A Mobile Health System for Neurocognitive Impairment Evaluation based on P300 Detection

D. DE VENUTO, V.F. ANNESE, G. MEZZINA, F. SCIOSCIA, M. RUTA, E. DI SCIASCIO,
Dept. of Electrical and Information Engineering, Polytechnic of Bari,
Bari, Italy

A. SANGIOVANNI VINCENTELLI, Department of Electrical Engineering and
Computer Science, University of California, Berkeley, USA

A new mobile healthcare system for neuro-cognitive function monitoring and treatment is presented. The architecture of the system features sensors to measure the brain potential, localized data analysis and filtering and in-cloud distribution to specialized medical personnel. As such it presents trade-offs typical of other Cyber Physical System, where hardware, algorithms and software implementations have to come together in a coherent fashion. The system is based on spatio-temporal detection and characterization of a specific brain potential, called P300. The diagnosis of cognitive deficit is achieved by analyzing the data collected by the system with a new algorithm called tuned-Residue Iteration Decomposition (t-RIDE). The system has been tested on 17 subjects (n=12 healthy, n=3 Mildly Cognitive Impaired (MCI) and n=2 with Alzheimer Disease (AD) involved in three different cognitive tasks with increasing difficulty. The system allows fast diagnosis of cognitive deficit, including mild and heavy cognitive impairment: t-RIDE convergence is achieved in 79 iterations (i.e., 1.95s) yielding an 80% accuracy in P300 amplitude evaluation with only 13 trials on a single EEG channel.

1. INTRODUCTION

Mobile healthcare is a very important application domain that is rapidly capturing the attention of industry and academia. Given the evolution of devices such as tiny sensors, of infrastructures such as the human intranet and of sophisticated data-analysis techniques, designing a mobile healthcare system exhibits the typical characteristics of CyberPhysical Systems. The importance of CPS techniques in this domain is only destined to grow as we close the loop between diagnosis and cure. In particular, systems that involve brain measurements have evolved beyond expectations. In this paper, we deal with the design of a system for the detection and analysis of brain activity focusing on Event-Related brain Potentials (ERP). ERP analysis is a widely used diagnostic tool for evaluating brain activity. The analysis is oriented to the detection of early neurodegenerative pathologies as well as for monitoring and rehabilitation. Pathologies of relevance here are Alzheimer Disease (AD) in adults and Attention-Deficit/Hyperactivity Disorder (ADHD) in children. The treatment for ADHD involves pharmacotherapy by psychostimulants, which directly affect the central nervous system, and consequently is associated with severe side effects [1].

ERPs are usually analyzed using electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS), the leading non-invasive neuro-imaging solution in terms of cost and portability [2]. While EEG offers a temporal resolution of about 0.05s and spatial resolution of ~10 mm, fNIRS provides less temporal resolution (~1s) but higher spatial resolution (~5mm) [3].

Most of the medical centers performing AD or ADHD diagnosis, base their analysis on the characterization of P300, a particular brain ERP that is usually elicited using

This paper has been supported in part by the Apulian regional technological cluster 'PERSON' and by the TerraSwarm Research center at the University of California, Bekreley. The authors would like to thank the Medical Center 'Giovanni XXIII' and the neurological research team headed by prof. De Tommaso and Ing. Sabino Loconte for his collaboration during data collection.

the oddball paradigm, where low-probability target items are mixed with high-probability non-target (or "standard") items. Many solutions have been already proposed in literature (e.g., see [4, 7] and references therein) for P300 extraction and detection starting from EEG raw data. fNIRS is used for drowsiness detection while driving [3], and EEG is used together with Bayesian Spatio-Spectral Filter Optimization (BSSFO) for Brain Computer Interfaces (BCI) [8]. Hybrid systems combining EEG and fNIRS have also been implemented to control a four directions mechanical arm [9]. These methods are not in general appropriate for diagnosis where a deeper data analysis is needed. The most commonly used algorithms to measure and characterize the P300 potential in a clinical environment are the Independent Component Analysis (ICA) [10], the Principal Component Analysis (PCA) [11] and the 'grand average'. However, these algorithms suffer drawbacks (as reported in Section II) which motivated us to develop a novel algorithm that overcomes these limitations. Furthermore, there are significant advantages if the measurement equipment is portable and can connect wirelessly with the Internet to allow remote monitoring and eventually treatment delivery [12, 18].

In this paper, we present a novel EEG-based mobile health (m-Health) system for neuro-cognitive impairment diagnosis. The architecture of the system features sensors to measure the brain potential, localized data analysis and filtering and in-cloud distribution to specialized medical personnel. As such it presents trade-offs typical of other Cyber Physical System, where hardware, algorithms and software implementations have to come together in a coherent fashion. In particular, the system uses P300 spatio-temporal characterization, based on a variation of the Residue Iteration DEcomposition (RIDE) approach [19]. It has been tuned for a P300 analysis called "tuned RIDE" (t-RIDE). t-RIDE allows extracting spatial (e.g., topography and source of ERP) and temporal (e.g., latency and peak) parameters to detect neuro-cognitive impairment. t-RIDE obviates the drawbacks of methods mentioned above. The proposed system has a number of advantages for both patient and medical personnel, given its remote monitoring capability that requires neither physical proximity of the patient and the doctor nor the hospitalization of the patient. To fully exploit the remote monitoring capabilities of the system, we note that:

- When analyzed in a medical environment, the P300 is measured with a high degree of reproducibility (distraction sources can be eliminated).
- However, when the test is carried out remotely in a domestic environment, the presence of factors related to the subject such as other diseases, visual disturbances, mood, and effects of drugs, as well as to the environment such as time of day, temperature, ambient light, disturbing sources as children or pets, can heavily affect the P300 measurements.

To identify and filter out these disturbances, we adopted knowledge representation and reasoning methods to yield an automated real-time assessment of the remote cognitive test outcome.

The paper is organized as follows. Section II presents most relevant P300 features and describes how automatic detection has been carried out, with particular attention to the RIDE method. Section III outlines the m-Health architecture, the t-RIDE algorithm and the semantic-based process improving the accuracy of the algorithm. Section IV presents the experimental results coming from in vivo measurements on 12 healthy subjects (age 26.5 ±3.5) and 5 patients with cognitive impairments of various degrees (age 64.9±10.9).

2. EVOKED RELATED POTENTIALS

2.1 Evoked Related Potential: the P300

The P300 is a particular ERP that is evoked when a subject is actively and cognitively engaged in the discrimination of a target stimulus with respect to non-target ones [20, 22]. In literature [21], 'stimulus' is intended a single external (e.g., audio, visual, or tactile) event delivered to the subject under test. The target stimulus is the event to be recognized among non-target ones. A game/task is an assembly of stimuli (target and non-target ones). The nature of the external event to cause P300 occurrence is not relevant (we used visual stimuli). However, in a single task, the probability of target occurrence has to be lower than the non-target one. This well-consolidated procedure is generally known as the "oddball" paradigm [21, 23]. The P300 characterization is based on latency, amplitude of the detected pulse, location and source. The P300 latency is heavily affected by trial-to-trial variability (P300 jitter) within a given experimental condition and, according to [24], ranges from 290ms to 447.5ms depending on the cognitive difficulty of the discrimination. Typically, the P300 is a positive deflection in the human brain event-related potentials (ERPs), see Fig. 1. P300 amplitude is the peak-to-peak amplitude between the previous deflection (N200) and the P300 maximum value (see Fig. 1).

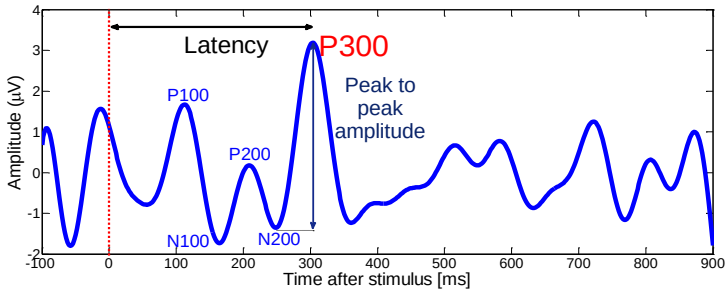

Fig. 1. Time-domain P300 waveform recorded on a healty subject on a sigle EEG channel (PZ.).

According to [23], P300 amplitudes can reach even 37.7μV depending on the age of the subject under test and on the rarity of the target stimulus. Generally, the P300 is more clearly detectable in the central parietal cortex [20]. The brain mapping of P300 is computed by a topography.

2.2 P300 as Biomarker for Cognitive Impairment Diagnosis

The EEG-based P300 characterization is a promising biomarker approach due to its non-invasiveness and portability compared to alternative diagnosis methods. The P300 latency and amplitude have been demonstrated to reflect the degree of cognitive decline in dementing illness [24]. A single P300 pulse is anticipated by other ERPs (i.e., P100, N100, P200 and N200 – see Fig. 1) which classify the cognitive process. The P300 characterization as biomarker for cognitive impairment is based on the simultaneous evaluation of amplitude and latency. We use a Fig. of Merit (FoM), defined as:

$$FoM = \frac{Peak\text{-}to\text{-}Peak\ (P300\text{-}N200)}{Latency}\ \left[\frac{\mu V}{ms}\right] \quad (1)$$

where the relative peak distance between P300 and N200 potentials are considered. According to [24], it is possible to extract threshold values for amplitude, latency and FoM to trigger the impairment classification. It is known that P300 in healthy subjects

has amplitude greater than 5.3 μV and latency lower than 349ms (calculated on 26 subjects aged 64.9 ± 10.9 years), i.e., FoM >0.01 μV/ms [24]. A subgroup of subjects, characterized by 0.008 μV/ms <FoM<0.01 μV/ms, includes potentially healthy subjects. In this case, it is necessary to rely on the expertise of the physician to finalize the diagnosis. Table I reports on threshold values according to [24].

