GRAPHONOMICS AND YOUR BRAIN ON ART, CREATIVITY AND INNOVATION

PROCEEDINGS OF THE 19TH GRAPHONOMICS CONFERENCE
(IGS 2019 – YOUR BRAIN ON ART)
Cancun Mexico, June 9-13 2019

Claudio De Stefano, Jose L. Contreras-Vidal
(Editors)

EUC - EDIZIONI UNIVERSITÀ DI CASSINO

COLLANA SCIENTIFICA
E-BOOK – 3

"Graphonomics and your brain on art, creativity and innovation".

A single track, international forum for discussion on recent advances at the intersection of the creative arts, neuroscience, engineering, media, technology, industry, education, design, forensics, and medicine.

The contributions reviewed the state of the art, identified challenges and opportunities and created a roadmap for the field of graphonomics and your brain on art.

The topics addressed include: integrative strategies for understanding neural, affective and cognitive systems in realistic, complex environments; neural and behavioral individuality and variation; neuroaesthetics (the use of neuroscience to explain and understand the aesthetic experiences at the neurological level); creativity and innovation; neuroengineering and brain-inspired art, creative concepts and wearable mobile brain-body imaging (MoBI) designs; creative art therapy; informal learning; education; forensics.

Centro Editoriale di Ateneo
Università degli Studi di Cassino e del Lazio meridionale
Campus universitario – Palazzo degli Studi – Località Folcara
03043 Cassino (FR), Italia

ISBN **978-88-8317-118-5**

Impaginazione a cura di EUC, Alfiero Klain

Progetto della copertina a cura di Edmondo Colella

Volume edito in versione digitale nel mese di dicembre 2022

COLLANA SCIENTIFICA
E-BOOK – 4

GRAPHONOMICS AND YOUR BRAIN ON ART, CREATIVITY AND INNOVATION

PROCEEDINGS OF THE 19TH GRAPHONOMICS CONFERENCE
(IGS 2019 – YOUR BRAIN ON ART)
Cancun Mexico, June 9-13 2019

Claudio De Stefano, Jose L. Contreras-Vidal
(Editors)

EDIZIONI UNIVERSITÀ DI CASSINO
Centro Editoriale di Ateneo – Università degli Studi di Cassino e del Lazio meridionale | 2022

Index

Preface

The International Graphonomics Society has a long history. The first conference was held in 1982 in Nijmegen, The Netherlands. In those years there was no internet, e-mail, facebook, twitter etc., and therefore the use of cursive or printed text as a means of communication was much more widespread than today: as a consequence, handwriting studies, offering the possibility to develop efficient tools for automatic handwriting recognition, represented an application field of great interest.
During the second international graphonomics conference, which was held in Hong Kong in 1985, the decision was made to establish the International Graphonomics Society (IGS). The basic motivation of this choice was to facilitate the organization of conferences and workshops and the publication of their proceedings, to stimulate communication and research contacts among researchers of different fields, and to maintain a graphonomics research directory.
From the scientific point of view, studies on the motor aspects of handwriting were of great interest for IGS members, as well as the analysis of both generation and control models of handwriting and obviously the techniques for automatic character and word recognition.

Since its beginning, one of the key issues of IGS was its multidisciplinary spirit, which allowed the interactions among researchers from different fields, different cultural areas and different points of view on many topics.
Coming to the present-day, which role can IGS play today in the digital communication era? I would point out that the basic research fields of our society, such as neuroscience, handwriting analysis and recognition, forensic applications, cultural heritage applications etc. are still of great interest together with other emerging fields such as those related to biomedical applications.
I believe that in a society where the convergence of information and communication technology are redefining the relation between individuals and the world, as well as the social interactions between individuals, there is the need of a deeper knowledge of the language used to express such relations, which is still based on signs: handwriting, signatures or drawings. The pervasive use of new devices is not only affecting the way these signs are produced, stored and communicated, but it is also extending the alphabet of signs that can be manipulated by them.

In such a scenario, graphonomics, because of its multidisciplinary and interdisciplinary nature, may provide a valuable framework for the social, scientific and technological challenges of the digital communication society.
Finally, I hope that many young researchers will approach our association and will share their enthusiasm and knowledge with us.

Cassino, October 2022

Claudio De Stefano
(IGS President)

Chapter 1

How the creative arts and aesthetic experiences
engage the human mind and promote creativity and innovation?

Characterizing the stages of creative writing from frontal and temporal mobile EEG data using Partial Directed Coherence

Jesus G. Cruz-Garza, Akshay Sujatha Ravindran, Cristina Rivera Garza, Jose L. Contreras-Vidal

Abstract - The development of mobile brain-body imaging technology provides the opportunity to study the human creative process outside of constrained laboratory settings. In this study, we used portable dry EEG systems (four channels: TP09, AF07, AF08, TP10, with reference at Fpz), coupled with video cameras, to record the brain activity of Spanish heritage students as they developed their creative writing skills over four months enrolled in an undergraduate course on creative writing in Spanish. The students recorded their own brain activity as they walked through and experienced areas in the city (Preparation phase), and while they worked on their creative texts (Generation phase). We measured Partial Directed Coherence (PDC) between the Preparation and Generation phases of their work. There was higher PDC in the Preparation Phase at a significance level of $p < 0.05$, from TP10 to AF7 among all frequency bands analyzed: 1-50 Hz. The opposite directionality was found for the Generation phase, in frequency bands: 13-50 Hz. Information transfer from temporal-parietal to anterior-frontal areas of the scalp may reflect sensory interpretation during the Preparation phase, while high frequency bands PDC directionality originating at the anterior-frontal areas during the Generation phase may reflect the final decision making process to translate the sensory experience into a tangible product: text.

Keywords: EEG, MoBI, creative writing, creativity, creative process, partial directed coherence.

1. INTRODUCTION

Recent advances in MoBI technology, de-noising algorithms, and data analytics allow the study of natural cognition and action in real-world complex environments. The development of user-friendly mobile Electroencephalography (EEG), synchronized with accelerometer, electrode-connection status, and context-aware video recording allow for the simultaneous recording of context-aware brain activity and body movements during mobile applications [1].

Writing involves embodied practices that physically connect us with our surroundings [2], [3]. We investigated creative writing as a bodily experience, in which the author's interaction with the world around them (physically, verbally, etc.) informs the cultivation and elaboration of their work. In this way, as an author engages in actively experiencing the world around them through their body, they may seek to achieve an aesthetic effect to aim for in their creative production. We integrated wearable MoBI technology into a creative writing course in Spanish, designed after the idea of approaching creative writing as a bodily experience. In the class, a creative writing professional served as the class instructor and relayed her creative methods on a creative writing workshop for 18 non-expert heritage Spanish speaking students.

In this report, we studied the process of creative writing on non-expert Spanish heritage speakers, as they engaged in the Preparation and Generation phases of their writing. The students were asked to walk through different areas of the city and experience their environment in a variety of settings, and use the experience to create an aesthetic effect in their texts.

Using fMRI to measure functional connectivity (FC) of subjects at rest, the resting state FC (rFC) between medial prefrontal cortex (mPFC) and the posterior cingulate cortex (PCC) [4] and medial temporal gyrus (mTG) [5] has been found to correlate positively with the individual's performance in creative problem solving tests. Lotze et al. [6] found decreased rFC between inter-hemispheric areas BA 44, and left area BA 44 with the left temporal lobe for individuals who scored higher in a verbal creativity index test.

Moving past correlating resting state brain dynamics to creativity scores, recent fMRI studies have analyzed the human creative process through its distinct stages of preparation, generation, and revision. Shah et al. [7] studied the Preparation and Generation phases, finding distinct cortical networks associated with each. Erhard et al. [8] found that experts had higher activation in prefrontal and basal ganglia areas. Liu et al. [9] studied the generation and revision phases; finding that the mPFC was active during both phases and the responses in dorsolateral prefrontal cortex (DLPFC) and Intraparietal sulcus (IPS) were deactivated during the Generation phase.

Although fMRI studies report involvement of the mPFC, phase-dependent and creative level-dependent activation of DLPFC, IPS, PCC, and basal ganglia, differences in brain activity for the distinct stages of the creative process remain mostly unexplored in the EEG domain; particularly for creative writing tasks. Martindale and Hasenfus [10] reported that highly creative individuals exhibited higher alpha indices during a creative inspiration (Preparation) than creative elaboration (Generation); which was not found in less creative subjects. Since 1978, there have been no quantitative EEG studies on the stages of creative writing nor during the ideation process for generating writing ideas.

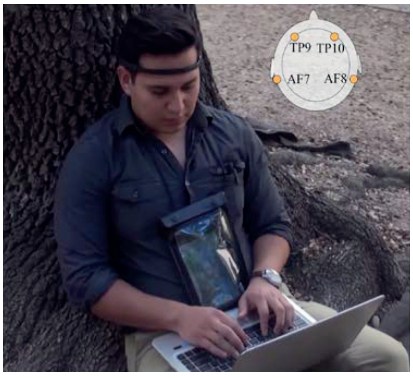


Figure 1. A participant engages in the creative writing Generation phase. The participant is shown wearing the EEG headset on his forehead, and a recording tablet with a video camera on his chest.

2. METHODS

Through readings and writing prompts, non-expert creative writing students were asked to acknowledge the physicality of the writing process and to relate it to the materiality of language. Prompts issued in the upper-division undergraduate creative writing workshops encouraged students to develop and record a series of specific writing preparation tasks (walking, running, climbing in different locations of Houston) as they completed the first phase of required assignments. Students also wore head devices as they sat down and completed their creative texts.

A. Task

Eighteen (7 males; 11 females) non-experts, heritage

Jesus G. Cruz-Garza, Akshay S. Ravindran, Jose L. Contreras-Vidal are with the Department of Electrical and Computer Engineering, University of Houston, Houston TX 77004 USA. Akshay S. Ravindran e-mail: asujatharavindran@uh.edu. Jose L. Contreras-Vidal e-mail: jlcontreras-vidal@uh.edu. Corresponding author: Jesus G. Cruz-Garza. Phone: 713-743-0796; e-mail: jgcruz@uh.edu.
Cristina Rivera Garza is with the Department of Hispanic Studies, University of Houston, Houston TX 77004 USA. criverag@uh.edu.

Spanish speakers, participated in a Spanish-language creative writing workshop at the University of Houston. The study was approved by the University of Houston Institutional Review Board. The participants provided Anonymous Informed Consent at the beginning of the workshop. They were trained to set up their own EEG headsets and body-mounted video cameras for the experiment. The participants were responsible for the collection of EEG data, video, and to keep a diary with notes on each recording session (Fig. 1).

The participants were asked to walk around the city in a variety of locations, and to use their experience to generate their creative texts, constrained only by a 3-5 page length suggestion (double space, 11 point font). The participants were instructed to use the EEG and video cameras during their walking activities and writing time. There could be more than one session of walking and writing times per prompt.

This experimental setup produced data in two phases of the creative process: the Preparation Phase and the Generation Phase. The Preparation Phase involved tasks such as walking, active observation of their environment, taking notes, and ideation. For the Generation Phase, the task involved reviewing their notes and typing their texts into a complete creative piece, with iterative revisions and modifications.

B. Measurement equipment:

EEG and head acceleration data were captured using Muse headsets (Interaxon, Toronto, Ontario, Canada). The headset has seven sensors, two out of these seven sensors were positioned at the frontal region (AF07 and AF08), two at temporalparietal region (TP09 and TP10), and the remaining three sensors served as electrical reference located at the center of the forehead (Fpz). The headset has an inbuilt accelerometer used to measure the head acceleration. EEG data for each channel were sampled at 220 Hz. The acceleration data was recorded at 50 Hz. Additionally, the data recordings contain a vector indicating contact quality for each electrode sampled at 10 Hz, rating contact quality as "indicator = 1: good", "indicator = 2: Ok", "indicator ≥ 3: bad".

The participants set up their own headset with a custom application given to them in a personal tablet, which recorded EEG and head acceleration data, and labeled it with identification numbers. The participants set up body-cameras (Conbrov, ShenZhen, China) to record their exploration (Preparation) and writing (Generation) sessions.

C. Data collection

The students were asked to make five writing exercises and collect their brain activity as they walked and observed their environment (Preparation phase), and created their texts (Generation phase). Only writing assignments that were submitted and accompanied by both video and EEG data were considered for the analysis. From the eighteen initial subjects, data from eleven subjects was discarded due to incomplete data (video or EEG missing) or assignments not submitted on time.

The Preparation and Generation phases for each writing exercise were done in several distinct recording sessions as each phase could take several time-separated recording sessions to compete. We kept each data recording as a separate session to analyze. Recording sessions were considered for analysis when all 4 electrodes had a "good" contact indicator for at least one continuous minute of data.

D. Pre-processing:

Data recordings with both video (context) and EEG data were considered for this analysis. An online notch filter was applied on the EEG data to remove the 60 Hz power line noise.

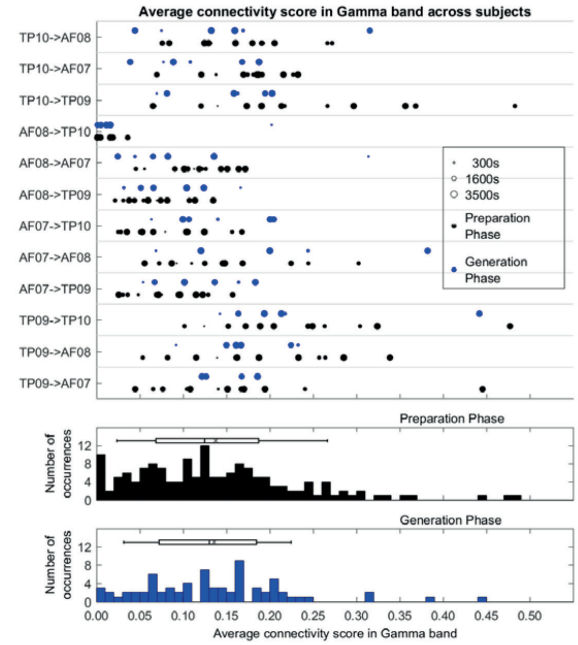


Figure 2. PDC in the gamma band. Average gamma-band PDC score for each directed connection pair, across EEG recording sessions marked as circles. Black: Preparation Phase. Blue: Generation Phase. A histogram and boxplot describing the distributions are shown in the lower panels.

We applied an offline 4th order, zero-phase Butterworth band-pass filter from 1 to 100 Hz. Artifact Subspace Reconstruction (ASR) [11] was used for the removal of short-time high-amplitude artifacts in the continuous data. Calibration data for each individual subject was taken from the entire length of the trial using automated methods. We used a cut-off threshold of ten standard deviations, a window length of 500 milliseconds, and a step size of 250 milliseconds. Among the data windows, channels having a PC loading to be greater than 0.75 were removed. We removed segments with any abrupt change of voltage greater than 100μV, or an absolute acceleration magnitude larger than 1 ms^{-2}.

E. Partial Directed Coherence for connectivity analysis:

FC is defined as the statistical association among two or more anatomically distinct time-series and can be assessed with EEG coherence measures or fMRI [12]. C analysis was performed upon our EEG channels by computing the PDC measure [13] over six-second time segments [14] applied to a Hamming window function with 25% overlap, known as the short-time PDC (ST-PDC) measure [15]. Partial coherence measures have been found to perform well with lowdensity EEG [16]. The PDC was computed for all pairs of electrodes in the frequency bands (Fig. 2): delta [1-4Hz], theta [4-8Hz], alpha [8-12Hz], beta [12-30Hz], gamma [30-50Hz].

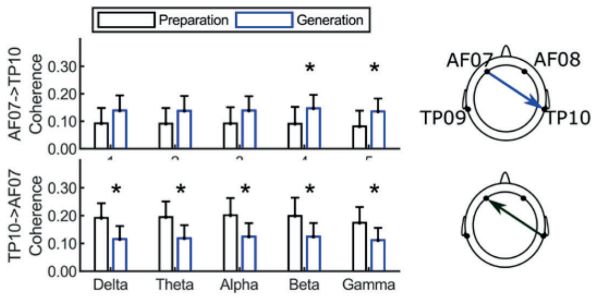


Figure 3. Directionality of PDC between the Preparation and Generation phases. There was higher PDC in the Preparation Phase at a significance level of $p < 0.05$, from TP10 to AF7 among all frequency bands analyzed: 1-50 Hz. The opposite directionality was found for the Generation phase, in frequency bands: 13-50 Hz.

3. RESULTS

The PDC between the Preparation and Generation of creative writing has opposite directionality between right temporal and left anterior frontal area. Fig. 3 illustrates the results for the connectivity between electrodes, using PDC, during the two stages of the creative writing process analyzed.

Preparation Phase: There was higher Partial Directed Coher-

ence in the Preparation Phase originating from TP10 towards AF7. The PDC difference between the Preparation and the Generation phases were statistically significant at a confidence level of $p<0.05$ for the frequency bands delta, theta, alpha, beta, and gamma. Fig. 2 shows the PDC scores, bounded between 0 and 1, for all pairs of electrodes.
Generation Phase: There was higher PDC in the Generation Phase originating from AF7 towards TP10. The PDC difference between the Preparation and the Generation phases were statistically significant at a confidence level of $p<0.05$ for the frequency bands beta and gamma.
The statistical difference in PDC and its opposite directionality when comparing the Preparation and the Generation Phases indicates that there was a strong relation between the left anterior frontal with the right temporal-parietal areas when the students engaged in the tasks.

4. DISCUSSION

The higher coherence values from the right temporal towards the left anterior frontal electrode during the Preparation phase, is potentially associated with the processing of sensory input [17], [18] and episodic emotional memory retrieval [19] in the temporal lobe as subjects explore their surroundings actively engaging the frontal cortex in integrating the experience. The opposite directionality between the same electrodes (Figure 3) reinforces this hypothesis in which processed input in the frontal areas is related back to sensory processes.
Our results show higher coherence values from the right temporal towards the left anterior frontal electrode during the Preparation phase for all frequency bands analyzed (1-50 Hz); and the opposite directionality for the Generation phase in higher frequencies (13-50 Hz). We did not find statistical differences between the Preparation and the Generation phases for Sample Entropy of frequency band power.
Overall, these findings suggest that ideation, exploration, and observation during the Preparation phase of a creative writing task can be characterized by a state of long-range cortico-cortical communication between multisensory integration brain areas (temporal regions) and high-order execution and planning areas of the brain (prefrontal regions), perhaps leading to selective storage of ideas, concepts or observations candidate for creating writing during the generation phase. We hypothesize this focal activity may be related to working memory, sequence production, and processing of filtered information from the Preparation Phase.
Information transfer from temporal-parietal to anterior-frontal areas of the scalp may reflect sensory interpretation during the Preparation phase, while high frequency bands PDC directionality originating at the anterior-frontal areas during the Generation phase may reflect the final decision making process to translate the sensory experience into a tangible product: text.

5. ACKNOWLEDGEMENTS

This research was supported by NSF award #BCS1533691, and IUCRC BRAIN award #1650536.

REFERENCES

[1] J. L. Contreras-Vidal, J. Cruz-Garza, and A. Kopteva, "Towards a whole body brain-machine interface system for decoding expressive movement intent Challenges and Opportunities," in *Brain-Computer Interface (BCI), 2017 5th International Winter Conference on*, 2017, pp. 1-4.
[2] C. Rivera, "Los muertos indóciles. Necroescrituras y desapropiación." Ciudad de México:

Tusquets, 2013.
[3] P. Lopate, *Portrait of my body*. Anchor Books, 1996.
[4] H. Takeuchi et al., "The association between resting functional connectivity and creativity," *Cereb. Cortex*, 2012.
[5] D. Wei, J. Yang, W. Li, K. Wang, Q. Zhang, and J. Qiu, "Increased resting functional connectivity of the medial prefrontal cortex in creativity by means of cognitive stimulation," *Cortex*, vol. 51, pp. 92-102, 2014.
[6] M. Lotze, K. Erhard, N. Neumann, S. B. Eickhoff, and R. Langner, "Neural correlates of verbal creativity: differences in resting-state functional connectivity associated with expertise in creative writing," *Front. Hum. Neurosci.*, vol. 8, p. 516, 2014.
[7] C. Shah, K. Erhard, H. J. Ortheil, E. Kaza, C. Kessler, and M. Lotze, "Neural correlates of creative writing: An fMRI Study," *Hum. Brain Mapp.*, 2013.
[8] K. Erhard, F. Kessler, N. Neumann, H. J. Ortheil, and M. Lotze, "Professional training in creative writing is associated with enhanced fronto-striatal activity in a literary text continuation task," *Neuroimage*, 2014.
[9] S. Liu *et al.*, "Brain activity and connectivity during poetry composition: Toward a multidimensional model of the creative process," *Hum. Brain Mapp.*, vol. 36, no. 9, pp. 3351-3372, Sep. 2015.
[10] C. Martindale and N. Hasenfus, "EEG differences as a function of creativity, stage of the creative process, and effort to be original," *Biol. Psychol.*, 1978.
[11] T. Mullen et al., "Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2013, pp. 2184-2187.
[12] K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. J. Frackowiak, "Functional connectivity: the principal-component analysis of large (PET) data sets," *J. Cereb. Blood Flow Metab.*, vol. 13, no. 1, pp. 5-14, 1993.
[13] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biol. Cybern.*, vol. 84, no. 6, pp. 463-474, 2001.
[14] M. Fraschini, M. Demuru, A. Crobe, F. Marrosu, C. J. Stam, and A. Hillebrand, "The effect of epoch length on estimated EEG functional connectivity and brain network organisation," *J. Neural Eng.*, vol. 13, no. 3, p. 36015, 2016.
[15] A. Omidvarnia, M. Mesbah, J. M. O'Toole, P. Colditz, and B. Boashash, "Analysis of the time-varying cortical neural connectivity in the newborn EEG: A time-frequency approach," in *International Workshop on Systems, Signal Processing and their Applications, WOSSPA*, 2011, pp. 179-182.
[16] E. Barzegaran and M. G. Knyazeva, "Functional connectivity analysis in EEG source space: The choice of method," *PLoS One*, vol. 12, no. 7, p. e0181105, 2017.
[17] C. Perrodin, C. Kayser, N. K. Logothetis, and C. I. Petkov, "Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices," *J. Neurosci.*, vol. 34, no. 7, pp. 2524-2537, 2014.
[18] A. C. Schapiro, E. Gregory, B. Landau, M. McCloskey, and N. B. Turk-Browne, "The necessity of the medial temporal lobe for statistical learning," *J. Cogn. Neurosci.*, vol. 26, no. 8, pp. 1736-1747, 2014.
[19] R. K. Lech and B. Suchan, "The medial temporal lobe: memory and beyond," *Behav. Brain Res.*, vol. 254, pp. 45-49, 2013.

How and What a Motivated-reinforcement-learning Theory of Aesthetic Values Learns

Norberto M. Grzywacz, Hassan Aleem

Abstract - We recently developed a neuroimaging-based computational theory for the learning of aesthetic values. This theory is based on the tenets of reinforcement learning with a modification to include the motivation to perform actions that lead to rewards. Another modification of the theory is that its sensory, reward, and motivation inputs are stochastic. We showmathematically that the theory learns at an optimal rate to minimize the error of aesthetic values.

Keywords: Aesthetic values, reinforcement learning, motivation, neuroaesthetics, computa-tional model, creativity.

1. INTRODUCTION

Aesthetic appraisal is central to both art production and art appreciation [1]. A recent meta-analysis of neuroimaging studies indicated that a small set of brain regions underlies aesthetic appraisal [2]. Almost all the regions uncovered by the meta-analysis participate in the processing of rewards, values (i.e., estimation of rewards), and reinforcement learning. To investigate the neurobiological mechanisms of aesthetical appraisal, we developed a theoretical framework based on these brain regions [3,4]. The core of the theory is the link between aesthetics and motivation-modulated reinforcement learning. The theory postulates that humans learn correlations between sensory inputs and possible rewards that one may get by choosing to perform certain actions. To take into account individuality, and social, cultural, and environmental variability, our theory makes probabilistic assumptions about the inputs, rewards, and motivations to act. Computer simulations have shown that the theory captures key aspects of the learning of aesthetic values. Here, we use mathematical analysis to probe how well the theory learns aesthetic values and how efficient the learning process is.

2. METHODS

A. Theory

We consider sensory inputs to be N dimensional, with components corresponding to variables used by the brain in its representation of the external world:

$$\vec{u}(t) = [u_1(t), u_2(t), \cdots, u_N(t)],$$

where the overhead arrow indicates a vector and t indicates time.

We assume a linear model for estimating reward, as is common in Reinforcement-learning models [5]. Thus, there exists a parameter vector of "synaptic" weights

$$\vec{w}(t) = [w_1(t), w(t), \cdots, w_N(t)],$$

such that the estimated reward is

$$v(t) = m(t)\, \vec{w}(t) \cdot \vec{u}(t), \qquad (1)$$

where $0 \leq m(t) \leq 1$ is the motivation function. If m were interpreted as the probability ofacting, then the received reward would be

$$r(t) = m(t)\, r^*(t), \tag{2}$$

where r^* would be the reward that a fully motivated person would get. Inserting m in the reward estimate (1) is necessary, because the reward varies with motivation (2).
Reinforcement-learning theories then use Temporal Difference [5] in their calculations:

$$\delta(t) = r(t) - v(t), \tag{3}$$

$$\frac{d\vec{w}(t)}{dt} = k\, \delta(t)\, \vec{u}(t), \tag{4}$$

where $k > 0$ is a constant. Inspection of (4) reveals that since motivation affects both v and r (as seen in (1) and (2)), it influences learning though the δ (3).
We assume that $\vec{u}$, m, and r^* are stochastic and thus, we must specify their statistical properties. Let us begin with the variables $\vec{u}$, and r^*. Both these variables come from the external world, but have different origins, with $\vec{u}$, being sensory, whereas r^* arising from performing an action. These variables are dependent and thus, possibly statistically correlated (the motivation to act depends on the sensory input). Hence, we must specify the probability density functions

$$P\left(\vec{I}_u \middle| \vec{B}\right),\ P\left(\left(\vec{u}(t), r^*(t)\right) \middle| \vec{I}_u\right), \tag{5}$$

where $\vec{B}$ is the vector of parameters characterizing the social and environmental background, and $\vec{I}_u$ is the vector of parameters characterizing an individual in this society.
Finally, we must specify the statistical properties of the motivation function m. Because this variable represents internal signals from the body related to motivation, m varies across individuals. However, the sensory input $\vec{u}$ also affects motivation.
If, say, a person is hungry, but the sensory input is not food, then the individual will not have motivation to act, that is, to eat.
The probability density functions of m is thus written as

$$P\left(\vec{I}_m \middle| \vec{B}\right),\ P\left(\left(m(t)\right) \middle| \vec{u}(t), \vec{I}_m\right), \tag{6}$$

where we insert $\vec{B}$ to indicate that individual motivation may depend on environmental and social backgrounds.

