International Journal of Neural Systems

EVIDENCE OF A TASK-INDEPENDENT NEURAL SIGNATURE IN THE SPECTRAL SHAPE OF THE ELECTROENCEPHALOGRAM

MARCOS DELPOZO-BANOS
Division of Digital Signal Processing, IDeTIC,University of Las Palmas de Gran Canaria, Las Palmas, 35017, Spain

CARLOS M. TRAVIESO
Division of Digital Signal Processing, IDeTIC,University of Las Palmas de Gran Canaria, Las Palmas, 35017, Spain
E-mail: ctravieso@ulpgc.es

JESUS B. ALONSO
Division of Digital Signal Processing, IDeTIC,University of Las Palmas de Gran Canaria, Las Palmas, 35017, Spain
E-mail: jalonso@ulpgc.es

ANN JOHN
College of Medicine, Swansea University
Swansea, SA2 8PP, Wales, UK
E-mail: a.john@swansea.ac.uk

Genetic and neurophysiological studies of electroencephalogram (EEG) have shown that an individual's brain activity during a given cognitive task is, to some extent, determined by their genes. In fact, the field of biometrics has successfully used this property to build systems capable of identifying users from their neural activity. These studies have always been carried out in isolated conditions, such as relaxing with eyes closed, identifying visual targets or solving mathematical operations. Here we show for the first time that the neural signature extracted from the spectral shape of the EEG is to a large extent independent of the recorded cognitive task and experimental condition. In addition, we propose to use this task-independent neural signature for more precise biometric identity verification. We present two systems: one based on real cepstrums and one based on linear predictive coefficients. We obtained verification accuracies above 89% on 4 of the 6 databases used. We anticipate this finding will create a new set of experimental possibilities within many brain research fields, such as the study of neuroplasticity, neurodegenerative diseases and brain machine interfaces, as well as the mentioned genetic, neurophysiological and biometric studies. Furthermore, the proposed biometric approach represents an important advance towards real world deployments of this new technology.

Keywords: Electroencephalogram; biometry; task-independent; neural signature.

1. Introduction

As early as 1936 – only 12 years after Hans Berger recorded the first human electroencephalogram (EEG)[1] – twin EEG research by H. and P. Davis evidenced the existence of a brain activity inheritance model.[2] They concluded that the posterior rhythm of the resting EEG between twins were as similar as recordings from an individual across time. Many studies followed, including analyses of EEG recorded from twins reared apart to isolate exogenous factors,[3] from families to evaluate the continuity of the phenotypic range,[4] and from datasets recorded across long periods of time to assess the inheritance of maturation processes.[5] They revealed that the EEG follows elaborated models of inheritance affecting a wide range of properties, especially the power and peak frequency of the alpha rhythm over occipital regions.[3, 6, 7]

These results laid the foundation for the first attempts at automatic EEG-based biometric identification.[8, 9] The field has progressed substantially since then. Some recent works include the use of functionality connectivity between brain regions,[10]

spectral coherence,[11] wavelet package decomposition,[12] similarity-based approaches[13] and features of the N400.[14] Refs.15–17 are detailed reviews of the state of the art, while Ref.18 provides a fundamental study of the properties of this neural signature.

To date, EEG genetic, neurophysiological and biometric studies have assessed numerous recording conditions, including relaxed and engaged states and stimulus-elicited activity, but always in an isolated manner. This should come as no surprise given that, historically, functional brain research has mainly studied the differences across task-related or condition-related activity. The only multitask studies come from the biometric field,[19–21] and are aimed at finding the most favourable condition performance-wise.

Here we propose a fundamentally different approach. We hypothesize the existence of a neural pattern homogeneous across cognitive tasks and recording conditions, which is concomitant to the subject's identity. We call this 'task-independent neural signature'. Specifically, this paper presents two novelties in this field. First, we provide for the first time evidence of the existence of such task-independent signature. Second, we propose for the first time a biometric identity verification system that relies on the task-independent EEG signature.

To understand the magnitude of the difference between this (task-independent) and previous (task-specific) studies, let us consider fingerprints and written signatures respectively as analogous examples. While the former uses properties 'inherent'[a] to the individual, the latter focuses on how they perform a given task. Hence, we argue that task-specific studies are closer in nature to describing behaviour (idiosyncratic activity during cognitive processing), while the current research tries to describe identity in and of itself.

In the remainder of this paper, we will first introduce the 6 publicly available databases used during experimentation, as well as their preparation (section 2). We will continue detailing the experimental methods, describing algorithms, evaluation procedures and experiment designs (section 3). We will then present all the obtained results (section 4), followed by the discussion, including a comparison of our results with those of the state of the art, some hypotheses of the physiological source of the signature, and a discussion of the advantages of the proposed task-independent approach (section 5). To finalise, we will clarify the limitations of this work (section 6) and summarize its conclusions (section 7). When necessary, the reader is referred to the supplementary material for additional results and discussion, as well as for the result of statistical tests.

[a] Strictly speaking, fingerprints are not inherent parts of an individual, as a subject can still *be* without them.

2. Materials

In a bid to gather enough evidences, we tested our hypothesis on an array of 6 publicly available data bases of different nature. To remove uninteresting differences across them, we applied a common preprocessing stage. This stage filtered out unwanted or contaminated frequencies from the EEG signal; rejected highly noisy channels, trials, sessions and subjects; interpolated rejected channels to retrieve the full original set of sensors; normalized the sampling frequency across databases to 128 Hz; and selected a common subset of 19 channels evenly spread throughout the scalp. For a detailed description of this preprocessing, the reader is referred to Ref.18

As a result of the preprocessing and after preparing each dataset we obtained the following databases used during experimentation (name code in parenthesis):

(B) BCI2000[22, 23] contains data from 100 subjects (from the 109 original ones) while performing 6 different tasks, including two 1-minute baseline runs of resting with eyes open (REO) and closed (REC), and three 2-minute runs of 2 motor and 2 motor-imagery tasks: (T1 and T2) a target appears on either the left or the right side of the screen, the subject opens and closes (T1) or imagines opening and closing (T2) the corresponding fist until the target disappears; and (T3 and T4) a target appears on either the top or the bottom of the screen, the subject opens and closes (T3) or imagines opening and closing (T4) either both fists (if the target is on top) or both feet (if the target is on the bottom) until the target disappears. In total, 11 different tasks/conditions were identified: left/right in T1 and T2, and top/bottom in T3 and T4 were differentiated –, and 9 or 10 4-seconds trials per task/condition were extracted from the data.

(D) Dataset for Emotion Analysis using EEG, Physiological and Video Signals (DEAP)[24, 25] was originally collected to study emotional responses. It contains 20 subjects (from the 32 original ones – 50% males, aged between 19 and 37, mean age 26.9) while

they visualized 40 60-second music videos which elicited different emotions. Before each session, a 2-minute baseline was recorded with subjects REO. In total 5 tasks/conditions were identified, i.e. REO and the four quadrants of the valence-arousal representation of emotions, and we extracted 5 to 10 20-seconds trials per task/condition.

(K) Keirn's dataset[26, 27] contains data from 5 subjects (from the 7 original ones – 6 males and 1 female between the ages of 21 and 48) recorded during 2 different-day sessions. Subjects performed 5 tasks: (T1) relax, (T2) mentally solve non-trivial multiplication problems, (T3) mentally rotate 3-dimensional complex objects, (T4) mentally write a letter to a friend or a family member and (T5) visualize numbers being written on a blackboard sequentially. Each of these tasks was repeated 5 times under both REC and REO on every session. We extracted 8 to 10 2-seconds trials per task. Only 6 channels were available: O1/O2, P3/P4, C3/C4.

(P) P. Ullsperger's dataset[28] contains Auditory Evoked Potentials (AEPs) recorded from 5 subjects. Auditory stimuli (words) were presented to the participants, who had to classify each of the stimuli as synonyms or non-synonyms of a given target. The number of trials varied across subjects. Inter-stimulus time between trait and test stimulus was set to 1 second. We extracted 180 4.1-seconds trials per condition.

(Y) Yeom's dataset[29, 30] contains Visual Evoked Potentials (VEPs) from 10 male subjects (from the 11 original ones); including 1 pair of monozygotic twins, with ages between 20 and 29 years old (mean 26.67). Self and non-self images were presented to the subjects in 2 different-day sessions. Each session consisted of 20,000 trials divided into 2 runs (with a short break in between), 50 blocks per run, and 20 trials per block (10 self and 10 non-self stimuli). We extracted 900 1-second trials per condition. Only 18 channels were available spread throughout the scalp.

(Z) Zhang's dataset[31–33] contains VEPs of 30 subjects (from the 37 original ones) exposed to black and white images taken from the set of Snodgrass and Vanderwart[34] in an identification problem. Subjects were asked to determine whether the first stimulus was the same as the second stimulus. In some cases, only one stimulus was presented. Forty trials were recorded from each subject. We identified 3 conditions (references, targets and lures) and extracted 15 to 20 1-second trials per condition.

Together, the above datasets encompass the following tasks[b]: relaxed states, VEPs, AEPs, motor-tasks, intellectual tasks and elicited emotions.

To reduce bias from artefacts during exploration of the signature, an artefact free version of the databases was computed during preprocessing. Specifically, we applied ADJUST[35], a tool based on the Independent Component Analysis (ICA) representation of the EEG. ADJUST automatically identifies artefactual independent components from time and topological features by means of an unsupervised classification method. We slightly modified the algorithm to account for missing features by simply ignoring them from the calculations. Due to the requirements of ICA, ADJUST was only applied to datasets with more than 20 channels (i.e. all but Keirn's and Yeom's). See Ref. 18 for details on how ADJUST was applied within the preprocessing.

3. Methods

We followed an experimentation approach composed of two stages. First, we focussed our efforts on obtaining firm evidence of the existence of a task-independent neural signature. Second, we designed a biometric identification system capitalizing on such task-independent neural signatures.

3.1. *Evidence of a task-independent neural signature*

This experimentation stage builds on the results and methods of Ref.18. The reader is referred to Ref.18 for any further details.

3.1.1. *System*

We made use of a supervised classification system. The system is first fed with a set of sample-identity pairs (training set). Subsequently, the system is presented with test samples and asked to find the matching identities.