Table I. Clinical P300 Reference for Diagnosis

	Amplitude [μV]	Latency [ms]	FoM [μV/ms]
Healthy	>5.3	< 349	≥ 0.01
MCI[1]	1.4 < A < 3.1	>389	0.003≤FoM≤0.008
HCI[2]	< 1.4	>400	FoM≤0.003
[1] Mild Cognitive Impairment; [2] Heavy Cognitive Impairment			

2.3 Brief Review of the Automatic P300 Detection Methods

The EEG signal is often corrupted by noise and motion artifacts. Moreover, the background brain activity often hides ERPs. A number of methods have been presented in the literature for P300 detection in single-trial and averaged-trials environments [3-10, 25, 26] to overcome these problems. Among these methods, the use of conventional ERP averaging ('grand average'), often in use in hospital labs, is not effective since the intrinsic variability of the ERP leads to the erroneous latency evaluation, reduction in maximum amplitude (peak) and a broadening of the component.

Woody in [25] suggested a method for single-trial ERPs, based on an iterative strategy. At first, the latencies are estimated from the cross-correlation between the 'grand average' (first template) and each trial. Then, all single-trials are aligned to the estimated latency and averaged again, leading to a second template. These steps are iterated until the templates converge.

The main limitation of this approach is the hypothesis that the ERP is monolithic (only latency jitter without shape distortion), an assumption that is not verified in reality. Other latency detection methods (i.e., peak-detection) face the same problem. Different approaches such as the Independent Component Analysis (ICA) [8] and the Principal Component Analysis (PCA) [9] need starting assumptions on the amplitude and latency value and to monitor a high number of EEG electrodes. ICA assumes that there are independent sources generating signals that are projected to the scalp [8]. However, the source of P300 is not known a priori thus invalidating the use of this method. To separate the ERP components as a projection on the scalp, the ICA needs spatially limited and temporally independent neural activation pools (sources), which can be related to each other in a linear sum. In addition, to separate the components, ICA uses a high number of trials. Generally, finding N stable components (data provided by N channels) requires more than kN^2 data sample points (for each channel), where N^2 is the number of weights in the unmixing matrix that ICA is trying to learn and k is a multiplier [8]. The value of the multiplier k increases as the number of channels increases, but more channels ensure a proper decomposition (without noisy components) of neural pool sites [8]. The large amount of data to be analyzed excludes, in computational terms, the ICA approach as diagnostic tool that could speed up the test times. Moreover, the high number of brain loci (and thus of electrodes) to be monitored make this method not suitable for everyday wearable diagnostic equipment. However, to reduce the amount of data required (e.g. reduction in the number of component), a PCA approach can be used instead. PCA separates the signal into

orthogonal components. The method works under the assumption of amplitude variations within trials and of no latency jitter [9]. However, PCA will slightly corrupt the data by adding nonlinearities [27]. A further class of methods are based on deconvolution. These approaches attempt to separate stimulus-locked and response-locked ERP components assuming a linear model of ERP interaction. Specifically, the ERP is de-convoluted into at least two ERP components, one stimulus-locked *s*(t) and one response-locked *r(t)*. The main limitation of this model is the undesired amplification of slow noise components (≈1Hz). Since ERP alignment is performed by using the physical response of the subject (go-tasks, i.e., tasks where the subject is asked to perform a motor action) to the target as the time-base, de-convolutive methods fail to find latency jittered components in tasks with no external response (no-go tasks). The method adopted in this paper is a tuned version of the Residue Iteration DEcomposition (RIDE) which is a hybrid approach based on linear superposition and iterative residual calculation [19]. The RIDE allows detecting spatio-temporal ERP characteristics with no limitations in terms of number of electrodes and number of target stimuli. RIDE considers a linear superposition model of single-trial ERPs. Except for the noise ε, the single-trial EEG is decomposed into two components: the stimulus-locked (S) and cognitive-locked (C) components. In a go-task, a third component response-locked (R) has to be considered (but this is not our case). A single-trial EEG, including EEG background activity and noise, can be expressed as:

$$EEG_i(t) = S(t) + C(t + \tau_i) + \varepsilon(t) \quad (2)$$

Where the stimulus-locked component *S(t)* is supposed to be deterministic in each trial, τi is the latency of component C in the *i*-th trial and is characterized by a distribution ρ(t) assumed to be Gaussian (this is not a limiting hypothesis), ε(t) is the noise supposed to have a uniform distribution. A conventional average over N trials would result in:

$$ERP = \mathrm{E}[EEG_i(t)] = \mathrm{E}[S(t)] + \mathrm{E}[C(t + \tau_i)] + \mathrm{E}[\varepsilon(t)] = S(t) + \int C(t + \tau)\, \rho(\tau)\, d\tau + \frac{\varepsilon(t)}{\sqrt{N}} = S(t) + C * \rho + \frac{\varepsilon(t)}{\sqrt{N}} \quad (3)$$

Equation 3 shows that, although noise is reduced, the average creates a broadening of the *C* component, which is convolved with its distribution. Neglecting ε, it is possible to consider the residues in single-trial:

$$Res_i(t) = EEG_i - ERP = C(t + \tau_i) - C * \rho \quad (4)$$

If the residues are aligned to their τ_i through cross-correlation jitter-latency estimation and averaged again, the distortions are reduced and a first estimate of *C* is computed as:

$$C_1(t) = \mathrm{E}[Res(t)] = C(t) - (C * \rho) * \rho \quad (5)$$

By replacing C_1 (5) in (3) it is possible to obtain a first estimate of S_1. The procedure is then iterated using a first ERP estimate $[ERP_1 = S_1(t) + C_1(t+\tau_1)]$ leading at the end to a more precise *S* and *C* estimate. After the *n*-th iteration, the components C_n and S_n are given by:

$$C_n(t) = C - C * \rho_0 * \rho_1 \ldots * \rho_n \rightarrow C \quad (6)$$

$$S_n(t) = S - S * \rho_0 * \rho_1 \ldots * \rho_n \rightarrow S \quad (7)$$

Since the iterative convolution by ρ approaches zero, C_n and S_n converge to *C* and *S* after n iterations. Differently from similar iterative methods (i.e., Takeda et al. [26]), the RIDE method does not introduce systematic artifacts and its convergence is fast (≈ 10 iterations). The RIDE algorithm has been tested for different ρ distributions and it has been verified to be robust and accurate [19]. Due to its qualities that include low

number of target stimuli, no-go task applicability, few electrodes needed, good accuracy, and information regarding single-trial, the RIDE method was selected as core for our diagnostic method, called tuned RIDE (t-RIDE) and described in Section III.

3. SYSTEM ARCHITECTURE

The system is based on a client/server configuration and consists of two parts corresponding to the patient and the medical staff respectively, who communicate using cloud technologies grounded on TCP/IP. Fig. 2 summarizes the overall architecture. In the implemented solution, the patient who is wearing a wireless EEG cap, can perform autonomously at home three different oddball tasks of increasing cognitive difficulty on a PC, tablet or smartphone. The tests are completely driven by plug-and-play software and no user intervention is needed. The oddball protocol (described in detail in the next sections) and the t-RIDE parameters are based on a configuration cloud-shared with the medical center in order to be eventually modified by the physician. EEG data are immediately processed in-loco by t-RIDE. t-RIDE outcomes are subsequently interpreted by a semantic-based matchmaking to relate P300 amplitude and latency to the particular conditions where the test was performed and with the clinical history of the patient. The consequent medical report is created and cloud-shared in real-time with the physician. The report contains the spatio-temporal P300 characterization and the semantic annotations. Hence, the physician, performs the final diagnosis based on the data and the clinical history of the patient. In case of cognitive impairment monitoring, the physician, for instance, could remotely verify the effectiveness of a specific drug treatment. In case of periodical analysis for a predisposed subject, the physician can detect the early presence of cognitive impairment. It is important to remark that the m-Health system performs data processing and semantic matchmaking, but the final diagnosis is left to the medical staff.

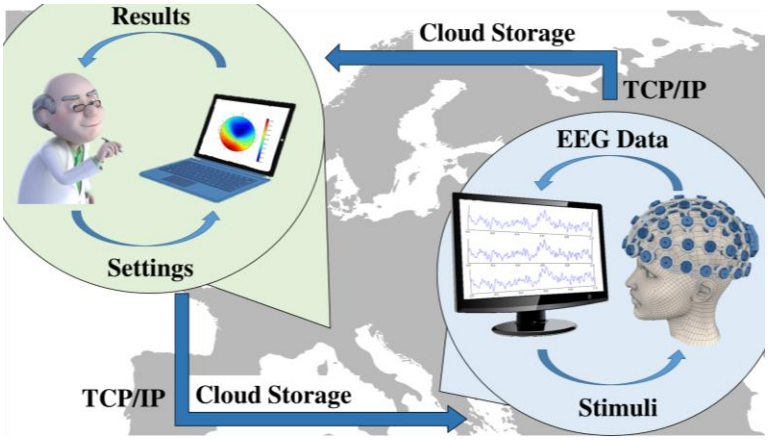

Fig. 2. Overall Architecture of the m-Health service proposed.

