Expert Systems With Applications 64 (2016) 261–268

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Detection of attention in multi-talker scenarios: A fuzzy approach

Jesus Minguillon[a,b,*], M. Angel Lopez-Gordo[c,d], Francisco Pelayo[a]

[a] Department of Computer Architecture and Technology, University of Granada, C/ Periodista Daniel Saucedo Aranda, s/n, 18071 Granada, Spain
[b] Research Centre for Information and Communications Technologies (CITIC), University of Granada, C/ Periodista Rafael Gomez Montero, 2, 18014 Granada, Spain
[c] Department of Signal Theory, Telematics and Communications, University of Granada, C/ Periodista Daniel Saucedo Aranda, s/n, 18071 Granada, Spain
[d] Nicolo Association, Churriana de la Vega, Spain

ARTICLE INFO

Article history:
Received 29 January 2016
Revised 29 July 2016
Accepted 30 July 2016
Available online 30 July 2016

Keywords:
Brain–computer interface
EEG, attention detection model
Multi-talker attention detection
Fuzzy

ABSTRACT

The automatic and online detection of auditory attention in multi-talker scenarios (e.g., cocktail party paradigm) is a current topic in electroencephalography (EEG)-based brain-computer interfaces (BCIs). Recent works have demonstrated a way to make it possible by means of a model based on an m-ary phase shift keying (m-PSK) detector. However, this attention detection model lacks of relevant information such as the non-stationary nature of EEG signals, the neuro-plasticity/habituation effects or the nonlinearities of the attention. In this paper we propose an enriched version of the attention detection model constituted by an automatic adaptive m-PSK detector implemented on fuzzy logic. In it, the relevant information mentioned before is modeled as two inputs that feed the fuzzy-based attention detection model. The output provides the detection. Our enriched model outperformed the results of previous works in terms of mean information transfer rate (ITR) (4-PSK: 5.41 bpm; 6-PSK: 6.03 bpm) and accuracy (4-PSK: 0.54; 6-PSK: 0.39) after only 4.63 (4-PSK) and 2.93 (6-PSK) seconds of processing. The proposed model for the automatic detection of auditory attention can have relevant impact on several areas such as education, public transport, jobs, industry, attention disorders, ubiquitous systems, sports and art.

1. Introduction

During the last years, brain-computer interface technology (BCI) has been used in multitude of applications, for instance in visual spellers (Birbaumer et al., 1999; Ron-Angevin, Varona-Moya, da Silva-Sauer, & Carrion-Robles, 2014) for wheelchair control (Li et al., 2013), for simple binary volition detectors (e.g., *yes*/*no*) (Hill et al., 2014), for classification and detection of covert visual attention (Lopez-Gordo, Pelayo, & Prieto, 2010) and even in the auditory modality (Höhne & Tangermann, 2014; Halder et al., 2010). In this context, new uses and applications for BCIs such as detection of selective attention to auditory sources have emerged. This cognitive ability enables one to attend a target source and ignore the others by means of concomitant cognitive processes (Ikeda et al., 2010) such as in the case of a cocktail party phenomenon (Cherry, 1953). There are examples of it in BCI literature. For instance, in these studies (Kubanek, Brunner, Gunduz, Poeppel, & Schalk, 2013; Martin et al., 2014) two auditory sources were presented simultaneously to the participants. The authors evidenced that the reconstruction of the envelope of the attended speech can be obtained by direct analysis of that of the gamma band of electrocorticography (ECoG) signals.