We used the Sort-Time Fourier Transform (STFT) for the computation of the EEG's time-frequency representation. We applied a 2-seconds long Hamming window during STFT segmentation to avoid edge leakage, and a 75% overlap between windows. We

[b] For simplicity, we will use the term 'tasks' to refer to changes in conditions, states or cognitive tasks accordingly for each database.

computed 256 Power Spectral Density (PSD) coefficients for each window and concatenated those between 1 and 40 Hz from all EEG channels to build the feature vector. This vector was fed to a Gaussian Naive Bayes classifier with uniform prior distribution across classes, which classified each sample into one of the registered subjects. Results from all windows were added together to generate the final response for each sample. Because Yeom's and Zhang's datasets had 1-second long recordings, a 0.5-seconds window was used and 64 coefficients calculated during STFT in these cases (results highlighted in grey).

We evaluated the following 3 different system conditions/configurations:

- *Raw system*: The system as described above fed with the preprocessed databases.
- *ADJUST processed databases*: The system as described above fed with the artefact free version of the databases.
- *rNorm systems*: A robust normalization method, which has been found to reduce the effect of artefacts on the spectral shape,[18] is applied to the PSD coefficients (*H*) of each window. This method is defined as

$$rNorm(H) = \frac{H - median(H)}{iqr(H)}, \qquad (1)$$

where *iqr* is the interquartile range function. Normalizing factors are computed from the training data.

3.1.2. *Evaluation*

The performance of the above system was measured through a multi cross-validation (CV) approach, applying stratified K-Folds within 20 Monte Carlo (MC) iterations. This K-Fold + MC design benefits from the stability (lack of bias) of the former and the low-variance of the latter.[36] To maintain the testing parameters across databases as similar as possible and reduce the computational time, the number of subjects considered was limited to 20 for BCI2000 and Zhang's databases – the only sets with more than 20 subjects. These were randomly selected for each MC iteration.

We first computed accuracies within MC repetitions. Let *M* be a 4D matrix with dimensions $C \times K \times N_s \times N_s$, where *C* is the number of MC iterations, *K* is the number of folds in K-Folds and N_s is the number of subjects in the experiment. For simplicity, let $M_{(i,j)}$ be the sub-matrix of *M* corresponding to the $N_s \times N_s$ confusion matrix of the i-th MC repetition and j-th K-Fold iteration. Then, $M_i = \sum_j M_{(i,j)}$ is the aggregated confusion matrix $N_s \times N_s$ for the i-th MC iteration. Let A_i be the corresponding mean accuracy rate defined as the proportion of correctly classified samples:

$$A_i = \frac{\sum_{k=1}^{N_S} M_i(k,k)}{\sum_{k=1}^{N_S}\sum_{m=1}^{N_S} M_i(k,m)}. \qquad (2)$$

Because at this stage it was not necessary to have absolute accuracy rates, we transformed A_i into the more easily interpreted Percentage Reduction of Error (PRE)[37] comparing the system with a random process (i.e. chance classification accuracy $1/N_S$). Formally,

$$PRE_i = 100\frac{A_i - 1/N_S}{1 - 1/N_S}\%. \qquad (3)$$

As a result, PRE is 0% if the system performs at chance level, and 100% if it has perfect accuracy, regardless of the number of users used during experimentation (N_S). The mean PRE (μ_{PRE}) and 95% Confidence Intervals (CI) were finally computed across all MC iterations.

3.1.3. *Experiments*

Within this phase, we executed 3 experiments:

(Task-CV) We began by asserting that the EEG contains task-independent discriminant information. Tasks were crossed between training and testing sets. The system was trained with samples from some tasks and evaluated with samples from other tasks (e.g. trained with tasks A and B and tested with tasks C and D). Crucially, we executed this segmentation individually for each subject. As a result, in a given iteration, a task may be used to train some subjects and test others (e.g. subject 1 trained with task A and tested with task B; subject 2 trained with task B and tested with task A). This forces the system to use task-independent characteristics. Hence, a PRE above 0% will suggest that the proposed hypothesis is true.

(*Single-Task* and *Bal-CV*) We then assessed how much of the identity information within the EEG is task-independent and how much is task-specific. We compare *Task-CV* results with *Single-Task* and *Bal-CV* ones. In *Single-Task*, the system was fed with the EEG from each task individually – as it is common in the

state of the art. In other words, the system was trained and tested with a single task each time. To compute a single PRE, we aggregated the confusion matrices obtained for each task. During *Bal-CV,* all tasks were used for training and testing, with samples from each task evenly distributed between training and testing sets (i.e. balanced CV). In this case, a PRE on *Single-Task* or *Bal-CV* similar to that on *Task-CV* will suggest that the task-specific signature contains little extra information over and above the task-independent signature.

(*Sess-CV*) Finally, we assessed the permanence of the task-independent property across time using Keirn's and Yeom's datasets – the only sets with 2 recording sessions. We repeated the previous experimental modes while crossing sessions between training and testing sets (i.e. training the system with samples from one session and testing it with samples from the opposite session). For example, in the *Sess-CV* + *Task-CV* mode, the system was trained with samples from a set of tasks in one session (tasks A and B, session 1), and tested with samples from the opposite tasks and session (tasks C and D, session 2). This experiment also serves to assure that the performance is not due to idiosyncrasies of the setup, such as the exact location and impedance of EEG channels. In this case, a PRE above 0% will suggest that results of our previous experiments are not due to idiosyncrasies of the data, and that the task-independent neural signature is stable across time[c].

Recall that the system always aims at recognising the identity of the subject. EEG from different cognitive tasks are strategically used here as a mean to answer our questions. The system is, at all times, unaware of what tasks correspond to the fed samples.

We run *Sess-CV* experiments with Keirn's and Yeom's databases, and 2 folds (from K-Folds CV). For the other experiments, we distributed samples from each session evenly between training and testing sets, and set the number of folds to the minimum between 5 and the number of tasks in each dataset.

3.2. Biometric identity verification

During the second stage, we assessed the potential of a biometric identity verification system based on the task-independent neural signature.

[c] Due to the limited data, we cannot extract strong conclusions about the permanence of the neural signature. See section 6 for more details.

3.2.1. *System*

In a verification system, users request access by providing a 'user name' (i.e. their identity) and a password (in our case, their EEG). The system then analyses this information and concludes whether both pieces match (genuine user) or not (impostor). This paradigm is arguably more suitable for the biometry in hand than that of classification, as the user's consent and collaboration will always be required to proceed, more so than in other biometries, since users need to wear an EEG device during the process.

We followed a step-by-step approach were we started with a baseline design similar to that used in phase 1 (section 3.1.1) and introduced small changes to each of its components. We will only present the two best performing systems: one based on Real Cepstrums (RCeps) and one on Linear Prediction Coefficients (LPC). Short descriptions of all the systems evaluated can be found in Appendix A.

Cepstral coefficients have been extensively used on signal processing problems.[38] The RCeps (*C*) – usually just called '*cepstrums*' – are defined as

$$C(q) = \left| FFT^{-1}\left(\log\left(\left|FFT(X(t))\right|^2\right)\right)\right|^2 \qquad (4)$$

where *X(t)* is any signal in the time domain and *q* is a quefrency index with time units. Here, we delimited the application of the FFT^{-1} operator within the [1, 40] Hz range, and set it to compute the same number of coefficients as *FFT* (i.e. 64 for Yeom's and Zhang's datasets and 256 for the rest). The cepstral space codes the broad shape of the spectrum in the lower quefrencies (first coefficients) and its details and periodicity in higher quefrencies.

The coefficients of an Auto-Regressive model or LPC have been a popular choice for subject characterization within the EEG biometric identification literature.[15] Such a model predicts samples of a time series (*X(t)*) as a function of the past *N* observations, where *N* is the order of the system. This is typically defined as

$$X(t) = c + \sum_{i=1}^{N} \varphi_i X(t-i) + \varepsilon_t \,, \qquad (5)$$

with *c* a constant, φ_i the LPC and ϵ_t white noise. The above can also be seen as the output of an all-pole Infinite Impulse Response system with noise presented at its input. Therefore, the LPC describe the spectral

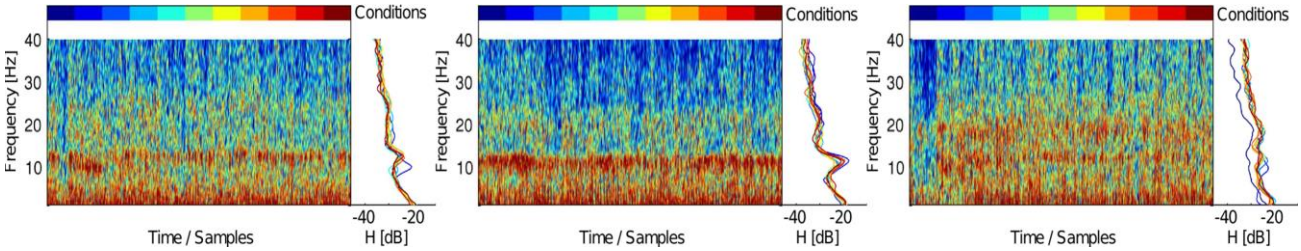

Figure 1: Neural signature of three subjects from BCI2000 database, extracted from channel C3. Images are spectrograms from multiple trials stacked to generate a piece-wise continuous in time representation of the signature. The median spectrum curve can be seen to the right of each spectrogram. Different tasks are coded by the top colour bars. These colours match the corresponding spectrum curve at the right of the spectrograms. Conditions include different real and imaginary motor tasks as well as REO and REC conditions.

shape of the modelled signal. The higher the order of the model, the more detailed the description.

In both systems, the EEG signal was again segmented into 2-second windows 50% overlapped, with RCeps and LPC computed for each window. We concatenated coefficients from all EEG sensors to build the feature vector. A Linear Discriminant Classifier (LDC) evaluated this vector. Results from all windows were added together to generate the final response for each sample.

We defined the RCeps system to use the first *P*% of all the computed cepstral coefficients ($\text{Rceps}_{P\%}$), and the LPC system to have a order *N* (LPC_N). Based on previous results,[18] we aimed at capturing the overall shape of the spectrogram and discarded irrelevant details. We ran a battery of experiments with increasing *P* (between 5% and 100%) and *N* (between 2 and 50), and found $\text{RCeps}_{20\%}$ (i.e. $P = 20$) and LPC_8 (i.e. $N = 8$) to be the overall optimal points across datasets (Appendix B). Thus, the feature vectors had length $N_{Ch} \times (0.2 \times N_q)$ for $\text{RCeps}_{20\%}$ and $N_{Ch} \times 7$ for LPC_8, with N_{Ch} the number of channels (i.e. 6 for Keirn's dataset, 18 for Yeom's and 19 for the rest) and N_q the number of cepstral coefficients (i.e. 64 for Yeom's and Zhang's datasets and 256 for the rest).