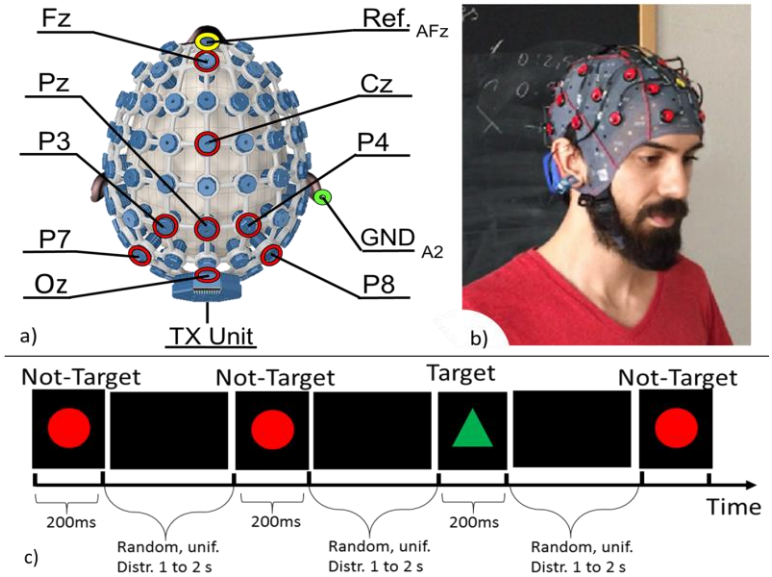

Fig. 3. a) Channels of interest (in red) according to the international system 10-20. In yellow the nasion reference and in green the ear-lobe ground. b) Demonstrative picture of the EEG wireless headset. c) time diagram of task A.

3.1 The Hardware

The patient side equipment is: i) a EEG wireless headset (sensors and wireless communication interface), ii) a PC or tablet or smart phone, which interfaces to the headset and to the Internet with TCP/IP, iii) the test game (stimulus delivery) defined by the medical center and iv) a user-friendly software, which organizes the data, processes them (by t-RIDE), performs the semantic-based correction and uploads them on the cloud. The patient side performs data collection and P300 detection: EEG data collected by the wireless EEG headset are sent to the gateway (i.e., PC) which delivers the video game/test and performs EEGs processing. Once processed, the results are uploaded on the cloud and made available to the medical center. In our experiments, the EEG headset is a 32-channels wireless recording system with active electrodes (conditioning integrated circuit are embedded in the electrode performing amplification, filtering and digitalization). According to the international 10-20 system for EEG data acquisition, eight channels are considered (Fz, Cz, Pz, Oz, P7, P3, P4, P8 – in red in Fig. 3.a) referenced to AFz electrode (in yellow in fig. 3.a) while the right ear lobe (A2) is used as ground (in green in fig. 3.a). The EEG signals are recorded during the test and are synchronized with the delivered stimuli by the gateway, which drives the test. The gateway is a PC (we used an Intel i5 CPU, RAM 8 GB, 64 bit configuration) with proper wireless communication interfaces for the BLE link to the headset and an efficient wide-area communication interface (i.e., TCP/IP). EEG data collection, the game/stimuli generation and subsequent data processing are performed with Simulink. Once the P300 processing is completed, outputs are delivered both to the patient (immediate response) and to the physician who, by visual inspection, can perform a remote diagnosis based on t-RIDE-processed measures.

The medical center side of the system consists of: i) a personal electronic device (PC, tablet, or smart phone), ii) the configuration setting for the signal processing results to be loaded from the cloud and iii) the clinical Knowledge Base (KB) storing information describing the medical history of the patient. The physician can interact with the system in two ways: he can modify the cloud-shared parameters downloaded

by the gateway before each test and containing directives for tasks and for signal processing (e.g., window of interest and channels); he can consult the results of the test. The results are the P300 spatial (topography, i.e., brain mapping) and temporal (latency, peak) characterization and the semantic annotations. The physician has always access to the clinical records of the patient to evaluate his medical history. He can then evaluate the evolution of the neuro-cognitive impairment. Data security can be guaranteed by proper-compounded authentication systems (i.e., fingerprint or double password). In the next section, a detailed description of each block of the proposed m-Health system is presented.

3.2 t-RIDE: Spatio-Temporal P300 Characterization

A dedicated Simulink block for data collection managing the API of the headset has been developed. Data are recorded by a wireless EEG commercial headset communicating to the gateway trough the Bluetooth Low Energy (BLE) protocol. The hardware used for data collection is the same as the one used in [28-35]. However, in [29, 34], the EEG channels were the motor-cortex ones, whereas in this paper the main area of interest is the central-parietal cortex. We use eight EEG electrodes (*Fz, Cz, Pz, Oz,* P7, P3, P4, P8) which are recorded at 500*Hz*, with 24-bit resolution, input range ±187.5mV and filtered using a bandpass (Butterworth, 8th order 0.5-100*Hz*) and a notch (Butterworth, 4th order 48-52*Hz*) filter [36, 41]. These filters are embedded into the signal conditioning circuit of the EEG electrodes. The recording scheme is monopolar and the buffer size for wireless transmission is 192 bits (i.e. 8 samples).
The algorithm running on the PC for the P300 extraction is a tuned version of RIDE. RIDE overcomes the limitation of conventional stimulus-locked average-based methods, which consists of a blurred ERP because the sub-components are not locked to stimulus-onsets. RIDE consists of separating different ERP sub-components with different latency variability and then rebuilding the ERP. However, the following limitations should be addressed:

- RIDE is a generic approach for ERP extraction. When it is used for a specific ERP such as P300, some a priori information about the signal shape (*e.g.*, "the expected latency is in the neighborhood of 300ms") should be used to increase accuracy.
- RIDE requires the user (e.g., physician, medical staff, caregivers) to select the unique computation window. The window is usually chosen by visual inspection often with no formal consideration or specific ERP knowledge. Additionally, the latency of the subject is not known a priori and its variability from test to test generates jitter: the same window may be suitable for the extraction of P300 related to a particular stimulus but it can be totally inadequate for the next one. Thus, the RIDE method on the same dataset may return very different results with small variation of the calculation window, i.e., the RIDE method lacks the needed robustness.
- RIDE does not perform signal pre-processing. For this reason, RIDE may fail in presence of a negative offset in the EEGs [24].
- RIDE does not allow the automatic spatial characterization (topography) that can give an additional diagnostic support to the physician.

For these reasons, we have "tuned" the RIDE algorithm to increase its accuracy for P300. The tuned RIDE (t-RIDE) includes the following features:

Pre-processing. This stage is intended to reduce the sources of noise, artifacts such as eye movements and head movements preserving the P300 and eliminating the critical

issue affecting RIDE. Pre-processing is performed for each monitored channel and for signals obtained by averaging different electrodes. For the sake of simplicity, only a single-channel processing chain will be described: the acquired signal is low-pass filtered (Butterworth, 6th order, *fstop = 15Hz*) and aligned to the stimulus signal. Subsequently, the EEG signal is decomposed in epochs of $1s$. Each epoch starts $100ms$ before the rising edge of the stimulus (target and non-target) and ends $900ms$ after. Epochs are fitted in a 6th order polynomial[1]. The resulting fitted curve is subtracted from the EEG signal, which is then centered (offset cancellation) and normalized. In order to preventively exclude invalid signals, the epochs exhibiting a maximum peak higher than a predefined threshold (in this paper the threshold is $50\mu V$) are automatically excluded from the next processing stages. The choice of this threshold depends on the assumption that the highest peak detectable in each valid epoch is the P300 (Fig. 1 represents the physiognomy of a valid epoch), which - as mentioned in Section 2.1 - is limited in amplitude [23].

Window optimization. Starting from a default window, t-RIDE automatically optimizes it. The source of P300 is not known a priori. Hence, at first, only one signal derived from the average of *Pz* and *Cz* is considered for window optimization. A general pseudocode describing this stage is presented below:

```
LOAD CONFIGURATION.TXT
WINDOW <- READ WINDOW
 WHILE I< NIT                  # WINDOW OPTIMIZATION PROCESS
        RUN RIDE(CZPZ,WINDOW)              # RIDE CALL
      # RIDE CALL EXTRACTION: TIME-DOMAIN P300 WAVEFORM
        READ ERP
      # STORE P300 AMPLITUDE COMPUTED ON THE I-TH WINDOW
        PEAK(I) ← MAX(ERP)|WIN
        IF MAX < PEAK(I)
        # DYNAMIC STORING OF THE OPTIMIZED WINDOW
                OPTIMIZED_WINDOW ← I
        END
        I ← I+1
        # SHIFT THE WINDOW AND REPEAT THE CYCLE
        WINDOW ← [TS,WIN+(TSH,L*I), TE,WIN+(TSH,R*I)]
END
# THE OPTIMIZED WINDOW HAS BEEN SELECTED: RUN RIDE ON ALL THE CHANNELS
WINDOW ← [tS,WIN+(tSH,L* OPTIMIZED_WINDOW), tE,WIN+(tSH,R* OPTIMIZED_WINDOW)]
RUN RIDE((FZ, CZ, PZ, OZ, P7, P3, P4, P8,WINDOW)
```

