* (figure 19 and table C.20). The use of rNorm had opposite effects on both databases, increasing the difference by ∼10 percentage points in the former and reducing it by a similar amount in the latter.

Discussion. Genetic and neurophysiological studies have described changes in human EEG activity across maturation [55]. However, these are long term effects that are disproportionally relevant to younger ages (until approximately 19–20 years old). In the short term, the PSD appears to be relatively stable, comparable with other biometric modalities.

The above results are compatible with those obtained by analogous experiments [8, 20, 27, 29, 32]. However, a limitation of all of these studies (including the present one) is the low number of subjects and sessions in the used databases. Having said that, although a drop in performance is anticipated when train and test sets drift further apart in time:

(1) this can be circumvented with a multi-session or a continuous training approach;

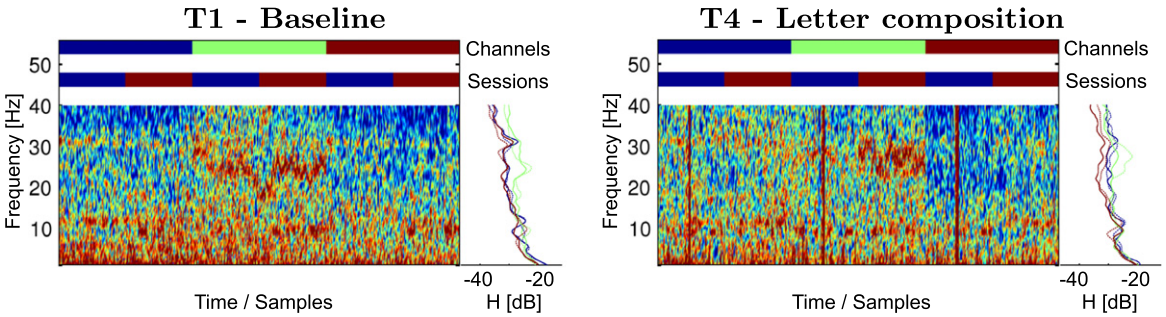

Figure 18. Qualitative analysis of permanence. PSD in dB of tasks T1 and T2 of one subject from Keirn's database. The colour bar-codes above each spectrogram represent changes in channels and sessions. PSD for each channel and session is attached to the right of the spectrograms, using the same colour scheme as the channel bar-codes.