3. RESULTS

A. Optimization of the Learning of Value

Computer simulations show that during learning, the "synaptic" weights first change quickly, but they then appear to "converge" around a cloud of values [3]. Why does this apparently stable cloud form? A simple answer would be that the weights gravitate around a fixed point, but do not converge to it, because the inputs are stochastic. However, our mathematical analysis shows that the answer is more complex and interesting. The analysis first demonstrates

N. M. Grzywacz is with the Departments of Physics and Neuroscience, the Interdisciplinary Program in Neuroscience, and the Graduate.
School of Arts and Sciences, Georgetown University, Washington, DC 20057 USA (202-687-5671; e-mail: norberto@georgetown.edu).
H. Aleem is with the Interdisciplinary Program in Neuroscience, Georgetown University, Washington, DC 20057 USA (e-mail: ha438@georgetown.edu).

that the learning process implements a gradient-descent optimization of the prediction of reward. Thus, value approaches reward as much as possible in a statistical sense (Claim 1 below – compare the green and blue lines in Fig. 1). If the process were not stochastic, then value would converge exactly to the reward (Claim 2).

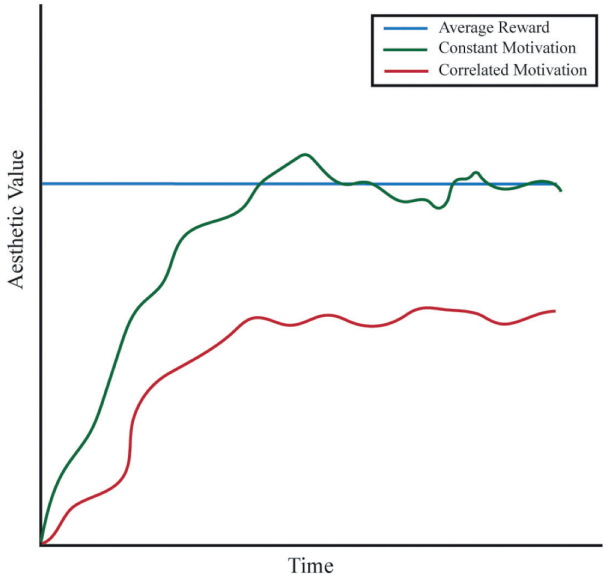


Figure 1. Schematics of Motivation-modulated Learning of Aesthetic Value. The example illustrated by the green line shows the temporal evolution of aesthetic value for a case in which the motivation function is constant. The aesthetic value "converges" statistically around the average reward (blue line). In turn, the example shown by the red line shows aesthetic-value progression for a case in which motivation is low most of the time, being only high when the sensory stimuli predict high rewards. Thus, aesthetic value undershoots average reward. This is optimal in the sense of minimization of (7), because most actions lead to less reward than average.

Claim 1: If for every τ there is a $t > \tau$ such that $m(t) > 0$, then the learning process minimizes

$$E(\vec{w}) = \langle m(t)\left(r^*(t) - \vec{w}(t)\cdot\vec{u}(t)\right)^2\rangle_t, \qquad (7)$$

where $\langle\ \ \rangle_t$ stands for time average.

Proof: The gradient of E with respect to the components of $\vec{w}$ is

$$\nabla_w E(\vec{w}) \propto -\langle m(t)\left(r^*(t) - \vec{w}(t)\cdot\vec{u}(t)\right)\vec{u}(t)\rangle_t,$$

or

$$\nabla_w E(\vec{w}) \propto -\langle \delta(t)\vec{u}(t)\rangle_t.$$

Therefore, the process governed by (4) minimizes $E\ (\vec{w})$ by performing a gradient descent [6]. The requirement that for every τ there is a $t > \tau$ such that $m(t) > 0$ is necessary to give the process enough time to reach optimization. If $m(t) = 0$ for every $t > \tau$ then the learning process freezes after τ as shown by (1), (2), (3), and (4).

Claim 2: If $\vec{u} \neq 0$, $m > 0$, r^* are constant, then

$$\lim_{t\to\infty} v(t) = r.$$

Proof: By combining (3) and (4), we get

$$\frac{d\vec{w}(t)}{dt} = k\left(r - v(t)\right)\vec{u} \qquad (8)$$

or

$$\frac{d\vec{w}(t)}{dt} = k\left(r - m\vec{w}(t)\cdot\vec{u}\right)\vec{u}. \qquad (9)$$

The existence of a stable fixed point depends on the properties of the second term of the right-hand side of (9) [7]. This term can be written [5] as

$$-m(\vec{w}(t)\cdot\vec{u})\vec{u} = -m\,U\,\vec{w}(t),$$

where U is the matrix operator

$$U = \begin{pmatrix} u_1u_1 & \cdots & u_1u_N \\ \vdots & \ddots & \vdots \\ u_Nu_1 & \cdots & u_Nu_N \end{pmatrix}.$$

Because this matrix is symmetric and positive definite [5], its eigenvalues are positive [7]. This shows after the multiplication by the negative $-m$ that the system has a stable fixed point [7]. This fixed point occurs when (8) converges to 0, that is, when $v(t)$ converges to r. The requirement that $\vec{u} \neq 0$ stems from the same reason why we need $m(t) \neq 0$. If either $\vec{u} = 0$ or $m(t) = 0$, then the learning process freezes.

B. Maximization of the Learning Rate

However, although value gravitates around a fixed point, the mathematical analysis shows that the weights do not. The same value can be obtained by multiple sets of weights (Claim 3 below). The analysis shows that this redundancy of weights has an important advantage, allowing the optimization of the learning rate (Claim 4 – Fig. 2).

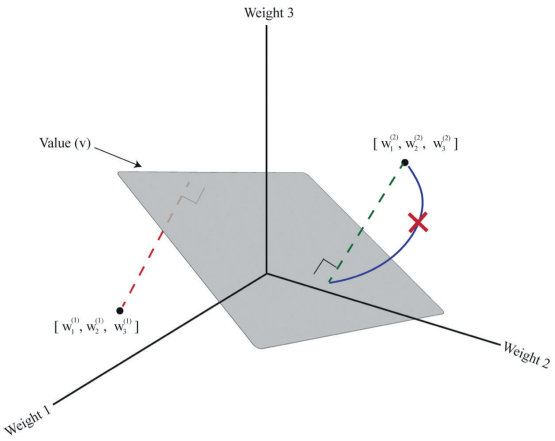


Figure 2. Schematics of Optimization of Learning Rate. In our theory, an aesthetic value corresponds to a plane as a function of aesthetic weights, as in (10). This plane is represented as the shaded area marked with value $\boldsymbol{v}$ in the figure. Consider an individual whose aesthetic weights $\boldsymbol{w_1^{(1)}}, \boldsymbol{w_2^{(1)}}, \boldsymbol{w_3^{(1)}}$ are such that the value is not $\boldsymbol{v}$. Let this person learn this value. Then, the learning trajectory is the shortest path to the plane. The same will happen to another individual with weights $\boldsymbol{w_1^{(2)}}, \boldsymbol{w_2^{(2)}}, \boldsymbol{w_3^{(2)}}$, whose value is not $\boldsymbol{v}$ (green line). Eventually, both individuals will learn the value $\boldsymbol{v}$, but their aesthetic weights will be different (i.e., they will be in different positions in the plane). That the trajectories are as short as possible mean that they tend to be perpendicular to the plane and not some curved trajectory (crossed-out blue line)

Claim 3: An infinite number of $\vec{w}(t)$ are compatible with any given $v(t)$.

Proof: We can rewrite (1) as follows

$$\sum_{i=1}^{N} a_i(t)\, w_i(t) = v(t), \qquad (10)$$

where the coefficients $a_i(t)$ are

$$a_i(t) = m(t)u_i(t).$$

The linear polynomial in (10) describes a time-dependent, N-1-dimensional hyperplane on the weight variables $w_i(t)$. Any point that belongs to this hyperplane is compatible with the value $v(t)$.

Claim 4: The trajectory of $\vec{w}(t)$ is the shortest possible towards achieving $v(t) \to r(t)$.

Proof: At time t, the $\vec{w}(t)$ are at a point of the weight space such that the value is $v(t) = m(t)\vec{w}(t) \cdot \vec{u}(t)$. However, the ideal value should be equal to the reward, that is, $r(t) = m(t)\vec{w}_I(t) \cdot \vec{u}(t)$, where $\vec{w}_I(t)$ are ideal weights. This equation describes a hyperplane over the variables $\vec{w}_I(t)$. All points in this hyperplane have an ideal value, namely, $v_I(t) = r(t)$. The shortest path from point $\vec{w}(t)$ to this ideal hyperplane is via a straight line perpendicular to the hyperplane passing through $\vec{w}$.
Vectors perpendicular to this hyperplane are parallel to $\vec{u}(t)$ [8]. Hence, the shortest path from point $\vec{w}(t)$ to the hyperplane must move in a direction parallel to $\vec{u}$. That is exactly how (4) moves the weights $\vec{w}(t)$.

4. DISCUSSION

The analysis of our theory shows that its outcome is an optimization of the aesthetic value through learning (compare the green and blue lines in Fig. 1). Thus, the system tends to learn minimizing the average error (7) in the estimation of reward given a sensory input. Strictly speaking, however, the system does not minimize this error but a version of it weighed by the statistics of motivation (7) (red line in Fig. 1). Consequently, if the inputs, rewards, or motivations are stochastic, the minimization is not absolute but statistical (Claim 1). Otherwise, the minimization of error in aesthetic value is absolute (Claim 2). Curiously, the aesthetic weights achieving the minimization are not unique (Claim 3). However, this lack of uniqueness is advantageous, because it leads to the maximization of the learning rate (Claim 4). If one thinks of aesthetic weights as synaptic weights, our results suggest an interesting learning function for the synaptic abundance in the brain [9]. The trajectory of populations of synapses would indicate optimal learning dynamics. Recording from each synapse in isolation would not teach us anything about its function.
According to our theory, the optimization of learning of aesthetics values is individualistic. The main reason for the lack of universality in this learning is the individuality of both mot vations (6), and exposures to sensory inputs and rewards (5). Elsewhere, we argue that this individuality is one of the roots of artistic creativity [3]. Therefore, artistic creativity may in part arise from a process of optimized stochastic learning. Each of us would learn aesthetic values optimally according to our experiences and motivated states. However, for a few of us, this learning process would take us to a unique region of the aesthetic space, making us creative.

5. ACKNOWLEDGMENTS

This research was supported in part by Georgetown University Provostial Funds to N.M.G.

REFERENCES

[1] H. Aleem, I. Correa-Herran, and N.M. Grzywacz, "Inferring master painters' esthetic biases from the statistics of portraits," *Frontiers in Human Neuroscience*, vol. 11, 94, 2017.

[2] S. Brown, X. Gao, L. Tisdelle, S. B. Eickhoff, and M. Liotti, "Naturalizing aesthetics: brain areas for aesthetic appraisal across sensory modalities," *Neuroimage*, vol. 58, pp. 250-258, 2011.

[3] H. Aleem, I. Correa-Herran, and N.M. Grzywacz, "A theory for how the brain learns aesthetic values and implications to creativity," *NeuroImage*, submitted for publication.

[4] H. Aleem, M. Pombo, I. Correa-Herran, and N.M. Grzywacz, "Is beauty in the eye of the beholder or an objective truth? a neuroscientific answer," in *Mobile Brain-Body Imaging and the Neuroscience of Art, Innovation and Creativity*, J. Contreras-Vidal, D. Robleto, J. G. Cruz-Garza, J. M. Azorin, and C. S. Nam Eds. Cham, Switzerland: Springer, 2019.

[5] P. Peter, and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: MIT Press, 2001.

[6] J. A. Snyman, and D. N. Wilke, *Practical Mathematical Optimization: Basic Optimization Theory and Gradient-based Algorithms*, vol. 133. Cham, Switzerland: Springer, 2018.

[7] M. W. Hirsch, S. Smale, and R. L. Devaney, *Differential Equations, Dynamical Systems and an Introduction to Chaos*. San Diego, CA: Academic Press/Elsevier, 2004.

[8] T. Shifrin, and M. Adams, *Linear Algebra: A Geometric Approach*, 2nd ed. New York, NY: W. H. Freeman, 2011.

[9] D. A. Drachman, "Do we have brain to spare?" *Neurology*, vol. 64, 2004-2005, 2005.

Body is Other:
Mediations and Translations in Creative Writing and Neuroaesthetics

Maria J. Delgadillo, Jesús G. Cruz Garza, Cristina Rivera Garza

Abstract - Through the use of MoBI and EEG, and in a collaborative setting for working, we're creating a new approach to creative writing, which incorporates data, transforming it in different outputs. This process of transforming the EEG recordings resulting of these creative practices, and working with them through different mediums is regarded as a translation, as it takes the language of EEG recordings, and by transforming it, makes it readily accessible to a wide array of audiences.

1. INTRODUCTION

We live in a world where our consumption of art, media, and communication is increasingly visual and mediated through technology which has allowed every person who possess a computer or mobile device to become a writer/author and speak, in their own words, about their realities and relationships. With the evolution of technology, the gatekeeping of information is no longer limited by geographical or linguistic borders. Because of this shift on the notion of authorship and access, both the exercises of reading and creative writing call for new understanding of how social media, blogs, images, e-books and so on, work, and what language can and cannot access (Rivera Garza, 2013 p. 22). With this background, and following the idea of artist and performer Laurie Anderson that "language is an elaborate code" and that every element of art and performance is a type of language that can be decoded and translated (Anderson, 2018, pps. 10-11) we felt compelled to find new ways of approaching the creative practices of writing, and test the limits of what a text or a written piece can be understood as. In this sense, we began by thinking of translation not only as a shift from one language to another, but as "displacement, invention, mediation, the creation of a link that didn't exist before and to some degree modifies two elements or agents" (Latour, 1994 p. 32), a shift between the many codes that compose our interactions. And if there has already been a push from biology to create a "biotranslation", that is, a system of translating codes that have no syntax but that can be distinguishable "transmission between umwelten" (Kull, Torop, 2003 pps. 318-320); then by the introduction of Mobile Brain-Body Imaging (MoBI) technologies such as mobile Electroencephalogram (EEG) and collaborative work with the I/UCRC BRAIN Center at the University of Houston, our approach to writing creative pieces was challenged and shifted into something far more complex than text on a page. Aligned with John Vincler, on his work reviewing Renee Gladman's *Prose Architectures*, where he asks "what are we reading or seeing when moving through books of writing containing only gesture and abstraction? What does it mean to write free from language?" (Vincler, 2018), the collaborative work began to question how language can be read and how writing could be decoded if we considered the recordings of the brain as a central part of the creative process.

By this, we are trying to translate, into an array of outputs, the moments of creative work that lie hidden behind the concept of inspiration. The unreadable pieces of data that are the very pillars of human creativity: movement, intent, language. The process involves several steps, beginning at the initial moment of composing text in diverse genres such as poetry, translation, and fiction. This writing is done while wearing an EEG cap which records the

brain activity; and will be used afterwards, after a serial processes of editions and translations, allowing this process to become a complex network of creative pieces connected by the concepts of mediation between the body that creates 1 and the machines that help in this creation. With the use of an algorithm, the EEG recordings resulting of this sessions of writing are transformed into sound which will also become subject of a series of editions –much as we do in the creative writing field–, to become melodies that will musicalize the written results. The recordings will also become part of videos and hand made embroidery, which will act as pieces of text narrating the experience, as much as the written word does, becoming pieces of code to be translated by the reader.

This multimedia collaborative project is titled Body is Other, and looks to question what can literature do and how it can benefit from performing along EEG recordings. Beyond being a collaboration that stays in the realms of what can be achieved aesthetically, the project has a unique urgency to understand reading and writing as a series of mediations and translation of codes: the written word, the brain signals, the movements of the body, sounds, images, and the intertextuality between them.

The work has also allowed us to generate new questions and pose new possibilities for the work and the advancement and development of a relationship between neuroscience and writing. By proposing that these recorded brain activity generates a language that speaks for itself, we are trying to become translators from data to art practice, nurturing both the advancement of the research on neuroaesthetics and creativity. We believe that the project advances into more than just written text, rather, it becomes literature that is highly imbued in the moments when it's being created. And, more than anything, it stands along Donna Haraway's Cyborg Manifesto in proposing that the limits of our skin, particularly when trying to create a space where an embodied experience of writing literature is exposed, are not the limits of our body. (Haraway, 1991, p. 178).

2. METHODS

Recording equipment

We collected brain activity, eye movement, heart beat, and skin conductance was measured together using a 64-channel ActiCHamp recording system (Brain Products GmbH, Germany) with Ag/AgCl active electrodes. The electrodes in the recording system were used for non-invasive electroencephalography (EEG) to record electrical activity from the brain at the surface of the scalp; electro-oculography sensors (EOG) to record eye motion; and an electrocardiogram sensor (EKG) to record their heartbeat. The system integrated a galvanic sensor response (GSR) unit to record skin conductance and a tri-axial head accelerometer to record head motion. The data was recorded at 500 Hz. The 64 electrodes were distributed as: 58 for EEG, 4 for EOG, and 2 for EKG.

A conductive electrolyte gel was applied at the electrode tips to reduce the interface impedance between the scalp and the electrodes. The biometric data were recorded using the BrainVision Recorder software (Brain Products GmbH, Germany). The 58 EEG channels were arranged according to the international 10–20 system.

Sonification

We generated the audio by mapping EEG power in some of the electrodes in the frontal, central, parietal, and occipital regions of the scalp. The power in the delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (30-50 Hz) frequency bands was mapped

to the amplitude of sine waves in the C minor 9th chord (C3, D#2, G3, B3, D4) pentatonic scale, with frequencies at 128.43, 152.74, 192.43, 242.45, 288.33 Hz, respectively to each frequency band.
Figure 1 shows the original EEG frequency activations in a spectrogram, and the mapping to the corresponding sinewaves produced from the power modulations in channel F2.

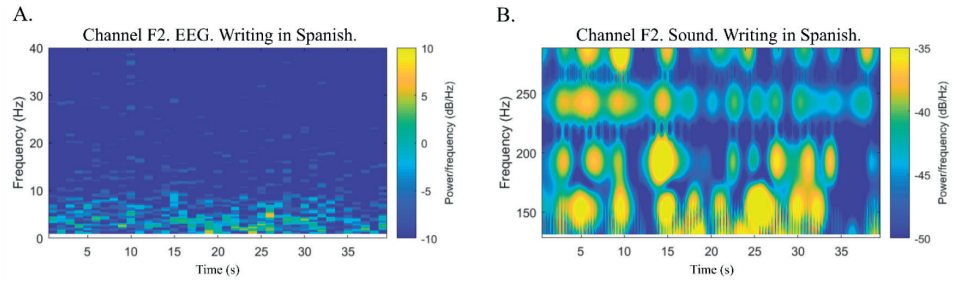


Figure 1. Spectrogram of the EEG data in F2 sonified as a mixture of sine waves. A. EEG activation of the F2 channel across the whole frequency spectrum analyzed. B. Spectrogram of the sine waves used to generate sound from the EEG channel.

3. RESULTS

Body is Other, as a whole, is a project that tries to understand creative writing as a process far beyond written text on a page. Echoing what Ulises Carrión proposed in The New Art of Making Books, "when writing the text, the writer followed only the sequential laws of language, which are not the sequential laws of books.[...] In the new art every page is different; every page is an individualized element of a structure (the book)wherein it has a particular function to fulfill", the project is trying to create a new kind of book and a new kind of reading experience that embodies, for the audience, the core of human creativity.
While based on the act of writing, the work we're creating is intending to question our relationships with language, the bridges we create between signifier and signified, the moments of inspiration, and the levels of separation between the bodies that create and the devices those bodies use in the acts of performativity.
In this sense, we're trying to emulate Anderson in her approach to art, making sure that we put language into art through all different possible media (Anderson, 2018, p.11). That is, the text, the melodies that are created through the EEG recordings, and the many layers of translation processes of data into visual arts are looking to perform, with the audience, an experience of reading that reaches beyond books, words, pages, experiences. By allowing for the audience to create their own stories and relationships with the piece.
Figure 2 show the early results of experimenting with the materials for the project.

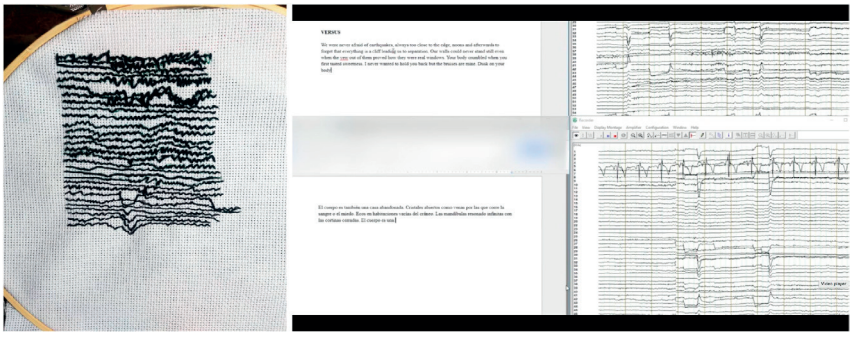


Figure 2: Handmade embroidery work for Body is Other and screenshot of a video made featuring both EEG recordings, bilingual creative writing in real time, and sonifications edited to become melodies.

REFERENCES

Books

Anderson, Laurie. *Things I Lost in the Flood* . First. New York, NY: Rizzoli Electa, 2018. Rivera Garza, Cristina. *Los Muertos Indóciles* . First. México: Tusquets, 2013.

Articles

Carrión, Ulises. "The New Art of Making Books | Poetry | Reality." *Kontexts* , no. 6-7, Maastricht, (1975). Kalevi Kull, and Peeter Torop. "Biotranslation: Translation between Umwelten." In *Translation, Translation* . Amsterdam - New York, NY: Rodopi, 2003. Latour, Bruno. "On Technical Mediation – Philosophy, Sociology, Genealogy." *Common Knowledge* 3, no. 2 (1994): 29-64.

Online Resources

Haraway, Donna. "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century." In *Simians, Cyborgs and Women: The Reinvention of Nature*. , 149-81. New York: Routledge, 1991. http://www.earthwidemoth.com/mt/archives/001923.html. Vincler, John. "Dwelling Places: On Renee Gladman's Turn to Drawing." The Paris Review (blog), August 28, 2018. https://www.theparisreview.org/blog/2018/08/28/dwelling-places-on-renee-gladmans-turn-to-drawing/.

Chapter 2

Noninvasive Brain-Machine-Interfaces:
From Basic to Clinical Applications

Feasibility and safety of EEG/EOG-based bilateral exoskeleton control to restore hand motor function

Marius Nann, Niels Peekhaus, Surjo R. Soekadar

Abstract - Cervical spinal cord injuries (SCIs) often lead to quadriplegia with loss of motor function in both hands and legs affecting autonomy and quality of life.While it was shown that unilateral hand function can be restored after SCI using a hybrid electroencephalography/electrooculography (EEG/EOG)-controlled hand exoskeleton, feasibility of restoring bilateral hand function using such paradigm, e.g., to eat with fork and knife, was not demonstrated yet. To test whether EEG/EOG signals allow for fluent and reliable as well as safe and user-friendly bilateral control of two hand exoskeletons, eight healthy participants (6 female, mean age 24±3 years) performed an audio-visual neurofeedback paradigm consisting of a complex sequence of bilateral grasping and releasing motions visualized on a screen. While grasping motions were controlled by voluntary EEG modulations related to motor imagery of left- or right-hand closing movements, EOG signals related to horizontal eye movements (HOVs) were translated into releasing motions as well as commands to interrupt or switch laterality. Fluent and reliable EEG control was defined as average 'time to initialize' (TTI) grasping motions below 4s with complete closing motions of at least 75% within 6s. Safety was defined as 'time to stop' (TTS) an ongoing grasping motion by performing a HOV within 1s. After the experiment, all participants were asked to rate the user-friendliness of the EOG control using Likert-scales. While average TTI of EEG-controlled operations ranged at 3.3±2.2s, 75% successful closings were reached at 5.3s. Safety requirements were met by an average TTS of 0.97±0.3s. All out of one participant rated the use of the system as very user-friendly (89±13% of maximum score). These results suggest that an EEG/EOG-based bilateral exoskeleton control is feasible and safe, paving the way to test such paradigm in patient populations performing bimanual tasks.

Keywords: brain-machine interface, BMI, bilateral exoskeleton control, bimanual tasks, EEG, EOG.

1. INTRODUCTION

Traumatic cervical spinal cord injuries (SCIs) often results in loss of motor function in all four extremities. Such severeparalysis in both arms and legs termed quadriplegia is accompanied by substantial constraints in autonomy and quality of life [1].While the inability ofwalking is usually compensated by using a wheelchair, restoration of hand function is still insufficiently solved and, thus, limits the autonomy of quadriplegics in conducting basal activities of daily living, e.g., to pursue grooming, cooking or eating. Depending on the specific SCI location, the degree of impairment and related motor inabilities can vary substantially. Approx. 30% of SCIs are located at the lower cervical vertebrae (C5-C8) and result in incomplete quadriplegia with residual motor capabilities [2]. In particular, injuries between C5 and C7 are characterized by remaining motor skills in the shoulder and arm, but lack of motor function in wrist and fingers [3]. For such specific cases, adding up to 2.500 to 4.500 newly affected persons every year in the U.S. [4], the restoration of hand function would be a substantial improvement to regain autonomy in daily life.