3.2.2. *Evaluation*

As before, experiments combined stratified K-Folds and MC CV methodologies. The number of folds (from K-Folds) was set to 2, and the number of MC iteration was set equal to the number of subjects in the dataset. In databases where the number of subjects was smaller than 20, we repeated the whole K-Fold MC process M times until at least 20 experiments were executed. For example, for Keirn's database (5 subjects) we repeated the whole process 4 times (4 x 5 = 20).

Because we are now interested in the absolute performance of the systems, instead of PRE, we computed and reported accuracy results (Acc.), defined as the average between the sensitivity (percentage of positive samples correctly classified), also known as Genuine Acceptance Rate (GAR), and the specificity (percentage of negative samples correctly classified).

$$\text{Acc.} = \frac{\text{sensitivity} + \text{specificity}}{2}. \qquad (6)$$

Receiver Operating Characteristic (ROC) curves, showing False Acceptance Rate (FAR) against GAR for different decision threshold, and optimal real ROC points (i.e. actually computed points rather than interpolated points) will also be provided.

3.2.3. *Experiments*

As we aimed at assessing the feasibility of using the task-independent neural signature, we evaluated the systems on Task-CV mode, except for Keirn's and Yeom's datasets for which we used *Sess-CV + Task-CV* mode (section 3.1.3).

Recall that these are verification systems. Within each MC iteration, a subject was treated as the positive class (registered user) and the remaining as impostors. Each subject is used as the registered user at least once in the whole CV process. In addition, we applied an open-set segmentation, where different sets of impostors were used for training and testing. Hence, impostors were segmented at subject level, using all data from an impostor (all tasks and sessions) either for training or testing – as opposed to the segmentation of the registered user data, which was done at sample level using *Task-CV* or *Sess-CV + Task-CV*.

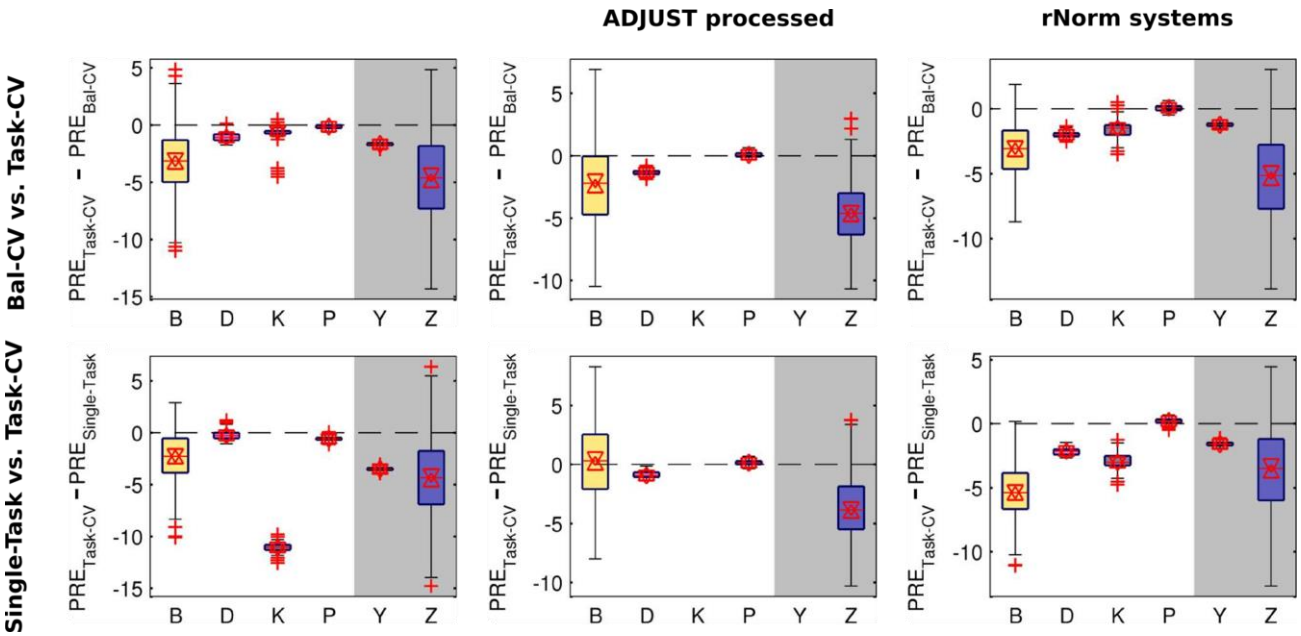

Figure 2: Difference between PRE of *Task-CV vs Single-Task* and *Task-CV vs Bal-Task* classification experiments. Box limits are 25 and 75 percentiles, while black bars shows maximum and minimum values after excluding outliers (red crosses). The red line within each box and triangle markers show median values and their 95% CI. The system uses a special configuration for Yeom's and Zhang's datasets (highlighted in grey).

Table 1: Results of *Task-CV* classification experiments. Mean PRE values and 95% confidence intervals obtained with each database (Dat.) and system configuration/condition. The system uses a special configuration for Yeom's and Zhang's datasets (highlighted in grey).

Dat.	-	ADJUST	rNorm
B	83.07 [82.13, 84.01]	87.77 [86.66, 88.89]	83.91 [83.13, 84.70]
D	97.03 [96.92, 97.14]	97.97 [97.90, 98.03]	96.26 [96.19, 96.32]
K	73.49 [73.40, 73.59]	-	93.31 [93.10, 93.52]
P	93.70 [93.65, 93.75]	96.84 [96.78, 96.90]	94.16 [94.11, 94.21]
Y	51.25 [51.18, 51.31]	-	49.79 [49.73, 49.86]
Z	74.63 [73.28, 75.98]	74.32 [73.45, 75.20]	68.74 [67.55, 69.93]

4. Results

4.1. *Evidence of a task-independent neural signature*

Looking at the visual representation of the EEG's spectrogram, the general spectral shape, although not constant, is remarkably stable across tasks and different across subjects (Fig. 1). At the same time, details such as the exact position, height and width of the spectral peaks and valleys are more sensitive to task variations.

The classification system described in section 3.1.1 was able to differentiate between subjects using task-independent identity information (Table 1). PRE values from *Task-CV* experiments are well above 0% (the point of chance) with Yeom's database yielding the lowest PREs around 50% (i.e. reducing chance's error by half). The use of ADJUST or rNorm had no major or homogeneous effects on PRE, except for an increase of 20 percentage points by rNorm in Keirn's database.

Next, we compared results between *Task-CV* and *Single-Task* or *Bal-CV* modes. PRE values of *Task-CV* was on average less than 5 percentage points lower than *Single-Task* and *Bal-CV* (Fig. 2 and Tables C.1 and C.2). The use of rNorm or ADJUST had no major effect on this relationship. PRE differences between *Task-CV* and *Single-Task* were substantially more variable across databases and configurations than between *Bal-CV* and *Task-CV*. [d]

[d] Comparisons between *Task-CV* and *Single-Task* experiments should be interpreted as hints of their real relationship, as the amount of data available to train each system (number of training samples) in the latter is lower than in the former. See section 3.1.3 for more details.

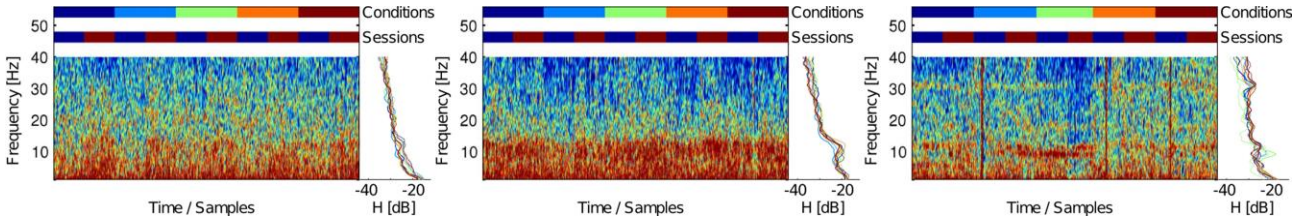

Figure 3: Neural signature from three subjects of Keirn's database, extracted from channel C3. Conditions include resting and different intellectual-tasks, such as letter-composing and arithmetic operations, recorded during two different-day sessions coded in the colour bars above each spectrogram. See the legend of Fig. 1 for more details.

Table 2: Results of *Sess-CV* classification experiments. Mean PRE values and 95% confidence intervals of *Sess-CV* + {*Task-CV; Single-Task; Bal-CV*} for Keirn's and Yeom's databases (Dat.). The system uses a special configuration for Yeom's dataset (highlighted in grey).

Dat.	Bal-CV	Sess-CV + Task-CV	Single-Task
	(a) Raw systems		
K	73.29 [72.66, 73.92]	69.07 [65.82, 72.31]	68.52 [67.72, 69.32]
Y	34.36 [33.73, 34.98]	33.11 [32.31, 33.91]	35.01 [34.52, 35.51]
	(b) rNrom systems		
K	82.18 [81.34, 83.02]	73.58 [70.20, 76.97]	78.14 [77.41, 78.87]
Y	42.60 [42.46, 42.74]	41.76 [41.38, 42.13]	42.97 [42.78, 43.16]

Finally, we assessed the permanence of the task-independent signature across time. Examining the spectrograms (Fig. 3), time fluctuations seemed more prominent than differences across tasks in some cases (Fig. 3 left panel) while on par with task variations in others (Fig. 3 right panel).

PRE results show a similar scenario than before, with performance during *Sess-CV + Task-CV* still well above chance levels (Tables 2 and C3). The decrease on PRE between *Sess-CV + Task-CV* and *Sess-CV + Bal-CV* was not greater than 10 percentage points for Keirn's database and not greater than 1 percentage point for Yeom's dataset. PRE differences between *Sess-CV + Task-CV* and *Sess-CV + Single-Task* were again highly variable. In this case, the drop was never higher than 5 and 2 percentage points for Keirn's and Yeom's databases respectively.