The software initializes test and computation by loading the configuration setting. It contains all the parameters that are useful to the tuning phase. The setting can be remotely customized by the physician. In particular, it contains the time limits of the first time-window and the time-step increments (*ms*) of the start and the endpoints of the window for their shift. The system also provides an automatic generator of the configuration file, which acquires via user interface the age of the patient, assessed cognitive level, and the presence of relevant diseases (e.g., aphasia, memory difficulty, and depression). The tool also generates the file in complete autonomy and determines the time increment granularity. In case the configuration file is not changed, a default window is defined as $250\text{-}400ms$ after the stimulus and the default time-step

[1] A 6th order polynomial has been selected as it is the highest that eliminates the slow trends without modifying the ERP patterns.

increments are set to 4ms and 8ms for starting point and end-point of the window, respectively. The procedure for window optimization is as follows. The number of iterations for an optimal tuning phase is given by:

$$n_{IT} = \left\lceil \frac{t_{lim} - t_{e,win}}{t_{sh,r}} \right\rceil \quad (8)$$

where t_{lim} is the upper limit proposed in literature [21], $t_{e,win}$ is the selected window end time , $t_{sh,r}$ is the right-shift parameter. Therefore, nIT different windows are considered, that sweep the entire time slot where the P300 is expected. By default configuration, the iteration cycle starts from a fixed window, then its begin/end points are progressively right shifted by 4ms and 8ms respectively. Thus, the last window considered in the computation is 278ms–456ms. For each considered window, an early estimation of latency C is performed using Woody's Method [20]. Based on template matching, the cross-correlation between the template of P300 and a single trial EEG is performed and residuals are calculated separating the S and C components. The procedure is iterated until C does not change in every single trial. The P300 is rebuilt attaching the $S(t)$ component to the target stimulus rising edge while the $C(t+\tau^*)$ component is queued to $S(t)$ using τ^* which is the average latency estimated on the target stimuli. Results for this window are stored and the above process is iterated again with a different windowed EEG signal. At the end, the algorithm selects the optimized window i.e., the one with the highest P300 amplitude. User interaction is possible but not strictly needed: t-RIDE includes default settings allowing signal processing even by individuals with no background knowledge on ERP.

Result Extraction. t-RIDE calculation is performed on all channels [8]. A statistical analysis of data provides the physician with medium latency and peak for each channel. In addition, latency and peak are available for every trial. The P300 characterization outputs are automatically stored on the cloud and consist of:

i. *Diagrams* showing the time-domain waveforms of the target stimuli compared to not-target ones for each channel and task;
ii. *Maps* showing information on the P300 generation and propagation for each task (topography);
iii. An *automatically generated table* that expresses presence/absence of P300, peak and latency values for each channel and task.

The results can be visualized *in loco* from the caregivers or familiars. No formal diagnosis is provided leaving this task to the physician.

3.3 The Cognitive Tasks

The remotely performed cognitive tasks are based on the oddball paradigm [19] and are delivered by the gateway through visual stimuli (e.g., a videogame). The patient performs three different no-go cognitive tasks (task A, B and C) of increasing difficulty, where he has to recognize a rare target stimulus among the non-target ones. During the tasks, the patients are asked to reduce eye movements, blinking, head and body movements, as well as jaw contractions, in order to reduce artifacts. Before each task, no stimulus is presented for a $20s$ slot to allow the filter effect to disappear.

Task A. On a black screen (15"), a red circle and a green triangle are repeatedly, randomly flashed with non-uniform probability. The subject is asked to count the occurrence of the less frequent target stimulus (the green triangle). The target stimulus probability is 20%. The inter-stimuli time is randomized and has a uniform distribution ranging from 1s to 2s. Each visual stimulus persists on the screen for

$200ms$. The subject distance from the screen is approximately $1.5m$. The time length of task A is $127s$ (in this interval approximately 25 target stimuli are presented). Since task A involves both chromatic and geometrical mental classification, the P300 is expected to be more pronounced. In Fig. 3.c a time-diagram of task A is shown.

Task B. In this task, the chromatic difference between stimuli is eliminated: a red triangle (target stimulus with 20% probability of occurrence) and a red circle (non-target stimulus) are randomly shown to the patient. The cognitive difficulty for task B is increased since the human brain classification is now based on geometrical shape alone.

Task C. Task C preserves the configuration of task B but the classification based on the geometrical shape is made more difficult by the presence of stimuli with very similar shapes (non-target: red circle; target: red ellipse).

3.4 Semantic-based Automated Test Assessment and Correction

The P300 spatio-temporal characterization (latency and amplitude for each monitored channel and task) obtained with t-RIDE is the input to a semantic-based reasoning system that provides an automated real-time assessment of the cognitive test outcome and possibly identifies factors to improve it. In a generic knowledge-based framework grounded on Description Logics formal languages [42], semantic matchmaking allows ranking the best matching resources R_i for a given request Q through proper deductive inferences. In this case, concept contraction and concept abduction [43] are used, as implemented by the Mini-ME embedded reasoning engine [44].

Adapting these algorithms to the problem at hand requires building on the fly a semantic test description acting as request for the matchmaking problem and matching with a set of outcome "templates", representing the available resources. Concept abduction can be used to discover both the closest outcome type and the most suitable corrective actions for a given subject in a given context. If Q does not match R_i fully, concept abduction extracts which hypothesis H should be made about R_i to obtain a full match and associates a semantic distance score.

A suitably devised ontology in the *Attributive Language with unqualified Number restrictions (ALN) Description Logic* provides the conceptual vocabulary to model the neurocognitive test context (time of day, features of the test environment, etc.), subject's conditions (including age, sex, diseases, ongoing treatments and sensory impairments) and finally the test response P300 amplitude and latency. Threshold values have been set in consultation with domain experts (namely the Neurology team of Bari General Hospital) and analyzing with the physicians the experimental results. A set of reference semantic annotations, together with this conceptualization, forms a knowledge base of cognitive test outcome "templates" in a semantically rich formalism. The templates include five outcome classes – from "excellent" to "critical" – for assessing the subject's performance, and two descriptions of corrective actions on the outcome related to subject's conditions and test context respectively. Abstracting from raw data to semantic annotations consists of mapping each attribute of the current test to a class in the ontology and assembling the test description as their logical conjunction. We modeled both ontology and annotations, and leveraged the strict collaboration with neurologists and physiologists to evaluate the neurocognitive impairment, The semantic distances allowed detecting the test outcome type automatically, while the hypotheses return the corrective factors to improve the test results.

4. RESULTS

The dataset is based on 12 healthy subjects with age between 23 and 30, and a group of 5 Cognitive impaired subjects (n=3 MCI, n=2 early AD stage) in an age range from 55 to 74. Data are acquired with a wireless equipment and supported by highly specialized medical staff. Each group was selected in consideration of a certain degree of homogeneity in terms of age and level of education. Recordings were performed in a controlled environment. Subject were asked to perform task A, B and C. Table II recaps main characteristics of the dataset, including a brief summary of the performed tasks. After the tests, the subject expressed an opinion about difficulty of the tasks. This test confirmed the increasing complexity of the tasks to be performed.

4.1 P300 Spatio-Temporal Characterization

Fig. 4 shows the amplitude topography (for both target and not-target stimuli), the latency topography and the FoM topography for the task A. It is noteworthy the difference between the P300 amplitudes associated to target and not-target detection. This voltage variation is, in average, $\Delta V = 2.95\ \mu V \pm 1\ \mu V$ (on *Pz*). In the 100% of the recordings, P300 had an amplitude in presence of the target higher than for the not-target one. In particular, for the not-target, the P300 highest voltage levels (in the target amplitude topography) are concentrated in the center-parietal region (*Cz, Pz, P3, P4*). The time instant when P300 reaches its maximum is related to the electrodes position.

Table II. Protocol Dataset

Subjects	12 healthy subjects (age 26.5 ±3.5) , 5 diseased subjects (age 64.9 ±10.9)		
EEG setting	Fz, Cz, Pz, Oz, P7, P3, P4, P8; reference: nasion; gnd: ear-lobe		
EEG data	500Hz ; 24-bit resolution		
Processing	Pre-processing; t-RIDE; P300 characterization		
	Task A	Task B	Task C
Type of Stimulus	Visive (red circle and green triangle)	Visive (red circle and red triangle)	Visive (red circle and red oval)
target Distribution	Random, 20% probability	Random, 20% probability	Random, 20% probability
Inter-stimuli time	Random, 1-2s	Random, 1-2s	Random, 1-2s
Stimulus permanence	200ms	200ms	200ms
N° of targets (average)	25	25	25
Classification	Chromatic & geometric	Geometric (easy, different shapes)	Geometric (difficult)
Subject Opinion on the difficulty	3 – very easy	5 – moderate	7.5 - difficult

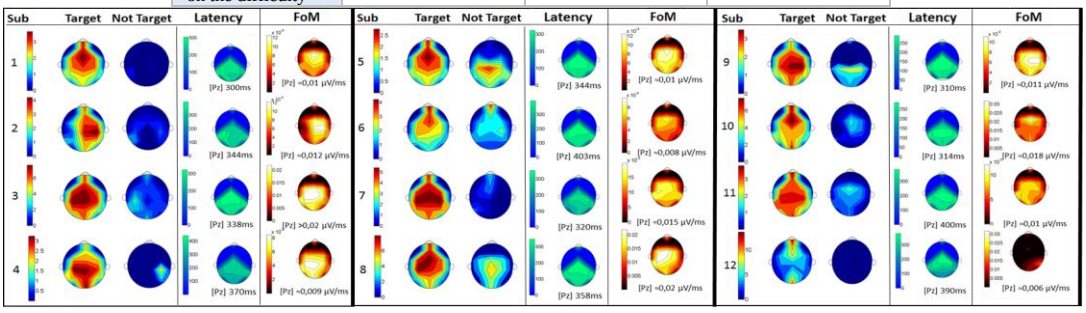

Fig. 4. t-RIDE results for each healthy subject. From left to right: target and non-target amplitude topography, latency and FoM. P300 was detected on the 100% of the subjects. P300 is more evident in the central-parietal electrodes. The amplitude difference between target and non-target response was ΔV=2.95 µV ± 1 µV (on Pz).