Brain machine interfaces (BMIs) have become a promising approach to restore motor function by translating voluntary modulation of electroencephalography (EEG) sensorimotor-rhythms (SMR, 8-15Hz) into control signals, e.g. for a hand exoskeleton [5]. Beside motor restoration, repeated BMI-use in incomplete quadriplegia has additionally the potential to trigger neural recovery [6]. As control feature, event-related desynchronization (ERD) of SMR (SMR-ERD), associated with motor imagery or the attempt to move the paralyzed fingers, is commonly used. However, the inherent non-stationarity and low signal-to-noise ratio especially of non-invasive EEG makes BMI control susceptible to false classifications and, thus, limits reliability inferring unacceptable safety issues especially in unpredictable real-life environments. To improve BMI control, electrooculography (EOG) - based maximal horizontal oculoversions (HOVs), eye movements that – on the contrary to vertical eye movements – rarely occur in everyday life routines, were integrated as a further control signal to reduce false classifications to a minimum [7]. Recently, it has been demonstrated that in case of incomplete quadriplegia a hybrid EEG/EOG-controlled hand exoskeleton restores unilateral hand function, e.g. allowing to use cutlery for eating [8].While exoskeleton closing motions were controlled by SMR-ERD related to intended grasping movements, HOVs were translated into releasing motions or commands to interrupt unintended closing motions due to wrong classification resulting in increased reliability and safety in daily-life scenarios.

However, up to now, only unilateral BMI-based restoration of motor function was achieved, i.e., only one hand has been mobilized for grasping. While this constitutes a significant improvement for hemiplegics, e.g. after stroke, in case of quadriplegia unilateral systems provide only limited advantage since most activities of daily living, e.g. eating a full meal with fork and knife, require bimanual object manipulation. Therefore, an EEG/EOG-based bilateral control paradigm for quadriplegics would be desirable.

For unilateral restoration of hand function, it has been shown that Laplace-filtered SMR-ERD from the hand knob area of the contralateral primary motor cortex allows for specific detection of hand movement intention [9, 10]. For the implementation of a bilateral control paradigm, an option would be to evaluate the respective contralateral SMR-ERD from each hemisphere independently. However, this would be very unspecific in terms of laterality since imagination of hand movement or the attempt to move the paralyzed hand elicits not only contralateral but also ipsilateral SMR-ERD [11]. Even though contralateral modulations are usually more distinct, reliable bilateral BMI-classification either of left- or right-hand grasping intentions remains a challenge. Although the achieved classification accuracy has been shown to be high for offline analyses (around 70-80%), online detection methods do not achieve such classification rates and, thus, are not sufficiently reliable for daily live applications.

To still allow for reliable and safe BMI-based bilateral hand exoskeleton control, a novel paradigm was implemented enabling only one hand exoskeleton at once to be controlled by contralateral SMR-ERD while ipsilateral modulations remains unevaluated. Such a mechanism ensures that false classifications due to motor-related bilateral brain activation is eliminated. To switch laterality, prolonged HOV, defined as left or right HOV exceeding 1s, enables the respective exoskeleton. To test feasibility and safety of such novel EEG/EOG-based

Marius Nann, Niels Peekhaus and Surjo R. Soekadar are with the University Hospital of Tübingen, Tübingen, Germany. Surjo R. Soekadar is additionally with the Charité – Universitätsmedizin Berlin, Berlin, Germany (phone: +49-30-450-528 153; e-mail: surjo@soekadar.com).

bilateral exoskeleton control, eight healthy participants performed an audio-visual neurofeedback paradigm consisting of a complex sequence of bilateral grasping and releasing motions of two virtual exoskeletons visualized on a screen. Feasibility was assessed in terms of fluency, defined as average 'time to initialize' (TTI) EEG-controlled operations, as well as reliability, defined as successful EEG-based closing motions of at least 75%. Safety was evaluated on 'time to stop' (TTS) unintended closing motions by using EOG. Moreover, user-friendliness of prolonged versus short HOV was assessed by a Likert-scale.

2. METHODS

A. Participants

Eight healthy participants (6 females, age 24.1±3.2 years) were invited to a single-session experiment at the University Hospital of Tübingen. Before entering the study, all participants provided written informed consent. The study protocol complied with the Declaration of Helsinki and was approved by University of Tübingen's local ethics committee.

B. BMI system and bilateral control commands

EEG was recorded from nine conventional recording sites (F3, T3, C3, P3, F4, T4, C4, P4, Cz according to the international 10/20 system). Two additional electrodes were placed at the left and right outer canthus for EOG recordings. Reference and ground electrode were placed at FCZ and FpZ, respectively. All biosignals were sampled at 1kHz and amplified by using a wireless active-electrode EEG system (actiCAP®, LiveAmp®, Brain Products GmbH, Gilching, Germany) (Fig. 1). For online processing and classification, the BCI2000 software platform was used. EEG signals were first bandpass-filtered at 1-30Hz to remove drifts and high frequency noise. Afterwards, surface Laplacian filters were applied to increase signal-to-noise ratio of the targeted electrodes at C3 and C4, respectively [9]. A surface Laplacian filter was shown to be effective in detecting specific SMR-ERD while suppressing distant sources (e.g. eye blinks) without the need of complex models, e.g. accounting for volume conduction. To compute SMR-ERD related to right- or left-hand motor imagery, the power method by Pfurtscheller and Lopez da Silva was applied [12]. In order to attenuate eye-blinks, bipolar EOG signal was generated by subtracting left from right EOG. To remove low-frequency drifts as well as high frequency noise, bipolar EOG signal was bandpass-filtered at 0.02-3Hz.

Four EEG/EOG-based control commands allowed for bilateral control of two virtual exoskeletons presented on a screen in front of the participants (Fig. 1). While grasping motions were controlled by voluntary SMR modulations related to motor imagery of left- or right-hand closing movements, EOG signals related to horizontal eye movements (HOVs) were translated into releasing motions as well as commands to interrupt an ongoing closing motion (both short HOVs) or to switch laterality (prolonged HOVs > 1s). To enhance EOG control, auditory feedback was provided to confirm successful execution of short as well as for prolonged HOVs. A detailed overview on the BMI's bilateral control commands and their corresponding EEG/EOG-based triggers is summarized in Table I.

TABLE I. BMI Control Commands

BMI control commands	*EEG/EOG trigger*
Close virtual hand exoskeleton	SMR-ERD at contralateral motor cortex
Open virtual hand exoskeleton	Short HOV towards direction of activated virtual hand exoskeleton
Stop ongoing closing motion	Short HOV towards any direction
Switch active exoskeleton	Prolonged HOV (> 1s) towards desired virtual hand exoskeleton

B. BMI calibration and study protocol

After the EEG/EOG recording system was mounted, participants were instructed how to perform HOVs as well as imagination of left and right grasping motions. For EOG calibration, participants performed 5 short and 3 prolonged HOVs to each side, respectively, to compute HOV detection thresholds for each movement direction at two thirds of maximum HOVs. EEG calibration was performed according to Soekadar, Witkowski [8].

After successful calibration and familiarization with the bilateral control paradigm, participants performed twice a pseudo-randomized sequence of 35 sub-tasks consisting of all four control commands required for bimanual control (see Table I). Each sequence lasted approximately 5 minutes. The sequence included sub-tasks to either close a virtual exoskeleton (SMR-ERD), to open a closed virtual exoskeleton or to stop immediately an ongoing closing motion to simulate unintended hand exoskeleton motions or unexpected incidents (both requiring short HOV). In case a sub-task required to switch laterality, e.g. when an instruction to close left exoskeleton is given but right exoskeleton is still enabled, participants needed to first independently perform a prolonged HOV to the respective side followed by left or right motor imagination to close the virtual exoskeleton. The time between sub-tasks varied randomly between 5-7s. In case no valid SMR-ERD was elicited, sub-task was aborted after 10s. At the end of the session participants rated the user-friendliness of EOG control by using a five-level Likert-scale questionnaire. Since user-friendliness of hybrid ('short') EOG/EEG-control was already proven [8], the questionnaire focused on the newly implemented prolonged EOG command.

C. Outcome measures

'Time to initialize' (TTI) EEG-controlled operations were evaluated as time between visual indication of task instruction and SMR-ERD exceeding detection threshold. Fluent control was assumed when average TTI ranged below 4s. Even though 3s were reasoned to be conceived as fluent operation in a previous work [13], we extended the accepted TTI threshold by one second due to the more complex control paradigm, i.e. switching laterality before initializing closing motion, requiring high concentration. Reliable EEG control was defined when at least 75% successful, i.e. complete, closing motions were performed within 6s. Since a full closing motion lasted 2s in the case SMR-ERD detection threshold was continuously exceeded, this is an appropriate duration when maximal TTI of 4s is considered. Safety was assumed when average TTS was below 1s meaning that closing motion was interrupted before virtual exoskeletons were half closed. This value was chosen as most objects of daily life are smaller than the half of a full hand span. Moreover, user-friendliness was met when the majority of all participants rated EOG control as comfortable and easy to apply.

3. RESULTS

A. Feasibility

The average TTI (±s.d.) of all EEG-controlled closing operations ranged at 3.3±2.2s documenting fluent bilateral hand exoskeleton control (Fig. 2 left). Moreover, reliable control was achieved, since 75% of all left and right closing motions were successfully performed within 5.3s.

B. Safety

The participants were able to stop an ongoing closing motion of the virtual exoskeletons by a short HOV with an average TTS (±s.d.) of 0.90±0.33s (Fig. 2 right). Thus, safety requirements were met as TTS ranged below 1s.

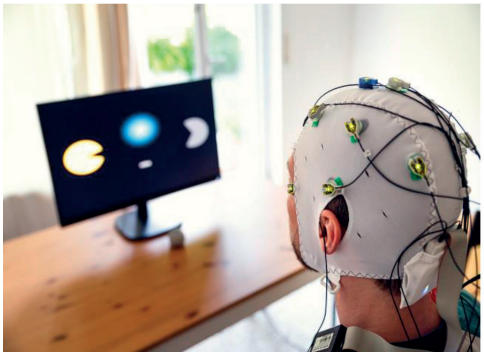


Figure 1: Experimental set-up. Participants were equipped with a wireless active-electrode EEG/EOG system (actiCAP®, LiveAmp®, Brain Products GmbH, Gilching, Germany) and comfortably seated in front of a screen receiving visual and auditory feedback.

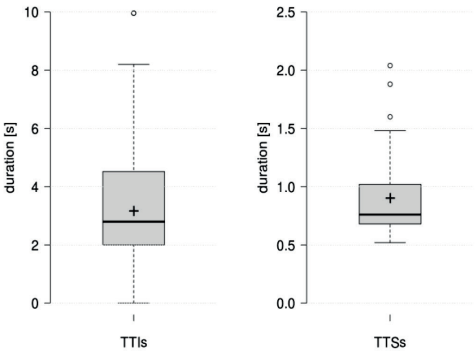


Figure 2: Left: 'Time to initialize' (TTI) EEG-controlled closing motions of left or right virtual hand exoskeleton. Average TTI ranged below 4s documenting fluent control. Right: 'Time to stop' (TTS) an ongoing closing motion by using short HOVs. Since average TTS is below 1s, safety requirements were met. Centerlines of boxplot show median, while crosses show the mean. Box limits indicate the 25th and 75th percentiles.

C. User-friendliness

With 89±13% of the maximum achievable score, participants rated EOG control as user-friendly. Only one person reported discomfort while performing HOVs.

4. DISCUSSION

The presented study demonstrates feasibility and safety of a novel EEG/EOG-based paradigm for bilateral control of two hand exoskeletons. We showed that eight healthy participants were able to perform a complex sequence of sub-tasks with four EEG/EOG-based control commands (Table I). Fluent and reliable control was documented by an average TTI (3.3±2.2s) of all EEG-controlled operations below 4s and by achieving 75% of successful closing motions (reached at 5.3s) within 6s. These results are comparable to a previous study, in which a whole-arm exoskeleton was controlled [13]. Moreover, average TTS (0.90±0.33s) below 1s substantiates safety. All out of one participant reported no discomfort and ease of use in performing HOVs. These results show for the first time that a bilateral control based on EEG/EOG signals is feasible and safe and can be applied directly after a short familiarization without long training. The motivation was to investigate whether restoration of bilateral motor function in quadriplegics, e.g. allowing them to eat with knife and fork by using two hand

exoskeletons, is possible. Regaining motor function in both hands would significantly increase their autonomy and quality of life.
The here presented EOG control paradigm consisting of short and prolonged HOVs required only a short preparation (only two electrodes at the outer canti) and calibration phase, and was reliably used by the participants. Since EEG control is generally more effortful, one could argue that also the sub-task of exoskeleton closing could be performed with a short HOV resulting in a faster overall control as well as less preparation and calibration time. However, contrary to eye movements, EEG-based control was shown to be very intuitive since exoskeleton closing is directly linked to the attempt to move the paralyzed fingers [8]. Moreover, there is increasing evidence that, in case of incomplete quadriplegia, EEG-controlled movement initiation can trigger neural recovery [6, 14]. Both aspects, the intuitiveness of grasping as well as the potential to trigger neurorestoration justifies the additional cost and effort of EEG recordings.
To reliably detect prolonged HOVs, introduced in this study as additional EOG control command to switch laterality between the two exoskeletons, bipolar EOG signal needed to contain low-frequency information requiring specific bandpass-filtering (lower cutoff frequency at 0.02Hz). However, low frequencies are prone to be susceptible to movement artefacts, e.g. head movements, as well as settling time of such filters is relatively long. While under lab condition this issue could be controlled, EOG control needs to be tested during daily life routines in possibly noisy and unpredictable environments. A possible solution could be to use other EOG-related signal features decreasing dependency on low-frequency information content.
Also, an even more fluent bilateral control as demonstrated here would be desirable. In this context, investigating whether lateralization of motor-related brain activation can be trained over several sessions allowing for higher classification accuracies would be important. Since it was showed that SMR-ERDs are generally more pronounced over the contralateral hemisphere [11], it might be possible to determine the side of the intended movement by assessing such lateralized activity only.
Our results suggest that an EEG/EOG-based bilateral exoskeleton control is feasible and safe, paving the way to test such paradigm in patient population performing bimanual tasks in everyday life environments.

5. ACKNOWLEDGMENTS

This research was supported by the Baden-Württemberg Stiftung (NUE007/1) and the European Research Council (ERC-2017-STG-759370).

REFERENCES

[1] Campbell, M.L., D. Sheets, and P.S. Strong, *Secondary health conditions among middle-aged individuals with chronic physical disabilities: implications for unmet needs for services*. Assist Technol, 1999. 11(2): p. 105-22.

[2] Devivo, M.J., *Epidemiology of traumatic spinal cord injury: trends and future implications.* Spinal Cord, 2012. 50(5): p. 365-72.

[3] Ahuja, C.S., et al., *Traumatic spinal cord injury*. Nat Rev Dis Primers, 2017. 3: p. 17018.

[4] DeVivo, M.J., B.K. Go, and A.B. Jackson, *Overview of the national spinal cord injury statistical center database*. J Spinal Cord Med, 2002. 25(4): p. 335-8.

[5] Wolpaw, J.R., et al., *Brain-computer interfaces for communication and control*. Clin Neurophysiol, 2002. 113(6): p. 767-91.

[6] Donati, A.R., et al., *Long-Term Training with a Brain-Machine Interface-Based Gait Protocol Induces Partial Neurological Recovery in Paraplegic Patients*. Sci Rep, 2016. 6: p. 30383.

[7] Soekadar, S.R., et al., *An EEG/EOG-based hybrid brain-neural computer interaction (BNCI) system to control an exoskeleton for the paralyzed hand*. Biomed Tech (Berl), 2015. 60(3): p. 199-205.

[8] Soekadar, S.R., et al., *Hybrid EEG/EOG-based brain/neural hand exoskeleton restores fully independent daily living activities after quadriplegia*. Science Robotics, 2016. 1(1).

[9] McFarland, D.J., *The advantages of the surface Laplacian in brain-computer interface research*. Int J Psychophysiol, 2015. 97(3): p. 271-6.

[10] Neuper, C., M. Wortz, and G. Pfurtscheller, *ERD/ERS patterns reflecting sensorimotor activation and deactivation*. Prog Brain Res, 2006. 159: p. 211-22.

[11] Nikulin, V.V., et al., *Quasi-movements: a novel motor-cognitive phenomenon*. Neuropsychologia, 2008. 46(2): p. 727-42.

[12] Pfurtscheller, G. and F.H. Lopes da Silva, *Event-related EEG/MEG synchronization and desynchronization: basic principles*. Clin Neurophysiol, 1999. 110(11): p. 1842-57.

[13] Crea, S., et al., *Feasibility and safety of shared EEG/EOG and vision-guided autonomous whole-arm exoskeleton control to perform activities of daily living*. Sci Rep, 2018. 8(1): p. 10823.

[14] Wagner, F.B., et al., *Targeted neurotechnology restores walking in humans with spinal cord injury*. Nature, 2018. 563(7729): p. 65-71.

Preliminary Assessment of hands motor imagery in theta- and beta-bands for Brain-Machine-Interfaces using functionalconnectivity analysis

J. A. Gaxiola-Tirado*, Eduardo Iáñez, Mario Ortiz, D. Gutiérrez, José M. Azorín

Abstract - The use of time- and frequency-based features has proven effective in the process of classifying mental tasks in Brain Computer Interfaces (BCIs). Still, most of those methods provide little insight about the underlying brain activity and functions. Thus, a better understanding of the mechanisms and dynamics of brain activity, is necessary in order to obtain useful and informative features for BCIs. In the present study, the objective is to investigate the differences in functional connectivity of two motor imagery tasks, through a partial directed coherence (PDC) analysis, which is a frequency-domain metric that provides information about directionality in the interaction between signals recorded at different channels. Four healthy subjects participated in this study, two mental tasks were evaluated: Imagination of the movement of the right hand or left hand. We carry out the differentiation of these tasks through two different approaches: on one hand, the traditional one based on spectral power; on the other hand, an approach based on PDC. The results showed that EEG-based PDC analysis provides additional information and it can potentially improve the feature selection mainly in the beta frequency band.
Keywords: PDC, brain connectivity, motor imagery, hand MI, BCI.

1. INTRODUCTION

A brain-computer interface (BCI) is a communication system that allows the user to communicate with the external world through their brain's activity without the assistance of peripheral nerves and muscles [1]. The generation of successful control commands in these systems is achieved through five consecutive steps: signal acquisition, preprocessing, feature extraction, classification and the association with a control command [2]. The key problem in the analysis of EEG signals is the feature extraction process. In this framework, various techniques such as time-domain analysis, power spectral estimation, and wavelet transform have been investigated for channel selection in the processing of EEG signals.

Most training methods for non-invasive electroencephalogram (EEG)-based BCI systems involve performing a particular cognitive task, such as motor imagery (MI). In this context, classification of different mental tasks such as left/right hand MI, left-right leg IM or cognitive tasks have been investigated [3]. MI causes distinctive patterns in the electrical activity of the sensory-motor cortex, mainly in alpha (8–12 Hz) and beta (13–30 Hz) frequency bands. Then, mental tasks differentiation has been employed to control systems such as wheelchairs or a planar robot [4].

In this paper, we present the differentiation of left/right hand MI based on traditional approach based in EEG power. Then, we introduce a complementary study of the brain connectivity presented in these two tasks. Thus, a better understanding of the mechanisms and dynamics of brain activity is useful in order to obtain informative features for BCI. Therefore, we use Partial Directed Coherence (PDC) analysis, which is a frequency domain metric that provides information about directionality in the interaction between EEG signals recorded at different channels.

2. METHODS

A. Partial Directed Coherence

The partial directed coherence (PDC) is a frequency domain measure of the relationships (information about directionality in the interaction) between pairs of signals in a multivariate data set for application in functional connectivity inference in neuroscience [5]. If one assumes a set $S = \{x_m(n), 1 \le m \le M\}$ of M EEG signals (simultaneously observed time series), is adequately represented by a multivariate autoregressive (MVAR) model of order *p*, or simply MVAR(*p*):

$$\mathbf{x}(n) = \sum_{k=1}^{p} \mathbf{A}_k \mathbf{x}(n-k) + \mathbf{e}(n) \tag{2}$$

Where $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_p$ are the coefficient matrices (dimensions M×M), containing the coefficients $a_{ii}(k)$ which represent the linear interaction effect of $x_j(n-k)$ onto $x_i(n)$ and $\mathbf{e}(n) = [e_1(n), e_2(n), \ldots, e_M(n)$, is the noise vector (uncorrelated error process). A measure of the direct causal relations (directional connectivity) of xj to xi is given by the PDC defined by [5]

$$\pi_{i \leftarrow j} = \frac{A_{ij}(f)}{\sqrt{\mathbf{a}_j(f)\mathbf{a}_j^T(f)}}, \tag{3}$$

where $A_{ij}(f)$ and $\mathbf{a}_j$ are, respectively, the *i,j* element and the *j*-th column of

$$\boldsymbol{A}(f) = \boldsymbol{I} - \sum_{k=1}^{p} \boldsymbol{A}_k \, e^{-2i\pi fk} \, . \tag{4}$$

PDC values range between 0 and 1; $\pi_{i \leftarrow j}$ measures the outflow of information from channel xj to xi in relation to the total outflow of information from xj to all of the channels.

B. EEG acquisition

The brain activity was recorded using an EEG array of 16 electrodes (gUSBAmp from g.tec) placed on the scalp following the International 10/20 System (FC5, FC1, FC2, FC6, C3, CZ, C4, CP5, CP1, CP2, CP6, P3, PZ, P4, PO3 and PO4) at a sampling frequency of 1200 Hz.

C. Experimental procedure

Four healthy subjects (labeled as S1, S2, S3, and S4) participated in this study, all men, with ages between the 22 and 26 years old. Subjects were sitting in a comfortable chair in front of a screen that provided instructions while their EEG signals were being recorded. Visual cues indicated which of the following two mental tasks they should have performed: Imagination of the movement of the right hand (denoted as Class 1) or Imagination of the movement of the left hand (Class 2). The task the user should have thought was displayed for 2 seconds. Each one of the two tasks had a related image that identified them. Finally, the user had to think the specific task for a period of 8-10 seconds. This process was repeated 5 times per each two tasks.

J. A. Gaxiola-Tirado and D. Gutierrez are with the CINVESTAV, Monterrey's Unit, Vía del Conocimiento 201, PIIT, 66600 Apodaca NL, Mexico. jgaxiola@cinvestav.mx, dgtz@ieee.org.
J. A. Gaxiola-Tirado, E. Iánez, M. Ortiz and J. M. Azorín are with the Brain-Machine Interface Systems Lab. Systems Engineering and Automation Department at Miguel Hernandez University of Elche. Avda. de la Universidad s/n. Ed. Innova, 03202 Elche (Alicante), Spain.
jorgen.gaxiola@goumh.umh.es, eianez@umh.es, mortiz@umh.es, jm.azorin@umh.es.

D. Preprocessing

The methods presented in this paper were implemented in the Matlab package ARfit. Signals were processed in 1s epochs (30 epochs for each subject and Class). A digital band-pass filter between 5 and 50 Hz, a notch filter with 50 Hz cut-off frequency, were applied to the data.

3. TRADITIONAL APPROACH BASED ON SPECTRAL POWER

Signals were processed in 1s epochs (30 epochs for each subject). For each epoch, we estimated the Power Spectral Density (PSD) in the range from 8 to 30Hz using Burg's method (function pburg in Matlab) with an autoregressive model. In order to determine the model order, we utilized the reflection coefficients (function arburg in Matlab). Thus, we used the matlab function arburg with the order set to 20 to obtain the reflection coefficients. Then, the reflection coefficients were plotted, we observed that the reflection coefficients decayed to zero after order 12. This indicated that an order 12 was the most appropriate.

Once the PSD was calculated for all frequencies, channel and class. Using all epochs of the recording session, the R-squared value was calculated for each channel and frequency as the proportion of total variance due to the power difference between the two classes [6]. A higher r2 value is related to a higher discrimination between classes [7].

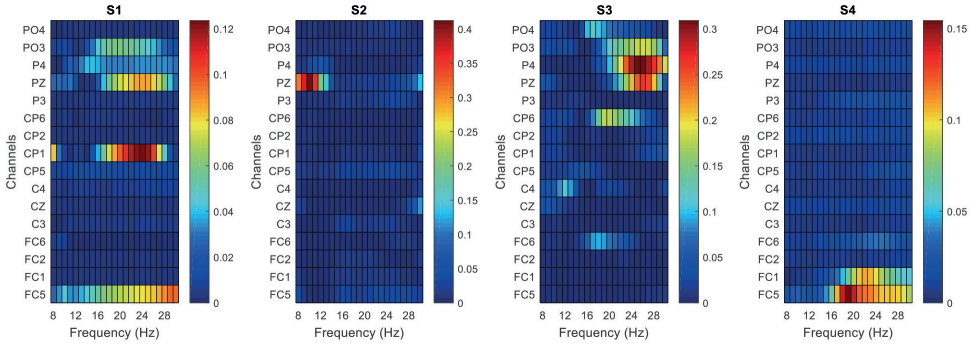


Figure 1. r2 maps of Class 1 vs Class 2 for all subjects, channels and frequencies.

Under these conditions, the r2 maps obtained for each subject are shown in Figure 1. Then, the channel and frequencies with the strongest r2 (±1 Hz) were selected as optimal features. Once the features were selected, feature vectors containing PSD from the selected channels were used as input in a classifier based on Support Vector Machine (SVM) with a radial basis function as kernel (C = 512 and γ = 0.002). The parameters C and γ used in this study have been selected in function of a previous study [8]. The classifier was trained with 50% of the trials available and tested with the remaining trials. This process was repeated 100 times in a cross-validation scheme in which trials were randomly assigned to the training and testing data sets. In Table I, the channel selected, and the classification average accuracy for each subject are shown.