Table 1: Results of $RCeps_{20\%}$ and LPC_8 verification experiments. Mean accuracy and 95% CI. The systems use a special configuration for Yeom's and Zhang's datasets (highlighted in grey).

System	B	D	K
$RCeps_{20\%}$	95.40 [94.70, 96.10]	**97.59 [96.12, 99.05]**	**80.06 [73.97, 86.16]**
LPC_8	**95.42 [94.79, 96.05]**	96.82 [95.66, 97.97]	79.10 [72.03, 86.18]
System	**P**	**Y**	**Z**
$RCeps_{20\%}$	89.19 [85.55, 92.83]	73.94 [69.04, 78.84]	91.58 [89.30, 93.87]
LPC_8	**89.88 [85.49, 94.26]**	**74.73 [69.57, 79.88]**	**93.24 [91.53, 94.94]**

Table 4: Optimal real ROC points (i.e. points obtained from testing, not interpolations) of $RCeps_{20\%}$ and LPC_8 experiments. The systems use a special configuration for Yeom's and Zhang's datasets (highlighted in grey).

System	B	D	K	P	Y	Z
$RCeps_{20\%}$	96.85	97.99	80.71	94.42	76.19	93.21
LPC_8	96.06	96.89	79.36	94.46	75.61	94.02

4.2. *Biometric identity verification*

Finally, we assessed the feasibility of using the task-independent neural signature for biometric verification. Results for $RCeps_{20\%}$ and LCP_8 systems were virtually identical, with accuracies above 89% in all cases except Keirn's and Yeom's datasets, which yielded accuracies of 80% and 74% respectively (Table 3).

After inspecting the ROC curves (Table 4, Figs. B.2 and B.4), both systems obtained accuracy values close to their optimal ROC points – the only notable exception was P. Ullsperger's dataset, which underperformed on both systems by ~5 percentage

points. The obtained accuracy corresponds to a sensitivity of 99% and a specificity of 78% for $RCeps_{20\%}$ and 79% for LPC_8.

On BCI2000 and DEAP databases, $RCeps_{20\%}$ and LPC_8 systems showed a strong performance for lower FARs. GAR was above 88% for an FAR of 1%, and above 82% for an FAR of 0.5%. Experiments with Zhang's database resulted on a GAR above 63% for 1% FAR and 49% for 0.5% FAR. The remaining datasets had relatively poor performances at these FAR values. Overall, $RCeps_{20\%}$ seemed to work marginally better than LPC_8 under low FAR conditions, except for P. Ullsperger's and Zhang's datasets.

5. Discussion

5.1. *Evidence of a task-independent neural signature: discussion.*

Results from some previous EEG studies already hint that part of the EEG neural signature is indeed task-independent. First, the performance of a subject classification system has been shown to increase when fed with labelled data from multiple tasks.[21] Second, poor results have been obtained when trying to identify tasks with the same features used to identify individuals.[39] Finally, remarkably similar behaviours to variations in some of the system's parameters has been reported across systems, tasks and databases.[18] However, we present here for the first time, a direct evaluation of the existence of such task-independent neural signature within the EEG.

Performance in *Task-CV* experiments was well above chance levels in all cases, showing that inter-individual inter-task variability is greater than the intra-individual inter-task one. PRE values should have been close to 0% if differences in brain activity across tasks were predominant over the subjects' signatures. Furthermore, the small drop in performance observed between *Task-CV* and *Single-Task* or *Bal-CV* modes suggests that most of the identity information within the spectral shape of the EEG is, in fact, task-independent.

The above is reinforced by results of *Sess-CV* experiments. If our first results had been due to peculiarities of the EEG set-up (i.e. variations in the exact sensor location, sensor-scalp impedance and signal quality) rather than to real subject features within the signal, PRE values should have slumped to zero under *Sess-CV* modes. Results also suggest that the neural signature is relatively stable across time. However, this should be interpreted carefully (section 6).

All this suggest that previous studies have probably relied, inadvertently, on task-independent traits. The staggering similarity observed between spectrograms of different tasks/conditions, together with the differences between individuals, indicates that the spectral shape of continuous brain activity, as recorded by an EEG device, is more defined by the individual's identity than it is by the performed task or experimental condition.

This is not to say that the EEG is constant across tasks. In fact, on the spectrograms showed here (Figs. 1 and 3), variations between tasks can be appreciated. Cognitive tasks have a modulating effect on the signature, introducing dynamics that (A) coexist with task-independent features and/or (B) are of smaller magnitude than differences across individuals. Thus, under the described experimental conditions, inter-task variations have a smaller magnitude than inter-individual ones. The use of different algorithms may invert the situation (i.e. inter-task variations magnified while inter-individual ones minimized).

5.2. *Biometric identity verification: discussion*

We found that $RCeps_{20\%}$ and LPC_8 are the best performing characteristics from those tested. Adding any extra feature and fusing $RCeps_{20\%}$ and LPC_8 in various ways had a neutral or negative effect on the system's accuracy, as did the application of non-linear classification algorithms such as Artificial Neural Networks and Support Vector Machines with a Gaussian Kernel (see Appendix A). Therefore, we may resolve that:

(1) The information extracted by $RCeps_{20\%}$ and LPC_8 is highly correlated, as evidenced by the fact that fusing them had no effect on the system's performance.

(2) The problem of verification based on $RCeps_{20\%}$ and LPC_8 features is a linear problem, as evidenced by the fact that the evaluated non-linear classifiers only equalled the performance of LDC.

(3) $RCeps_{20\%}$ and LPC_8 can encode most of the discrimination power of the EEG's spectral shape, as evidenced by the fact that all the systems tested performed worse than, or similarly to, those based solely on $RCeps_{20\%}$ or LPC_8 – including systems that fused different measurements of the spectral data.

Overall, we obtained relatively good and stable accuracy results across systems and databases. $RCeps_{20\%}$ and LPC_8 systems maximized their potential, with performances similar to their optimal ROC points. The poor GAR for lower values of FAR in all but BCI2000 and DEAP datasets, and the relatively lower performance observed in some cases when compared with other works (section 5.3), may be due to the following reasons:

Keirn's and Yeom's databases were tested under the Sess-CV + Task-CV paradigm, which is arguably more difficult than the Task-CV one. This is especially true when considering that a single session is used for training, which in turn leaves the system prone to detecting the idiosyncrasies of that specific recording session. We can expect higher accuracies when more sessions are available for training.[40–43]

Performance on Keirn's and P. Ullsperger's databases was certainly affected by the reduced number of available subjects (only 5 subjects). Following the described experimentation methodology, 1 subject is designated as the registered user and only 2 subjects are used as impostors for training, with the other 2 utilized for testing. With such a small number of impostors, we can expect the system's performance to be more sensitive to the peculiarities of each CV partition.

The above should also be considered, to a lesser degree, for Yeom's database, which only has 10 subjects. In addition, this dataset contains EEG from two MZ twins. How these subjects are distributed among the registered and impostor sets should be presumed to have an impact on the test result.

P. Ullsperger's and Yeom's datasets contained only two experimental conditions. This left the system with a single condition for training, which, as described before, is not the ideal scenario to account for the dynamics of the neural signature described in section 5.1.

Finally, in the case of Yeom's and Zhang's databases, the systems' performance was certainly compromised by the length of the EEG segments (only 1 second). This forced us to use a window length of 0.5 seconds, as opposed to the optimal 1 or 2 seconds.

Table 5: Comparison of results obtained with BCI2000 database. Columns correspond to the publication reference (Ref.), the number of subjects used (# Subj.), the task used (Task), the CV method applied (CV), and the success rate (Succ.). For completeness, classification results are also shown, which can be deduced applying eq. 3 to the PRE results of Table 1.

Ref.	# Subj.	Task	CV	Succ.
Classification experiments				
11	109	REO & REC	5 K-Folds	100%
44	18	Task 4	3 K-Folds	96%
This	20	All	●	85%
Verification experiments				
10	109	REO	N.A.	96%
This	100	All	◊	95%

● Task-CV; 5 K-Folds + 20 MC.
◊ Open-set; Task-CV; 2 K-Folds + 20 MC.
Task 4: A target appears on either the top or the bottom of the screen. The subject imagines opening and closing either both fists (if the target is on top) or both feet (if the target is on the bottom) until the target disappears. Then the subject relaxes.

5.3. *Comparison with the state of the art*

Although BCI2000, Keirn's, Yeom's and Zhang's databases have also been used by other authors, a direct comparison between our results and those within the state of the art is impossible due to numerous reasons (tables 5 to 8). In addition to differences in the number of subjects and tasks used and in the applied experimentation methodology, the following two factors should be considered:

Firstly, the design proposed here is a generalised one. Unlike most of the systems presented by other authors, we have designed the system using multiple databases and selected the collective optimal configurations – these configurations were not necessarily optimal for individual datasets. Hence, our system is not tuned to maximize the performance within a single case, but to work well across multiple scenarios. In fact, the performance of the three largest databases are within 6 percentage points within each other, suggesting that this level of discrimination is indeed a property of the neural signature and not due to idiosyncrasies of the data.

Table 6: Comparison of results obtained with Keirn's database. For completeness, classification results are also shown, which can be deduced applying eq. 3 to the PRE results of Table 1. Refer to table 6 for a description of the columns and symbols ● and ◊.

Ref.	# Subj.	Task	CV	Succ.
Classification experiments				
20	5	All on MTL	2 K-Folds	100%
This	5	All	●	89%
Verification experiments				
19, 45	5	Task 3	2 K-Folds	100%
10	5	Task 1	2 K-Folds	80%
This	5	All	◊	80%

Task 1. Subjects were asked to relax.
Task 3. Subjects were presented with an image of a 3-D complex object and were asked to mentally rotate it.

Secondly, our proposed approach on a task-independent neural signature represents a more complex problem than that of task-specific identification via EEG. As highlighted by our comparison between *Task-CV* and *Single-Task* or *Bal-CV* modes, task specific neural activity carries some extra discriminant information. Systems from other architectures may potentially exploit this information, hence obtaining higher overall performances.

In addition, on Keirn's and Yeom's datasets, sessions were also crossed between training and testing sets (*Sess-CV* + *Task-CV*), a practice that was not generally followed in the literature.