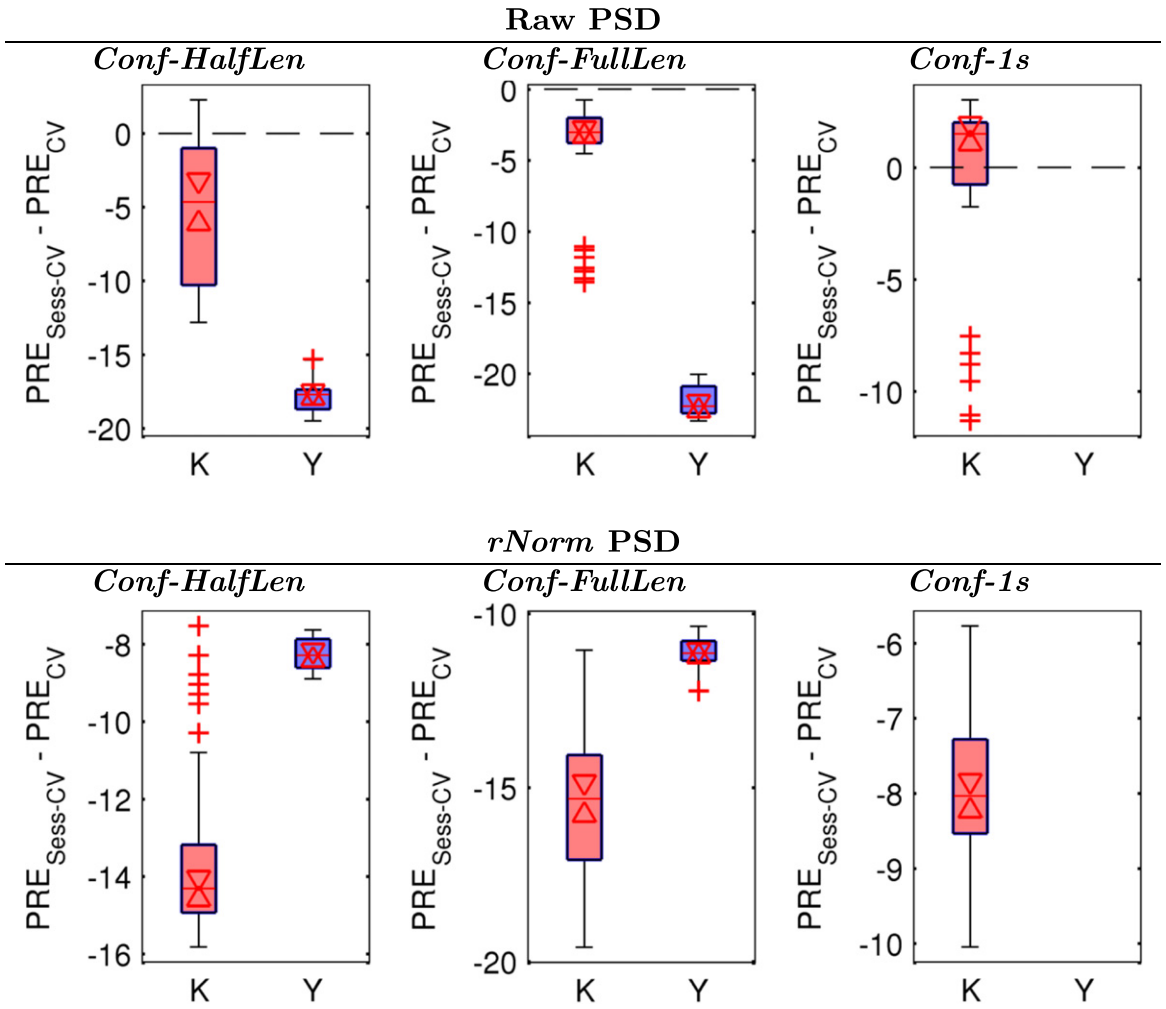

Figure 19. Quantitative analysis of permanence. Relative PRE values between normal and *sess-CV* experiments, obtained with the raw (top) and the rNorm (bottom) PSD of Keirn's (K) and Yeom's (Y) databases. Note that for Keirn's data set, *Conf-HalfLen* is equivalent to *Conf-2s*. Refer to the caption of figure 11 for details on box markings within the image.

(2) results suggest that the decline may pose a problem no worse than the one observed on other biometric modalities; and
(3) this drop would be expected to be lower than the results presented here—with a single training session, the system is unable to model variations across time, leaving it more vulnerable to time effects.

6.5. Conclusions

Along this experimentation block, we have defined various properties of the discriminant information within the EEG PSD, some of which were obscured by the lack of consensus across previous research. Specifically, our results suggest that there is no clear location outperforming others systematically

across systems, databases and cognitive tasks. Hence, we conclude that the relative performance of individual sensor locations appears to be largely driven by idiosyncrasies of the recordings set-up and by the characteristics of the system.

Looking at the frequency distribution of the discriminant information, results hint at the existence of a performance peak within the alpha rhythm. In addition, the beta rhythm, up to 40 Hz, seemed as much or more discriminant than lower frequencies. Finally, it should be noted that an extra performance peak may arise below 5 Hz.

Vitally, our results suggest that this subject-specific information is 'unique enough' to discriminate between a high number of users (>100) and is relatively constant along short periods of time. Nevertheless, additional experimentation with larger databases recorded on multiple sessions—with greater temporal distances between each session—are needed to make stronger assertions in this regard.

7. Summary and overall conclusions

In this work, we have presented the results of an extensive study of the individual's discriminant information within the time-frequency representation of EEG signals. In doing so, we have used six databases (divided into eight data sets) with different recorded cognitive tasks and states. This, together with the performed complementary qualitative and quantitative analyses, allowed us to distinguish inherent characteristics of the EEG signal from idiosyncrasies of individual data sets. Specifically, we ran three experimentation blocks, each with a specific goal, which resulted in the following synthesized recommendations and conclusions:

(1) Configuration of the PSD: recommendations.
   (a) Record, at least, 5 s of EEG to perform the identification.
   (b) Divide the EEG into segments between 1 or 2 seconds long.
   (c) If data volume and computational speed is not an issue, use some degree of window overlap.
   (d) Compute the spectral representation for each window, using a number of spectral coefficients similar to the number of samples within the window to maintain the spectral resolution.
   (e) Retain a bandwidth from the lowest frequencies to 30 or 40 Hz.
   (f) Perform classifications for each window individually and generate a single response by fusing scores.
(2) Representation of the time and frequency domains: recommendations.
   (a) Use *AvgMnt* as the default montage.
   (b) Consider using *BIHMnt* in cases where processing time or data volume is a concern.
   (c) As a substitute for complex artefact rejection methods, you may normalize the spectral coefficients with a method robust to outliers.
(3) Properties of the discriminant information: conclusions
   (a) There seems to be no best performing sensor location found across systems, databases and/or cognitive tasks.
   (b) In terms of frequency distribution, there is a performance peak within the alpha rhythm. Frequencies within the beta rhythm (up to 40 Hz) are much or more discriminant than lower bands. Frequencies below 5 Hz may also contain an important amount of discriminant information.
   (c) Subject traits within the EEG activity are 'unique enough' to discriminate 100+ subjects when an appropriate system configuration is used.
   (d) Subject-specific EEG spectral patterns seem to be 'permanent enough' to use them as a biometric modality.

In future works, we will use all of the above recommendations and conclusions to (1) design an EEG based subject identification system and (2) further explore the properties of the discriminant information within the EEG.

References

[1] Abdullah M, Subari K, Loong J and Ahmad N 2010 Analysis of effective channel placement for an EEG-based biometric system *Proc. of the IEEE EMBS Conf. on Biomedical Engineering and Sciences* p 303–306
[2] Abdullah M, Subari K, Loong J and Ahmad N 2010 Analysis of the EEG signal for a practical biometric system *World Acad. Sci. Eng. Technol.* **4** 1123–7
[3] Abo-Zahhad M, Ahmed S and Abbas S 2015 State-of-the-art methods and future perspectives for personal recognition based on electroencephalogram signals *IET Biometrics* **4** 179–90
[4] Abo-Zahhad M, Ahmed S M and Abbas S N 2015 A novel biometric approach for human identification and verification using eye blinking signal *IEEE Signal Process. Lett.* **22**
[5] Armstrong B C, Ruiz-Blondet M V, Khalifian N, Kurtz K J, Jin Z and Laszlo S 2015 Brainprint: assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics *Neurocomputing* **166** 59–67
[6] BCI2000. Bci2000 public database. www.bci2000.org, 2000. Online; accessed 1 February 2014
[7] Berger H and das Über 1929 Electroencephalogram des Menschen *Arch. für Psychatrie und Nervenkrankheiten* **87** 527–70
[8] Brigham K and Kumar B 2010 Subject identification from electroencephalogram (EEG) signals during imagined speech *Proc. of the 4th IEEE Int. Conf. on Biometrics: Theory Applications and Systems* pp 1–8
[9] Campisi P, Scarano G, Babiloni F, DeVico Fallani F, Colonnese S, Maiorana E and Forastiere L 2011 Brain waves based user recognition using the 'eyes closed resting conditions' protocol *Proc. of the IEEE Int. Workshop on Information Forensics and Security* pp 1–6
[10] Collura T 1993 History and evolution of electroencephalographic instruments and techniques *J. Clin. Neurophysiol.* **10** 476–504
[11] Davis H and Davis P 1936 Action potentials of the brain: In normal persons and in normal states of cerebral activity *Arch. Neurology Psychiatry* **36** 1214–24
[12] DEAPdataset (http://eecs.qmul.ac.uk/mmv/datasets/deap/)