User	Accuracy (%)	SD	Channel selected	Frequency (Hz)
S1	67.20	4.25	CP1	23-25
S2	78.50	2.45	PZ	9-11
S3	80.30	3.45	P4	24-26
S4	58.00	5.26	FC5	19-21

Table I. Average accuracy based on psd features for each subject in terms of the percentage of correct classification between class 1 and class 2.

4. COMPLEMENTARY APPROACH BASED ON PARTIAL DIRECTED COHERENCE

A. Complementary approach based on Partial Directed Coherence

Once preprocessing was performed, we analyzed the directed interconnections in the set of M=16 (FC5, FC1, FC2, FC6, C3, CZ, C4, CP5, CP1, CP2, CP6, P3, PZ, P4, PO3 and PO4) electrodes. In order to compute the PDC, the signals were fitted with a MVAR, where the model order was determined by the Akaike Information Criterion [9]. We analyzed the frequency range of 8 to 30 Hz. For the given set of frequencies, the PDC values from electrode j to electrode i (i=1, 2,…, 16; j=1, 2,…, 16) were obtained trough (3) for each 1s epoch (30 epochs per subject).

For each subject, we analyzed the PDC values at the frequency bands alpha (8-12 Hz) and beta (13-30 Hz). In all cases (frequency-band and direction), differences between Class1 and Class2 were tested using the Wilcoxon rank-sum test, under the null hypothesis that data in Class 1 and Class 2 have equal medians, against the alternative that they were not. The null hypothesis was rejected with $p<0.001$. Once differences were determined, the median over the trials for each frequency band are calculated with the purpose of comparing their magnitudes and thereby determine if the corresponding direction is predominant of Class 1 or Class 2. In Figure 2 is shown the results of this process for each of them.

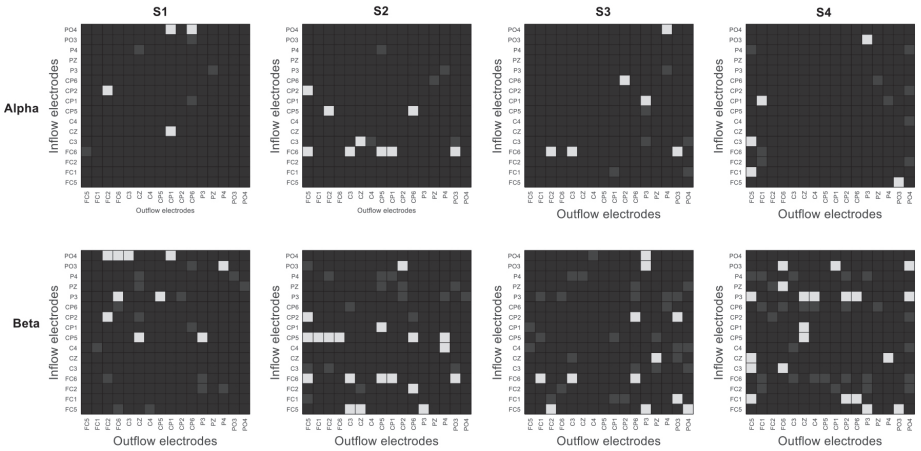


Figure 2. Interconnection directions that characterize the difference between Class 1 and Class 2 ($p<0.001$), the predominant directions of Class 1 are shown in brown, while the predominant directions in Class 2 are shown in green.

B. Analysis of information flows

According to the definitions of information inflow and outflow from directed transfer function analysis [10]. From the PDC values obtained in section IV.A, the outflow information related to certain EEG channel was determined by summing-up the intensity of information flow propagated (the square of all PDC values) to other EEG channels. Meanwhile, the inflow information related to certain EEG was determined by summing-up the intensity of information flow received from other EEG channels.

The distribution of outflow and inflow related to each electrode per subject, at beta frequency band is shown in Figure 3. It can be observed the flows are different for each subject, which

allows to obtain personalized information about the brain activity when performing both mental tasks.

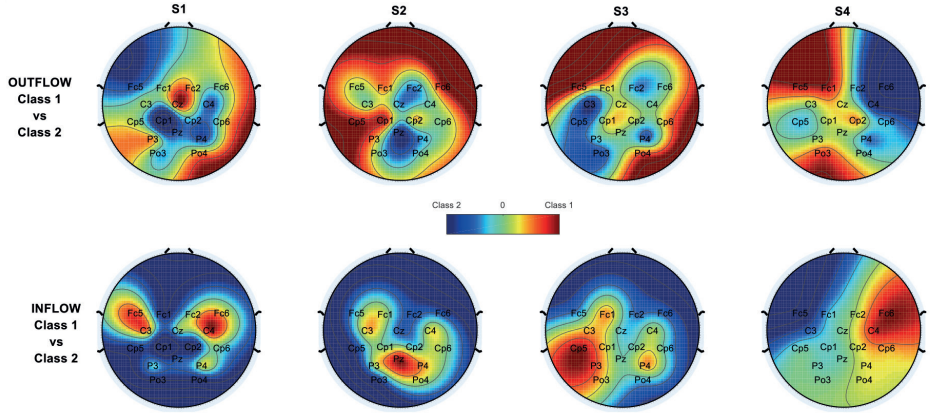


Figure 3. Outflow and inflows maps obtained through PDC when comparing Class1 and Class2.

5. DISCUSSION

From Table 1 in section III, it can be observed that for three subjects (S1, S3 and S4) the frequencies selected are within beta frequency band. Nevertheless, this approach is based on individual electrodes, which gives poor information about the brain dynamic. So when analyzing the brain connectivity using the PDC in section IV.A. , it can be observed that during the beta band there is a greater number of statistically significant directions than in alpha band, particularly for Class 1, which indicates that the brain connectivity is characterized in a better way in the beta band, for which any of these significant directions could be used in the process of feature selection for the classification process.

With respect to information flows obtained through the PDC, this way of visualizing brain connectivity allows to define regions of interaction that it is not possible to define through traditional approaches based on EEG power. For instance, for subject S1, it can be observed that the outflow is predominant for channel Cz for Class 1, while for Class 2, the predominant channels are CP1, P4, Pz. Concerning to the inflows, the predominant electrodes for Class 1 are C4, FC5 and CZ, which indicates the flow of information goes from Cz to these electrodes. On the other hand, it can be observed that for the subject S4, who obtained the lowest classification percentage (58%) the inflow and outflow from Figure 3 are more dispersed in comparison with those obtained for the subjects S1, S2 and S3. Thus, these relations of inflows and outflows can be useful in order to obtain informative features for BCI.

In conclusion, in this preliminary study was demonstrated that EEG-based PDC analysis is able to detect differences in functional connectivity presented in two motor imagery tasks. Furthermore, this analysis provides additional information and could potentially improve the feature selection process. Our future work will include a more rigorous assessment of our EEG-based connectivity analysis and extending our approach to the study of resting-state brain networks.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," Clinical Neurophysiology, 113(6), pp.767–791, 2002.

[2] B. He, S. Gao, H. Yuan, and J. R. Wolpaw, "Brain–computer interfaces," in Neural Engineering, pp. 87–151. Springer, 2013

[3] R. Chaudhari, and H. J. Galiyawala, "A review on motor imagery signal classification for BCI," Signal Process. Int. J.(SPIJ), 11(2), 16, 2017

[4] E. Hortal, A. Úbeda, E. Iánez and Azorin, J. M. "Control of a 2 DoF robot using a Brain–Machine Interface," Computer methods and programs in biomedicine, 116(2), 169-176, 2014

[5] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," Biological cybernetics, 84(6), pp. 463-474, 2001

[6] A. H. Murphy, "The Coefficients of Correlation and Determination as Measures in Performance in Forecast Verification," Weather and forecasting, 10(4), pp. 681–688, 1995.

[7] J. R. Wolpaw and D. J. McFarland, "Multichannel EEG-based BrainComputer Communication," Electroencephalography and clinical Neurophysiology, vol. 90, pp. 444–449, 1994

[8] F. Florez, J.M. Azorín, E. Iánez, A. Ubeda and E. Fernández, "Development of a Low-cost SVM-based Spontaneous Brain-computer Interface," IJCCI (NCTA), pp. 415-421, 2011.

[9] Akaike, H.: "A new look at the statistical model identication," IEEE transactions on automatic control 19(6), 716-723, 1974

[10] B. He, Y. Dai, L. Astolfi, F. Babiloni, H. Yuan, and L. Yang, "eConnectome: A MATLAB toolbox for mapping and imaging of brain functional connectivity," J. Neurosci Methods, 195(2), pp. 261–269, Feb. 15, 2011

Chapter 3

The neuroscience of artistic and contemplative arts
(drawing, music, painting, creative handwriting)

Frequency Analysis of EEG Activity during Artistic Creative Tasks

M. Ortiz, E. Iáñez, J. Gaxiola, J. M. Azorín

Abstract - The present paper shows a preliminary study of the electroencephalography (EEG) activity of different subjects during artistic creative and contemplative tasks. The main objective of the paper is to show if differences can be assessed in the comparison of creative vs. contemplative tasks. A test protocol and a processing algorithm are proposed. The results reveal that during creative tasks there is a gamma band decrement of power with respect to rest state. However, differences between creative and contemplative tasks are very subject dependent.
Keywords: Brain-computer interface, creativity test, electroencephalography, time-frequency analysis.

1. INTRODUCTION

The concept of creativity or creative thinking is hard to define. The more basic definition it would be the ability of creating new ideas or concepts based on already known ideas, with the intention of creating original solutions. Creativity is usually assessed by creativity tests. One example would be the creativity thinking test of Torrance [1]. This test is used for the detection of exceptional gifted children. However, the multidimensional nature of creative thinking makes necessary the support of additional tests, such as academic results or intelligence qualification (IQ) tests, in order to obtain a more accurate evaluation [2].

If we are talking about creative thinking in art, it is even more difficult to define it, as a new variable is added to the equation: aesthetics. Thus, an artistic creative act could be defined as an act that combines originality and aesthetics. However, originality and aesthetics can be very subjective concepts and their assessment is far from the scope of this paper.

The question to the relationship between brain and artistic creation is a current research topic. For instance, related to music in [3], the authors indicate that "Changes in prefrontal activity during improvisation were accompanied by widespread activation of neocortical sensorimotor areas (that mediate the organization and execution of musical performance) as well as deactivation of limbic structures (that regulate motivation and emotional tone). This distributed neural pattern may provide a cognitive context that enables the emergence of spontaneous creative activity". Another study using functional magnetic resonance imaging (fMRI) during RAP improvisation states that "improvised performance is characterized by dissociated activity in medial and dorsolateral prefrontal cortices". However, the relationship is more complicated. As it is indicated by a review of 63 articles by [4]: '*creative thinking does not appear to critically depend on any single mental process or brain region, and it is not especially associated with right brains, defocused attention, low arousal, or alpha synchronization, as sometimes hypothesized. To make creativity tractable in the brain, it must be further subdivided into different types that can be meaningfully associated with specific neurocognitive processes*'.

Therefore, in order to have non-confusing results, the activity must be clearly defined. As the artistic creation process consists of, a first stage of improvisation followed by a second phase of analysis and revision [5], the research will be focused on the first stage of improvisation to

enclose the mental task assessed, evaluating the electroencephalography (EEG) activity of the subjects during a mental creative task in comparison to a contemplative or blank state mental task.

2. METHODS

A. Brain-computer interface (BCI)

A brain-computer interface (BCI) allows to capture the EEG signals of a subject during a certain action in order to transmit and interpret them by a computer. In this research, the EEG signals were sampled and captured at 500 Hz thanks to 31 non-invasive scalp electrodes placed on the skin with the help of a cap of Brain Products GmbH (Germany). The reference electrode was placed in the right earlobe.

The signals were amplified by an actiChamp box and transmitted by wire to the computer for its recording and pre-processing in Brain Vision (Brain Products GmbH, Germany) and posterior analysis in a PC by Matlab.

B. Protocol

The experiment was supervised by the ethics committee of the Miguel Hernández University of Elche obtaining signed consent for all the volunteers based on the agreement of Helsinki declaration. For the experiment, 5 adults (3 male, 2 female) with ages 32.8±4.6 participated in 4 different events per session. All the individuals started the trials with the eyes closed, and after one beep opened them for the rest of the event. The first 2s were neglected to avoid evocated potentials and no movement was allowed.

• Event 1 (E1): Improvisation based on an abstract stimulus. A black ink abstract sheet was shown to the subjects during 17s, trying to imagine which composition they could develop based on the sheet stimulus.

• Event 2 (E2): Improvisation based on simple geometric figures. A simple figure (for instance, two parallel lines) was shown to the subjects during 17s, trying to imagine which composition they could develop based on the geometric figure.

• Event 3 (E3): Contemplation of painted art.

A non-known by the subject painted work was shown during 17s using a high quality printed book. If the picture was known by the subject, a different work was chosen, and the experiment repeated. During the event, the subject focused on the shapes, colors and sensations, the picture produced to them. It was important that the picture was not known for the individual, in order to have a similar 'surprise' effect to the two previous events. Subjects 1, 2, 4 and 5 used the same work: Medusa of Lucien Lévy-Dhurmer (see Fig. 1), while subject 3 used a different work due to its previous knowledge of the work.

• Event 4 (E4): Blanked mind state. During 17s the subject tried to focus on a blanked mind state looking at the surface of the table.

After the tests the subjects carried out the imagined paintings using watercolor for the evaluation of the creativity of the works by an artist.

M. Ortiz, E. Iáñez and J. M. Azorín are with the Brain-Machine Interface Systems Lab, Miguel Hernández University of Elche, Av. Universidad S/N Ed. Innova. 03202 Elche, Spain (phone: +34 965 222198; fax: +34 966 658979; e-mail: mortiz@umh.es, eianez@umh.es, jm.azorin@umh.es).

J. Gaxiola is a member of the "Biomedical Signal Processing Laboratory" at Cinvestav, Monterrey's Unit (Mexico) and visitor Ph.D student in the Brain-Machine Interface Systems Lab, Miguel Hernández University of Elche, Spain (jgaxiola@cinvestav.mx).

Figure 1. Example of E4 user. The subject observes the high quality representation of Medusa by Lévy-Dhurmer.

C. Processing

Built-in filters by manufacturer.

Brain Vision software allows to apply different filters to the signal before its transfer to Matlab. In the case of this research, a band-pass filter within 0.5Hz and 100Hz in combination with a notch filter at 50Hz to avoid the network component in Spain were applied.

Software filters

After that, the signals were processed offline applying other filters. First of them was a Laplatian spatial filter based on Euclidean distance between electrodes i and j, being V the voltage of the electrode.

$$V_i^{Lp} = V_i - \sum_{i \neq j} g_{ij} \cdot V_j \quad (1)$$

with

$$g_{ij} = (1/d_{ij}) / (\sum_{i \neq j} 1/d_{ij}) \quad (2)$$

Being d_{ij} the euclidean distance between electrodes i and j.

Second one was a wavelet filter based on the discrete approximation of Meyer [6].
Signal was decomposed in 6 detail levels extracting the level 6 approximation. This way, possible moving artifacts due to spurious blinks were eliminated from the signal in a better way than by a high-pass filter.

Normalization

Normally, analysis and modelling usually was applied personalized by subject. Due to the nature of this research, comparison needs to be done between subjects. As EEG voltages can be different between subjects, electrodes and even different days for the same person and electrode, a standardization that keeps the differences but allows to compare them in a same level was needed. The standardization used was based on the Maximum Visual (MV) threshold introduced in [7]. MV thresholds are computed as in (3) for the electrode *i*, based on the voltage *V* for the *N* epochs of a trial with a window length (*L*) of 500 samples as:

$$MV^i = 1/N \sum^N_{m=1} [\max(\mathrm{abs}(V^i_{(m-1)\cdot L+1:m\cdot L}))] \quad (3)$$

Once thresholds were computed for each electrode the voltage was standardized as (4) indicates.

$$SV^i(t) = V^i(t) / (1/31 \sum^{31}_{j=1} [MV^j] \quad (4)$$

Feature extraction

Feature extraction was based on the frequency analysis of the components of each event. For the feature extraction the power spectrum density via Yule-walkers method was used. As Fig. 2 shows, there were mainly two dominant components in alpha and low gamma band which were the features considered for analysis. Therefore, each event provided two standardized power values related to the 4-15 Hz and 20-50 Hz peaks.

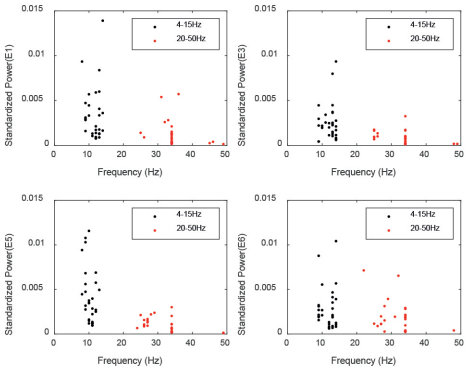


Figure 3. . Frequency distribution of the 2 dominant peaks for the 31 electrodes considered. Subjec S3.

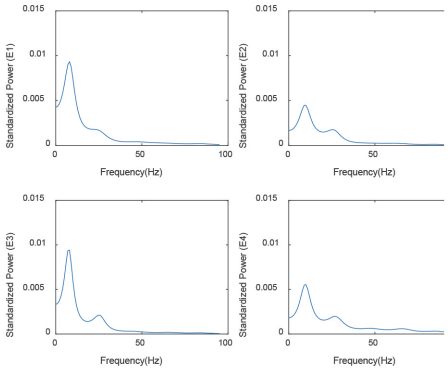


Figure 2. Power spectral density for electrode Fz for subject S3.

3. RESULTS

There are 2 different questions to assess in this research. First one is: There are differences between creative events (E1 and E2), contemplative (E3) and reference (E4)? Second one is: Is there any difference between the subjects while they conduct the tasks?

In order to see the distribution of the dominant frequencies both peaks of the Fig. 2 are represented in Fig. 3 for the 31 electrodes. As it can be seen, the distribution is centered around 11Hz and 35Hz for most of the electrodes. However, gamma band has a wider deviation. In order to see the differences between events, to the power of each electrode was divided by the power of the rest state of each subject. This way a new standardization was done:

$$P_{Sj}^{i}(\mathrm{E_m}) = P_{Sj}^{i}(\mathrm{E_m}) \,/\, P_{Sj}^{i}(\mathrm{E_4}) \qquad (5)$$

Where i=1:31 electrodes, j=1:5 subjects, m=1:3 events.

Then, the resulting value was logarithmic scaled for a better visualization on the topoplot maps that can be seen in Fig. 4 for the 5 subjects and the events 1 to 3. A bluest color means that the power was much lower than the one acquired during rest event, while a yellow color indicated that the power was much bigger. When the color turns to green it meant that the power was similar.

4. DISCUSSION

From the subjects, only S3 had previous experience in art as amateur painter. The other four subjects had no proven expertise in arts although S2 indicated that had used watercolors before.

Lower gamma band (20-50 Hz)

Fig. 4 shows that all the subjects showed a clear lower level of power during the 3 events under test with minor differences in external electrodes for lower gamma band (blue color in right images). However, there were not important differences in gamma band between creativity events (E1 and E2) and contemplative ones (E3) (similar behavior for the E3 regarding E1 and E2). Therefore, it can be concluded that there was a clear desynchronization in lower gamma band during creative and contemplative tasks, but there were not differences between contemplative and creative tasks.

Tetha and alpha band (4-15 Hz)

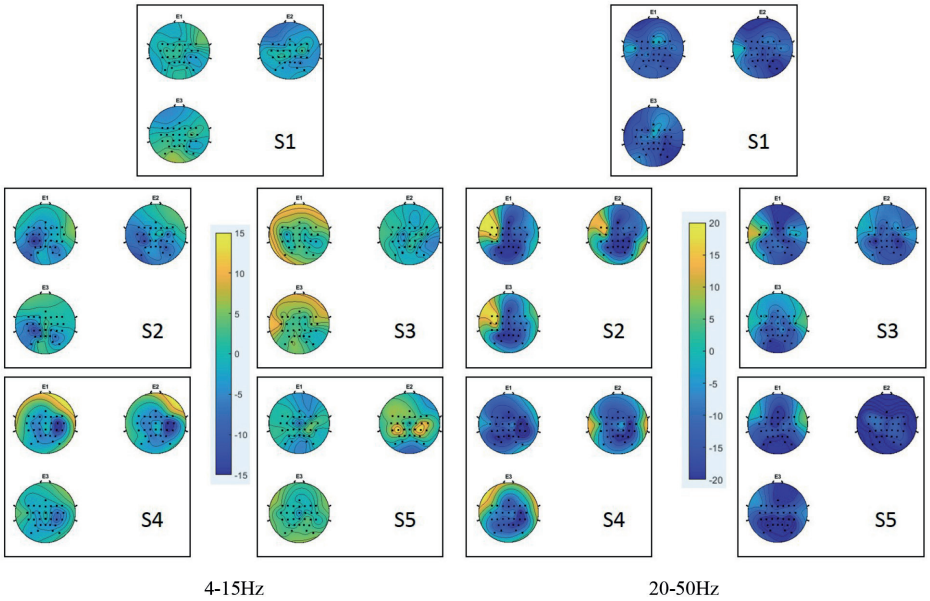


Figure 4. Topoplots for the standardized power of the peaks related to rest event E4.

In the case of 4-15Hz frequencies, mainly alpha band as Fig. 3 shows, the behavior was less homogeneous. S1 and S2 showed almost no differences between the three events (similar topoplots for E1-3). S3 and S5 showed similar patterns for E1 and E3, while S4 showed a similar behavior for the events E1 and E2.

Therefore, only S4, who was the subject who developed the most creative paintings based on the opinion of a consulted artist, showed a different pattern between the creative events and contemplation (similar E1 and E2 and different E3). S3 showed a different result for the geometric creative event in comparison to the painting mental and contemplative tasks (similar E1 and E3 but different E2). This could be related to its previous expertise in art, but this is an assumption which would need a broader analysis and tests with artist subjects. Therefore, results were not conclusive for the alpha band as there was a high dependency on the subject.

In conclusion, a clear desynchronization was detected for lower gamma band for creative and contemplative tasks, but no clear differences were assessed between creative and contemplative tasks

5. ACKNOWLEDGMENTS

The acquisition wireless system of EEG signals with 32 channels from Brain Products was partially funded by the European Union (P.O. FEDER 2007/2013), with the management of Generalitat Valenciana (Spain).

REFERENCES

[1] [1] The Alberta Tearchers' Association, Torrance Tests of Creative Thinking (TTCT): Figural & Verbal, vol. 6, no. 00001. Scholastic Testing Service, 2007.

[2] [2] K. H. Kim, "Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT)," Creat. Res. J., vol. 18, no. 1, pp. 3–14, Jan. 2006.

[3] [3] C. J. Limb and A. R. Braun, "Neural substrates of spontaneous musical performance: An fMRI study of jazz improvisation," PLoS One, vol. 3, no. 2, p. e1679, Feb. 2008.

[4] [4] A. Dietrich and R. Kanso, "A review of EEG, ERP, and neuroimaging studies of creativity and insight," Psychol. Bull., vol. 136, no. 5, pp. 822–848, 2010.

[5] [5] M. Ellamil, C. Dobson, M. Beeman, and K. Christoff, "Evaluative and generative modes of thought during the creative process," Neuroimage, vol. 59, no. 2, pp. 1783–1794, Jan. 2012.

[6] [6] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 11, no. 7, pp. 674–693, Jul. 1989.

[7] [7] Á. Costa et al., "Decoding the attentional demands of gait through EEG gamma band features," PLoS One, vol. 11, no. 4, p. e0154136, Apr. 2016.

Screens vs Galleries: A Neural Study on the Importance of Museums

Jesús Tamez-Duque[2], Fernanda Zapata-Murrieta[1,2], Rodrigo Peimbert[1], Memo Santos[2], Jose Luis Contreras-Vidal[1,3], Rogelio Soto[1]

Abstract - Brain Response is compared during Art Viewing in Non-Laboratory & Quasi-Lab-oratory conditions, using Screen-Viewing and Gallery-Viewing of specific pieces in an exhib-it. After Offline Statistical Analysis of Power Spectral Density in EEG Signals, specific neural patterns were found, focused on the Beta-Band (15-25 Hz) and related to an increase in power only when subjects were physically exposed to the pieces.
Keywords: Neuroscience, EEG, Neural activity, Museum, non-Laboratory, art, Neuroaesthetics

1. INTRODUCTION

During the last decade, a new wave of the digital age has started a shift of diverse physical experiences into virtual contexts [1], exploiting the growing technologies of Augmented and Virtual Reality. Virtual social interactions, entertainment and tourism have been proposed for the near future [2,3].

In an effort to determine the implications of such substitution of traditional experiences by those in virtual environments, studies have recently explored and compared psychological and physiological reactions in people as they various activities in virtual and non-virtual contexts; however, there are still important potential contributions to be made by the field of neurosciences, where defining the neural effects of virtual and non-virtual aesthetic stimulation could prove critical in defining whether virtual tourism and virtual visual entertainment would or would not be equivalent to the traditional experience [3-5].

In this study, an initial effort is thereby proposed by recording and analyzing brain activity from visitors to Museo de Arte Contemporáneo (MARCO) in Monterrey, México, as they viewed pieces from an exhibit by Martin Scorsese in two different states: (1) in non-laboratory conditions within the main gallery and (2) in quasi-laboratory conditions [6,7] within an independent, closed room with pieces shown in a screen. For direct comparison of pleasing and non-pleasing experience, subjects were asked to choose their Favorite (FP) and Non-Favorite (NFP) piece, after the aesthetic experience. Additionally, all subjects had a Baseline (BL) condition recorded, which was devoid of artistic stimuli and which took place before the viewing of the pieces.

2. MATERIALS AND METHODS

A. Subject preparation

The subjects were previously informed about their participation and signed voluntarily an informed consent approval to do the study. It was asked to complete a questionnaire that included information about gender, age, occupational status, and art consumption. To create a neural-data baseline(BL) that is to say, a non aesthetic stimulation moment before the exposure of art, the participants stared at a white wall for a minute on the exhibit room and they stared at a black screen on the classroom (Aula Terra) for 45 seconds.