5.4. *Source of the task-independent neural signature*

Given its independence of the recorded condition, we may associate the EEG neural signature with unconscious processes working uninterruptedly in the background, similar to M.E. Raichle's et. al. (2001) concept of a '*default mode of brain function*'.[52–54] Such a concept hypothesises the existence of an intrinsic activity which "*instantiates the maintenance of information for interpreting, responding to and even predicting environmental demands*". Moreover, "*its functions are spontaneous and virtually continuous, being attenuated only when we engage in goal-directed actions*",[53] which coincides with the described dynamic nature of the signature across tasks. The task-independence property arises from the fact that fluctuations have a smaller magnitude than differences across subjects. Hence, we could interpret these fluctuations as task specific activity superimposing the default mode.

Table 7: Comparison of results obtained with Yeoms's database. For completeness, classification results are also shown, which can be deduced applying eq. 3 to the PRE results of Table 1. Refer to table 6 for a description of the columns and symbols ● and ◊.

Ref.	# Subj.	Task	CV	Succ.
Classification experiments				
This	10	All	●	53%
Verification experiments				
30	10	All	10 K-Folds	86%
This	10	All	◊	75%

Table 8: Comparison of results obtained with Zhangs's database. For completeness, classification results are also shown, which can be deduced applying eq. 3 to the PRE results of Table 1. Refer to table 6 for a description of the columns and symbols ● and ◊.

Ref.	# Subj.	Task	CV	Succ.
Classification experiments				
46, 47	20	All (ignored)	Leave-One-Out	100%
48-50	20	All (ignored)	2 K-Folds	99%
51*	20	All (ignored)	3 K-Folds	93%
This	20	All	●	76%
Verification experiments				
This	10	All	◊	93%

* 10 healthy and 10 alcoholic users.

The neural signature could also be due to the structure and disposition of the underlying neural networks, and nothing to do with their cognitive processes. Indeed, fMRI results of Ref.55 support this hypothesis. Given the nature of the electrical fields and their propagation through the skull, two networks with identical functionality but different organizations may produce distinct EEG signals. The EEG inverse problem describes how a recorded signal cannot be assigned to a unique disposition of sources within the skull.[56] Accordingly, the signature would be broadly defined by the layout of networks within the subject's brain, with cognitive processes playing a modulating role.

Each of these proposed explanations by themselves are unlikely to account for the neural signature. While we cannot deny the effect of network organization on the EEG, especially on the light that these networks can also be subject specific, evidence of continuous activation of some of these networks strongly suggest that underlying neural activity also plays an important part. Therefore, the solution is likely to fall somewhere in between the two proposals, with the distribution of neural networks and background brain activity working in conjunction to generate the signature.

Regarding the linkage between the task-independent EEG signature and genetics, although Vogel's work supports the genetic basis of such signature (encoded in the spectral shape of the EEG)[3], it is still unclear if, how and to what extent this is the case.

5.5. *The task-independent EEG biometric approach*

To date, the approach followed in the literature has been based on the analysis of EEG from isolated conditions. Even when the system was fed with signals from multiple tasks, they were usually labelled with the task itself, so that systems could differentiate amongst them and exploit task-specific information (Multi Task Learning[21]).

Ref.16 expands on the idea of an acquisition protocol, where users were asked to perform a particular task while their EEG was recorded for identification/verification. Specifically, the authors have focused their efforts mainly on REC and REO conditions:[11, 45, 57, 58] "*Within this paradigm subjects are typically seated in a comfortable chair with both arms resting, in a dimly lit or completely dark room. Generally, external sounds and noise are minimized to favour the relaxed state of the subjects. Participants are asked to perform a few minutes of resting state with eyes closed or eyes open, avoiding any focusing or concentration, but staying awake and alert.*"

Ref.59 goes a step further and proposed a system which assigns specific tasks to groups of subjects. By identifying the performed task during verification, they can effectively reduce the problem's complexity by a factor of N, where N is the number of considered tasks.

It is undeniable that performing these tasks during the verification of a user's identity is, in many real scenarios, cumbersome. If this modality is to be integrated within a system such as the biometric passport, performing "*a simple 'resting state' protocol*" is utterly impractical. In addition, this biometric security will almost certainly find an application within other BCIs, which were originally intended for different purposes other than identity verification. For example, a recent patent issued by Google embeds the identification of the user within a multi-sensor diagnostic system.[60]

To overcome these difficulties, we have proposed a new approach where subjects are not asked to perform any specific task. Instead, the system is tuned to extract the subject's task-independent neural signature. This will leave the user free to perform any other operations. For example, in an airport border control, the EEG activity could be collected while users present their passport, introduce any required information, and/or provide any supplementary biometry. In a general purpose BCI, the verification of the user's identity could take place in the background invisibly. As a result, the security procedure will not interfere with the user experience of the system. Moreover, verification checks could happen continuously or periodically, again, without interfering in the operation of the device.

6. Limitations

The biggest limitation of this work, and those of the state of the art, is that of the data. To obtain stronger evidence and draw bolder conclusions we need a complex database containing the EEG of more than 100 subjects, recorded with different equipment and under different conditions, over a span of years, while performing highly different cognitive tasks.

We have tried to alleviate this by using 6 different databases to support our conclusions. Unfortunately, only Keirn's and Yeom's had data from different days, but even then it was only 2 sessions from no more than 10 subjects. This hinders the interpretation of the *Sess-CV* experiments.

Equally, all the datasets were recorded with medical or research equipment. To properly assess the feasibility of a real world application of the proposed biometric system, we need to test its performance on data recorded by commercial EEG equipment. Although the literature already contains examples of this[42], complex databases such as the one described are needed to draw stronger conclusions.

7. Conclusions and implications

In this work, we have presented for the first time evidence supporting the existence of a neural signature within the EEG that is independent of the performed task or recording condition. Specifically, we observed this task-independent signature across motor tasks, relaxation and resting states (BCI2000 dataset), emotional states and REO (DEAP dataset), problem solving tasks and relaxation (Keirn's dataset), synonyms and non-synonyms AEPs (P. Ulssperger's dataset), self and non-self VEPs (Yeom's dataset) and target and non-target VEPs (Zhang's dataset).

We have also proposed to use this task-independent neural signature for biometric identity verification. We found that a feature vector composed of the first 20% of the RCeps or the order 8 LPC encodes most of the discrimination power within the spectral shape of the EEG in a linearly separable problem. Both systems yielded verification accuracies above 89% on 4 of the 6 databases used, and between 73% and 80% when training and testing sessions were recorded in different days.

We anticipate the finding of a task-independent signature will create a new set of experimental possibilities within brain research fields. Genetic and neurophysiological studies could use this neural signature to further the understanding of the EEG inheritance model, differentiating between task-independent and task-dependent activity. Through targeted experiments, it could also be used in the study of neuroplasticity,[61] neurodegenerative diseases[62] or the default mode of brain function. It will also be interesting to test whether existing neural models or models of EEG activity[63, 64] produce the described signature.

At the same time, EEG have the advantage of being a more affordable and easier to use tool than other neuroimaging techniques, especially after the recent proliferation of consumer EEG devices. As a result, while findings on other modalities tend to be confined in the first place within the neuroscience field, we expect our results on EEG to have direct implications and applications within other disciplines. We have outlined here the advantages of using the task-independent neural signature for biometric identity verification. At the same time, the complete understanding of the EEG genotype-phenotype map will ultimately allow the development of a major and inexpensive instrument for the improved understanding and diagnosis (early and new) of many diseases, especially those affecting the brain. Mainly because an instrument based on the quantitative measure of EEG properties will be closer to gene function than the traditional interpretation of cognitive tests.[7, 65]

Overall, this work represents a step forward in the understanding of the individual differences in brain activity, which will, in turn, help in the understanding of the commonalities, and in the design of practical biometric systems based on neural activity.

References

1. T. Collura, History and evolution of electroencephalographic instruments and techniques, Journal of Clinical Neurophysiology 10(4) (1993) 476–504.
2. H. Davis and P. Davis, Action potentials of the brain: In normal persons and in normal states of cerebral activity, Archives of Neurology And Psychiatry 36(6) (1936) 1214–1224.
3. F. Vogel, Genetics and the Electroencephalogram (Springer Verlag, 2000).
4. S. Eischen, J. Luckritz and J. Polich, Spectral analysis of EEG from families, Biological Psychology 41(1) (1995) 61–8.
5. M. Näpflin, M. Wildi and J. Sarnthein, Testretest reliability of resting EEG spectra validates a statistical signature of persons, Clinical Neurophysiology 118 (November 2007) 2519–2524.
6. C. van Beijsterveldt and G. van Baal, Twin and family studies of the human electroencephalogram: a review and a meta-analysis, Biological Psychology 61 (October 2002) 111–138.
7. B. Zietsch, J. Hansen, N. Hansell, G. Geffen, N. Martin and M. Wright, Common and specific genetic influences on EEG power bands delta, theta, alpha, and beta, Biological Psychology 75 (January 2007) 154–164.
8. H. Stassen, Computerized recognition of persons by {EEG} spectral patterns, Electroencephalography and Clinical Neurophysiology 49(12) (1980) 190–194.
9. M. Poulos, M. Rangoussi and E. Kafetzopoulos, Person identification via the EEG using computational geometry algorithms, Proceedings of the Ninth European Signal Processing, EUSIPCO'98 , 1998, pp. 2125–2128.
10. M. Fraschini, A. Hillebrand, M. Demuru, L. Didaci and G. Marcialis, An EEG-based biometric system using eigenvector centrality in resting state brain networks, Signal Processing Letters, IEEE 22(6) (2015) 666–670.
11. D. La Rocca, P. Campisi, B. Vegso, P. Cserti, G. Kozmann, F. Babiloni and F. De Vico Fallani, Human brain distinctiveness based on EEG spectral coherence connectivity, Biomedical Engineer- ing, IEEE Transactions on 61(9) (2014) 2406–2412.
12. Q. Gui, Z. Jin and W. Xu, Exploring EEG-based biometrics for user identification and authentication, Signal Processing