[13] del Pozo-Baños M, Alonso J B, Ticay-Rivas J R and Travieso C M 2014 Electroencephalogram subject identification: a review *Expert Syst. Appl.* **41** 6537–54
[14] Eischen S, Luckritz J and Polich J 1995 Spectral analysis of EEG from families *Biol. Psychol.* **41** 61–8
[15] Fraschini M, Hillebrand A, Demuru M, Didaci L and Marcialis G 2014 An EEG-based biometric system using eigenvector centrality in resting state brain networks *IEEE Signal Processing Lett.* **22** 666–70
[16] Goldberger A L, Amaral L A N, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C-K and Stanley H 2000 Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals *Circulation* **101** e215–e220
[17] Gui Q, Jin Z, Blondet M V R, Laszlo S and Xu W 2015 Towards eeg biometrics: similarity-based approaches for user identification *Proc. of the IEEE Int. Conf. on Identity, Security and Behavior Analysis* pp 1–6
[18] Gui Q, Jin Z and Xu W 2014 Exploring EEG-based biometrics for user identification and authentication *Proc. of the IEEE Signal Processing in Medicine and Biology Symp.* (*SPMB*) (Piscataway, NJ: IEEE) pp 1–6
[19] Gupta C, Palaniappan R and Swaminathan S 2008 On the analysis of various techniques for a novel brain biometric system *Int. J. Med. Eng. Inf.* **1** 266–73
[20] Hu B, Liu Q, Zhao Q, Qi Y and Peng H 2011 A real-time electroencephalogram (EEG) based individual identification interface for mobile security in ubiquitous environment *Proc. of the IEEE Asia-Pacific Services Computing Conf.* pp 436–441
[21] Judd C M, McClelland G H and Ryan C S 2011 *Data Analysis: A Model Comparison Approach* (New York, NY: Routledge)
[22] Keirn Z 2012 (www.cs.colostate.edu/eeg/main/data/2011-12_BCI_at_CSU)
[23] Keirn Z and Aunon J 1990 A new mode of communication between man and his surroundings *Biomed. Eng.* **37** 1209–14
[24] Kennet F June 2008 Brain wave based authentication *Master's Thesis* Gjøvik University College, Department of Computer Science and Media Technology
[25] Koelstra S, Muhl C, Soleymani M, Lee J-S, Yazdani A, Ebrahimi T, Pun T, Nijholt A and Patras I 2012 Deap: a database for emotion analysis; using physiological signals *IEEE Trans. Affective Comput.* **3** 18–31
[26] Kohavi R 1995 Wrappers for Performance enhancement and oblivious decision graphs *PhD Thesis* Department of Computer Science, Stanford University
[27] Kostílek M and Šťastný J 2012 EEG biometric identification: repeatability and influence of movement-related EEG *Proc. of the Int. Conf. on Applied Electronics* (*AE*) pp 147–150
[28] La Rocca D, Campisi P and Scarano G 2012 EEG biometrics for individual recognition in resting state with closed eyes *Proc. Int. Conf. of the Biometrics Special Interest Group* (*BIOSIG*) pp 1–12
[29] la Rocca D, Campisi P and Scarano G 2014 Stable EEG features for biometric recognition in resting state conditions *Biomedical Engineering Systems and Technologies* (Berlin: Springer) pp 313–330
[30] La Rocca D, Campisi P, Vegso B, Cserti P, Kozmann G, Babiloni F and De Vico Fallani F 2014 Human brain distinctiveness based on EEG spectral coherence connectivity *IEEE Trans. Biomed. Eng.* **61** 2406–12
[31] Libenson M H 2012 *Practical Approach to Electroencephalography* (Philadelphia, PA: Elsevier Health Sciences)
[32] Marcel S and Millan J April 2007 Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation *IEEE Trans. Pattern Anal. Mach. Intell.* **29** 743–52
[33] Mognon A, Jovicich J, Bruzzone L and Buiatti M 2010 ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features *Psychophysiology* **48** 229–40
[34] Mohammadi G, Shoushtari P, Ardekani B and Shamsollahi M February 2006 Person identification by using ar model for EEG signals *World Acad. Sci. Eng. Technol.* **11** 281–5
[35] Muthukumaraswamy S D 2013 High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations *Frontiers Hum. Neurosci.* **7** 128
[36] Nolan H, Whelan R and Reilly R 2010 FASTER: fully automated statistical thresholding for EEG artifact rejection *J. Neurosci. Methods* **192** 152–62
[37] Palaniappan R 2004 Method of identifying individuals using vep signals and neural network *Proc. of the IEEE on Science, Measurement and Technology Conf.* **vol 151,** pp 16–20
[38] Palaniappan R 2005 Identifying individuality using mental task based brain computer interface *Proc. of the 3rd Int. Conf. on Intelligent Sensing and Information Processing* pp 238–242
[39] Palaniappan R, Gosalia J, Revett K and Samraj A 2011 Pin generation using single channel EEG biometric *Advances in Proc. of the Computing and Communication* (*Communications in Computer and Information Science* vol 193) pp 378–385
[40] Palaniappan R and Raveendran P 2002 Individual identification technique using visual evoked potential signals *Electron. Lett.* **38** 1634–5
[41] Palaniappan R and Ravi K 2006 Improving visual evoked potential feature classification for person recognition using pca and normalization *Pattern Recognit. Lett.* **27** 726–33
[42] Paranjape R, Mahovsky J, Benedicenti L and Koles Z 2001 The electroencephalogram as a biometric *Proc. of the Canadian Conf. on Electrical and Computer Engineering* vol 2, pp 1363–1366
[43] Poulos M, Rangoussi M, Alexandris N and Evangelou A 2001 On the use of EEG features towards person identification via neural networks *Med. Inf. Internet Med.* **26** 35–48
[44] Poulos M, Rangoussi M and Kafetzopoulos E 1998 Person identification via the EEG using computational geometry algorithms *Proc. of the 9th European Signal Processing Conf.* pp 2125–2128
[45] Power A, Lalor E and Reilly R 2006 Can visual evoked potentials be used in biometric identification? *Proc. of the 28th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society* pp 5575–5578
[46] Ravi K and Palaniappan R 2005 Leave-one-out authentication of persons using 40 HZ EEG oscillations *Proc. of the Int. Conf. on Computer as a Tool* vol 2, pp 1386–1389
[47] Singhal G and RamKumar P 2007 Person identification using evoked potentials and peak matching *Biometrics Symp.* pp 1–6
[48] Stassen H 1980 Computerized recognition of persons by EEG spectral patterns *Electroencephalogr. Clin. Neurophysiol.* **49** 190–4
[49] Stassen H, Bomben G and Hell D 1998 Familial brain wave patterns: study of a 12-sib family *Psychiatric Genetics* **8** 141–54
[50] Tangkraingkij P, Lursinsap C, Sanguansintukul S and Desudchit T 2009 Selecting relevant EEG signal locations for personal identification problem using ica and neural network *Proc. of the 8th IEEE/ACIS Int. Conf. on Computer and Information Science* pp 616–621
[51] Tangkraingkij P, Lursinsap C, Sanguansintukul S and Desudchit T 2010 Personal identification by EEG using ica and neural network *Proc. of the Int. Conf. on*