B. Data Acquisition

Participants were asked to attend two different days to get the analysis, The half of the volunteers attended first to the Aula Terra of the MARCO and then attended to the exhibit room, the other half attended first to the exhibit room and on their second visit they went to the classroom. Given this, the subjects were divided on two groups: Aula First (AF) and Exhibit First (EF), referring to the order of the visits.

On each group, the subjects were divided on groups of three, the task asked on Aula Terra was to stare at the television screen that was in front of them. The participants were asked to stare at the same pictures on the exhibit room as well.

The series of pictures that the subjects stared at belongs to the exposition of Martin Scorsese, a series of photographs taken by Brigitte Lacombe as seen in Fig. 1. There were 8 pictures and the time of staring was controlled for every participant, on the Exhibit room, the subjects stared at every picture for one minute checked by the person who was guiding meanwhile, on the Aula Terra they stared for 30 seconds with a break of 15 seconds with a black screen between every picture. The break for the exhibit room depended on how much time the subjects walked to the next picture. The participants were asked to choose their favorite painting (FP) and their non-favorite (NFP) at the end of the session, resulting on having three events for every subject on the experiment: Baseline, Favorite Painting and Non-Favorite painting.

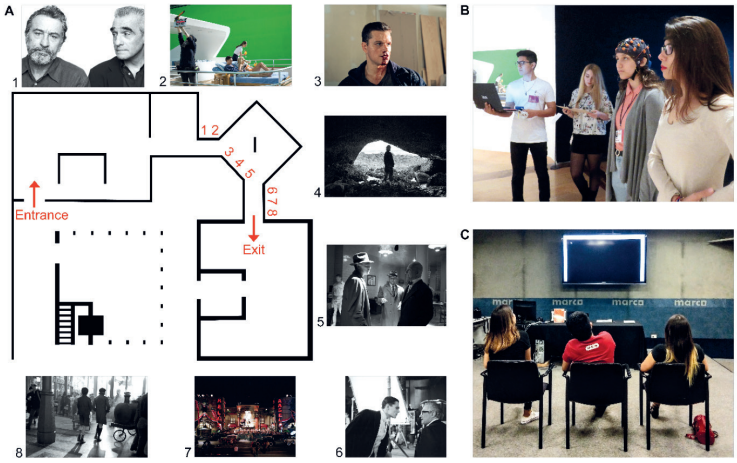


Figure 1. Museum layout and experimental setup of Gallery and Screen areas. A. Painting locations within non-laboratory experiments. Paintings are marked in the order in which they were seen by all subjects. B and C. subjects during data acquisition. B corresponds to non-laboratory conditions at Martin Scorsese's exhibition at MARCO museum. C corresponds to quasi-laboratory conditions at MARCO's Aula Terra.

The EEG data acquisition was taken with the Muse headband, which is a non-invasive device that includes four channels as dry electrodes on the locations of TP9, TP10, AF7 and AF8 according to the 10-20 system [5]. The headband has a sampling frequency of 220 Hz and was connected to a HP mini-tablet which had the software of Muse-IO, MuseSaver and MuseTimeTracer V2 to record the signals in real time.

C. Data Analysis: Preprocessing and Denoising

1 Tecnológico de Monterrey, National Robotics Laboratory, School of Engineering and Sciences, Monterrey, México.
2 INDI Ingeniería y Diseño S.A.P.I. de C.V., Monterrey, México.
3 Laboratory for Non-invasive Brain-Machine Interface Systems, Department of Electrical and Computer Engineering, University of Houston, Houston, TX, United States.

The muse files obtained were converted from .muse to .m format with the command window features. A zero-phase Butterworth bandpass filter for frequencies from 1-60 Hz was applied in order to get only the frequencies we were interested on and to avoid easily identifiable artifacts. The EEG signal was divided by channel and segmented in times to match the moment when the subject was on Baseline section, at the time that it was watching its favorite painting and its non favorite painting. The signals were divided in windows of 256 samples (windows of 1.16 seconds) in order to get epochs, this resulted as approximately 80 epochs per event (BL, FP, NFP). The implemented process details are described in Fig. 2.

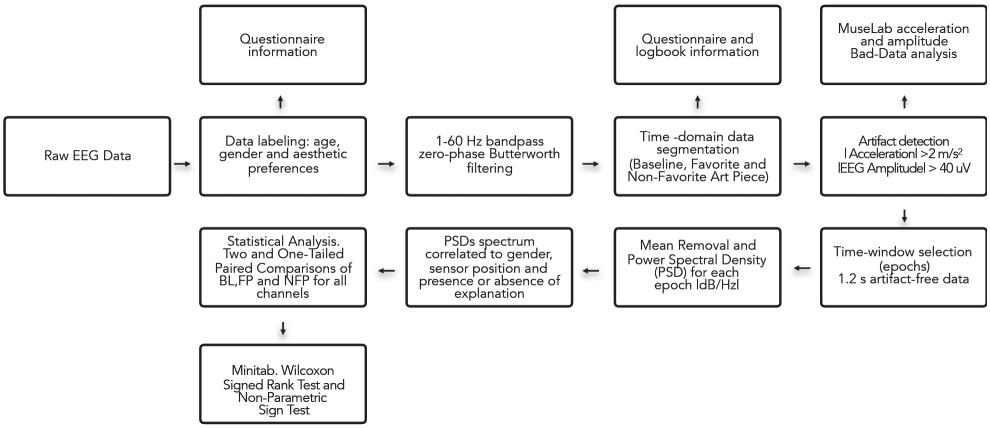


Figure 2. Preprocessing and processing flowchart including artifact removal, feature extraction and statistical analysis. Artifacts were considered when accelerations were 2 m/s2 over and when amplitudes over 40uV were measured. Frequency-domain features were used for neural-activity comparison. Normal-distribution tests were performed, and Non-Parametric statistical analysis strategies were used.

D. Data Analysis

The parameter used to do the analysis was the Power Spectral Density (PSD) of the EEG signals. The PSDs for every event and every subject was visually inspected in order to identify abnormalities. Previous to this step, a code was applied to the signal to identify and exclude the elements that were in an odd shape from the analysis.

For the analysis, a Non-parametric test of signs was applied due to the distribution of the data and the Probability value method(p-value) was used to analyze the difference between the events, the discrepancy between BL and FP, BL and NFP, NFP and FP. Fig. 3 presents the statistical analysis performed.

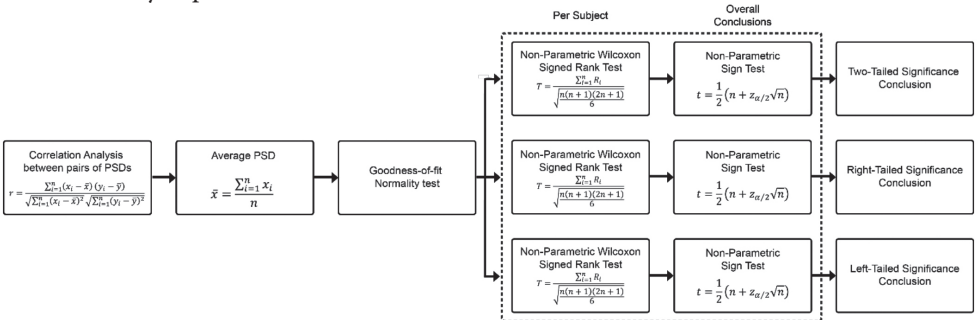


Figure 3. Statistical analysis was performed using paired comparisons between Screen and Gallery for each individual subject. A Non-Parametric Wilcoxon Signed Rank Test was performed for every event and every subject in Gallery and Screen variants. Non-Parametric Sign tests were performed to obtain overall conclusions.

3. PRELIMINARY RESULTS

A. Demographic Analysis

The demographic results shown in Fig.4 presents data obtained from the questionnaire that was applied to the participants. The study was made with a total of 60 voluntary participants in the ages between 17 and 34 and a cohort of 13 subjects was selected to analyze the brain signals obtained. The vast majority were students, science and engineering professionals and artrelated people. The cohort showed that the subjects were mostly students, science and engineering and social sciences as occupation. The results also show that subject population consume art once or twice every two months.

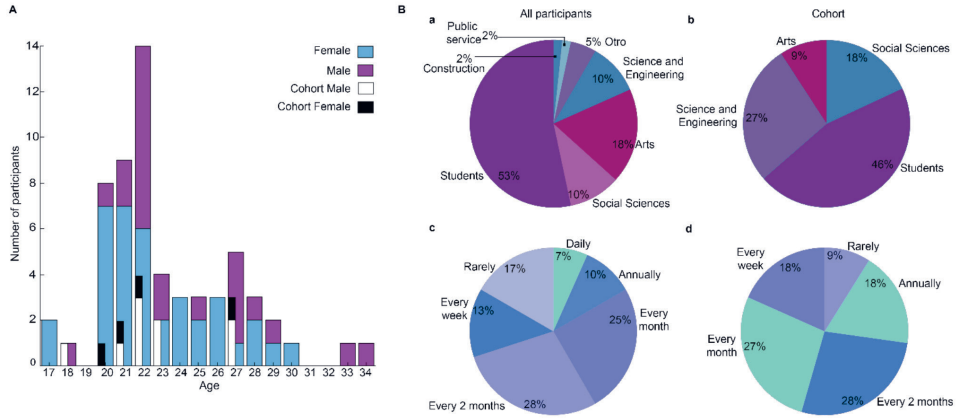


Figure 4. Demographics, Art Consumption levels and Occupation Distribution of total subjects and cohort. A. Total population N=60; cohort n=13. B.a and B.b. Most of the tested subjects were students and around one-fifth had a profession related to arts. B.c and B.d. Most subjects consumed art every 2 months.

The picture that was considered the most favorite painting considering the opinion of all the subjects was picture 4 as shown in Fig 5, and the most considered as Non-Favorite painting was the picture 7 and 8. The image that stood up on the Aula Terra as FP was the picture 4 and the one that was most pointed as NFP was the picture 7, whilst on the Exhibit room, the FP was picture 1 and NFP was picture 8. Additionally, the painting most chosen as favorite and less favorite changes when the exposure to art is different, the most favorite was painting 1 on the Gallery and painting 4 on the Screen whilst the painting pointed as less favorite on the Screen was number 2 and on the Gallery was number 3.

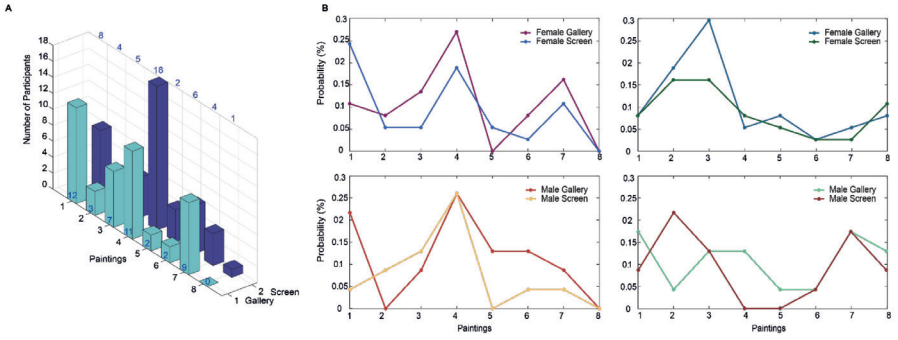


Figure 5. Comparison between aesthetic preference on Gallery and Screen. A. There was a difference of aesthetic preference, painting number 4 was chosen as most favorite on the Screen meanwhile painting number 1 was chosen on the Gallery. B. The probability of preference according to gender show that there's a similar tendency to choose the same paintings, the election of the Non-favorite painting is more varied.

B. Statistical Analysis

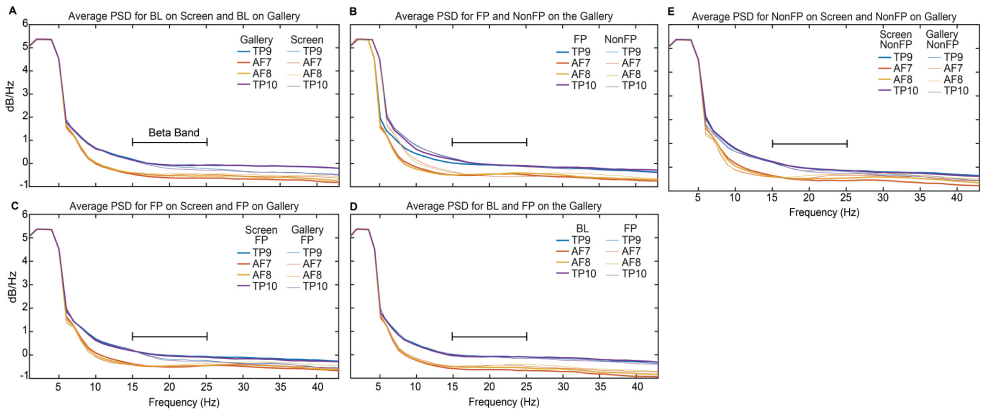


Figure 6. Comparison of average Power Spectral Densities. The focus of interest is on the Beta Band range, which goes from 15-25 Hz.

Results from a Non-Parametric Sign Test showed that aesthetic stimulation generated specific neural patterns of beta-band power-increase (P-value = 0.005) when subjects were exposed to the physical pieces within the main gallery of the museum, and that such neural pattern did not appear when the viewing of the pieces was done on a screen. When comparing the PSD (Fig. 6) between screen and gallery, it was found that there is no significant difference on Baselines but there is between FPs and NonFPs. Additionally, total power of neural activity was also found to be higher when experiencing art in the exhibit room, compared to when being experienced on screen.

4. REFERENCES

[1] Monteiro, R. X., & Barranha, H. (2018). What Can and Cannot Be Felt: The Paradox of Affectivity in Post-Internet Art. Journal of Science and Technology of the Arts, 10 (1), 2-03.

[2] Deloitte (2018). Realidad Digital. Retrieved December 2018, from https://www2.deloitte.com/content/dam/Deloitte/co/Documents/ technology/Realidad digital 2018.pdf

[3] Wagler, A., & Hanus, M. D. (2018). Comparing Virtual Reality Tourism to Real-Life Experience: Effects of Presence and Engagement on Attitude and Enjoyment. Communication Research Reports, 1-9.

[4] Siri, F., Ferroni, F., Ardizzi, M., Kolesnikova, A., Beccaria, M., Rocci, B., ... & Gallese, V. (2018). Behavioral and autonomic responses to real and digital reproductions of works of art. In Progress in brain research (Vol. 237, pp. 201-221). Elsevier.b

[5] Herrera-Arcos, G., Tamez-Duque, J., Acosta-De-Anda, E. Y., Kwan-Loo, K., De-Alba, M., Tamez-Duque, U., ... & Soto, R. (2017). Modulation of Neural Activity during Guided Viewing of Visual Art. Frontiers in human neuroscience, 11, 581.

[6] Kontson, K., Megjhani, M., Brantley, J. A., Cruz-Garza, J. G., Nakagome, S., Robleto, D., ... & Contreras-Vidal, J. L. (2015). 'Your Brain on Art': Emergent cortical dynamics during aesthetic experiences. Frontiers in human neuroscience, 9, 626.

[7] Cruz-Garza, J., Kopteva, A., Paek, A., and Contreras-Vidal, J. L. (2016). Your Brain on Art: Examining the Neural Substrate of Creativity in the Arts Using Mobile Brain-Body Imaging (MoBI). San Diego, CA: Society for Neuroscience

Chapter 4

Novel approaches to the diagnosis of neurological disorders and other medical applications of graphonomics

Handwriting Analysis for the Diagnosis of Alzheimer's Disease: A Preliminary Study

Nicole Dalia Cilia, Claudio De Stefano, Francesco Fontanella, Alessandra Scotto di Freca

Department of Electrical and Information Engineering (DIEI),
University of Cassino and Southern Lazio Via G. Di Biasio, 43 03043 Cassino (FR),
Italy (nicoledalia.cilia, destefano, fontanella,a.scotto)@unicas.it

Abstract - Alzheimers disease (AD) is one of the most common neurodegenerative diseases of the elderly, which produces progressive cognitive, functional and behavioural changes, whose eects are very serious not only for patients but also for their family members and for all the people which interact with them. Unfortunately, these diseases cannot be cured, but their progression can besignicantly delayed if they are diagnosed in a very preventive way, before their eects from a cognitive or neuromotor point of view become evident. In this context, it has been observed that handwriting is one of the motor activities signicantly altered by AD, being the result of a complex network of cognitive, kinaesthetic and perceptive-motor skills. In particular, researchers have showed that the patients aected by these diseases exhibit alterations in the spatial organization, and poor control of movement. From the literature reviewed, most of the studies have been conducted in the medical and psychology elds, where typically statistical tools are used to investigate the relationship between the disease and each of the variables taken into account to describe the patients' handwriting. Moving from these consideration, we present the preliminary results of a study in which an experimental protocol (including copy of words, letters and sentence task) has been used to assess both kinematic and pressure features of on-line handwritings and drawings. Such features have been used to implement a classication system for detecting people aected by AD. The obtained results are very encouraging and seem to conrm the hypothesis that machine learning-based analysis of handwriting can be protably used to support AD diagnosis.
Keywords: Alzheimer's Disease Diagnosis, Handwriting Analysis, Classication Algorithms.

1. INTRODUCTION

Many people in the world suer from neurodegenerative diseases (ND), which are becoming increasingly common and for which it is estimated that their incidence on the world population will grow signicantly in the coming years. In this eld Alzheimers disease (AD) is the most prevalent brain neurodegenerative disorder, progressing to severe cognitive impairment and loss of autonomy in older people (Jelic et al., 1988; Price, 2000; Kang et al., 2015). Unfortunately most of these diseases cannot be cured, but their progression can be signicantly delayed if they are diagnosed in a very preventive manner, before their eects become evident from a cognitive or neuromotor point of view. Criteria for clinical diagnosis of AD were proposed in 1984 (McKhann et al., 1984) by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and by the Alzheimers Disease and Related Disorders Association (ADRDA). According to these criteria, the diagnosis of AD needs histopathologic conrmation (i.e., microscopic examination of brain tissue) in autopsy or biopsy (Babiloni et al., 2016). Nonetheless, an early diagnosis would greatly improve the eectiveness of available treatments: this aspect explains the growing scientic interest in the study of early diagnosis methods, which still represent a very challenging task.

To date, clinical diagnosis of neurodegenerative diseases are performed by physicians and may be supported by tools such as imaging (e.g. magnetic resonance imaging), blood tests and lumbar puncture (spinal tap). Recently, researchers have showed that the patients aected by these diseases exhibit alterations in the spatial organization and poor control of movements. This implies that, at least in principle, some diagnostic signs of AD should be detectable by motor tasks. In this context, alterations in the ability of writing are considered very signicant, since writing skill is the result of complex interactions between the biomechanical parts (arm, wrist, hand, etc.) and the control and memorization part of the elementary motor sequences used by each individual to produce the handwritten traces (Tseng & Cermak, 1993).
For example, in the clinical course of AD, dysgraphia occurs both during the initial phase, and in the progression of the disorder (Yan et al., 2008). The alterations in the form and in the characteristics of handwriting can therefore be indicative of the onset of neurodegenerative disorders, helping physicians to make an early diagnosis. In this eld many published studies have been conducted in the areas of medicine and psychology, where typically standard statistical tools (with ANOVA and MANOVA analysis) are used to investigate the relationship between the disease and the variables taken into account to describe a patients handwriting. Conversely, as we have shown in (De Stefano et al., 2018), few studies have been published, which use classication algorithms to detect people aected by AD from their handwriting.
Moving from these considerations, in previous studies we tried to analyse the correlation between onset and progression of AD, and writing deterioration in order to develop simple and eective systems for supporting AD diagnosis. To this purpose, we proposed (Cilia et al., 2018) a protocol consisting of 25 tasks (copy, reverse copy, free writing, drawing etc.) to verify the impact of dierent tasks and dierent motor skills on AD patient performance. In this paper we present the results of a preliminary study in which we have considered only a subset of the tasks included in the above protocol, collecting the data produced by 130 subjects with a graphic tablet. The rationale of our approach is twofold: on one hand, to asses the kinematic and pressure properties of handwritings by using some standard features proposed in the literature (De Stefano et al., 2018), such as Pressure, Velocity, Acceleration, and Jerk, both on-paper and on-air traits. On the other hand, to test the discriminative power of such features to distinguish patients from healthy controls: for this last task we considered two eective and widely used classication methods, namely Random Forest and decision trees. The obtained results are very encouraging and seem to conrm the hypothesis that machine learning-based analysis of handwriting can be protably used to support the diagnosis of AD.

2. MATERIALS AND METHODS

In the following subsections, the dataset collection procedure and the protocol designed for collecting handwriting samples, are detailed.

2.1 Dataset collection

The 130 subjects who participated to the experiments, namely 66 AD patients and 64 healthy controls, were recruited with the support of the geriatric ward, Alzheimer unit, of the "Federico II" hospital in Naples. As concerns the recruiting criteria, we took into account clinical tests (such as PET, TAC and enzymatic analyses) and standard cognitive tests (such as MMSE). In these tests, the cognitive abilities of the examined subject were assessed by using questionnaires including questions and problems in many areas, which range from orientation to time and place, to registration recall. As for the healthy controls, in order to have a fair comparison, demographic as well as educational characteristics were considered and

matched with the patient group. Finally, for both patients and controls, it was necessary to check whether they were on therapy or not, excluding those who used psychotropic drugs or any other drug that could in uence their cognitive abilities.

2.2 Protocol

The aim of the protocol is to record the dynamics of the handwriting, in order to investigate whether there are specic features that allow us to distinguish subjects aected by the above mentioned diseases from healthy ones. The six tasks considered for this study are arranged in increasing order of diculty, in terms of cognitive functions required (see table 1). Their goal is to test the patients abilities in repeating complex graphic gestures, which have a semantic meaning, such as letters and words of dierent lengths and with dierent spatial organizations. The description of the tasks is the following:
(1) In the rst task the subjects must copy three letters. The letters (l, m, p) were chosen so that they had dierent graphic composition and presented ascender and descender in the stroke (Werner et al., 2006; Lambert et al., 2007). The copy tasks allow to compare the variations of the writing respect to dierent stimuli.
(2) The second task consists in copying four letters (n, l, o and g) on adjacent rows. The aim of the cues is to test the spatial organization abilities of the subject (Onofri et al., 2013).
(3-4) The task 3 and 4 require the participants to write continuously for four times, in cursive, the letter l and the bigram le, respectively (Slavin et al., 1999; Impedovo & Pirlo, 2018). These letters have been chosen because they have variable heights, can be done with a single continuous stroke and contain ascenders, descenders and loops. These characteristics allow the testing of the motion control alternation.
(5-6) The tasks 5 and 6 implies word copying, which is the most explored activity in the analysis of hand- writing for individuals with cognitive impairment (Onofri et al., 2015; Werner et al., 2006; Impedovo & Pirlo, 2018). These tasks, involving dierent pen-ups, allow the analysis of air movements, which it is know to be altered in the AD patients. Moreover, to observe the variation of the spatial organization, we have introduced the copy of the same word without or with a cue. We have chosen the words foglio (sheet in English) because it contains several ascenders and descenders and allows testing ne motor control capabilities.

#	Description	References
1	Copy the letters 'l','m' and 'p'	(Werner et al., 2006; Lambert et al., 2007)
2	Copy the letters on the adjacent rows	(Onofri et al., 2013)
3	Write exactly four joined lowercase cursive letter 'l', in a single smooth movement	(Slavin et al., 1999; Impedovo & Pirlo, 2018)
4	Write exactly four joined lowercase cursive letter 'le', in a single smooth movement	(Slavin et al., 1999; Impedovo & Pirlo, 2018)
5	Copy the word "foglio"	(Onofri et al., 2015; Werner et al., 2006)
6	Copy the word "foglio" above a line	(Onofri et al., 2013, 2015)

Table 1. The tasks.

Furthermore, in order to evaluate patient responses under dierent fatigue conditions, these tasks should be provided by varying their intensity and duration.

2.3 Segmentation and feature extraction

The features extracted during the handwriting process have been exploited to investigate the presence of neurodegenerative diseases in the examined subjects. We used the MovAlyzer tool to process the handwritten trace, considering both on-paper and on-air traits. Their segmentation in elementary strokes was obtained by assuming as segmentation points both pen up and pen down, as well as the zero crossing of the vertical velocity prole. The feature values were computed for each stroke and averaged over all the strokes relative to a single task: we considered for each feature both the mean value and the maximum value for that task. Note that, as suggested in (Werner et al., 2006), we have separately computed the features over on-paper and on-air traits, because the literature shows signicant dierences in motor performance in these two conditions.

Summarizing, in our experiments we have obtained the following features for each considered task, separately computed for both on-paper and on-air traits: (i) Number of Strokes; (ii) Absolute Jerk, that is the Root Mean Square (RMS) value of the jerk across all the strokes (the jerk is dened as the third derivative of the movement); (iii) Average Pen Pressure; (iv) Peak Vertical Velocity, that is the maximum of vertical velocity values; (v) Peak Vertical Acceleration, namely the maximum value of vertical acceleration values; (vi) Average Vertical Velocity; (vii) Age of the subject; (viii) Education level of the subject.

3. EXPERIMENTAL RESULTS

We designed the experiments organizing the data in three groups: data obtained by extracting on paper features, data related to on air features and data including both types of features. Thus, for each task, we generated three dierent datasets, each relative to one of the above groups and containing the samples derived for the 130 subjects. As for the classication phase, we used two dierent classication schemes, namely the Random Forest (RF) and the Decision Trees (DT) with J48 algorithm. For both of them, 500 iterations were performed and a 5 fold validation strategy was considered. In practice, according to this strategy, we divided the whole data set into ve subsets of equal size, and performed ve experiments using each time a dierent subset for evaluating the results, and the remaining ones for the training.

