Expert Systems with Applications 41 (2014) 6537–6554

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Review

Electroencephalogram subject identification: A review

Marcos Del Pozo-Banos, Jesús B. Alonso, Jaime R. Ticay-Rivas, Carlos M. Travieso *

University of Las Palmas de Gran Canaria, Institute for Technological Development and Innovation in Communication (IDeTIC), 35017 Las Palmas de Gran Canaria, Spain

ARTICLE INFO

Article history:
Available online 22 May 2014

Keywords:
Electroencephalogram
Biometrics
Genetics
Identification

ABSTRACT

This is, to the best of the authors knowledge, the first complete research on the state of the art on EEG based subject identification. As well as covering the full story of this field (from 1980 to 2013), an overview of the findings made in genetic and neurophysiology areas, from which it is based, is also provided. After a comprehensive search, 109 biometric publications were found and studied, from which 88 were finally included in this document. A categorization of papers is proposed based on the recording paradigm. The most used databases, some of them public, have been identified and named to allow the comparison of results from these and future works. The findings of this work show that, although basic questions remain to be answered, the EEG, and specially its power spectrum in the range of the alpha rhythm, contains subject specific information that can be used for classification. Moreover, approaches such as a multi-day-session training, the fusion of information from different electrodes and bands, and Support Vector Machines are recommended to maximize the system's performance. All in all, the problem of subject identification by means of their EEG is harder than initially expected, as it relies on information extracted from complex heterogeneous EEG traits which are the results of elaborated models of inheritance, which in turn makes the problem very sensitive to its variables (time, frequency, space, recording paradigm and algorithms).

1. Introduction

The genetic traits on the human electroencephalogram (EEG) have received great attention from the scientific community almost since the very beginning of the human EEG recordings by Hans Verger in 1924 (Collura, 1993). This genotype–phenotype map will make a major and inexpensive tool for understand, diagnose and early diagnose many diseases, specially those affecting the brain (Begleiter & Porjesz, 2006; Zietsch et al., 2007). Mainly because a tool based on the quantitative measure of EEG properties will be closer to gene function than the traditional interpretation of cognitive tests (Begleiter & Porjesz, 2006).

Biometric systems based on the EEG, as a non-invasive and relatively inexpensive window into the human brain, have received special attention within the scientific community. Most of these efforts have focussed on the development of diagnosis and monitoring tools for conditions such as sleep apnea, schizophrenia or epilepsy (Sabeti, Katebi, Boostani, & Price, 2011; Song & Zhang, 2013; Tagluk & Sezgin, 2011) and on the creation of Brain Machine Interfaces (BMIs) to assist disabled people (Blasco, Iez, Beda, & Azorn, 2012). Applications in other, perhaps more exotic fields, such as marketing, has also been explored (Khushaba et al., 2013).

EEG based subject identification is a relatively new biometric modality which finds its origins in the advances of such human genetics and clinical neurophysiology studies. Its relevance relies mainly in the prospects of high quality and robustness. Passwords will be harder to steal, as users do not need to perform any revealing action. Even if stolen, the system can be tuned to respond not to the passwords semantic meanings but to the subjects specific EEG patterns, which are extremely hard to reproduce, if at all possible. Furthermore, if a user is forced to enter their password, their high stress level could be detected by the system, forbidding the access. This property is referred to as "*circumvention*" within the biometry field.

On the other hand, EEG based biometric systems face their most obvious drawback in the inconvenience of the recording method. Although the sensor technology has given giant steps forward in EEG machines, users still need to have contact with them, and the preparation time is longer than other modalities and require of qualified staff. Moreover, the vast majority of these devices still rely on conductive gel to decrease the impedance between the scalp and the electrode, and obtain quality signals. All in all, EEG based biometric is a modality that promises to deliver real high security tools in the future.

* Corresponding author. Address: University of Las Palmas de Gran Canaria, Institute for Technological Development and Innovation in Communication (IDeTIC), 35017 Las Palmas de Gran Canaria, Spain. Tel.: +34 928452864; fax: +34 928451243.

E-mail address: carlos.travieso@ulpgc.es (C.M. Travieso).

http://dx.doi.org/10.1016/j.eswa.2014.05.013
0957-4174/

This work is a comprehensive research on the state of the art on EEG based biometric systems from its origins until the end of 2013. In particular, it aimed to answer the following questions:

- What are the subject specific traits of the EEG?
- Where are these traits in terms of space and frequency?
- Are they constant across recording paradigms and time?
- Which are the best techniques to extract and evaluate such traits?

Although there exist overviews on the field, they focussed on some of the algorithms applied over a particular approach for EEG based biometric systems (Khalifa, Salem, Roushdy, & Revett, 2012; Revett, 2010, 2012; Singh, Singh, & Ray, 2012). On contrary, this is a broad study considering all publications on the matter, and therefore fully covering all techniques and strategies as well as their relationships. In addition, it provides an overview of genetic and neurophysiology findings, so that a link between them and biometric studies can be draw. Hence, this is, to the best of the authors knowledge, the first complete research on the state of the art on EEG based biometric systems.