A different approach for the detection of attention in dichotic listening tasks is based on digital modulation of electroencephalography (EEG) signals. In this model, speeches are barely perturbed to evoke a constellation of binary phase shift keying (BPSK) signals (Lopez-Gordo, Fernandez, Romero, Pelayo, & Prieto, 2012; Lopez-Gordo, Pelayo, Prieto, & Fernandez, 2012; Lopez-Gordo & Pelayo, 2013). The two counter-phased symbols of the BPSK constellation correspond to the conditions *attend* and *ignore* and the attentional cognitive effort can be robustly detected by means of a BPSK receiver. In these experiments, they obtained an accuracy of 88% with binary detection and an information transfer rate (ITR) of 2 bits per minute (bpm) approximately. Other studies tried to develop a more complexed model for multi-talker scenarios. In (Lopez-Gordo, Pelayo, Fernandez, & Padilla, 2015), authors digitally modulated sentences taken from a speech corpus used for the assessment of speech intelligibility in military communication. Particularly, from the Coordinate Response Measure speech corpus (Bolia, Nelson, Ericson, & Simpson, 2000). They used 4-PSK and 6-PSK modulations to detect the attended speech among 4 and 6 concurrent speeches. Although this experiment probed that detec-

* Corresponding author.
E-mail addresses: minguillon@ugr.es (J. Minguillon), malg@ugr.es (M.A. Lopez-Gordo), fpelayo@ugr.es (F. Pelayo).

http://dx.doi.org/10.1016/j.eswa.2016.07.042
0957-4174/

tion of attention in multi-talker scenarios was possible by means of m-ary phase shift keying (m-PSK) modulation scheme, the results were poor (an accuracy of 0.47% with four-symbols detection and an ITR of 1.25 bits/m approximately). As far as we know, the m-PSK is the only BCI approach capable to detect attention in multi-talker scenarios (up to six concurrent speeches).

In the studies mentioned in the previous paragraph, the attention is detected by a pure signal processing approach. However, EEG signals and attentional paradigms do not meet general assumptions that are typically considered under a signal processing approach. For instance (i) EEG signals are not stationary; (ii) the brain structure that generates them cannot be considered as a linear time-invariant system due to the neuro-plasticity, habituation and other cognitive and physiological factors. Improved models and novel expert systems must be investigated in order to cope with these problems. A way to improve the performance of the automatic detection of attention is by using an enriched model that incorporates the nature of EEG signals and the attentional effects.

In this work we aim to improve the performance of detection of attention in multi-talker scenarios by means of an automatic adaptive m-PSK detector based on fuzzy logic. The relevant information mentioned in the previous paragraph is modeled as two inputs that feed the fuzzy-based attention detection model. The output provides the detection. The full description of the model is reported in Section 2.4. Our fuzzy model improves the m-PSK-based detection by modelling both the m-PSK detector and the effects of attention on EEG signals during the execution of the attentional paradigm. In this study we have worked with the same data set used in (Lopez-Gordo et al., 2015) for a rapid comparison of performances. The benefit of our more efficient fuzzy approach can be used to implement an automatic and online detector of auditory attention in multi-talker scenarios. There are many potential scenarios of use of such a system. For instance, for training/assessment of attention in the presence of distracters in education or sports, for the automatic filtering of auditory distractors or auditory channels in music, for the diagnosis in attention disorders, or as promising neuro-marketing tool in radio/audio advertisement.

2. Detection of auditory attention in multi-talker Scenarios: the fuzzy approach

In this section we describe (2.1) the fundament of detection of auditory attention in multi-talker scenarios based on the m-PSK approach as described in (Lopez-Gordo et al., 2015), (2.2–2.3) psycho-physiological aspects of the evocation of EEG signals in attentional paradigms and finally 2.4) describe an automatic adaptive m-PSK detector based on fuzzy logic that outperforms the standard m-PSK detector.