in Medicine and Biology Symposium (SPMB), 2014 IEEE , IEEE2014, pp. 1–6.
13. Q. Gui, Z. Jin, M. V. R. Blondet, S. Laszlo and W. Xu, Towards EEG biometrics: pattern matching approaches for user identification, IEEE International Conference on Identity, Security and Behaviour Analysis (ISBA), IEEE2015.
14. B. C. Armstrong, M. V. Ruiz-Blondet, N. Khalifian, K. J. Kurtz, Z. Jin and S. Laszlo, Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics, Neurocomputing (2015).
15. M. del Pozo-Baños, J. Alonso, J. Ticay-Rivas and C. Travieso, Electroencephalogram subject identification: A review, Expert Systems with Applications 41 (May 2014) 6537–6554.
16. P. Campisi and D. La Rocca, Brain waves for automatic biometric based user recognition (2014).
17. M. Abo-Zahhad, S. Ahmed and S. Abbas, State-of-the-art methods and future perspectives for personal recognition based on electroencephalogram signals, IET Biometrics 4(3) (2015) 179–190.
18. M. DelPozo-Baños, C. Travieso, C. Weidemann and J. Alonso, EEG biometric identification: a thorough exploration of the time-frequency domain, Journal of Neural Engineering 12 (Septem- ber 2015) p. 056019.
19. R. Palaniappan, Identifying individuality using mental task based brain computer interface, Third International Conference on Intelligent Sensing and Information Processing, ICISIP'05 , 2005, pp. 238–242.
20. R. Palaniappan, Electroencephalogram signals from imagined activities: a novel biometric identifier for a small population, Proceedings of the 7th international conference on Intelligent Data Engineering and Automated Learning, IDEAL'06 , 2006, pp. 604–611.
21. S. Sun, Multi-task learning for EEG-based biometrics, 19th International Conference on Pattern Recognition, ICPR'08 , 2008, pp. 1–4.
22. BCI2000, Public database www.bci2000.org, (2000).
23. A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng and H. Stanley, Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals, Circulation 101(23) (2000) e215–e220.
24. DEAP, Public database http://www.eecs.qmul.ac.uk/mmv/datasets/deap/, (2012).
25. S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt and I. Patras, Deap: A database for emotion analysis using physiological signals, IEEE Transactions on Affective Computing 3(1) (2012) 18–31.
26. Z. Keirn, Public database http://www.cs.colostate.edu/EEG/main/data/1989_Keirn_and_ Aunon, (1990).
27. Z. Keirn and J. Aunon, A new mode of communication between man and his surroundings, Biomedical Engineering 37(12) (1990) 1209–1214.
28. P. Ullsperger, Public database http://sccn.ucsd.edu/wiki/Chapter_02:_STUDY_Creation, (2014).
29. S. Yeom, H. Suk and S. Lee, EEG-based person authentication using face-specific self representation, Proceedings of the Korea Computer Congress 382011, pp. 379–382.
30. S. Yeom, H. Suk and S. Lee, Person authentication from neural activity of face-specific visual self-representation, Pattern Recognition 46(4) (2013) 1159–1169.
31. X. Zhang, Public database http://archive.ics.uci.edu/ml/datasets/EEG+Database, (1995).
32. X. Zhang, H. Begleiter, B. Porjesz, W. Wang and A. Litke, Event related potentials during object recognition tasks, Brain Research Bulletin 38(6) (1995) 531–8.
33. X. Zhang, H. Begleiter, B. Porjesz and A. Litke, Electrophysiological evidence of memory impairment in alcoholic patients, Biol Psychiatry 42(12) (1997) 1157–71.
34. J. Snodgrass and M. Vanderwart, A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity, Journal of experimental psychology Human learning and memory 6(2) (1980) 174–215.
35. A. Mognon, J. Jovicich, L. Bruzzone and M. Buiatti, ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features, Psychophysiology 48(2) (2010) 229–240.
36. R. Kohavi, Wrappers for performance enhancement and oblivious decision graphs, PhD thesis, Department of Computer Science, Stanford University (1995).
37. C. M. Judd, G. H. McClelland and C. S. Ryan, Data analysis: A model comparison approach (Routledge, 711 Third Avenue, New York, NY 10017 USA, 2011).
38. J. B. Alonso, La evaluación acústica del sistema fonador (Universidad de Las Palmas de Gran Canaria, Vicerrectorado de Calidad e Innovacion Educativa, 2008).
39. F. Kennet, Brain wave based authentication, Master's thesis, Gjøvik University College, Department of Computer Science and Media Technology (June 2008).
40. S. Marcel and J. Millan, Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation, IEEE Transactions on Pattern Analysis and Machine Intelligence 29(4) (2007) 743–752.
41. K. Brigham and B. Kumar, Subject identification from electroencephalogram (EEG) signals during imagined speech, Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), IEEE2010, pp. 1–8.
42. B. Hu, Q. Liu, Q. Zhao, Y. Qi and H. Peng, A real-time electroencephalogram (EEG) based individual identification interface for mobile security in ubiquitous environment, IEEE Asia-Pacific Services Computing Conference, APSCC , 2011, pp. 436–441.
43. M. Kostílek and J. Stastny, EEG biometric identification: Repeatability and influence of movement-related EEG,

International Conference on Applied Electronics (AE), 2012, pp. 147–150.
44. S. Yang and F. Deravi, On the effectiveness of EEG signals as a source of biometric information, Third International Conference on Emerging Security Technologies, EST'12 , 2012, pp. 49–52.
45. D. La Rocca, P. Campisi and G. Scarano, Stable EEG features for biometric recognition in resting state conditions, Biomedical Engineering Systems and Technologies, Communications in Computer and Information Science 452 (Springer, November 2014), pp. 313–330.
46. R. Palaniapan and K. Ravi, Improving visual evoked potential feature classification for person recognition using PCA and normalization, Pattern Recognition Letters 27(7) (2007) 726-733.
47. K. Ravi and R. Palaniappan, Leave-one-out authentication of persons using 40 Hz EEG oscillations, The International Conference on Computer as a Tool, EUROCON'05, 2005, pp. 1386-1389.
48. R. Palaniappan, Method of identifying individuals using VEP signals and neural network, IEEE Prceedings on Science, Measurement and Technology, 151, 2004, pp. 16-20.
49. R. Palaniappan and D. Mandic, Energy of brain potentials evojed during visual stinulus: a new biometric?, Proceedings of the 15th Intenational Conference on Artificial neural Networks: formal models and their applications, ICANN'05, 2, (Springer-Verlag, 2005), pp. 735-740.
50. R. Palaniappan and D. Mandic, EEG based biometric framework for automatic identity verification, The Journal of VLSI Signal Processing 49(2) (2007) 243-250.
51. P. Nguyen, D. Tran, X. Huang and D. Sharma, A Proposed Feature Extraction Methods for EEG-based Person Identification, Proceeding on the International Conference on Artificial Intelligence (ICAI). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2012.
52. M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard and G. L. Shulman, A default mode of brain function, Proceedings of the National Academy of Sciences 98(2) (2001) 676–682.
53. D. A. Gusnard and M. E. Raichle, Searching for a baseline: functional imaging and the resting human brain, Nature Reviews Neuroscience 2(10) (2001) 685–694.
54. M. E. Raichle and A. Z. Snyder, A default mode of brain function: a brief history of an evolving idea, Neuroimage 37(4) (2007) 1083–1090.
55. E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris and R. T. Constable, Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity, Nature Neuroscience (2015).
56. R. D. Pascual-Marqui, Review of methods for solving the EEG inverse problem, International journal of bioelectromagnetism 1(1) (1999) 75–86.
57. P. Campisi, G. Scarano, F. Babiloni, F. DeVico Fallani, S. Colonnese, E. Maiorana and L. Forastiere, Brain waves based user recognition using the "eyes closed resting conditions" protocol, IEEE International Workshop on Information Forensics and Security, WIFS'11 , IEEE2011, pp. 1–6.
58. D. La Rocca, P. Campisi and G. Scarano, EEG biometrics for individual recognition in resting state with closed eyes, Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG), 2012, pp. 1–12.
59. T. Pham, W. Ma, D. Tran, P. Nguyen and D. Phung, Multi-factor EEG-based user authentication, Neural Networks (IJCNN), 2014 International Joint Conference on , IEEE2014, pp. 4029–4034.
60. H. Kayyali, Medical device and method with improved biometric verification (2014), US Patent 8,679,012.
61. L. Carrillo-Reid, V.G. Lopez-Huerta, M. Garcia-Munoz, S. Theiss and G.W. Arbuthnott, Cell assembly signatures defined by short-term synaptic plasticity in cortical networks, International Journal of Neural Systems, 25(7) (2015) 1550026.
62. F. C. Morabito, M. Campolo, D. Labate, G. Morabito, L. Bonanno, A. Bramanti, S. de Salvo, A. Marra and P. Bramanti, A longitudinal EEG study of Alzheimer's disease progression based on a complex network approach, International Journal of Neural Systems, 25(2) (2015) 1550005.
63. M. Koppert, S. Kalitzin, D. Velis, F. Lopes Da Silva and M.A. Viergever, Dynamics of collective multi-stability in models of multi-unit neuronal systems, International Journal of Neural Systems 24(2) (2014) 1430004.
64. Kalitzin, S. and Koppert, M. and Petkov, G. and Da Silva, F.L., Multiple oscillatory states in models of collective neuronal dynamics, International Journal of Neural Systems, 24(6) (2014) 1450020.
65. H. Begleiter and B. Porjesz, Genetics of human brain oscillations, International Journal of Psychophysiology 60 (May 2006) 162–171.
66. R. Palaniappan, Two-stage biometric authentication method using thought activity brain waves, International Journal of Neural Systems 18(1) (2008) 59–66.
67. R. Paranjape, J. Mahovsky, L. Benedicenti and Z. Koles, The electroencephalogram as a biometric, Proceedings on Canadian Conference on Electrical and Computer Engineering, 22001, pp. 1363–1366.
68. R. Scheeringa, P. Fries, K.-M. Petersson, R. Oostenveld, I. Grothe, D. G. Norris, P. Hagoort and M. C. Bastiaansen, Neuronal dynamics underlying high- and low-frequency EEG oscillations contribute independently to the human BOLD signal, Neuron 69(3) (2011) 572 – 583.