Finally, we averaged the results of these experiments. The tables shown below summarize the values of Recognition Rate (RR) and False Negative Rate (FNR) for each task. In each table, the rst column reports the types of features used, the second one the classier employed, while the following columns report, for each task, the value of RR and FNR, respectively. The False Negative Rate is very relevant in medical diagnosis applications, since it indicates the ability of correctly identifying the patients, thus allowing their inclusion in the appropriate therapeutic pathway. The preliminary results seem to encourage use of classication algorithms to diagnose of AD. From the tables shown below (Tables 2-4) we can point out that: rstly, for each task the maximum value (in bold) of the recognition rate is over 70%, reaching the best value in the fth task equal to 76.57%. Secondly, we can observe that, in the large majority of cases, emerges a better classication using the RF classier compared to DT. This is easily justiable considering that Random Forest, unlike DT, is an ensemble of classiers. However, as reported in the last column, FNR is lower using DT classier in most of the cases. In particular, the best result of FNR occurs using on-paper features in the second task, with a value equal to 8.82%.

4. CONCLUSIONS AND OPEN ISSUES

In this paper, we presented a novel solution for early diagnosis of Alzheimer disease by ana-

lyzing features extracted from handwriting. The preliminary results obtained are encouraging and the work is in progress to increase general performance. The next steps to be taken will include: combination of all tasks taken into account in a suitable way (the task itself can be used as a new feature) (De Stefano et al., 2011; Cordella et al., 2013); introduction of a reject option in order to reduce false negative rate; introduction of new features evaluated on slant, loop surface, horizontal size and vertical size, etc; reduction of the unbalancing dataset caused by diculties in recruiting young patients and old people without any cognitive disease.

		Task 1		**Task 2**	
Features	**Classifier**	**RR**	**FNR**	**RR**	**FNR**
All	RF	71.96	28.79	66.41	33.82
	DT	66.66	19.70	66.41	36.76
On paper	RF	**72.72**	28.79	**72.38**	23.53
	DT	65.90	27.27	67.16	36.76
On air	RF	71.21	22.73	60.44	33.82
	DT	68.18	**18.18**	70.89	**8.82**

Table 2. Classication results of tasks 1 and 2.

		Task 3		**Task 4**	
Features	**Classifier**	**RR**	**FNR**	**RR**	**FNR**
All	RF	**70.09**	44.90	**71.42**	36.00
	DT	61.81	40.82	63.81	30.00
On paper	RF	68.18	44.90	66.66	38.00
	DT	63.63	30.61	67.62	34.00
On air	RF	67.27	44.90	61.90	42.00
	DT	66.36	**22.45**	63.80	**18.00**

Table 3. Classication results of tasks 3 and 4.

		Task 5		**Task 6**	
Features	**Classifier**	**RR**	**FNR**	**RR**	**FNR**
All	RF	75.67	25.45	69.29	39.22
	DT	66.66	38.18	64.91	41.18
On paper	RF	**76.57**	**23.64**	68.41	43.14
	DT	67.56	41.82	**71.93**	**21.57**
On air	RF	64.68	34.55	64.03	45.10
	DT	61.26	38.18	63.15	39.22

Table 4. Classication results of tasks 5 and 6.

REFERENCES

[1] Babiloni, C., Triggiani, A., Lizio, R., Cordone, S., Tattoli, G., Bevilacqua, V., & al. (2016). Classication of single normal and alzheimer's disease individuals from cortical sources of resting state eeg rhythms. *Frontiers in Neuroscience*, 10, 1-18.

[2] Cilia, N. D., De Stefano, C., Fontanella, F., & Scotto di Freca, A. (2018). An experimental protocol to support cognitive impairment diagnosis by using handwriting analysis. *Procedia Computer Science*, 141, 466 - 471.

[3] Cordella, L., De Stefano, C., Fontanella, F., & Scotto Di Freca, A. (2013). A weighted majority vote strategy using bayesian networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Articial Intelligence and Lecture Notes in Bioinformatics), 8157 LNCS*, 219-228.

[4] De Stefano, C., Fontanella, F., Folino, G., & Scotto Di Freca, A. (2011). A bayesian approach for combining ensembles of gp classiers. *Lecture Notes in Computer Science (including subseries Lecture Notes in Articial Intelligence and Lecture Notes in Bioinformatics)*, 6713 LNCS, 26-35.

[5] De Stefano, C., Fontanella, F., Impedovo, D., Pirlo, G., & Scotto di Freca A, A. (2018). Handwriting analysis to support neurodegenerative diseases diagnosis: a review. *Pattern Recognition Letters*, 121, 37-45.

[6] Impedovo, D., & Pirlo, G. (2018). Dynamic handwriting analysis for the assessment of neurodegenerative diseases: a pattern recognition perspective. *IEEE Reviews in Biomedical Engineering*, (pp. 1-13).

[7] Jelic, V., Dierks, T., Amberla, K., Almkvist, O., Winblad, B., Nordberg, A., & Tsukahara, N. (1988). Longitudinal changes in quantitative eeg during long-term tacrine treatment of patients with alzheimers disease. *Neuroscience Letters*, 254, 8588.

[8] Kang, J., Lemaire, H. G., Unterbeck, A., Salbaum, J. M., Masters, C. L., Grzeschik, K. H., & al. (2015). The precursor of alzheimers disease amyloid a4 protein resembles a cell-surface receptor. *Nature*, 325, 733-736.

[9] Lambert, J., Giard, B., Nore, F., de la Sayette, V., Pasquier, F., & Eustache, F. (2007). Central and peripheral agraphia in alzheimer's disease: From the case of auguste d. to a cognitive neuropsychology approach. *Cortex*, 43, 935-951.

[10] McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. (1984). Clinical diagnosis of alzheimers disease: report of the nincdsadrda work group under the auspices of department of health and human services task force on alzheimers disease. *Neurology*, 34, 939-944.

[11] Onofri, E., Mercuri, M., Archer, T., Ricciardi, M. R., F.Massoni, & Ricci, S. (2015). Eect of cognitive fluctuation on handwriting in alzheimer's patient: A case study. *Acta Medica Mediterranea*, 3, 751.

[12] Onofri, E., Mercuri, M., Salesi, M., Ricciardi, M., & Archer, T. (2013). Dysgraphia in relation to cognitive performance in patients with Alzheimer's disease. *Journal of Intellectual Disability-Diagnosis and Treatment*, 1, 113-124.

[13] Price, D. L. (2000). Aging of the brain and dementia of the alzheimer type. *Princ. neural Sci.*, (p. 11491168).

[14] Slavin, M. J., Phillips, J. G., Bradshaw, J. L., Hall, K. A., & Presnell, I. (1999). Consistency of handwriting movements in dementia of the alzheimer's type: a comparison with huntington's and parkinson's diseases. *J Int Neuropsychol Soc.*, 5, 20-25.

[15] Tseng, M. H., & Cermak, S. A. (1993). The in uence of ergonomic factors and perceptual{motor abilities on handwriting performance. *American Journal of Occupational Ther-*

apy, 47, 919-926.
[16] Werner, P., Rosenblum, S., Bar-On, G., Heinik, J., & Korczyn, A. (2006). Handwriting process variables discriminating mild alzheimer's disease and mild cognitive impairment. *Journal of Gerontology: PSYCHOLOGICAL SCIENCES*, 61, 228-36.
[17] Yan, J. H., Rountree, S., Massman, P., Doody, R. S., & Li, H. (2008). Alzheimer's disease and mild cognitive impairment deteriorate ne movement control. *Journal of Psychiatric Research*, 42, 1203-1212.

Deep Transfer Learning for Alzheimer's disease detection

Nicole Dalia Cilia, Claudio De Stefano, Francesco Fontanella, Claudio Marrocco, Mario Molinara

Department of Electrical and Information Engineering University of Cassino and Southerh Lazio
Via G. Di Biasio, 43
03043 Cassino (FR), ITALY
(nicoledalia.cilia, destefano, fontanella, c.marrocco, m.molinara)@unicas.it

Abstract - Early detection of Alzheimer's disease is essential in order to initiate therapies that can reduce the eects of the disease, extending the life expectancy of the patients. It has been shown that early signs are identiable in handwriting. In recent years many studies have tried to use classication algorithms as decision support systems for the diagnosis of Alzheimer. In this work we are going to present an approach based on dierent CNNs trained on ImageNet and applied to synthetic images obtained from a dataset including handwriting and drawing samples of about 150 subjects. The data were collected according to an experimental protocol which includes copy of words, letters and sentences as well as graphic tasks. In this study, we considered two graphic tasks, whose objective is to test the patient's ability in drawing circles, scaled in two dimensions. In the two tasks, subjects were asked to trace four times a circle continuously (i.e. without pen-ups). The trajectories of the handwriting have been transformed into images where the three RGB channels, representing pressure, velocity and jerk, respectively, have been encoded as colour intensities. The preliminary results (Recognition Rate 75.17% and False Negative Rate 17.23%) are very promising and seem to encourage the use of Deep Learning systems to support AD diagnoses.
Keywords: Deep Transfer Learning, Alzheimer's disease, Convolutional Neural Network.

1. INTRODUCTION

Early detection of Alzheimer's disease is essential in order to initiate therapies that can reduce the eects of the disease, extending the life expectancy of the patients. It has been shown that early signs are identiable in handwriting (De Stefano et al., 2018). So far, however, a standard experimental protocol, composed of an ensemble of handwriting tasks the subjects should perform and a well-designed data set, large enough to allow an eective training of machine learning techniques, are missing. To solve some of these issues, we suggested in (Cilia et al., 2018) a protocol consisting of 25 tasks to analyse the impact of dierent motor skills on AD patients' performance. The aim of this protocol is to record the dynamics of the handwriting, in order to investigate whether there are specic features that allow us to distinguish the subjects aected by the above mentioned diseases from the healthy ones. Moreover, in (Cilia et al., 2019) we selected a subset of these tasks and tested 154 subjects (both patients and healthy controls) employing four classication algorithms. The nine tasks considered in such study were arranged in increasing order of diculty, in terms of the cognitive functions required, with the goal of testing patients abilities in repeating complex graphic gestures, which have a semantic meaning, such as letters and words of dierent lengths and with dierent spatial organizations. The encouraging results shown in both studies have led us to exploit the information derived from the above data applying new techniques to dierentiate research methods and to get better performance.
On the other hand, in recent years, the use of Deep Neural Networks (DNN) is becoming

increasingly widespread, both for their performances and for their capability of automatically extracting features from raw data, without a real feature engineering phase. It is possible to identify many scientic papers that use Deep techniques for handwriting analysis and often the approach is based on a Convolutional Neural Network (CNN), that is a particular DNN able to directly analyse images (LeCun & Bengio, 1995)) In the context of medical images datasets, one of the main problems is the high cost of data acquisition that leads to relative small size sets of labelled data. For this reason it is very dicult to employ DNN because very large datasets are needed for their training. Taking into account these considerations, more and more often, the Transfer Learning (TL) technique is adopted (Bria et al., 2018). In TL the starting point of the training phase is not a newly initialized network, but a network pre-trained on generic problems characterized by a very large set of data in terms of both samples and classes. The use of TL typically requires to modify the network structure by changing the number of outputs or, in some cases, substituting the selected classier with a dierent one. While TL can be considered as a dierent method for network initialization, the CNN can be considered as an initialization that includes the capability to extract a set of generalized features from images. This process typically consists in rening the training performed on a generic database like ImageNet by applying the following two steps: i) the so-called transfer learning, where the classier dened in the last layers is redened and then retrained; ii) the so-called ne tuning, where the convolutional section is partially or totally re-trained and the classication section restarts from the parameters evaluated during the step i).

In this work we present an approach based on dierent CNNs trained on ImageNet and applied to synthetic images generated from the dataset employed in (Cilia et al., 2018). In particular, we considered the data derived from two graphic tasks, whose objective is to test the patients ability in drawing circles, scaled in two dimensions. In these tasks, subjects are asked to trace a circle continuously for four times. The circle diameter is 6 cm for the rst task and 3 cm for the second task (see gure 1). The reason why we have included these tasks in the protocol can be found in (Schröter et al., 2003). In this study, we assume that that the writing frequency and the number of changes in direction and speed are good indicators of the automatism of movements, whereas the velocity variance measures regularity and coordination of sequences of movements. As concerns the smaller circle, we state that changing the size of the circle results in a re-adaptation of the motor control. We also considered the ability of CNNs to recognize the form of the object and we have chosen the tasks that have a xed form (with spatial cue) in order to decrease the variability among traits of the subjects. These features has been preliminary chosen because they are candidate to give the best performance, as shown in (De Stefano et al., 2018) in which dierent non-deep classier are presented. The written data were collected by using the WACOM Bamboo Folio graphic tablet, which allowed the subjects to write on standard A4 white sheets using a seemingly normal pen: such a pen produces both the ink trace on the sheet and the digital information, which are recorded on the tablet in the form of spatial coordinates and pressure for each point (x, y and z), acquired at a frequency of 200 Hz.

The tablet also records the in-air movements (up to a maximum of three centimetres in height). Under these conditions, the subject is not supposed to change his natural writing movements as it happens, for instance, when the writing is produced using an electronic stylus on the surface of a tablet. The 154 subjects who participated in the experiments, namely 85 Alzheimer Disease (Positive) patients and 69 Healthy Controls (Negative), were recruited with the support of the geriatric ward, Alzheimer unit, of the Federico II hospital in Naples.

As concerns the recruiting criteria, we took into account clinical tests (such as PET, TAC and enzymatic analysis) and standard cognitive tests (such as MMSE). Starting from the acquired traits and thanks to the Movalyzer software, several features characterizing the graphic traits, such as acceleration, velocity, slant, jerk, etc, have been calculated. The trajectories of the handwriting have been transformed into images where the three RGB channels representing velocity, jerk and pressure features have been encoded as color intensities. In the next paragraph, we are going to better explain how all the images have been created. In the third paragraph we will describe the four adopted models and in the nal paragraph we are going to show the results of employed deep learning techniques.

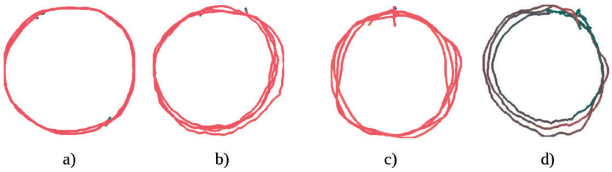


Fig. 1. Example of generated images: a) and b) are an examples of task 1 (Negative and Positive); c) and d) are an example of task 2 (Negative and Positive).

2. DATASET CREATION

From the acquisition phase the trajectories of handwriting, in terms of **x**, **y** coordinates, are available. For each acquired point a third information representing the pressure (coordinate **z**), is also provided. From these coordinates, we have calculated velocity **v** and jerk **j** features. Synthetic images have been generated starting from these values by considering: i) **x**, **y** as vertices of a polygonal (a curve); ii) **z**, **v** and **j** as RGB color components, respectively. Taking into account that each model automatically resize the input images to 256x256 for VGG19, 224x224 for ResNet50, 299x299 for InceptionV3 and InceptionResNetV2 respectively, the **x**, **y** coordinates have been resized into the range [0, 299] image by image, in order to provide ex ante images of suitable size and minimize the loss of information related to possible zoom in/out.

z, **v** and **j** have been normalized into the range [0, 255] so to match the standard 0-255 color scale by considering the minimum and the maximum value on the entire training set for these three quantities. An Example of a trait generated from these images is reported in Fig. 2. In the gure the colour of the rst trait corresponds to the triplet (z=127, v=127, j=0), while that of the second one to the triplet (z=127, v=127, j=127).

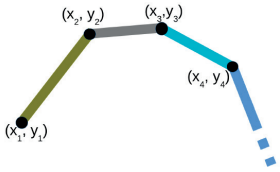


Fig. 2. Example of color encoding for the traits generation.

Model	Total	Transfer Learning (TL)		Fine Tuning (FT)
		Trainable	Non Trainable	Trainable
VGG19	25M	5M	20M	25M
ResNet50	32M	8M	24M	32M
InceptionV3	30M	8M	22M	30M
InceptionResNetV2	62M	7M	55M	62M

Table 1. Number of parameters for the adopted models in "Transfer Learning" and "Fine Tuning" congurations.

The data were produced by 154 subjects, each performing the 2 tasks illustrated in subsection 1. The training phase was conducted by using a validation set in order to reduce or avoid overtting of the network on training set. A 5-fold validation strategy has been considered where for each fold we included 70% of images for training, 10% for validation and 20% for test set.

3. TRAINING OF THE ADOPTED NETWORK

Deep transfer learning is becoming very popular in image classication problems and dierent CNN trained on public datasets like ImageNet (Deng et al., 2009) have obtained year by year the highest classication performance. In this paper four models have been adopted: VGG19 (Simonyan & Zisserman, 2015), ResNet50 (He et al., 2016), InceptionV3 (Szegedy et al., 2016b), InceptionResNetV2 (Szegedy et al., 2016a).

These models evolved in two dierent ways: the rst by introducing new structural elements (inception, residual, dropout); the second by increasing the number of layers. Consequently the number of parameters has been increased from 25 millions in VGG19 to 62 millions in InceptionResNetV2 (see Table 3.) and the number of layers has grown from the 19 of VGG to some hundred of InceptionResNetV2.

Generally speaking a CNN is composed of two parts: convolutional and classication. The rst part is conceived for feature extraction (FE) directly from images, while the second part for the classication (C).

The training of the network for each fold have been completed in two steps: Transfer Learning (TL) and Fine Tuning (FT). During the TL step, all the parameters of the FE part are frozen (see Table 3.) whereas the parameters of the classier C are randomly initialized and trained. During the FT step, both parts (FE and C) are involved in training and all the parameters are unfrozen (see Table 3.). The FT assumes that the parameters are initialized: with the ImageNet training for FE section, while the C module is initialized with values obtained during the previous TL step. In all adopted models, the original C layer has been replaced with a Multi-Layer Perceptron with two hidden layers instead of one, with 2048 neurons and a dropout between them. This replacement try to deal with the very dierent application domains (from ImageNet to Alzheimer's disease) with an increased complexity. This architecture has been assessed with a preliminary experimental phase by minimizing the mean Recognition Rate (RR) of all models for TL and selecting the following settings and hyper-parameters:

- Stochastic Gradient Descent (SGD) with learning rate = 0:001, momentum = 0:9;
- Batch size 16. Number of images from training set considered in each iteration.
- Max epochs equal to 2; 000. One epoch is one pass on the entire training set and contains a number of iterations equal to (trainingsetsize)=batch.
- Patience (a limits for epochs if validation loss does not improve for a while) equal to 200;
- RR (Accuracy) as a measure of performance.

4. RESULTS

The experiments were performed by using an Intel Core i7-7700 CPU @ 3.60 GHz 32GB of RAM with a GPU Titan Xp. In terms of framework, we used Keras 2.2.2 with TensorFlow 1.10.0 as backend. A time ranging from 2 hours to 6 hours was needed in the TL and FT phases with a maximum of 6 hours on InceptionResNetV2.
Table 2 are summarized the mean values evaluated on 5 folds of RR and False Negative Rate (FNR) for each task. In particular, the rst column reports the type of task, the second one the phase (TL or FT) of system, while the following columns report, for each task, the value of RR and FNR, respectively, for all models employed.
The RR is dened as the number of true positives (TP) and true negatives (TN) divided by the number of all samples. A TP or TN is a sample that the algorithm correctly classied as Positive or Negative, respectively. A false positive (FP) or false negative (FN), on the other hand, is a sample that the algorithm incorrectly classied. For example, if the algorithm classied a Negative sample as Positive, it would be a FP. The FN instead is an error in which a result improperly indicates no presence of disease (the result is Negative), when in reality it is present.
From the Table 2 we can point out that: rstly, for each task the maximum value (in bold) of RR is over 70% using the four models, reaching peaks in second task, exhibiting values of about 75% employing ResNet50. Secondly, we can observe that, on average, the ResNet50 model provides better results both in RR and in FRN. This outcome is in good accordance with the theory, considering that smaller networks show better results when the dataset is small. In other words, the absence of the inception module both in the VGG19 and ResNet50 models, as well as their reduced size in term of parameters, could represent the cause of the best performance considering the small size of the dataset.
As regards the tasks, we can claim that the second task shows high performance both in RR and in FNR.

		InceptionV3		InceptionResNetV2		ResNet50		VGG19	
		RR	**FNR**	**RR**	**FNR**	**RR**	**FNR**	**RR**	**FNR**
Task 1	**Transfer Learning**	67.74	27.06	63.34	40.00	51.61	52.94	66.45	09.41
	Fine Tuning	69.68	30.59	67.74	25.08	70.32	37.13	**72.26**	**24.31**
Task 2	**Transfer Learning**	69.08	43.70	61.98	41.51	68.08	23.78	64.76	28.32
	Fine Tuning	72.59	26.47	70.75	18.91	**75.17**	**17.23**	64.76	27.65

Table 2. Results of two tasks for all models employed on test set (mean on 5 -folds).

5. CONCLUSIONS

In this paper, we presented a novel solution for diagnosis of Alzheimer Disease by analyzing the images extracted from handwriting. The preliminary results are very promising and they seem to encourage the use of Deep Learning systems to support AD diagnoses. Currently, there are no analogous systems in the literature, which use deep type techniques for the diagnosis of Alzheimers disease from handwriting.
This means that this system represents the current state of the art in the eld at moment. The work is in progress to increase general performance and, in the next step, using the same approach, we will extent the dataset, including other images acquired with the same protocol.

6. ACKNOWLEDGMENT

This work is supported by the Italian Ministry of Education, University and Research (MIUR) within the PRIN2015-HAND project. Moreover, the authors gratefully acknowledge the support of NVIDIA Corporation for the donation of the Titan Xp GPUs.

REFERENCES

[1] Bria, A., Cilia, N., De Stefano, C., Fontanella, F., Marrocco, C., Molinara, M., Scotto di Freca, A., & Tortorella, F. (2018). Deep transfer learning for writer identication in medieval books. In *Proceedings of MetroArcheo 2018*. IEEE Computer Society - in press.

[2] Cilia, N., De Stefano, C., Fontanella, F., & Scotto di Freca, A. (2018). An experimental protocol to support cognitive impairment diagnosis by using handwriting analysis. *Procedia Computer Science, 141*, 466 - 471.

[3] Cilia, N., De Stefano, C., Fontanella, F., & Scotto di Freca, A. (2019). Handwriting analysis for the diagnosis of alzheimers disease: A preliminary study. *Submitted to International Conference on Computer Analysis of Images and Patterns - 2019*.

[4] De Stefano, C., Fontanella, F., Impedovo, D., Pirlo, G., & Scotto di Freca A, A. (2018). Handwriting analysis to support neurodegenerative diseases diagnosis: a review. *Pattern Recognition Letters, 121*, 37-45.

[5] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *CVPR* (pp. 248{255). IEEE Computer Society.

[6] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 770-778).

[7] LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time-series. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks*. Cambridge, MA, USA: MIT Press.

[8] Schröter, A., Mergl, R., Bürger, K., Hampel, H., Möller, H.-J., & Hegerl, U. (2003). Kinematic analysis of handwriting movements in patients with alzheimer's disease, mild cognitive impairment, depression and healthy subjects. *Dementia and geriatric cognitive disorders, 15*, 132-42.

[9] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *CoRR, abs/1409.1556*.

[10] Szegedy, C., Ioe, S., & Vanhoucke, V. (2016a). Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*.

[11] Szegedy, C., Vanhoucke, V., Ioe, S., Shlens, J., & Wojna, Z. (2016b). Rethinking the inception architecture for computer vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 2818-2826).

Explainable AI for automatic diagnosis of Parkinson's disease by handwriting analysis: experiments and findings

Antonio Della Cioppa, Giovanni Palladino, Antonio Parziale, Rosa Senatore, Angelo Marcelli

Natural Computation Lab
DIEM - University of Salerno - Via Giovanni Paolo II, 132 - 84084, Fisciano (SA), ITALY
adellacioppa@unisa.it, g.palladino14@studenti.unisa.it, anparziale@unisa.it, rsenatore@unisa.it, amarcelli@unisa.it

Abstract - Methods and tools for automatic diagnosis of Parkinson's disease adopting machine learning techniques exhibit impressive performance, but they are achieved by very sophisticated algorithms whose internal functioning hides both the features that are deemed as relevant for the purpose and the criteria by which the features are compared to reach the final decision. Such systems are perceived by the intended user as black-box to be trusted, without anyone able to account for their behaviour, and this is one of the main obstacles to their adoption in daily practice. In such a context we address the problem of managing the tradeoff between performance and explainability once the design is driven by the latter. For the purpose, we consider as machine learning tool the decision tree, as it provides the decision criteria in terms of both the features which are actually useful for the purpose among the available ones and how their values are used to reach the final decision, thus favouring explainability. On the other side, we consider the random forest, which is among the top performing machine learning tool, but whose decision criteria for both selecting the features and using then for the comparison are hidden into its internal structure. The performance of the two approaches has been evaluated on a public dataset, and the results show that the system based on the decision tree achieves comparable or better results on discriminating Parkinson's disease patients from healthy subjects in comparison with the random forest while providing a plain description of the decision criteria in terms of the observed features and their values.

1. INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative disorder that affects dopaminergic neurons in the Basal Ganglia, whose death causes several motor and cognitive symptoms. PD patients show impaired ability in controlling movements and disruption in the execution of everyday skills, due to postural instability, onset of tremors, stiffness and bradykinesia (1-4).

In the last decades the analysis of handwriting (or, more in general, the analysis of handwritten production) has bring many insights for uncovering the processes occurring during both physiological and pathological conditions (5-7) and providing a non-invasive method for evaluating the stage of the disease (8).