2.1. Modulation of the auditory sources

Detection of attention is performed in units called trials. In each trial, each of the m speeches is modulated by a barely-audible distortion. The distortion consisted of the amplitude modulation of the m speeches with m different pure-tones with the same frequency and different phases (phase shifts 0º, 90º, 180º and 360º for m = 4 and 0º, 60º, 120º, 180º, 240º and 300º for m = 6) as described in (1).

$$s_n(t) = \frac{1}{2}[1 + \sin(2\pi f_p 5 + \beta_n)]m_n(t) \quad (1)$$

Where $m_n(t)$ corresponds to one of the CRM speeches, with $n \epsilon \{0, 1, \ldots, m - -1\}$, β_n is the phase assigned to this message and equals $2\pi n/m$, f_p is the frequency of the pure tone and $s_n(t)$ corresponds to one of the modulated messages delivered to participants.

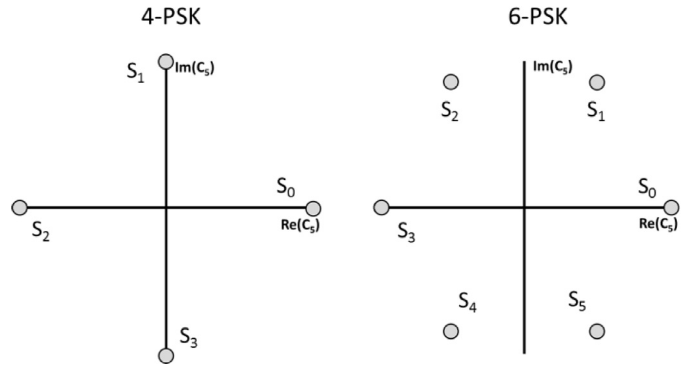

Fig. 1. The m-PSK constellations. The plots of this figure correspond to the representation of the spectral component at 5 Hz of the Fourier transform of the EEG signal. The Cartesian axes are the real and imaginary parts of the symbols of the constellation that are separated 90º or 60 º for 4-PSK or 6-PSK respectively.

Fig. 1 shows 4-PSK and 6-PSK constellations with phase shifts 90º and 60º respectively.

2.2. Physiological Response: the EEG constellation

The tone pure has a frequency of 5 Hz. This frequency was used because its period is 200 ms. This period fairly matches the time between the stimulus onset and two event-related potentials (ERPs) evoked by the modulated speeches (see Fig. 2). These two ERPs are N1 (negative deflection 100 ms after stimulus onset) and P2 (positive deflection approx. 200 ms after stimulus onset). The repetitive evocation of these two ERPs at a rate of 5 Hz causes a sinusoidal-shaped EEG signal of the same frequency (see Fig. 2). The concurrent presentation of the modulated speeches generates the simultaneous generation of the m symbols of the constellation depicted in Fig. 1. However, since symbols in the constellation are counter-phased, then the physiological response itself would give rise to their mutual cancellation. In other words, this scheme would outcome an m-PSK constellation without EEG signals. At this point is where attention plays a key role.

2.3. Cognitive Response: attentional effects on the physiological response

In the previous paragraph we concluded that a pure physiological response will evoke a zero-constellation of EEG signals. We must state that a well-known effect of selective attention on the attended stimuli is an enhancement of the energy of the corresponding ERP (see Fig. 2). Attentional effort causes modulation of the amplitudes of N1 and P2 and its rationale can be found in multitude of classical studies (Davis, 1964; Näätanen, 1975; Hillyard, Hink, Schwent, & Picton, 1973).

The net effect of selective attention on the constellation of m-PSK signals is that the response of the attended signal is no longer cancelled with its counter-phased symbol. Then, although all EEG physiological responses corresponding to modulated speeches are simultaneously evoked, only the one corresponding to the attended speech will give rise to a symbol in the m-PSK constellation.

We have stated that the standard m-PSK detector cannot cope with the non-linear effects of attention on ERPs, which in turns, is essential to build the constellation of EEG signals. As follows, we briefly present these cases that justify the use of a fuzzy approach as a better solution that just a standard m-PSK detector. They are illustrated in Fig. 3.

2.3.1. Lack of attention without artifacts

The constellation disappears due to mutual cancellation of counter-phased symbols. Then, the signal under detection has low