Appendix A: Other evaluated systems

A number of different architectures were evaluated in addition to the ones described in the main text. However, as they all yielded worse or equal performances with a more complex or computationally expensive design, their results are not reported in detail here. The designs are based in previous results reported in the literature as well as the authors' experience. In particular, the following were tested:

Time statistics of feature vectors: We computed several statistical measurements across time (across windows of the STFT) and used the results as inputs to the LDC. Specifically, we computed mean, std, kurtosis, skewness and percentiles 5, 25, 50 (median), 75 and 95 for each feature. Mean and median performed at a level similar to or worse than the original system, while the remaining measurements performed relatively poor. ROC curves showed an overall poorer behaviour away from the EER. Having said that, this method has the advantage of reducing the volume of data processed by the classifier, as all windows are merged (i.e. through average, median, std, etc) into a single vector. Thus, it may be considered in cases where the volume of data or the processing speed is a concern.

Feature statistics of feature vectors: We computed the statistics described in the previous point, but across features for each SFTF window instead of across time. Hence, the number of windows remained the same and the length of the feature vector was reduced to one (i.e. the statistical value). In this case, results were substantially worse than that of the original systems.

Power of bands: We divided the PSD into bands and computed the power within each band. Results were significantly worse than that of the baseline system; i.e. based on the full PSD vector.

Different forms of the LPC: We considered the use of other representation forms of the LPC. Specifically: Reflection Coefficients (RC) and Line Spectral Pairs (LSP), which have been identified as alternatives robust against noise. In addition, we also tested the performance of the model's fitting error (ϵ). Comparing the accuracy of 8-order LPC, RC, LSP and ϵ, only ϵ_8 performed substantially worse than the rest. LPC_8, RC_8 and LSP_8 resulted in virtually equivalent performances. Inspecting the ROC curves at lower FAR values, LPC8 produced, on average, best GAR.

Feature fusion: We fused $RCeps_{20\%}$ and LPC_8 in a single vector and fed it to the LDC. This fusion performed similarly to each feature individually, which evidenced the high level of correlation between the information extracted by both methods.

Fusion of statistical measurements: We combined the statistical measurements taken from $RCeps_{20\%}$ or LPC_8 in a single feature vector. The fusion of time-statistics performed similar to or worse than the individual mean and median vectors. On the other hand, the concatenation of the feature-statistics, i.e. taken within each vector instead of across time, produced a remarkable increase in accuracy compared to individual statistics. In some cases there were more than 10 percentage points of improvement. Therefore, we considered the fusion of feature-statistics to the original $RCeps_{20\%}$ or LPC_8 vectors. Based on Ref.66 results, we were expecting this to improve the system's accuracy. Nevertheless, once more, results were similar to or worse than those of the original systems. The enhancement observed by such work may therefore be due to a suboptimal extraction of the LPC coefficients (the authors used EEG segments of 0.5 seconds).

Score fusion of $RCeps_{20\%}$ and LPC_8 individual systems: This performed equal to, or worse than, the feature fusion version.

Multi-window length: We considered the possibility that different spectral widths could extract uncorrelated discriminant information. To test this, we created a system composed of multiple sub-systems with different window lengths fused at score level. Results were equal to or worse than those obtained with the original systems.

Projection methods: We applied Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA) to all the described architectures. This was not a bid to reduce the dimensionality of data, but rather an attempt to present the discriminant information to the classifier in a more suitable way, in the hope of boosting the performance and/or improve the stability of the system. We applied such projection techniques to the features of each sensor individually (as sensor experts) and to the vector containing the features from all sensors. In both cases, results were similar or worse than those of the original systems.

Score fusion of sensor experts: This architecture contained as many sub-systems as EEG sensors. The LDC scores obtained for all sensors were then averaged to build the final response. We evaluated this with all

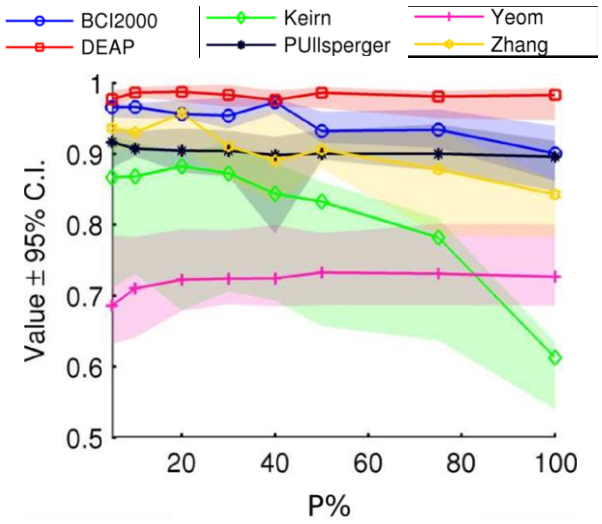

Figure B.1: Performance of $RCeps_{P\%}$ verification experiments with increasing number of coefficients. Mean accuracy values and 95% CI (shaded area) obtained with the first *P*% of the cepstral coefficients.

the described systems. In all cases, results were similar to those of the original systems.

Non-lineal classifiers: In the classification phase, we evaluated the performance of Support Vector Machines with Radial Basis Function kernel (RBF-SVM) and Artificial Neural Networks (ANN). The kernel and cost parameters of the former were optimized applying a CV procedure within the training data. To avoid any over fitting, this inner CV followed the same principle as the one used for testing, meaning different subjects were used as impostors during training and validation. Similarly, multiple configurations of the hidden layers were evaluated for the ANN. In both cases, we only managed to equal the results of LDC. Interestingly, the optimization of the RBF-SVM parameters showed a clear tendency to create a linear model, rather than a non-linear one.

Appendix B: Optimization of $RCeps_{P\%}$ and LPC_N

During the optimization of *P* and *N* from $RCeps_{P\%}$ and LPC_N, we used the same experiments as those described in section 3.2.3 with the exception of the number of MC iterations, which was truncated to 20 for BCI2000 and Zhang's datasets to reduce the computational time.

During optimization of the $RCeps_{P\%}$ system, maximum performance was achieved at different values of *P* across databases, with the majority of the configurations giving performances close to the maximum (Fig. B.1). On BCI2000, Keirn's and Zhang's databases, using too many cepstral coefficients translated on a loss of performance. This was especially

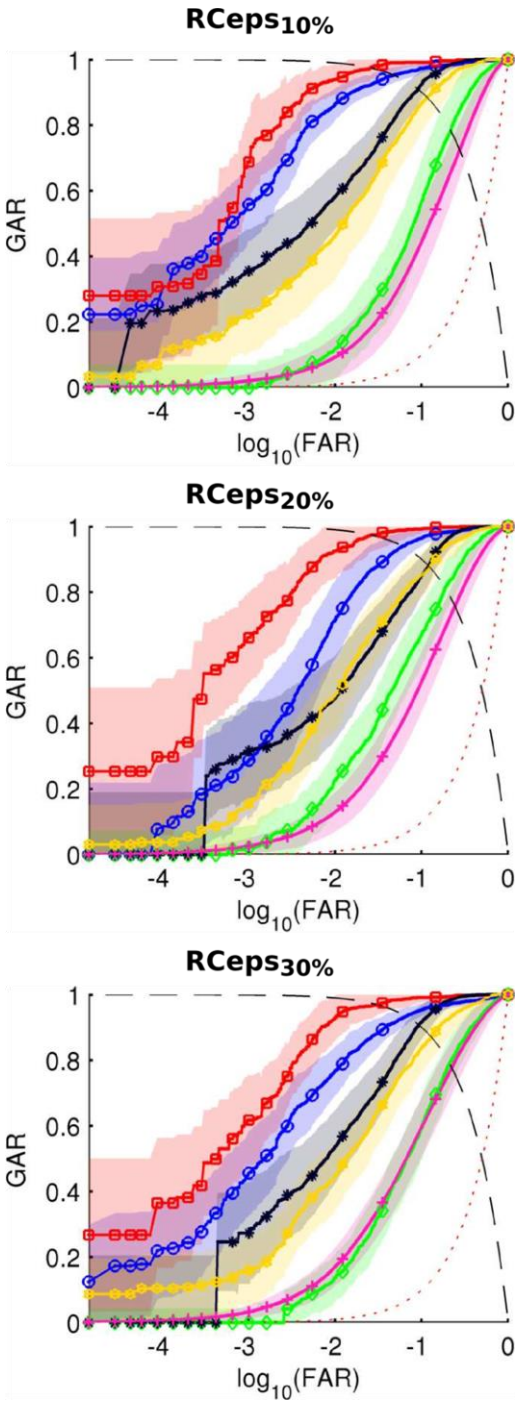

Figure B.2: ROC curves of $RCeps_{P\%}$ verification experiments with logarithmic FAR axis. Mean ROC curves and std (shaded area) obtained with different *P*. Refer to Fig. B.1 for legend details.

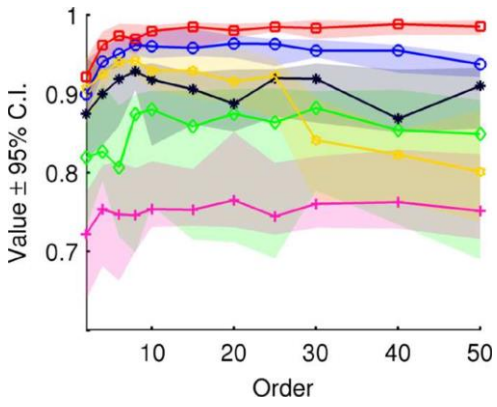

Figure B.3: Performance of LPC_N verification experiments with increased order. Mean and 95% CI (shaded area) verification accuracy obtained with different values of N. Refer to Fig. B.1 for legend details.

acute on Keirn's and Zhang's datasets, whose accuracy dropped ~20 percentage points. Looking at the ROC curves (Fig. B.2), we noticed a steady improvement of the GAR for lower FARs peaking at P = 20%. After that point, the GARs oscillated, regaining the maximum values in some cases but not in others.

Results for the LPC design showed a similar behaviour, with high variation of the optimal point and great number of configurations performing similar to the optimal (Fig. B.3). Accuracy within 1.5 percentage points to the maximum is reached at order 8. Orders above this had no effect on the performance of the system, except on Zhang's database, whose accuracy decreased abruptly passed order 25. The tendency was less clear when looking at GAR for lower FARs, with large variation across databases (Fig. B.4). Having said that, a steady increase in GAR was observed up to order 8 for all cases, followed by oscillating performances. In occasions, the maximum GARs were obtained with higher orders; e.g. LPC_{20} for BCI2000 database and LPC_{40} for DEAP dataset.