Table III. Results of t-RIDE

Dataset	Method	Subjects	P300 Peak	P300 Latency
Task A	t-RIDE, 25 Average targets, 8 channels	12, healthy (age 26.5 ±3.5)	Range : 3-8 μV Mean : 4.7 μV σ : 0.61	Range: 300-403 ms Mean : 349.25ms σ : 35.52
Task B	t-RIDE 25 Average targets, 8 channels	12, healthy (age 26.5 ±3.5)	Range : 3-6.2μV Mean : 4.4 μV σ : 1.28	Range : 340-410ms Mean : 363.3ms σ : 14.97
Task C	t-RIDE 25 Average targets, 8 channels	12, healthy (age 26.5 ±3.5)	Range : 2.8-4.9 μV Mean : 3.7 μV σ : 0.98	Range : 360-410ms Mean :378.46 ms σ : 14.91
Related work [21]	Mix, > 160 targets	75, healthy (age: 27,17± 19.16)	Range : 2.6-37.7 μV Mean: 10.4 μV σ : n.d.	Range : 290-447 ms Mean : 316.5 ms σ : n.d.
Task A	RIDE, 25 Average targets, 8 channels	12, healthy (age 26.5 ±3.5)	Range : 2.2 - 7 μV Mean : 4.1 μV σ : 1.43	Range: 302-387 ms Mean : 351.4ms σ : 36.13

The latency topography of all subjects shows that P300 is detected from the lateral mid-line electrodes (200-250ms at *P3* and *P4*) and the central electrodes (*Fz, Cz, Pz*) 300-400 ms after stimulus. The FoM analysis allows characterizing the subject both from amplitude and latency at the same time. The highest values of FoM are recorded in the parietal cortex (*Pz, P3, P4, P7 and P8*). Similar analysis have been computed on tasks B and C. The upper part of table III reports the results on the healthy subjects dataset for each task. For task A, the P300 amplitude ranges from 3-8μV with a mean value of 4.7μV and a 0.61 standard deviation; the P300 latency in task A is included in the range 300 – 403ms, with a mean value of 349.25ms and a standard deviation of 35.52ms.

4.2 Comparative analysis: t-RIDE Benefits with Respect to the State of the Art Methods

In Table III, a comparative analysis is reported. Authors in [21] reported that on 75 healthy subjects (age 27.17 ± 19.16) the P300 amplitude varied between 2.6μV and 37.7μV with a mean value of 10.4μV, while the P300 latency ranged from 290ms to 447ms with a mean value of 316.5ms. Note that [21] is a review of 75 different papers implementing several P300 extraction methodologies. For each task and subject, 100% of t-RIDE results were consistent with the reference, confirming the appropriateness of the approach. Comparing RIDE and t-RIDE results, in (Table III), the P300 peak calculated by t-RIDE is, in the average, +0.6μV (+12.76%) higher than the one calculated by RIDE. Comparing the standard deviations, t-RIDE performs with +57.34% higher accuracy: $\sigma_{t\text{-RIDE}} = 0.61$, $\sigma_{RIDE} = 1.43$. With respect to latency estimation, the two methods report very similar results. In Figs. 5, 6 and 7, a comparison between t-RIDE, RIDE, ICA and Grand Average (GA) is presented. ICA and GA are the most commonly used methods in specialized medical centers for P300 extraction.

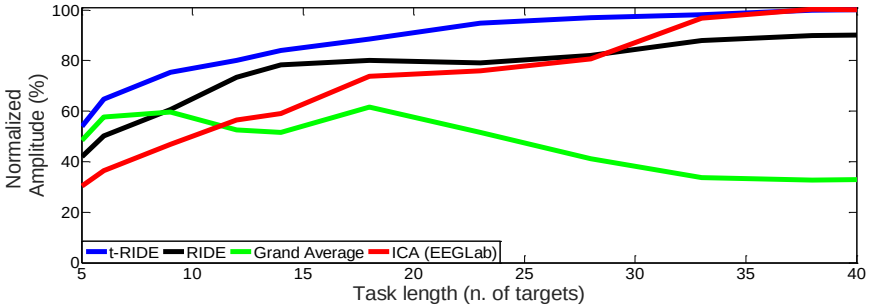

Fig. 5. P300 amplitude calculated using: t-RIDE (in blue) RIDE (in black), GA(in green), ICA (in red). The analysis is referred to identical dataset and on a single channel *Pz*.

The methods were applied on identical data stream (subject 1, Task A; 8 channels). In Fig. 5 and 6 the amplitude and latency were normalized as follows:

$$\varepsilon_{n,\%} = \frac{|x_N - \Delta x_n|}{x_N} \cdot 100; \quad \Delta x_n = |x_N - \hat{x}_n| \tag{9}$$

Where x_N is the P300 amplitude/latency evaluated with t-RIDE using a task with N= 40 target stimuli, x_n is estimated over n target stimuli, with $n < N$. x_N is the converged value of the method (steady state value), and can be used as a reference value to assess the accuracy of the calculation, when $n < N$ targets are used. Notice that, under this assumption, ε_n is the accuracy of the measurement.
Fig. 5 shows the convergence of P300 amplitude achieved by the different methods with increasing number of target stimuli within the same task.
The amplitude has been normalized to the "steady state" value achieved by t-RIDE using 40 targets, according to Equation (9). Fig. 5 shows that:

i. **t-RIDE and ICA converge to the same amplitude results** with an error of 0.1%;
ii. **t-RIDE is, on average, +12.3% more accurate than RIDE** although they exhibit the same convergence behavior;
iii. The GA converged value was 67% lower than the t-RIDE one;
iv. **For what concerns amplitude, t-RIDE is more accurate than the other methods.** t-RIDE showed the highest accuracy using 25 targets: t-RIDE = 96.05%; RIDE = 78%; ICA = 75.9%; GA = 51.6%;
v. **t-RIDE needs fewer targets to reach 90% amplitude accuracy if compared with competitors.** The number of target stimuli to reach 90% accuracy are: t-RIDE = 18; RIDE = 38; ICA = 30; GA = n.d. (GA never reaches 90% accuracy). Notice that for task A, the probability of target occurrence was 20%. Considering 1s inter-stimulus time, an approximate relation between number of targets and time duration of the task is: 1 target ≈ 5s. Using this relationship, we conclude that in order to extract the P300 amplitude with a 90% accuracy, the time duration of the task has to be: t-RIDE = 90s; ICA = 150s; RIDE = 190s.

A shorter task duration reduces the habit phenomenon (which degrades the P300), and at the same time, improves patient comfort.

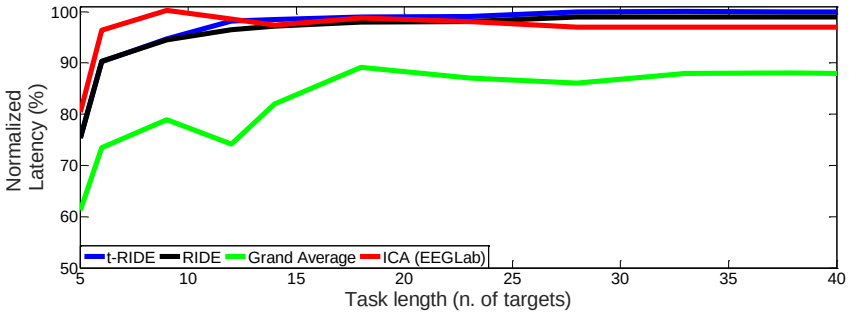

Fig. 6. P300 latency calculated using: t-RIDE(in blue) RIDE (in black), GA(in green), ICA (in red). The analysis is referred to identical data and on a single channel Pz.

Fig. 6 shows the convergence of P300 latency that was achieved by the various methods with increasing number of target stimuli within the same task. The latency has been normalized according to Equation (9). Fig. 6 shows that:

i. t-RIDE, ICA and RIDE converge to the same results;
ii. t-RIDE is, in average, +1% more accurate than RIDE +3% more accurate than ICA, although they exhibit the same convergence behavior;
iii. The GA converged value was 12% lower than the t-RIDE one;
iv. For latency, t-RIDE needs the same number of targets to reach 90% accuracy when compared to ICA and RIDE: t-RIDE = 6; RIDE = 6; ICA = 6. The number of targets to reach 90% accuracy for latency calculation with GA is 18.

t-RIDE analyzes the EEG channels individually: the minimum EEG channels for t-RIDE is 1. On the opposite, ICA requires a great number of electrodes: more than 32 channels are recommended [8]. The minimum number of EEG channels for ICA is 6. t-RIDE adoption allows wearing a more comfortable headset since there is no minimum required number of channels and 90% accuracy is reachable even with a single channel. Computationally, t-RIDE is 1.6 times faster than ICA. t-RIDE convergence is reached in 79 iterations (i.e., 1.95s) on a single EEG channel. With the same dataset, ICA converges to the same result in 216 iterations (i.e., 3.1s) giving 80% accuracy with 28 targets. Fig. 7 presents the time-domain P300 waveform on Pz calculated using the methods mentioned above using subject 1, task A, 25 targets.