In the field of graphonomics and pattern recognition, many studies proposed different AI based approaches for the automatic identification of tremor and micrographia phenomena in handwriting for supplying useful information to the clinician. A variety of tests have been proposed, requiring the subject to write simple letters patterns, geometric figures and so on, but it has been shown that the most suitable ones are those requiring the drawing of geometric shapes such as spirals and meanders with or without a reference pattern printed on the paper (Zham at al., 2017). Following these suggestions, Pereira and his collaborators have collected the NewHandPd data set, which include both off-line images and on-line signals of the traces

produced by the subjects while drawing 4 samples of spirals and 4 samples of meanders (Pereira et al., 2016). A variety of top performing machine learning algorithms, such as Convolutional Neural Networks (CNN), Support Vector Machine (SVM), Optimal Path Finder (OPF), Random Forest (RF) and Restricted Boltzman Machines (RBM) have been evaluated on both off-line samples and on-line samples, and the results show high level of accuracy in discriminating between samples produced by PD patients from those produced by healthy subjects, with the CNN achieving an accuracy greater than 87% on on-line samples (Pereira et al., 2016; Pereira et al., 2017).

The high level of accuracy exhibited by those methods, however, it is achieved by adopting algorithms whose internal functioning hides both the features that are deemed as relevant for the purpose and the criteria by which the features are compared to reach the final decision. They are perceived as black-box, whose decision criteria are inscrutable and therefore cannot be judged in clinical terms. It is not surprising, therefore, that their adoption as diagnostic tool is seriously questioned by the clinicians.

To overcome these drawbacks, we propose to approach the design of these tools having in mind the *explainability* of their functioning first, and then the overall *performance*. Along this line of thoughts, we propose to adopt two decision trees as the machine learning tool for deriving the decision criteria for spirals and meanders and a method to combine their outputs for making the final decision. The former assumption favours the understanding of the decision criteria, as they are represented in terms of *if-then* rules as in diagnostic process the clinicians are familiar with, while the latter is meant to improve the overall performance by combining the outputs of the decision trees on multiple samples of the same subject.

In the remaining of the paper, Section 2 describe the architecture of the tool we have developed and the main feature of its components, while Section 3 describes the experiments we have designed and performed for the automatic learning of the decision trees and reports the results obtained in terms of both sample classification, for comparison with the state of the art methods, and patient discrimination, as to show to which extent the proposed tool can be used by clinicians in their daily practice. Eventually, in the conclusion, we summarize the work that has been done, discuss the experimental results and outline our future investigations.

2. THE PROPOSED ARCHITECTURE

The architecture envisages an ensemble of classifiers, where each classifier is implemented by a Decision Tree (DT) aimed at discriminating the samples produced by each PD patients from those by healthy subjects. In particular, we use two different classifiers, the first one learned on spirals and the second one on meanders.

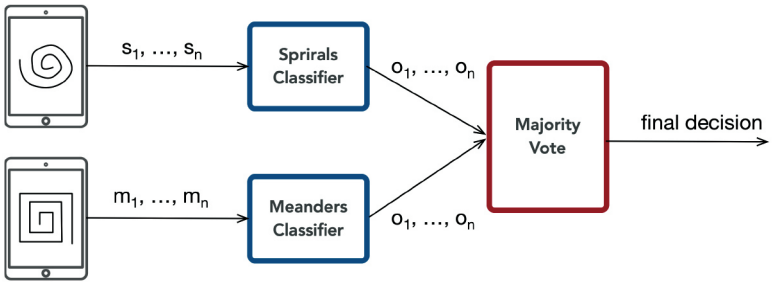


Figure 1. The proposed architecture. n represent to the number of tasks administered to each subject.

During the learning phase the samples of the Training set are provided to the classifier, and a DT is learned for discriminating between samples by PD patients and healthy subject (*Sample classification*). Once the learning phase is accomplished, the samples of both the spirals and meanders of a testing subject are provided to each classifier and the final decision is taken according to the majority vote rule (*Patient classification*). Figure 1 shows the whole architecture of the proposed approach.

3. EXPERIMENTS

The evaluate the performance of the proposed approach in providing explainable yet effective solutions, we performed a sample classification and a patient classification experiments, as described below.

3.1 The dataset

NewHandPD dataset contains handwritten data collected from graphical tests performed by 31 PD patients and 35 healthy subjects. Each subject produced 4 samples of spirals and meanders, and from each sample 9 features, reported in Table 1, were extracted. As a result, the dataset is composed of 264 spirals and 264 meanders drawn by the participants following a printed template on paper with a pen. As a consequence, we have two unbalanced datasets, i.e., spirals and meanders, each of which is composed of 124 samples belonging to PD patients and 140 belonging to healthy subjects. Figure 1 shows, together on one sample, the two main geometric entities, namely the distance between the centre of the template and the template/written trace (ET/HT radius), from which all the features are computed. Following (Pereira et al., 2016), we divided the dataset into a Training set and a Test set made of 75% and 25% of the original dataset, respectively, in such a way as to maintain the relative occurrence of patients and healthy subjects.

Feature	Description
x_0	RMS of the difference between HT and ET radius
x_1	Maximum difference between HT and ET radius
x_2	Minimum difference between HT and ET radius
x_3	Standard Deviation of the difference between HT and ET radius
x_4	Mean Relative Tremor
x_5	Maximum HT radius
x_6	Minimum HT radius
x_7	Standard Deviation of HT radius
x_8	Number of times the difference between HT and ET radius changes sign

Table 1. NewHandPD data set: the features list (HT: handwritten trace, ET: example template).

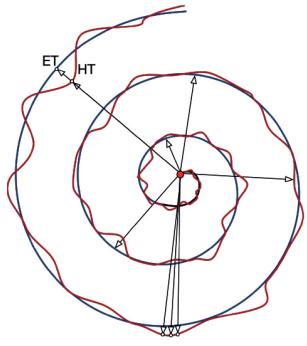


Figure 2. The feature extraction process for a spiral. The blue line is the ET, while the red line is the HT. The arrows indicate the radii for both ET and HT, the white circles indicate the intersection of a radius with the ET and the HT traces and the red circle represents the center of the ET. In order to compute all the features, the radius is shifted by using a predefined spanning angle.

3.2 Sample classification

In this experiment, we adopted a cross-validation procedure with 20 runs: at each run, samples were shaffled between training and test set in such a way to preserve the distribution among the subjects. In each run, the samples of the Testing

set were provided to the appropriate classifier and its output compared with the ground truth. From such a comparison we compute accuracy, sensitivity and specificity, the performance measures commonly adopted in medical test evaluation. For the sake of comparison with state-of-the-art methods, the accuracy, the sensitivity and the specificity were averaged over runs and compared with those obtained with four different algorithms, namely SVM, RF, OPF and CNN used in (Pereira et al., 2016). Table 2 reports the results obtained. They show that, as expected, CNN exhibits the highest accuracy, while RF, SVM and DT exhibits lower and very similar accuracy. However, it should be pointed out that none of the classifiers achieves the best performance on all the measure, and that only two of the best performance is bigger than 90%, which is the threshold above which a test is deemed as highly reliable. They also show all the classifiers perform better on meanders than on spirals.

	Meanders			***Spirals***		
	Accuracy	***Sensitivity***	***Specificity***	***Accuracy***	***Sensitivity***	***Specificity***
Random Forest	78.41	87.10	70.71	78.03	81.45	**75.00**
SVM	76.89	95.16	60.71	73.01	**91.94**	56.43
Decision Tree	77.27	89.51	66.43	65.91	75.81	57.14
OPF	66.72	**96.49**	36.94	64.23	62.91	65.56
CNN	**87.14**	91.25	**76.19**	**80.19**	87.95	59.52

Table 2. Sample-classification on Meanders and Spirals: best performance is in bold. Values are in percentage.

In order to understand the results provided by the DT, we have performed an analysis of its performance sampleby-sample: quite interestingly, we found that the fourth sample drawn by the subjects was responsible for most of the errors. We believe that psychomotor factors like tiredness (in case of PD patients) and boredom (in healthy subjects) had some influence on the subjects attention toward the end of the test. For this reason, the last sample produced by each subject was removed from the Test set before performing the patient classification experiment.

3.3 Patient classification

In this experiment, we replicated the clinical scenario, in which the test is administered to a subject that did not contribute to the learning phase. To implement such a scenario, we adopt a 20-fold validation procedure, as in the previous experiment. In each run, however, we removed from the training set the samples produced by a subject, performed the learning phase for building the classifiers, and then used them to classify the subject samples, as in the previous experiment. Eventually, we adopt a majority vote to combine the output of the two classifiers on the six samples for classifying the subject. Table 4 reports the final results provided by the RF and DT: the upper part of the table show the confusion matrix, the lower part the performance measures. In such an experiment we have considered only RF because it adopts the same classification paradigm as DT, but because it aims at maximizing the performance, it builds many DTs instead of one, so that it is extremely difficult to check the criteria adopeted for the final classification, as the DT to be considered depends on the features values and there is no way to figure out why many trees were built. Surprisingly, DT performs better than RF, while the overall performance of both the classifiers is much higher with respect to the one provided by the same classifiers using the Sample classification approach. Even more important, they show that the proposed architecture allows to implement a highly reliable test, as all the performance measure exceeds 90%.

	Random Forest			*Decision Tree*		
	Healthy	***Patient***	***Reject***	***Healthy***	***Patient***	***Reject***
Healthy	30	3	2	30	1	4
Patient	1	29	1	0	30	1
Sensitivity	96.67%			100%		
Specificity	90.91%			96.78%		
Accuracy	93.66%			98.37%		

Table 3. The results provided by RF and DT for the Patient classification experiment. Upper subtable: confusion matrix; lower subtable: performance measures

4. CONCLUSIONS

We have presented a method for discriminating between PD patients and healthy subjects by analysing their drawings of spirals and meanders. The experimental results show that, according to a procedure that mimic the clinical scenario, the proposed architecture leads to a test whose results are considered highly credible with respect to the current standards in medical test evaluation. Moreover, and even more important, they show that the results are achieved by selecting among the repertoire of machine learning approaches the decision tree, which is not the most performing one in classification problem, but offers the advantage of formalizing the decision process in terms of if-then rules chaining, so as to mimic very closely the inference process of medical diagnosis. All together, the results reported above, suggest that the proposed approach may be more easily adopted in daily clinical routine, as it is reliable and its decisions can be understood, and therefore evaluated, by the clinician for the final assessing of the case.

As the findings of this study have been obtained by using a dataset that contains samples produced by a population of about 30 subjects for each class, our current efforts are devoted to collect samples from larger populations in collaboration with two medical institutions. These samples will constitute the next Test set we will use for a further assessment of the system performance.

REFERENCES

[1] Jankovic, J. (2011). Parkinson's disease: clinical features and diagnosis, *Journal of Neurology, Neurosurgery, and Psychiatry*, 79(4), 368-376.

[2] Marsden, C.D. (1989). Slowness of movement in Parkinson's disease. *Movement Disorder*, 4, 26-37.

[3] Sheridan, M.R., Flowers, K.A., Hurrell, J. (1987). Programming and execution of movement in Parkinson's disease. *Brain*, 110, 1247-1271.

[4] Stelmach, G.E., Teasdale, N., Phillips, J., Worringham, C.J. (1989). Force production characteristics in Parkinson's disease. *Experimental Brain Research*, 165(1), 165-172.

[5] Broderick, M.P., Van Gemmert, A.W.A., Shill, H.A., Stelmach, G. (2009). Hypometria and bradykinesia during drawing movements in individuals with Parkinson's disease. *Experimental Brain Research*,197, 223-233.

[6] Van Gemmert, A.W.A., Adler, C.H., Stelmach, G.E. (2003). Parkinson's disease patients undershoot target size in handwriting and similar tasks. *Journal of neurology, neurosurgery, and psychiatry* 74, 1502-1508.

[7] Tucha, O., Mecklinger, L., Thome, J., Reiter, A., Alders, G.L., Sartor, H., Naumann, M., Lange, K.W. (2006). Kinematic analysis of dopaminergic effects on skilled handwriting movements in Parkinson's disease. *Journal of Neural Transmission*, 113, 609-623.

[8] Senatore, R., Marcelli, A.(2018). A paradigm for emulating the early learning stage of handwriting: Performance comparison between healthy controls and Parkinson's disease patients in drawing loop shapes. *Human Movement Science*, in press.

[9] Zham, P., Arjunan, S., Raghav, S., Kumar, D.K. (2017). Efficacy of guided spiral drawing in the classification of Parkinson's Disease. *IEEE Journal of Biomedical and HealthInformatics*, 22(5), 1648-1652.

[10] Pereira, C. R., Weber, S. A., Hook, C., Rosa, G. H., & Papa, J. P. (2016). Deep Learning-Aided Parkinson's Disease Diagnosis from Handwritten Dynamics. In *Proc. 29th SIBGRAPI Conference*, (pp.340-346). IEEE

[11] Pereira, C. R., Passos, L. A., Lopes, R. R., Weber, S. A., Hook, C., & Papa, J. P. (2017). Parkinson's Disease Identification Using Restricted Boltzmann Machines. *International Conference on Computer Analysis of Images and Patterns* (pp. 70-80). Springer.

An overview on handwriting tasks and features for the diagnosis of Alzheimer's Disease

Nicole Dalia Cilia, Claudio De Stefano, Francesco Fontanella, Alessandra Scotto di Freca

Department of Electrical and Information Engineering (DIEI),
University of Cassino and Southern Lazio
Via G. Di Biasio, 43 - 03043 Cassino (FR), Italy
(nicoledalia.cilia, destefano, fontanella, a.scotto)@unicas.it

Abstract - Alzheimer's disease (AD) is the most prevalent brain neuro-degenerative disorder, which leads to severe cognitive impairment and loss of autonomy (i.e., dementia) in old age. Unfortunately, these diseases cannot be cured, but an early diagnosis can help to better manage their symptoms and their evolution. These aspects explain the importance of developing support systems for the early diagnosis of neurodegenarative diseases. Handwriting is one of the abilities that is aected by AD. For this reason, researchers have also investigated the possibility of using handwriting alterations caused by AD as diagnostic signs. In this paper we present a review of the literature of writing protocols for supporting the diagnosis of Alzheimer's and mild cognitive impairments (MCI). We also detailed the most common and eective features previously used in the literature to represent handwriting dynamics of the subjects aected. Finally, open issues are also discussed to identify promising areas for future research.
Keywords: Alzheimer's Disease Diagnosis, Clinical Protocols, Handwriting Analysis.

1. INTRODUCTION

Since handwriting results from a complex network of cognitive skills, such as visual perception, motor planning and coordination abilities, AD causes signicant changes in writing performance (Kandel et al., 2000). Trying to understand the dimension of the phenomenon, many researchers have used dierent approaches, such as neuroimaging and intra-cranial registrations, clinical treatments, etc. However, the writing process can show diagnostic signs (Neils-Strunjas et al., 2006). Handwriting diculties were already reported by Alois Alzheimer in 1907 (Lambert et al., 2007) when he observed that patients reduplicated the same syllables and forgot others. Recently, several studies have analysed the writing process dynamics to identify and monitor AD, revealing that writing, persevering, substitutions, misalignments of strokes, link and spacing errors indicate deterioration of ne motor skills and coordination (Hayashi et al., 2011; Renier et al., 2016; Fernandes & Lopes Lima, 2017). In the literature reviewed, we found three review papers focusing on the handwriting analysis of AD patients (Croisile, 1999, 2005; Neils-Strunjas et al., 2006). (Croisile, 1999) reviewed some experimental works, and found that, in AD patients, writing disorders are more severe than language diculties, as can be seen in written descriptions of complex pictures and in lexical spelling. He also claims that spatial organization of handwriting is rapidly affected and therefore AD patients have mild diculties in maintaining a straight horizontal writing line, also leaving unnecessary gaps between words and letters. He concluded by saying that agraphia in AD patients is related to a disruption in the anatomic-functional cerebral network designed for writing processes, mainly in the parietal regions. The same author published a new and more extensive review in 2005 (Croisile, 2005). This review aimed to compare the handwriting anomalies due to ageing with those related to AD. The author reports that elderly people raise

their pens less often, also the pressure and width of writing decrease with age. As for people aected by AD, he reports that their handwriting gets progressively disorganized during the disease, whereas their spelling is altered by regularization errors which are evidence of lexical agraphia. The author concludes that agraphia of Alzheimer's disease comes from a progressive disorganization of the various components of language and writing, owing to brain lesions in several associative areas (parietal, temporal, occipital and frontal regions). From the Croisile's studies we can already draw some considerations about both the possible tasks to be submitted and about the features to be used to analyse the traces. As regarding the tasks, he reports a diculty of the patients in following a straight line, and in keeping a constant dimension in the produced traits. As we will show in the nal section, one of the tasks might require patient to write or copy a sentence on a line (cue). To analyse this sentence, Croisile suggests to use dimension of traits, number of pen up and pressure like features.
(Neils-Strunjas et al., 2006) presented a literature review of the research investigating the nature of writing impairment associated with AD. They reported that in most studies words are usually categorized in regular, irregular, and non-words. Orthographically regular words have a predictable phoneme-grapheme correspondence (e.g., cat), whereas irregular words have atypical phoneme-grapheme correspondences

		Statistical studies		Classification studies	
#	**Tasks**	**Features**	**References**	**Best Features**	**References**
1	• Straight line; • Loops or cursive L; • Single circle or continuos drawing.	• Dimension; • Distance from cue; • Velocity and peak; • Number of strokes; • Smoothness and jerk;	(Bellgrove et al., 1997); (Slavin et al., 1999); (Schröter et al., 2003); (Yan et al., 2008); (de Paula et al., 2016);		
2	• Copy; • Spontaneous writing; • Dictated words.	• Repetitions, omissions and substitutions; • Tremor; • Slant; • Pressure;	(Platel et al., 1993); (Pestell et al., 2000); (Luzzatti et al., 2003); (Hayashi et al., 2011); (Impedovo et al., 2014b); (Renier et al., 2016); (Fernandes & Lopes Lima, 2017); (Onofri et al., 2015); (Müller et al., 2017).	• Pressure; • Time on air; • Velocity; • Combination of features.	(Werner et al., 2006); (Pirlo et al., 2015); (Impedovo et al., 2014a); (Garre-Olmo et al., 2017).
3	• Name drawings of objects; • Picture description; • Spelling of name. • Memory.	• Word fluency; • Repetitions, omissions and substitutions.	(Small & Sandhu, 2008); (Groves-Wright et al., 2004); (Onofri et al., 2013).		

Table 1. Summary of task and features of statistical and classication studies

(e.g., laugh). Non-words or pseudo-words, instead, are non-meaningful pronounceable letter strings that conform to phoneme-grapheme conversion rules, and are often used to assess phonological spelling.
From the reviewed papers the authors conclude that writing impairment is heterogeneous in AD patients, aecting wors, sentences and discourse levels of written language production.

2. TASKS AND FEATURES OF STATISTICAL STUDIES

As shown, the idea of detecting AD by writing task is not new and, in the last few decades, several experimental studies have been proposed with this aim. These studies usually using a recording system for handwriting. However in this overview we divided these studies into two groups. In the rst group we included the studies that usually use statistical, or elementary, analysis, e.g. ANOVA and MANOVA, or compute correlation coecients to assess the statistical signicance of their results. Within these groups, we have ordered them according to the diculty of the tasks: from those with less cognitive load to those with greater cognitive load. In the Table 1 the papers have been organized according to the task diculty criterion.
Starting with the rst row of Table 1, we present the studies focused on elementary graphic gestures. In (Bellgrove et al., 1997) two groups are considered: Dementia of Alzheimer Type (DAT) patients and Healthy Controls (HC). The task to be performed consisted in connecting four targets, placed on a digitizing tablet, by a series of alternating horizontal drawing movements with a non-inking pen, in response to light stimuli. The task was executed in two dierent conditions: the next target was indicated before (cue condition) or after (no cue condition) the movement initiation. The cue condition allowed the subject to program the next movement, whereas the no cue condition forced the participant to reprogram the movement online. The authors found that DAT patients had programming decits, taking longer to initiate movements, particularly in the absence of cues. (Slavin et al., 1999) asked a group of AD patients to write four consecutive, cursive letter 'l's, on a graphic tablet, with four dierent visual conditions. These conditions were: a baseline condition with feedback of movement but not of output (no ink), with the presence of external cues (lines), with no visual feedback of the performed movements, and with feedback of output (ink). The authors found that AD patients had writing strokes of signicantly less consistent lengths than controls, and were disproportionately impaired by reduced visual feedback. Moreover, AD patients' strokes were of signicantly less consistent duration, and had signicantly less consistent peak velocity than controls, independently of the feedback conditions.
(Schröter et al., 2003) analysed handwriting kinematic to quantify dierences in ne hand motor function in patients with probable AD and MCI compared to depressed patients and healthy controls. The protocol consisted of two tasks: drawing concentric superimposed circles and drawing concentric superimposed circles simultaneously performing an additional distraction task (pressing a counting device as often as possible). The drawn circles were segmented into strokes, corresponding to the vertical up and down movements. The authors, used features such as: arithmetical mean of the velocity peaks of all strokes, the standard deviation of the intra-individual velocity prole, writing frequency, the number of strokes per second, number of changes of direction of velocity and relative velocity. They found that both patients with MCI and patients with probable AD exhibited loss of ne motor performances and that the movements of AD patients were signicantly less regular than those of the healthy controls.
In the study presented in (Yan et al., 2008), instead, the patients performed four types of handwriting movements on a digitizer: the up-down vertical movements that required the

nger joint movements; the left-right horizontal movements that primarily required wrist joint movements; the forward-slanted and the backward slanted movements that required the coordination of both nger and wrist movements. Movement time (MT) and smoothness were analysed between the groups of patients taken into account (probable AD, MCI and HC) and across the movement patterns. Kinematic proles were also compared among the groups, using MT and jerk as dependent measures. AD and MCI patients exhibited slower, less smooth, less coordinated, and less consistent handwriting movements than their healthy counterparts. More recently, the goal of the research of (de Paula et al., 2016) was to evaluate ne motor dexterity performance of MCI and AD patients and to investigate its association with dierent aspects of activities of daily living. The subject must put nine pegs in nine holes organized on a small board and subsequently remove them, as fast as possible. The authors nd that patients with AD or multiple-domain aMCI had slower motor responses when compared to controls. AD patients were slower than those with single-domain aMCI. As regards the studies belonging to the second subgroup (second row of Table 1), they analysed the handwriting of dierent kind of copied, spontaneous writing or dictated words or sentences. (Platel et al., 1993) described the evolution of agraphic impairments in DAT patients including lexicosemantic disturbances at the beginning of the disease, with impairments becoming more and more phonological as the dementia becomes more severe. They proposed a writing test from dictation to 22 patients twice, with an interval of 9-12 months between the tests. They found that the agraphic impairment evolved through three phases in patients with AD. The rst one is a phase of mild impairment (with a few possible phonologically plausible errors). In the second phase non-phonological spelling errors predominate, phonologically plausible errors are fewer and the errors mostly involve irregular words and non-words. The last phase involves more extreme disorders that aect all types of words. They observed many alterations due to impaired graphic motor capacity and concluded that grapho-motor impairments come in addition to the lexical and phonological impairments.

The study of (Pestell et al., 2000) investigated handwriting performance on a written and oral spelling task. The authors selected thirty-two words from the English language: twelve regular words, twelve irregular words and eight non-words. The study aims to nd logical patterns in spelling deterioration with disease progression. The results suggested that spelling in individuals with AD was impaired relative to HC but the comparison between those with mild AD and moderate AD failed to nd evidence of a logical pattern of deterioration.

(Luzzatti et al., 2003) used a written spelling test made up of regular words, non-words and words with unpredictable orthography. The purpose of the study was to test the cognitive deterioration from mild to moderate DAT. The authors found little correlation between dysgraphia and dementia severity. Thus they found that the hypothesis of a progressive deterioration, initially aecting semantic aspects, then lexical ones, and nally more surface abilities, would not appear to be generally applicable to all patients aected by AD. On the contrary, the data from this study conrmed that DAT is a mosaic of circumscribed cognitive decits.

In (Hayashi et al., 2011) the authors investigated the correlation between writing ability and regional cerebral blood ow (rCBF) in Japanese patients with mild AD compared to control group, using single photon emission computed tomography (SPECT). The task consisted in the writing of fty words under dictation, equally divided between concrete words and abstract words. Through an error analysis they found that, compared with control subjects, Kana writing to dictation and copying Kanji words were preserved in AD patients, but writing to dictating Kanji words was impaired. The authors concluded that the impaired recall of Kanji words in early AD is related to dysfunctional cortical activity, which appears to be predominant in

the left frontal, parietal, and temporal regions. As for the analysis of the handwriting of simple words, it is worth mentioning the work of (Impedovo et al., 2014b), in which the authors investigated the handwriting kinematics of the word "mamma" (mom in Italian) in AD patients. They motivated the choice of this word stating that it is one of the rst learned in speaking and writing. The authors examined the velocity proles and observed that in healthy person the maximum speed values were almost regular in height, whereas this regularity was strongly reduced at the beginning of the disease and progressively lost as the disease advanced. Also tasks like the copying of a shopping list or of a letter were considered. (Onofri et al., 2015) presented a protocol, organized in sessions performed in three days, which includes a copy of a shopping list and a letter. They found that on the third day there were graphic diculties and alterations in spatial organization accompanied by poor control of the movement.
The text appeared inconclusive, with a change between cursive and print and the tract was discontinuous between the letters. Recently, also signatures and spontaneous writing have been investigated for early diagnosis of AD. (Renier et al., 2016), for example, in their study recruited participants with diagnosis of MCI and with the diagnosis of initial dementia. For each subject they collected two samples of signature (an actual and a older one) and an extract of spontaneous writing. Furthermore, they administered a neuropsychological test battery to investigate the cognitive functions involved in decision-making. They found signicant correlations between spontaneous writing indexes and neuropsychological test results but the index of signature deterioration did not correlate with the level of cognitive decline.
In (Fernandes & Lopes Lima, 2017), instead, the authors compared the signatures of AD patients and healthy controls. The methodology used to examine the samples of the handwritten signatures involved the analysis of two categories of handwriting features: general features (legibility, tremor and line quality, level of connection between letters, velocity, pressure, slant, curvature, overall dimension, spacing between words and shape and direction of the baseline) and constructional features (shape and letter formation, as well as unusual features in the letters design). These features were statistically analysed and repetitions, omissions and substitutions were observed as indications of cognitive deterioration.
It is worth noting that also (Pirlo et al., 2015) investigated how signatures are aected by AD, but their study is reviewed below, with the other classication studies.