The results obtained with LPC contradict some of the conclusions in the literature. Ref. 67 concludes that an increase in the order of the LPC model is necessary to bear with the rise in the number of users. Even with only 5 subjects, they reported an increase in classification accuracy of 7 percentage points when moving from order 9 to 15. Furthermore, Refs. 45 and 57 found RC to outperform LPC, whereas we did not observed any improvement by using the more computationally expensive RC (Appendix A). These discrepancies may be due to differences in the

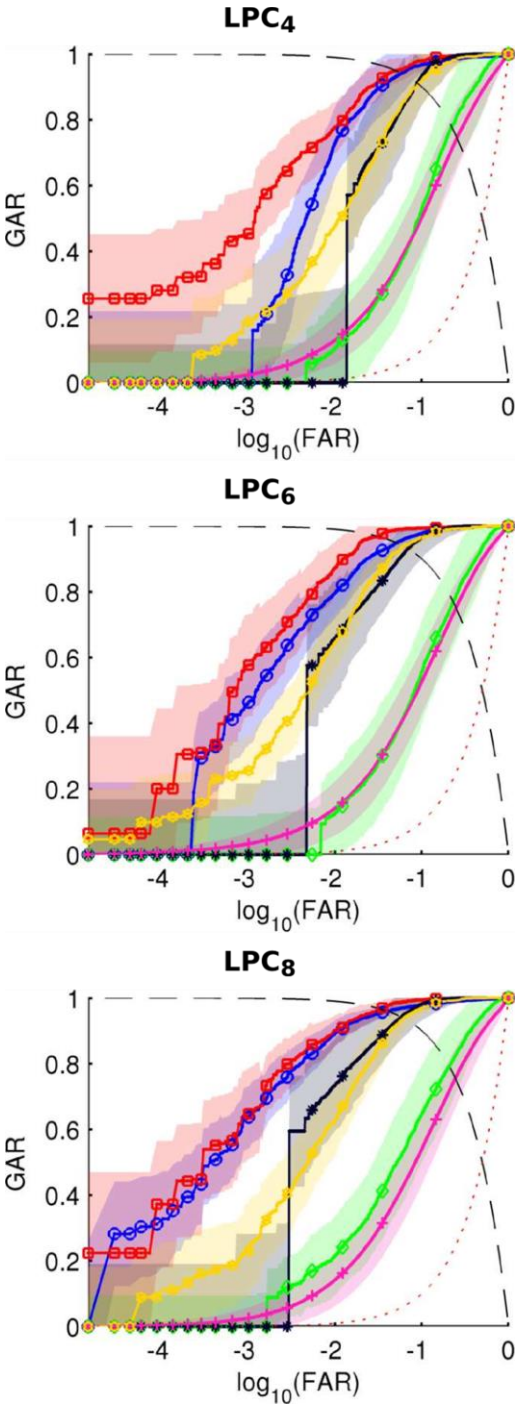

Figure B.4: ROC curves of LPC_N verification experiments with logarithmic FAR axis. Mean ROC curves and std (shaded area) obtained with different N. Refer to Fig. B.1 for legend details.

experimentation methodology: classification versus verification experiments, differences in the systems' architectures, and/or idiosyncrasies of the databases.

Table C.1: Independent t-test comparison of *Task-CV* and *Bal-CV* experiments. The test corresponds to the results of Figure 2. Null and alternative hypothesis were H_0: $\mu_{Bal-CV} - \mu_{Task-CV} = 5$ and H_1: $\mu_{Bal-CV} - \mu_{Task-CV} < 5$. P-values were adjusted with the BHFDR method. Yeom's and Zhang's databases are special cases of each system and are therefore highlighted in grey.

Dat.	t	df	p-value	SE	r
(a) Preprocessed datasets					
B	-3.06	38	< 0.01	0.62	-0.44
D	-50.19	38	< 0.001	0.08	-0.99
K	-21.96	38	< 0.001	0.19	-0.96
P	-124.02	38	< 0.001	0.04	-1.00
Y	-93.57	38	< 0.001	0.04	-1.00
Z	-0.47	38	0.37	0.87	-0.07
(b) ADJUST preprocessed datasets					
B	-3.65	38	< 0.001	0.76	-0.50
D	-84.59	38	< 0.001	0.04	-1.00
P	-109.08	38	< 0.001	0.05	-1.00
Z	-0.81	38	0.25	0.56	-0.13
(c) rNorm systems with preprocessed dataset					
B	-3.97	38	< 0.001	0.48	-0.53
D	-62.60	38	< 0.001	0.05	-0.99
K	-23.45	38	< 0.001	0.14	-0.97
P	-90.54	38	< 0.001	0.06	-1.00
Y	-112.36	38	< 0.001	0.03	-1.00
Z	0.30	38	0.68	0.81	0.05

Table C.2: Independent t-test comparison of Task-CV and Single-Task experiments. The test corresponds to the results of Figure 2. Null and alternative hypothesis were H_0: $\mu_{Single-Task} - \mu_{Task-CV} = 5$ and H_1: $\mu_{Single-Task} - \mu_{Task-CV} < 5$. P-values were adjusted with the BHFDR method. Yeom's and Zhang's databases are special cases of each system and are therefore highlighted in grey (Methods).

Dat.	t	df	p-value	SE	r
(a) Preprocessed datasets					
B	-4.79	38	< 0.001	0.56	-0.60
D	-48.51	38	< 0.001	0.10	-0.99
K	58.91	38	1.00	0.10	0.99
P	-90.36	38	< 0.001	0.05	-1.00
Y	-42.27	38	< 0.001	0.04	-0.99
Z	-0.74	38	0.28	0.88	-0.12
(b) ADJUST preprocessed datasets					
B	-7.10	38	< 0.001	0.75	-0.75
D	-74.86	38	< 0.001	0.06	-1.00
P	-97.82	38	< 0.001	0.05	-1.00
Z	-2.21	38	< 0.05	0.60	-0.33
(c) rNorm systems with preprocessed dataset					
B	0.66	38	0.85	0.47	0.10
D	-51.56	38	< 0.001	0.06	-0.99
K	-14.90	38	< 0.001	0.13	-0.92
P	-98.61	38	< 0.001	0.05	-1.00
Y	-95.22	38	< 0.001	0.04	-1.00
Z	-1.79	38	< 0.05	0.78	-0.27

Overall, for both designs, RCeps and LPC, results were not fully homogeneous across databases and systems, hindering the selection of the optimal configuration point. Based on the ROC curves, we chose to retain 20% of the cepstrums ($RCeps_{20\%}$) and use LPC of order 8 (LPC_8).

Appendix C: Statistical tests

To avoid cluttering the main text, we have only presented the main results. Here, we provide supplementary statistical tests. Note that the p-values presented here have been adjusted using the Benjamini-Hochberg False Discovery Rate (BHFDR) method.

Appendix D: Task-independent subject-specific information within the fMRI.

Interestingly, a similar experimentation paradigm showed that connectivity profiles recorded by Functional Magnetic Resonance Imaging (fMRI) are also subject specific and relatively homogeneous across tasks.[55] It is important to note that this is fundamentally different research. fMRI connectivity profiles measure the distribution of networks within the brain through blood-oxygen-level fluctuations, whereas EEG measures neuronal activity in and of itself. In other words, fMRI registers location and brain activity levels indirectly through metabolic changes while EEG measures the electric fields directly generated by specific, synchronous firing patterns of millions of

Table C.3: Independent t-test comparison of Sess-CV + Task-CV and Sess-CV + Bal-CV (Single-Task) experiments. The test corresponds to the results of Figure 1. Null and alternative hypothesis are H_0: $\mu_A - \mu_B = C$ and H_1: $\mu_A - \mu_B > C$, with A the Sess-CV + Bal-CV (Single-Task) condition, B the Sess-CV + Task-CV condition, and C a set threshold. P-values were adjusted with the BHFDR method. Yeom's database is a special case of each system and is therefore highlighted in grey (Methods).

System.	**t**	**df**	**p-value**	**SE**	**r**
(a) Sess-CV + TaskCV vs. Sess-CV + Bal-CV					
Keirn's dataset results; C = 10					
-	-3.65	38	1.00	1.58	-0.50
rNorm	-0.84	38	0.91	1.66	-0.13
Yeom's dataset results; C = 1					
-	0.51	38	0.45	0.49	0.08
rNorm	-0.82	38	0.91	0.19	-0.13
(b) Sess-CV + Task-CV vs. Sess-CV + Single-Task					
Keirn's dataset results; C = 10					
-	-3.47	38	1.00	1.60	-0.48
rNorm	-0.27	38	0.81	1.65	-0.04
Yeom's dataset results; C = 1					
-	-0.21	38	0.81	0.45	-0.03
rNorm	-3.93	38	1.00	0.20	-0.53

neurons. EEG is therefore closer to the recording of our 'thoughts', than it is to the spatial distribution of brain systems. Despite described correlations between them,[68] differences in the nature and source of these two signals inevitably led to the following dissimilarities in the properties and possibilities of the described signatures:

First, while Finn's et. al. fMRI signature relied on signals recorded during "several minutes", the EEG neural signature assessed here has been processed in segments of just 2 seconds (even 0.5 seconds for Yeom's and Zhang's datasets). Far from being arbitrary, in a previous study we have identified segments between 1 and 2 seconds to be optimal for the extraction of subject traits, and that 4 seconds of EEG is enough to obtain 90% of the potential discrimination power.[18]

Second, where Finn's et. al. found the frontoparietal networks to be the most discriminative ones, we found no consistent most discriminative region on the mentioned previous study. Best performing areas varied not only across databases and tasks, but also across system configurations within the same database and task. While EEG is notorious for its low spatial resolution, the observed high variability and sensitivity hints that the mentioned discrepancy probably roots not only on resolution differences.

Finally, Finn's et. al. experienced slumps in performance as large as 40 percentage points of PRE when moving from Sess-CV + Single-Task to a Sess-CV + Task-CV experiments, compared to the less than 10 percentage points of drop reported here. This difference can be explained by the paradigm itself. Because Finn's et. al. trained their system with a single task, the system lacks information about the natural variability of the signature across cognitive tasks/conditions, and is therefore unable to build a reliable model of the subject's signature.

Ultimately, these two studies are different enough to account for the described differences, and the fact that both modalities (fMRI and EEG) yielded analogous results in terms of task-independence emphasizes the scale of inter-individual differences in brain anatomy and activity.