4.3 Cognitive Impairment Detection

Fig. 8 shows the P300 degradation according to the task complexity: increasing the task difficulty (task A is the easiest, task C the most difficult) the P300 amplitudes decrease while latency times increase, as already presented in table III. A slight reduction is physiological, but strong variations are indicative of neuro-cognitive impairment. The joint analysis of amplitude and latency is performed comparing the FoM, introduced in Equation (1), according to the reference values reported in Table I. The 100% of the healthy subjects reported FoM reduction with the increased difficulty of the task. In particular, FoM decreases from $0.0135 \mu V ms$ ± 0.005 (task A) to $0.012 \mu V ms$ ± 0.004 (task B) until 0.011 ± 0.003 (task C).

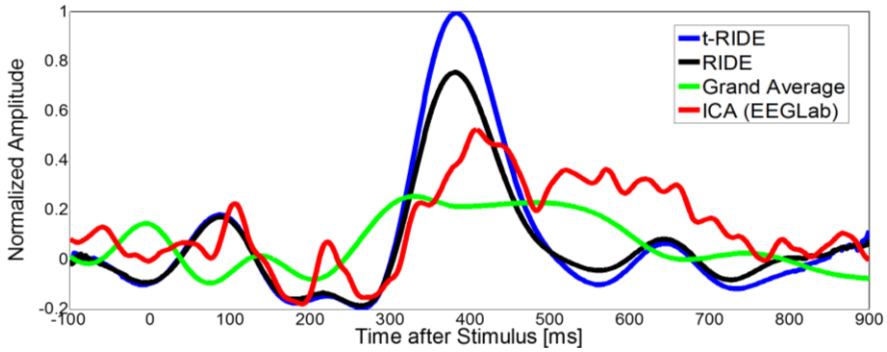

Fig. 7. P300 extraction from subject 1 during task A using t-RIDE (in blue) RIDE (in black), Grand average (in green), ICA (in red). To simplify the plot only *Pz* is shown. 25 targets were considered.

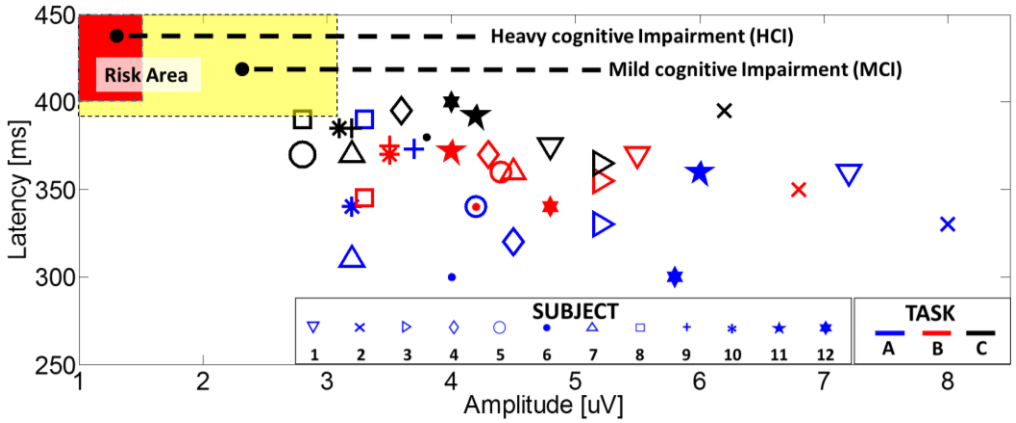

Fig. 8. Amplitudes and latencies for the complete healthy subjects dataset .

Fig. 8 also shows the risk area. This area is divided in two subgroups: MCI and HCI (AD patients) referring to latency/amplitude limits described in Table I. The five cognitive impaired subjects fall in the risk area. Fig. 9 zooms in this area, magnifying the amplitude and latency values of interest in case of cognitive impaired people (MCI and AD subjects). Typical trends in amplitude reduction and latency delay determine a FoM reduction with the increased difficulty of the task. FoM decreases from $0.0069 \mu V/ms \pm 0.001$ (task A) to $0.0059 \mu V/ms \pm 0.0009$ (task B) until 0.0045 ± 0.0002 (task C).

The selected thresholds are based on clinical evidence as reported in [21]. Indeed, the proposed CPS implements fixed clinical references for diagnosis. In the dataset analyzed and proposed in this paper, the CPS was able to correctly identify healthy subject from heavy and mild cognitive impaired patients (despite the presence of 'borderline' situations shown in Fig. 8) in 96% of the cases. On the overall dataset (n=12 healthy, n=3 MCI and n=2 AD – 3 tasks for each) only 4% of the measures returns a false positive situation (Sub.1 – Fig. 9 Task A and B). In this case, the medical staff identified that subject as MCI. In this context, further examination of the false positive identified in the evaluated subject an alarming latency (Task A: $436 ms$ Task B: $432 ms$). The latency zone where experiments did not find healthy subjects is identified in Fig. 9 with a gray semitransparent area, limited by a red dotted line. We reiterate that the system is a diagnosis tool for medical staff and is not intended to provide automatic definitive diagnosis. Human intervention is especially needed in borderline cases. In general, the detection of low amplitudes and heavy latency variability indicates the progression of neurocognitive impairment [24].

The *in vivo* measures in Fig. 10 show the relative error given by the t-RIDE (Fig. 10.a) and RIDE methods (Fig. 10.b), in the extraction of a known P300 waveform, when its amplitude goes below $3.1 \mu V$ (upper limit of the risk area) and the inter-trial latency variability grows (up to $100 ms$ from the τ_i in Equation (2)). The reconstruction error is defined as:

$$\varepsilon_r(\%) = \frac{1}{n_s} \sum_{i=1}^{n_s} \frac{|rP3(i) - P3(i)|}{\Delta P3} \cdot 100 \qquad (10)$$

where n_s is the number of samples constituting the ERPs (500 Sa), $rP3$ is the reconstructed P300 waveform, obtained by 25 trials, by using t-RIDE and RIDE on a known waveform drowned in background EEG; P3 is the known P300 waveform and $\Delta P3$ is the difference between the minimum and the maximum amplitude of the ERPs.

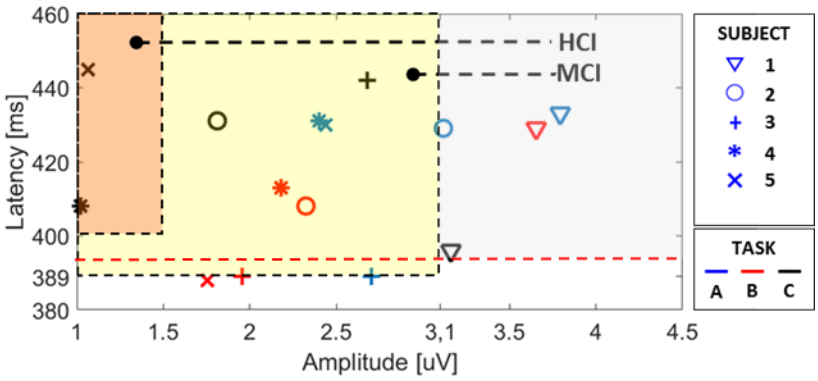

Fig. 9. Amplitudes and latencies with focus on risk area

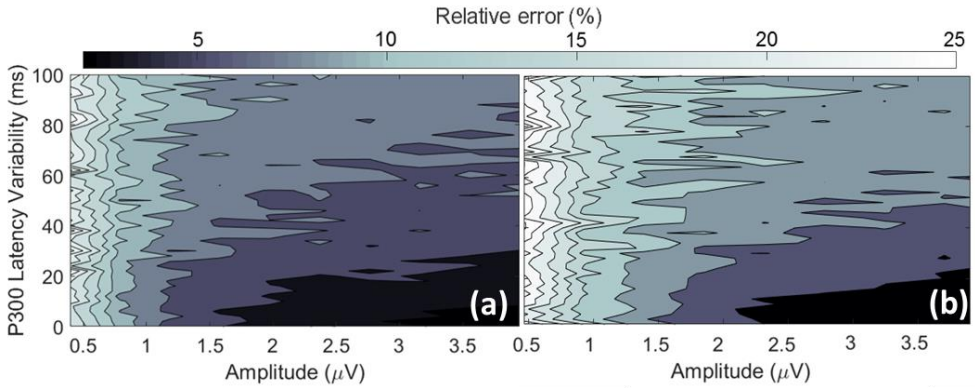

Fig. 10. 3D plot of P300 Relative error versus amplitude and inter-trial latency variability for (a) t-RIDE and (b) RIDE approaches

P3 is rewritten in Equation (11), starting from Equation (2) as:

$$P3 = a \cdot (S(t) + C(t + (\tau_i + l)) + \varepsilon(t) \quad (11)$$

where a modulates the amplitude that ranges between 0.3μV to 4μV of the known waveform, and l shifts the inter-trial latency of the P300.