The remaining of this document is organized as follows. First, a detailed description of the research methodology is given in Section 2. A brief presentation of the major findings on the genetic traits of the human EEG is then given in Section 3. Section 4 enumerates the main EEG databases used in the biometric field. Section 5 provides detailed descriptions of the most important results in each of the studied approaches. An extensive discussion is provided in Section 6, where findings are put together in a global picture. Finally, the extracted conclusions are presented in Section 7.

2. Research methodology

The present study was executed in two phases. First, a search in the genetic and neurophysiology fields regarding the phenotypic features of the EEG was made. This allowed the establishment of a scientific base in the matter and provided a non-technical point of view of the problem. A search with the criterion "genetic", "EEG" and "subject specific" was carried. Here, a study of the level of the biometric field was not intended, and therefore only the most relevant publications and reviews were revised. A total of 17 of these works were included here. Of special relevance is the book "Genetics and the Electroencephalogram" by Vogel, a key figure in the genetics field, which is a comprehensive compilation of findings on the genetic, clinical and neurophysiological aspects of the EEG (Vogel, 2000).

The second phase comprised the search in the biometric field. In this case, the keywords "EEG", "brainwaves", "biometric", "subject", "identification" and "verification" combined in numerous ways were used for the search. References of found articles were also scrutinized. After a comprehensive exploration, 108 works were finally found between 1998 and 2013. This emphasizes the novelty of the method, and explains the relatively small amount of publications when compared to older research lines. Therefore, instead of filtering the outcome to keep only journals; as it is usually the case in reviews, all 108 works were considered for this study, from which 87 were finally included. Fig. 1 shows the amount of publications per year.

The following classification of the literature based on the recording paradigm is proposed:

- **REC and REO:** A big part of the articles relies on EEG data recorded while subjects were resting with eyes closed (REC) or resting with eyes open (REO).
- **ERP:** Event related potentials (ERP) have also been used to identify users. Even though until the date only visual evoked potentials (VEP) have been used for this purpose, a global category name is proposed so that it will be able to accommodate possible future works on other ERPs.
- **Multi-task:** Some works used EEG recorded under different mind tasks such as mathematical operations, writing letters and imagined movements. Usually, this works study the differences in performance obtained by different recording paradigms.
- **Indirect:** Other researchers have tried to identify users by recognizing a thought password rather than subject-specific EEG traits.
- **Others:** This category includes reviews, dissertations, reports and any other published work that is related to the subject but does not propose any system architecture or experiment.

A further differentiation is proposed based on the hardware used to record the database. This consideration was taken as both **medical** and **consumer** equipments have been used. The later represent a cheaper alternative that does not require conductive gel and is considerably easier to use. However, all these come in detriment of the quality of the signal, providing lower signal to noise ratios and sensitivity. Accordingly, it seems fair to keep in mind which hardware has been used in each case when comparing results. Conveniently, all these categories can also be used to classify genetic and neurophysiology works in the matter.

Figs. 2 and 3 show the distribution of publications across each category. Note that these figures do not represent percentage values, as some works can fall in several categories. It can be seen that the great majority of studies have focussed on REC and VEP (ERP) modalities. This is consistent with genetic and neurophysiology studies. Multi-task studies have also received special attention, as authors tried to find the best suiting paradigm for their systems. In addition, as commercial EEG hardware have just recently appeared, the proportion of studies using them is obviously lower than those using medical equipment.

To maximize the understanding of the progress made in the field, publications where tagged and clustered in teams. This also helped to identify the databases used in each publication, which was specially important as most of these databases did not have a name to refer them and tended to be difficult to track. Once they where named, it was possible to compare results reported by different works.

3. EEG genetics

The uniqueness of individuals can be greatly attributable to their genetics. Thus, if a system tries to identify a subject, it is actually trying to identify their phenotypes as well as characteristic effects of exogenous factors. Moreover, if such machine is based on information regarding an organ as complex and unknown as the human brain, findings on the genotype–phenotype map of the EEG become vital.

Identify the genetic traits of the EEG has proven to be an arduous task. These are complex heterogeneous traits, as they are the result of elaborated models of inheritance. For example, some evidences suggest that some genes have different effects at different brain areas and EEG frequencies. In addition, exogenous factors have also been proven to influence the human EEG and have to be considered when evaluating the results (Zietsch et al., 2007).

Twin studies have proven of great help on the understanding of this genotype–phenotype relationship. Davis and Davis were the first to study the EEG on twins (Davis & Davis, 1936). Evaluating a number of EEG traits; mainly based on distinguishing marks of