Computational Science and Its Applications—Part III (*Lecture Notes in Computer Science* vol 6018) (Berlin: Springer) pp 419–430

[52] Ullsperger P (http://sccn.ucsd.edu/wiki/Chapter_02:_STUDY_Creation)

[53] van Beijsterveldt C and van Baal G 2002 Twin and family studies of the human electroencephalogram: a review and a meta-analysis *Biol. Psychol.* **61** 111–38

[54] Vogel F 1970 The genetic basis of the normal human electroencephalogram (EEG) *Hum. Genetic* **10** 91–114

[55] Vogel F 2000 *Genetics and the Electroencephalogram* (Berlin: Springer)

[56] Yang S and Deravi F 2012 On the effectiveness of EEG signals as a source of biometric information *Proc. of the 3rd Int. Conf. on Emerging Security Technologies* pp 49–52

[57] Yeom S, Suk H and Lee S 2011 EEG-based person authentication using face-specific self representation *Proc. of the Korea Computer Congress* vol 38, pp 379–382

[58] Yeom S, Suk H and Lee S 2012 Person authentication from neural activity of face-specific visual self-representation *Pattern Recognit.* **46** 1159–69

[59] Zhang X 1999 (http://archive.ics.uci.edu/ml/datasets/EEG+Database)

[60] Zhang X, Begleiter H, Porjesz B and Litke A 1997 Electrophysiological evidence of memory impairment in alcoholic patients *Biol. Psychiatry* **42** 1157–71

[61] Zhang X, Begleiter H, Porjesz B, Wang W and Litke A 1995 Event related potentials during object recognition tasks *Brain Res. Bull.* **38** 531–8

[62] Zietsch B, Hansen J, Hansell N, Geffen G, Martin N and Wright M 2007 Common and specific genetic influences on EEG power bands delta, theta, alpha, and beta *Biol. Psychol.* **75** 154–64