As shown in Fig. 10.a, when the amplitudes decrease, the relative error provided by the t-RIDE increases, reaching 25% (the system does not discriminate well a known P300 from the EEG background) when the P300 detected amplitude is about $0.5\mu V$. A similar behavior is achieved by the RIDE approach. A relative error of 25% occurs when the P300 amplitude is about $0.75\mu V$, which is higher than the minimum signal reliably detected by t-RIDE.

On the other hand, considering a reference amplitude of 3μV, if the inter-trial latency variability increases, the t-RIDE error remains under 10% for a latency jitter < $100ms$. Instead, the RIDE method, as shown in Fig.10.b, overcomes the 10% error already when the latency jitter is around $80ms$.

4.4 Case Study

Rose – a fictional name, for privacy – is an 85-year-old woman with dementia, anxiety disorder, osteoporosis and poor hearing. She is taking the neurocognitive test in the morning at home (a pleasant and familiar environment) while in an agitated state. Rose performs the test and t-RIDE extracts a P300 amplitude of 3.1μV and latency of

401ms. This corresponds to the following test description with respect to the devised domain ontology:

Q: ∃ has_Sex ⊓ ∀ has_Sex.Female ⊓ ∃ has_Age ⊓ ∀ has_Age.Old ⊓ ∃ has_Disorder ⊓ ∀ has_Disorder.(Osteoporosis ⊓ Dementia ⊓ Anxiety_Disorder) ⊓ ∃ has_Hearing ⊓ ∀ has_Hearing.Poor_Hearing ⊓ ∃ has_Eyesight ⊓ ∀ has_Eyesight.Average_Eyesight ⊓ ∃ has_Day_Time ⊓ ∀ has_Day_Time.Morning ⊓ ∃ has_Mood ⊓ ∀ has_Mood.Agitated ⊓ ∃ has_Location ⊓ ∀ has_Location. (Home ⊓ ∃ has_Location_Property ⊓ ∀ has_Location_Property.Pleasant_Location) ⊓ ∃ has_Amplitude ⊓ ∀ has_Amplitude.Poor_Amplitude ⊓ ∃ has_Latency ⊓ ∀ has_Latency.Critical_Latency

The five test outcome profiles take into account mainly P300 response; in the case study, matchmaking assigns the lowest semantic distance to the *Critical* outcome description (full results not reported for the sake of brevity). In order to detect the impairing factors, the above *Q* is further compared with the following two descriptions, referring to environmental and subjective factors respectively.

R1: ∃ has_Time_Time ⊓ ∀ has_Day_Time.Morning ⊓ ∃ has_Feeding_Status ⊓ ∀ has_Feeding_Status.Well_Fed ⊓ ∃ has_Mood ⊓ ∀ has_Mood.Calm ⊓ ∃ has_Location ⊓ ∀ has_Location.(∃ has_Location_Property ⊓ ∀ has_Location_Property.(Familiar_Location ⊓ Pleasant_Location ⊓ Mild_Temperature_Location))

R1 models favorable test conditions, including morning time, a well-fed state, calm mood and a familiar, pleasant and mild-tempered location.

R2: ∃ has_Age ⊓ ∀ has_Age.(Young ⊓ Adult) ⊓ ∀ has_Disorder.(∃ treated_With ⊓ ∀ treated_With.(Not_Antidepressant ⊓ Not_Anxiolytic ⊓ Not_Mood_Stabilizer)) ⊓ ∃ has_Eyesight ⊓ ∀ hasEyesight.Good_Eyesight

R2 models the ideal subject's conditions, such as young-adult age, good eyesight and no psychiatric medications undermining focus and attention. Computed hypotheses are:

H1: ∃ has_Mood ⊓ ∀ has_Mood.Calm ⊓ ∃ has_Feeding_Status ⊓ ∀ has_Feeding_Status.Well_Fed ⊓ ∃ has_Location ⊓ ∀ has_Location.(∃ has_Location_Property ⊓ ∀ has_Location_Property.Mild_Temperature_Location)
H2: ∃ has_Age ⊓ ∀ has_Age.(Young ⊓ Adult) ⊓ ∃ has_Disorder ⊓ ∀ has_Disorder.(∃ treated_With ⊓ ∀ treated_With.(Not_Anxiolytic)) ⊓ ∃ has_Eyesight ⊓ ∀ hasEyesight.Good_Eyesight

In *H1* only the mood status is different from *Q*, while other reported elements are just missing in *Q*: since semantic matchmaking is based on the Open World Assumption, they are not in contrast but they are regarded just as unspecified information. Conversely, *H2* is practically equal to *R2*, showing that all subjective criteria are adverse in this case. In particular, definitions in the ontology (not reported here due to space constraints) model anxiety disorder as requiring treatment with anxiolytics, which can impair neurocognitive performance. Finally, in this setting the semantic distance score (semDist) can be interpreted as a measure of the amount of corrective actions to be taken – if possible – for improving the test outcome. Intuitively, semDist

= 0 in case of a full match (no correction needed), whereas it takes the greatest value when comparing *Q* with the most generic concept (*Top*) of the ontology (root of the taxonomy). Therefore:
Correction(*Q, Ri*) = 100% semDist(*Q, Ri*) / semDist(*Q, Top*)
Results for the case study are:
Correction(*Q, R1*) = 58.8%
Correction(*Q, R2*) = 100%
Performance evaluation executed on an HTC-Google Nexus 9 tablet (with Nvidia Tegra K1 system-on-chip, comprising a 2.3 *GHz* dual-core CPU and 2 GB LPDDR3 RAM) are intended to demonstrate the feasibility of the proposed approach with devices comparable to wearable test equipment. Results reported in Table IV are the average of ten runs: each run loaded the KB and matched three test descriptions – referring to different subjects and conditions – against each of the seven test outcome templates. Loading the Knowledge Base required 134ms on average, but this is needed only once in the lifetime of the software process using the embedded reasoning system.

Table IV. Semantic matchmaking time (in ms)

Step	Average	Standard deviation	Worst
KB loading	133.745	31.454	227.118
Request 1	0.397	0.068	0.580
Request 2	0.493	0.052	0.604
Request 3	1.335	0.547	2.415

Matchmaking required less than 1.5ms on average for every request, and less than 2.5ms in the worst case. Results show the proposed approach is suitable for integration in a real-time feedback loop to suggest – and apply automatically, whenever possible – corrective actions for improving neurocognitive test performance.

5. CONCLUSION

We presented an m-Health cyber physical system for neuro-cognitive impairment monitoring based on P300 spatio-temporal characterization. To the best of our knowledge, this system is the first remote system that performs an analysis of this kind. The architecture is supported by a new method for P300 analysis (t-RIDE) which overcomes the limitations of previous approaches (RIDE, ICA, PCA, Grand Average). The t-RIDE method has been validated on a dataset of 17 subjects (12 healthy, 3 MCI, and 2 AD) performing three different cognitive tasks of increasing difficulty. A dedicated FoM outlines that bigger complexity of the cognitive task leads to a higher latency (latency average increment in task A/C: +7.7%) and a lower P300 peak (average peak decrement in task A/C: -21.28%). This result indicates that the P300 modulation on different cognitive tasks can be an important tool for neuro-cognitive diagnosis. The algorithm is efficient: convergence is reached in 79 iterations in 1.5*s*. Its robustness has been tested by decreasing the number of trials to be taken into account. The proposed system allows remote monitoring of neuro-cognitive impairment through a 'plug and play' application, while physician customization and data collection are enabled by cloud bridging.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the Bari General Hospital Neuroscience department and particularly Prof. Giovanni Defazio for his valuable help for framework definition and assessment.

REFERENCES

[1] Alzheimer's Disease International (ADI). World Alzheimer Report 2015: The Global Impact of Dementia. 2015.

[2] Khan, M. et al. "Passive BCI based on drowsiness detection: an fNIRS study." Biomedical optics express 6.10: 4063-4078. 2015.

[3] Khan, M. et al. "Decoding of four movement directions using hybrid NIRS-EEG brain-computer interface." Front. Hum. Neurosci 8.244: 10-3389. 2014.

[4] Lotte, Fabien, et al. "A review of classification algorithms for EEG-based brain–computer interfaces." Journal of neural engineering 4.2: R1. 2007.

[5] De Venuto, Daniela, and Alberto Sangiovanni Vincentelli. "Dr. Frankenstein's dream made possible: implanted electronic devices." Proceedings of the 2013 Design, Automation and Test in Europe Conference and Exhibition, DATE 2013.

[6] D. De Venuto, V. F. Annese and G. Mezzina, "An embedded system remotely driving mechanical devices by P300 brain activity," Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017, Lausanne, 2017, pp. 1014-1019.doi: 10.23919/DATE.2017.7927139aa

[7] D. De Venuto, V. F. Annese and G. Mezzina, "Remote Neuro-Cognitive Impairment Sensing Based on P300 Spatio-Temporal Monitoring," in IEEE Sensors Journal, vol. 16, no. 23, pp. 8348-8356, Dec.1, 2016. doi: 10.1109/JSEN.2016.2606553

[8] Suk, Heung-Il, et al. "Predicting BCI subject performance using probabilistic spatio-temporal filters." PloS one 9.2: e87056. 2014.

[9] B. Martínez-Pérez, et. Al. Mobile apps in cardiology: Review, JMIR Mhealth Uhealth 1 (2) e15. 2013

[10] Makeig, Scott, et al. "Blind separation of auditory event-related brain responses into independent components." Proceedings of the National Academy of Sciences 94.20: 10979-10984. 1997.

[11] Friston, K. J., et al. "Functional connectivity: the principal-component analysis of large (PET) data sets." Journal of Cerebral Blood Flow & Metabolism 13.1: 5-14. 1993.

[12] M. Kirwan, C. Vandelanotte, A. Fenning, J.M. Duncan, Diabetes selfmanagement smartphone application for adults with type 1 diabetes: randomized controlled trial, J. Med. Internet Res. 15 (11) e235.2013.

[13] Giovanni Chiarini, Pradeep Ray et al. mHealth Technologies for Chronic Diseases and Elders: A Systematic Review. IEEE Journal on Selected Areas in Communications/Supplement. Vol.31, no. 9, Sept. 2013.

[14] De Venuto, D., Annese, V.F., Ruta, M., Di Sciascio, E., Sangiovanni Vincentelli, A.L. "Designing a Cyber-Physical System for Fall Prevention by Cortico-Muscular Coupling Detection". IEEE Design and Test, 33 (3), art. no. 7273831, pp. 66-76. DOI: 10.1109/MDAT.2015.2480707. 2016.

[15] Annese, V.F., Crepaldi, M., Demarchi, D., De Venuto, D. "A digital processor architecture for combined EEG/EMG falling risk prediction". Proceedings of the 2016 Design, Automation and Test in Europe Conference and Exhibition, DATE 2016, art.7459401, pp. 714-719. 2016.

[16] Annese, V.F., De Venuto, D. "The truth machine of involuntary movement: FPGA based cortico-muscular analysis for fall prevention". Proceedings - IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2015, art. 7394398, pp. 553-558. DOI: 10.1109/ISSPIT.2015.7394398 2015.

[17] De Tommaso, M., Vecchio, E., Ricci, K., Montemurno, A., De Venuto, D., Annese, V.F. "Combined EEG/EMG evaluation during a novel dual task paradigm for gait analysis".Proceedings- 2015 6th IEEE International Workshop on Advances in Sensors and Interfaces, IWASI 2015, art. 7184949, pp. 181-186. DOI: 10.1109/IWASI.2015.7184949 2015.

[18] De Venuto D., Annese V.F., de Tommaso M., Vecchio E., Sangiovanni Vincentelli A.L. (2015) Combining EEG and EMG Signals in a Wireless System for Preventing Fall in Neurodegenerative Diseases. In: Ambient Assisted Living. Biosystems & Biorobotics, vol 11. Springer, Cham. ISBN 978-3-319-18373-2

[19] Ouyang, Guang, Werner Sommer, and Changsong Zhou. "A toolbox for residue iteration decomposition (RIDE) — A method for the decomposition, reconstruction, and single trial analysis of event related potentials." Journal of neuroscience methods (2014).

[20] C.C. Duncan, et al. "Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300 and N400", Clinical Neurophys., v.120, n.11, pp. 1883-1908, 2009.

[21] van Dinteren, Rik, et al. "P300 development across the lifespan: a systematic review and meta-analysis." PloS one 9.2: e87347. 2014.

[22] D. D. Venuto, V. F. Annese, G. Mezzina, M. Ruta and E. D. Sciascio, "Brain-computer interface using P300: a gaming approach for neurocognitive impairment diagnosis," 2016 IEEE International High Level Design Validation and Test Workshop (HLDVT), Santa Cruz, CA, 2016, pp. 93-99. doi: 10.1109/HLDVT.2016.7748261

[23] Polich, John, et al. "P300 latency reflects the degree of cognitive decline in dementing illness." Electroencephalography and clinical neurophysiology 63.2: 138-144. 1986.

[24] T Frodl, H Hampel, G Juckel, K Bürger, F Padberg, RR Engel, et al. "Value of event-related P300 subcomponents in the clinical diagnosis of mild cognitive impairment and Alzheimer's diseasePsychophysiology, 39 (2) (2002), pp. 175–181
[25] Woody, C.D. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Medical and Biological Engineering, 5, 539-553. 1967.
[26] Takeda,Y. et al. A generalize method to estimate waveforms common across trials from EEGs. NeuroImage, 51, 629-641. 2010.
[27] Scholz, Matthias, Martin Fraunholz, and Joachim Selbig. "Nonlinear principal component analysis: neural network models and applications." Principal manifolds for data visualization and dimension reduction. Springer Berlin Heidelberg, 2008. 44-67.
[28] Annese, V.F., De Venuto, D. "Fall-risk assessment by combined movement related potentials and co-contraction index monitoring". Proceedings - IEEE Biomedical Circuits and Systems Conference: Engineering for Healthy Minds and Able Bodies, BioCAS 2015 - Proceedings, art. 7348366. DOI: 10.1109/BioCAS.2015.7348366 2015.
[29] V. F. Annese, D. De Venuto. "FPGA based architecture for Fall-Risk Assessment during Gait Monitoring by synchronous EEG/EMG". Proceedings- 2015 6th IEEE International Workshop on Advances in Sensors and Interfaces, IWASI 2015. DOI: 10.1109/IWASI.2015.7184953. ISSBN: 978-1-4799-8981-2.
[30] Annese, V.F., De Venuto, D. "Gait analysis for fall prediction using EMG triggered movement related potentials". Proceedings - 2015 10th IEEE International Conference on Design and Technology of Integrated Systems in Nanoscale Era, DTIS 2015, art. no. 7127386. DOI: 10.1109/DTIS.2015.7127386. 2015.
[31] D. De Venuto, E. Stikvoort, D. Tio Castro and Y. Ponomarev, "Ultra low-power 12-bit SAR ADC for RFID applications," 2010 Design, Automation & Test in Europe Conference & Exhibition (DATE 2010), Dresden, 2010, pp. 1071-1075.doi: 10.1109/DATE.2010.5456968
[32] De Venuto, D., Annese, V. F., Defazio, G., Gallo, V. L. and Mezzina, G. "Gait analysis and quantitative drug effect evaluation in Parkinson disease by jointly EEG-EMG monitoring," 2017 12th International Conference on Design & Technology of Integrated Systems In Nanoscale Era (DTIS), Palma de Mallorca, Spain, 2017, pp. 1-6. doi: 10.1109/DTIS.2017.7930171
[33] G. Mezzina, V. L. Gallo and D. De Venuto, "Real-time muscle fiber conduction velocity tracker for diabetic neuropathy monitoring," 2017 7th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI), Vieste, 2017, pp. 127-132. doi: 10.1109/IWASI.2017.7974232
[34] V. F. Annese, G. Mezzina, V. L. Gallo, V. Scarola and D. De Venuto, "Wearable platform for automatic recognition of Parkinson Disease by muscular implication monitoring," 2017 7th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI), Vieste, 2017, pp. 150-154. doi: 10.1109/IWASI.2017.7974236
[35] V. F. Annese, G. Mezzina and D. De Venuto, "Towards mobile health care: Neurocognitive impairment monitoring by BCI-based game," 2016 IEEE SENSORS, Orlando, FL, 2016, pp. 1-3. doi: 10.1109/ICSENS.2016.7808745
[36] D. De Venuto, D. T. Castro, Y. Ponomarev and E. Stikvoort, "Low power 12-bit SAR ADC for autonomous wireless sensors network interface," 2009 3rd International Workshop on Advances in sensors and Interfaces, Trani, 2009, pp. 115-120.doi: 10.1109/IWASI.2009.5184780
[37] D. De Venuto, E. Stikvoort, D. T. Castro and Y. Ponomarev, "Novel low-power 12-bit SAR ADC for RFID tags," 2010 11th International Symposium on Quality Electronic Design (ISQED), San Jose, CA, 2010, pp. 532-537.doi: 10.1109/ISQED.2010.5450523
[38] D. De Venuto, G. Marchione and L. Reyneri, "A codesign tool to validate and improve an FPGA based test strategy for high resolution audio ADC," Sixth international symposium on quality electronic design (isqed'05), 2005, pp. 440-447. doi: 10.1109/ISQED.2005.3
[39] M. Blagojevic, M. Kayal, M. Gervais and D. De Venuto, "SOI Hall-Sensor Front End for Energy Measurement," in IEEE Sensors Journal, vol. 6, no. 4, pp. 1016-1021, Aug. 2006.doi: 10.1109/JSEN.2006.877996
[40] D. De Venuto, M. J. Ohletz and B. Ricco, "Automatic repositioning technique for digital cell based window comparators and implementation within mixed-signal DfT schemes," Fourth International Symposium on Quality Electronic Design, 2003. Proceedings., 2003, pp. 431-437.doi: 10.1109/ISQED.2003.1194771
[41] De Venuto, Daniela, Michael J. Ohletz, and Bruno Riccò. "Digital window comparator DfT scheme for mixed-signal ICs." Journal of Electronic Testing 18.2 (2002): 121-128. DOI: 10.1023/A:1014937424827
[42] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, The Description Logic Handbook: Theory, Implementation, Applications, New York, NY, USA: Cambridge Univ. Press, 2003.
[43] M. Ruta, E. Di Sciascio, and F. Scioscia, "Concept abduction and contraction in semantic-based P2P environments," Web Intell. Agent Syst., vol. 9, no. 3, pp. 179–207, 2011.
[44] F. Scioscia et al., "A mobile matchmaker for the ubiquitous semantic web," Int. J. Semantic Web Inf. Syst., vol. 10, no. 4, pp. 77–100, 2014.