Another kind of task includes naming drawings of objects, picture description, spelling of names and memory of words tasks (third row of Table 1). (Small & Sandhu, 2008) analysed the handwriting of subjects belonging to three groups: younger adults, healthy older adults, and older adults with AD. Participants were asked to name drawings of objects in four conditions: dated unique, contemporary unique, dated common, and contemporary common. The results indicated that all participants named the items that were common to both episodic periods more successfully than the items unique to one period. An interaction was observed such that the healthy older and AD groups were more successful in retrieving names of objects presented in the dated compared to contemporary unique conditions, whereas the younger adults showed the reverse pattern.
In (Müller et al., 2017) the authors investigated movement kinematics of patients with early dementia due to AD, patients with amnestic mild cognitive impairment (aMCI), and cognitively healthy controlì (HC). Participants were asked to copy a three-dimensional house using a digitizing tablet. The results showed that time-in-air diered signicantly between patients with aMCI, AD, and HC.

Finally, the researches of (Groves-Wright et al., 2004) and (Onofri et al., 2013) investigates more complex writing tasks, requiring a high cognitive load (last row of Table 1). (Groves-Wright et al., 2004) used parallel measures (picture description, word uency, spelling to dictation, and confrontational naming) to compare verbal and written language of individuals with mild AD, moderate AD, and HC. The results showed that mild AD subjects diered from healthy controls only for verbal and written versions of the word uency task. Moderate AD subjects diered from mild subjects and controls for all written and verbal tasks.
(Onofri et al., 2013) presented a study with mild AD and healthy control group based on a protocol including mnemonic task concerning semantic knowledge and spatial and temporal orientation, which was a variation of Mini Mental State Examination (MMSE) test. The same task was repeated during the course of several days. The statistical analysis showed a marked deterioration in performance during the days.

2.1 Classication Studies

As previously said, in the second group we have collected the studies using classication algorithms to distinguish patients aected by dementia pathologies.
The results of the studies belonging to this second group are expressed in terms of classication performances, such as recognition rate, false acceptance rate and false rejection rate. (Werner et al., 2006) performed kinematic measures of the handwriting process of people with MCI compared with those with mild AD and healthy controls. The aim was to assess the importance of measures for the dierentiation of the groups and to assess the characteristics of the handwriting process across ve dierent, functional tasks of copying. In order to assess the independent eect of the MMSE score and of the kinematic measures, they examined three separate equations. In the rst equation, they entered the MMSE score as the only independent variable to assess its contribution to the correct classication of the diagnostic groups. In the second equation, they performed a stepwise discriminant analysis to assess the relative contribution of the ve kinematic measures assessed. Finally, in the third equation, they assessed the contribution of the MMSE score together with the kinematic variables that were found to be statistically signicant predictors in the second equation. The classication performances of the three equations considered and of the dierent tasks were computed in terms of recognition rate. The results showed that the kinematic measures together with the MMSE score were able to distinguish eectively the patients belonging to the dierent groups considered. As for the feature analysis, pressure and time-in-air obtained the best performances.
Also in (Impedovo et al., 2014a), the authors analysed the stability of the oine handwritten word "mamma" (mom in Italian) to distinguish AD patients from healthy controls. The stability of the word was computed by splitting its image in elementary parties and measuring the similarity of the adjacent parts. As classication algorithm the authors adopted the Yoshimura approach, based on the comparison of the stability features among the sample to be recognized and those of the training samples.
(Pirlo et al., 2015) presented a novel approach in which handwritten signatures were analysed for the early diagnosis of AD. Patients' signatures were represented by using the Plamondon's Sigma-Normal model, by means of twelve features. These features comprised, among others, the maximum speed of the signing divided by the time of writing, number of Log-Normal divided by the time and the number of peaks of the speed/time graph. The samples were classied by using three well-known classication algorithms: CART, bagging CART and SVM with linear kernel. The bagging CART outperformed signicantly both CART and the SVM classiers, in terms of False Acceptance Rate (FAR) and False Rejection Rate (FRR).

Finally, the goal of the work reported in (Garre-Olmo et al., 2017) was to distinguish participants belonging to three dierent groups (AD, MCI and HC) by comparing their handwriting kinematics. The authors used discriminant analysis as classication algorithm and adopted a protocol consisting of seven tasks, which included copying and drawing tasks. In the experiments, the authors, for the same task, investigated which were the most discriminating features and the best distinguished groups. They found that: (i) discriminating features depended on the type of group to be discriminated; (ii) some tasks, e.g. the clock drawing test, allowed some groups, e.g. AD vs. MCI, to be well discriminated (100% of specicity and sensitivity).

3. CONCLUSION AND OPEN ISSUES

From the literature examined it can be observed that the largest part of the studies uses standard statistical techniques. Furthermore, the few studies that use machine learning techniques do not have a varied protocol (set of tasks). If we consider the Tab. 1, it is evident that the classication studies use tasks that fall only in the second group of tasks. However, the features most commonly used in the classication studies can also be extracted from the other two groups of tasks (1 and 3). Extending the experimental protocol and collecting the features coming from several possible tasks would also make it possible to expand the experimental dataset and consequently increase the analysis power of the classication techniques.

REFERENCES

[1] Bellgrove, M. A., Phillips, J. G., Bradshaw, J. L., Hall, K. A., Presnell, I., & Hecht, H. (1997). Response programming in dementia of the alzheimer type: A kinematic analysis. *Neuropsychologia, 35* , 229-240.

[2] Croisile, B. (1999). Agraphia in alzheimer's disease. *Dement Geriatr Cogn Disord, 10*, 226-230.

[3] Croisile, B. (2005). ecriture, vieillissement, alzheimer. *Synthese, Psychol NeuroPsychiatr Vieil, 3*, 183-197.

[4] Fernandes, C., & Lopes Lima, J. M. (2017). Alzheimer's disease and handwriting-what do we know so far? *Proceedings of IGS 2017*, (pp. 130-134).

[5] Garre-Olmo, J., Faundez-Zanuy, M., de Ipi~na, K. L., Calvo-Perxas, L., & Turro-Garriga, O. (2017). Kinematic and pressure features of handwriting and drawing: Preliminary results between patients with mild cognitive impairment, alzheimer disease and healthy controls. *Curr Alzheimer Res, 14*, 1-9.

[6] Groves-Wright, K., Neils-Strunjas, J., Burnett, R., & O'Neill, M. J. (2004). A comparison of verbal and written language in alzheimer's disease. *Journal of Communication Disorders, 37*, 109-130.

[7] Hayashi, A., Nomura, H., Mochizuki, R., Ohnuma, A., Kimpara, T., Ootomo, K., Hosokai, Y., Ishioka, T., Suzuki, K., & Morio, E. (2011). Neural substrates for writing impairments in japanese patients with mild alzheimer's disease: A spect study. *Neuropsychologia, 49*, 1962-1968.

[8] Impedovo, D., Pirlo, G., Barbuzzi, D., Balestrucci, A., & Impedovo, S. (2014a). Handwritten processing for pre diagnosis of alzheimer disease. In *Proceedings of BIOSTEC 2014* (pp. 193-199). Portugal: SCITEPRESS.

[9] Impedovo, D., Pirlo, G., Mangini, F. M., Barbuzzi, D., Rollo, A., Balestrucci, A., Impedovo, S., Sarcinella, L., O'Reilly, C., & Plamondon, R. (2014b). Writing generation model for health care neuromuscular system investigation. In *Proceedings of CIBB 2013* (pp. 137-148). Springer.

[10] Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of Neural Science*. (4th

ed.). McGraw-Hill Medical.
[11] Lambert, J., Giard, B., Nore, F., de la Sayette, V., Pasquier, F., & Eustache, F. (2007). Central and peripheral agraphia in alzheimer's disease: From the case of auguste d. to a cognitive neuropsychology approach. *Cortex, 43*, 935-951.
[12] Luzzatti, C., Laiacona, M., & Agazzi, D. (2003). Multiple patterns of writing disorders in dementia of the alzheimer-type and their evolution. *Neuropsychologia, 41*, 759-772.
[13] Müller, S., Preische, O., Heymann, P., Elbing, U., & Laske, C. (2017). Diagnostic value of a tablet-based drawing task for discrimination of patients in the early course of alzheimer's disease from healthy individuals. *Journal of Alzheimer's disease, 55*, 1463-1469.
[14] Neils-Strunjas, J., Groves-Wright, K., Mashima, P., & Harnish, S. (2006). Dysgraphia in Alzheimer's disease: a review for clinical and research purposes. *J Speech Lang Hear Res, 49*, 1313-30.
[15] Onofri, E., Mercuri, M., Archer, T., Ricciardi, M. R., F.Massoni, & Ricci, S. (2015). Eect of cognitive fluctuation on handwriting in alzheimer's patient: A case study. *Acta Medica Mediterranea, 3*, 751.
[16] Onofri, E., Mercuri, M., Salesi, M., Ricciardi, M., & Archer, T. (2013). Dysgraphia in relation to cognitive performance in patients with Alzheimer's disease. *Journal of Intellectual Disability-Diagnosis and Treatment, 1*, 113-124.
[17] De Paula, J. J., Albuquerque, M. R., Lage, G. M., Bicalho, M. A., Romano-Silva, M. A., & Malloy-Diniz, L. F. (2016). Impairment of ne motor dexterity in mild cognitive impairment and alzheimer's disease dementia: association with activities of daily living. *Revista Brasileira de Psiquiatria, 38*, 235-238.
[18] Pestell, S., Shanks, M. F., Warrington, J., & Venneri, A. (2000). Quality of spelling breakdown in alzheimer's disease is independent of disease progression. *Journal of Clinical and Experimental Neuropsychology, 22*, 599-612.
[19] Pirlo, G., Cabrera, M. D., Ferrer-Ballester, M. A., Impedovo, D., Occhionero, F., & Zurlo, U. (2015). Early diagnosis of neurodegenerative diseases by handwritten signature analysis. In *ICIAP Workshops* (pp. 290-297).
[20] Platel, H., Lambert, J., Eustache, F., Cadet, B., Dary, M., Viader, F., & Lechevalier, B. (1993). Characteristics and evolution of writing impairmant in alzheimer's disease. *Neuropsychologia, 31*, 1147-1158.
[21] Renier, M., Gnoato, F., Tessari, A., Formilan, M., Busonera, F., Albanese, P., Sartori, G., & Cester, A. (2016). A correlational study between signature, writing abilities and decision-making capacity among people with initial cognitive impairment. *Aging Clin Exp Res., 28*, 505-511.
[22] Schröter, A., Mergl, R., Bürger, K., Hampel, H., Möller, H.-J., & Hegerl, U. (2003). Kinematic analysis of handwriting movements in patients with alzheimer's disease, mild cognitive impairment, depression and healthy subjects. *Dementia and geriatric cognitive disorders, 15*, 132-42.
[23] Slavin, M. J., Phillips, J. G., Bradshaw, J. L., Hall, K. A., & Presnell, I. (1999). Consistency of handwriting movements in dementia of the alzheimer's type: a comparison with huntington's and parkinson's diseases. *J Int Neuropsychol Soc., 5*, 20-25.
[24] Small, J., & Sandhu, N. (2008). Episodic and semantic memory in uences on picture naming in alzheimer's disease. *Brain Lang., 104*, 1-9.
[25] Werner, P., Rosenblum, S., Bar-On, G., Heinik, J., & Korczyn, A. (2006). Handwriting process variables discriminating mild alzheimer's disease and mild cognitive impairment. *Journal of Gerontology: PSYCHOLOGICAL SCIENCES, 61*, 228-36.
[26] Yan, J. H., Rountree, S., Massman, P., Doody, R. S., & Li, H. (2008). Alzheimer's disease and mild cognitive impairment deteriorate ne movement control. *Journal of Psychiatric Research, 42*, 1203-1212.

An "in the moment" assessment of physiological responses and subjective wellbeing in a 12 week participatory choral program for people with a dementia

Emilie V. Brotherhood, Emma Harding, Nicholas C. Firth, Esther Jones, Paul M. Camic, Sebastian J. Crutch, on behalf of the Created Out of Mind research team

Abstract - We investigated in-the-moment affective and cognitive responses in people living with a dementia over a 12-week participatory choral program. Visual Analogue Scale (VAS) data were collected before and after six choir rehearsals over the 12 weeks to establish levels of wellbeing. Physiological data were collected shortly before, during and after the same choral sessions to establish relative changes in heart rate, electrodermal activity (indexing autonomic arousal), skin surface temperature and movement. We observed significant increases in self-reported wellbeing following four sessions, with these changes being predictive of the differences observed in electrodermal activity during the rehearsals. Additional analyses revealed that this relationship could not be accounted for by changes in movement or particular rehearsal sessions. The study concludes that skin conductance may be a helpful measure to gauge responses of people living with a dementia within a choral program, noting that the complexity of such an environment will warrant further data collection employing multiple physiological recording techniques.
Keywords: Creative arts, dementia, choral singing, physiological measures, wellbeing

1. INTRODUCTION

Choral singing has been shown to have cognitive, emotional, wellbeing, quality of life and social benefits for people with dementias and their caregivers in a range of studies [1] [2]. Studies investigating these benefits have typically captured behavioral responses to choral rehearsals by administering Visual Analogue Scales (VAS) and other measures before and after the sessions have taken place [3]. Physiological measures, which could capture in real-time the ever-changing demands and delights when people with a dementia engage in a choral singing program, may enrich the evidence-base for the use of these programs across community-based and residential care settings [4].

Few studies have captured physiological measures, such as heart rate and skin conductance, to objectively quantify responses of people living with a dementia while participating in choral programs [5]. Variations in skin conductance, measured by changes in electrodermal activity, act as a sensitive marker of changes in autonomic arousal, which may offer some insight into the affective and cognitive responses to a choral program involving people living with a dementia [6]. Indeed, studies collecting physiological data alongside stress and wellbeing measures in a choral singing group in people with dementia and their caregivers have reported positive initial results [7].

To date, studies into choral programs have focused on choirs comprising people living with dementia and their relatives accompanying the singing. This type of group composition may be driving, or at least contributing to, an effect on increased wellbeing and reduced stress during the choral sessions. Further investigation is needed to establish if these results are replicable in choral programs whereby the membership is exclusively made up of people living with a dementia.

This study took place alongside filming for a British Broadcasting Corporation documentary about a choir of people living with a diagnosis of a dementia ('Vicky McClure: Our Dementia Choir' [8]), providing a complementary scientific perspective to the narrative of how people living with a dementia respond to music activities.

The study used passive physiological recording of heart rate, electrodermal activity, skin temperature and motion-based activity, alongside participants' subjective self-report ratings of wellbeing and stress, to elucidate a richer understanding of the moment-to-moment changes in response to both the positive and challenging aspects of participating in a choral singing program where people with a dementia participated unaccompanied by their relatives.

We anticipated that the behavioral findings in this investigation would replicate previous reports of higher post-choral session VAS scores across all wellbeing measures and reduced post-session stress reports compared with pre-session reports. We predicted a difference between the physiological responses observed in the choral session compared with the outputs observed 15 minutes preceding and following the choir sessions.

2. METHODS

We employed a quasi-experimental repeated-measures exploratory approach for this investigation, in a naturalistic setting.

A. Participants

20 people with a diagnosis of a dementia (either Alzheimer's Disease (AD) (n = 7), Frontotemporal Dementia (FTD) (n = 2), Vascular Dementia (VD) (n=1), mixed types (MTD) (n=6), Parkinson's Disease Dementia (PDD) (n=1), or undefined (n=3)) volunteered to join the choir. All were subsequently invited by university researchers to take part in the scientific aspect of the documentary filming. Participants were presented with a written information sheet which described the project. All choir members consented to participating in writing (11 males). Participants' ages were calculated on the last data collection point over 12 weeks, resulting in a mean overall participant age of 72.41 years (13.77). The study was approved by a Canterbury Christ Church University Research Ethics Committee (V:075\Ethics\2015-18) and adhered to British Psychological Society ethical standards.

B. Measures

Subjective wellbeing was measured using the Canterbury Wellbeing Scales (CWS), developed specifically for a dementia (early to middle stages) and dementia caregiver population [9]. This is an easily-completed VAS-style questionnaire that measures subjective 'in-the-moment' wellbeing across five subscales using a scale of 0-100 (happy/sad, well/unwell, interested/bored, confident/not confident and optimistic/not optimistic) in addition to a composite overall wellbeing score (0-500) [3]. An additional VAS to measure stress was administered alongside the CWS (relaxed/stressed) where higher scores indicate lower levels of stress.

Physiological data: Accelerometer (ACC), Blood Volume Pulse (BVP), infrared thermopile and electrodermal activity data were collected using an Empatica® E4 device. The Empatica E4 is a discreet, unobtrusive class IIa medical device (EU (CE Cert. No. 1876/MDD (93/ 42/EEC Directive))) designed to acquire real-time data on measures of: motion-based activity (ACC),

E. V. Brotherhood, E. Harding and S. J. Crutch are with University College London, England WC1N 3AR (Corresponding author e-mail: e.brotherhood@ucl.ac.uk)
E. Jones is with the National Youth Choirs of Great Britain, Durham, England DH1 5TS
P. M. Camic is with Canterbury Christ Church University, Canterbury, England TN1 2YG

heart rate (HR), skin temperature (TEMP) and sympathetic nervous system arousal (EDA) (from which levels of engagement, stress and excitement can be derived).

C. Procedure

Up to 17 participants were selected from the group at random each week to complete the CWS before the choral sessions (owing to number of devices available). There were minor fluctuations in attendance across the 12 week program owing to availability and minor illness, but good consistency of singers (minimum of 17 present at each rehearsal out of a total of 20 singers involved). The Empatica E4 devices were synchronized to Unix time, and participants were subsequently fitted with one Empatica device on their dominant hand for 15 minutes to establish a baseline period before the choral session began. During the baseline period, participants were seated around tables and interacted with each other and relatives. The Empatica E4s passively recorded HR, TEMP, ACC, and EDA outputs before, during and after the choral sessions. No specific instructions were given by the researchers during the choral singing activities. Following each session, the Empatica devices continued to record physiological responses for at least a further 15 minutes while participants interacted with relatives. The devices were then removed. Participants subsequently completed the CWS for a second time. This method was repeated throughout the choral program (a total of six data collection sessions on Weeks 1, 2, 4, 6, 8, and 10). Data were analyzed using STATA. Linear regression models and Wilcoxon signed-rank tests were used to interrogate the dataset and establish any differences between pre-session, in-session and post-session wellbeing and physiological measures.

3. RESULTS

A. Behavioral findings

Post-session composite wellbeing scores had a general tendency to be higher than pre-session scores, with these values reaching Bonferroni-corrected significance (α=0.001) at Weeks 1 ($p=0.001$), 4 ($p<0.001$), and 8 ($p=0.001$) (see Fig 1.). An increase in happiness levels post-rehearsal reached significance in all but two weeks (Week 2 ($p=0.347$), and Week 10 ($p=0.636$)). Overall, self-reported levels of stress followed a similar pattern to the composite wellbeing scores. Post-session indications of stress were significantly lower compared with pre-session scores at Weeks 1 ($p=0.004$), 4 ($p=0.047$), 6 ($p=0.016$) and 8 ($p=0.002$). Weeks 2 and 10 self-reported stress levels pre- and post-sessions were not significantly different ($p=0.979$ and $p=0.082$ respectively), although Week 10 showed a trend towards a decline in stress post-session.

To establish any differences in wellbeing across weeks, ranksum unmatched sample comparisons were generated between Week 1 vs Week 2, 4, 6, 8, 10, and Week 2 versus Week 4, 6, 8, 10. These comparisons revealed no significant differences, however Week 1 vs 2 post-rehearsal confidence showed a trend towards a decline ($p=0.076$).

B. Physiological findings

Separate simple linear regression models of the four types of Empatica E4 baselined, mean-centered values (with standard errors) for EDA, HR, ACC and TEMP were interrogated for differences across pre-session, in-session and post-session values at group level (see Fig. 2). Across all six data collection points, EDA readings were significantly elevated in-session compared with pre-session ($p=0.013$) and post-session ($p<0.001$) responses. Skin temperature similarly showed a significant increase in-session compared with pre-session readings (p

< 0.001), however the difference between in-session and post-session responses did not reach significance (p=0.750). Heart rate and motion-based activity showed an inverse pattern to the EDA findings, revealing significantly lower heart rate and motion-based activity in-session compared with pre- and post-session readings (p<0.001).

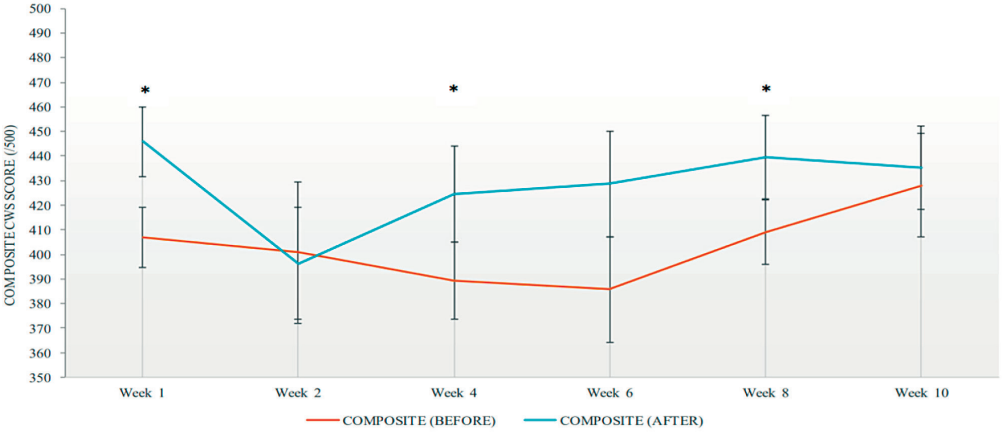


Figure 1. Overall composite CWS scores week-by-week. We observed significant increases in wellbeing following choral rehearsals at Weeks 1, 4 and 8. Week 6 reached significance at levels of less than α=0.05 (p = 0.007). Lower composite wellbeing scores were observed in Week 2, with participants reporting lower levels of interest, confidence, and optimism, however these pre- and post-rehearsal differences did not reach significance.

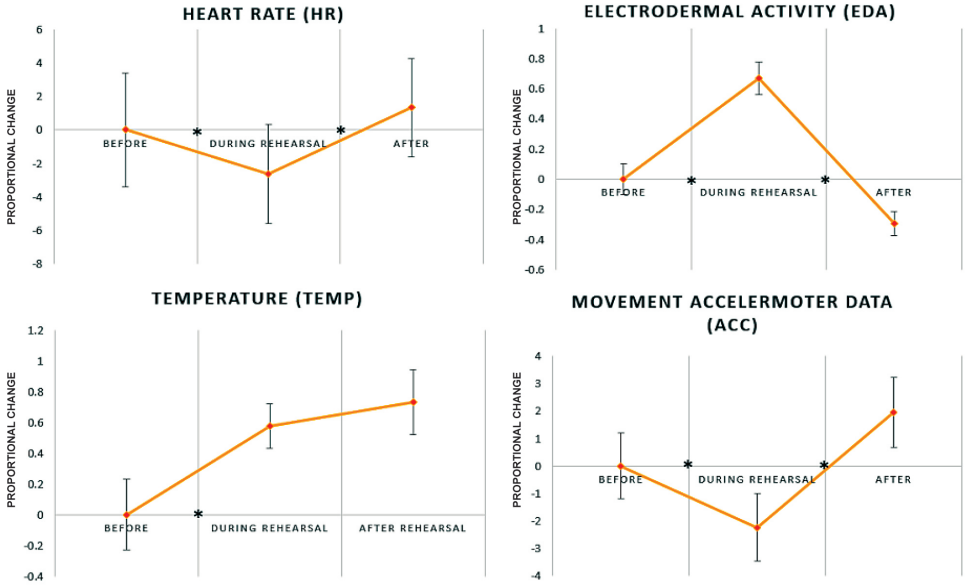


Figure 2. Heart rate (HR), electrodermal activity (EDA), temperature (TEMP) and accelerometer (ACC) readings before, during and after the choral rehearsals. Asterisks at the intersections of the X and Y axes indicate significant changes between conditions.

C. Physiological findings in relation to wellbeing scores

To account for potential confounds, three further linear regressions were undertaken to

analyze average EDA values during the rehearsal, controlling for session, participant and average ACC in the same time frame. There was no apparent relationship between change in EDA values and raw CWS composite scores taken before (p=0.21) or after (p=0.53) the rehearsal. However change in EDA was positively associated with change in CWS composite score ('after scores' minus 'before scores') (p=0.004) which does not appear to be attributable to any of the covariables.

4. DISCUSSION

To our knowledge, this is the first time a link between people's subjectively rated wellbeing and changes in electrodermal activity response has been shown in a participatory choral program with people living with a dementia. This is also the first publication describing physiological responses of people living with a dementia in a choral program without relatives accompanying them throughout the songs. Significant increases in self-reported wellbeing and a reduction in stress following four out of the six rehearsals are in line with initial predictions and replicate results from previous studies.

Understanding the difference in context across the choral program sessions may explain the unexpected behavioral findings at Weeks 2 and 10. For example, in Week 2, a rhythmically complex and unfamiliar song was introduced to the repertoire, which may account for the non-significant differences in pre- and post-rehearsal scores of relaxation, interest, confidence, optimism and composite wellbeing scores. Previous studies have demonstrated that when challenging music (e.g. singing in rounds) is introduced sensitively and slowly, these effects are eliminated [10]. This indicates that offering challenging music too quickly or novel pieces too close together may have a transient negative effect on wellbeing. While these variations in wellbeing and stress across the program were not in line with our initial predictions, they increase our confidence that the CWS and stress scale reports are an accurate reflection of the context and members' experience of each choral session. Although the primary purpose of the study was to examine associations between subjective and objective indices of wellbeing within singers with dementia, future studies involving separate or co-singing healthy participants can examine the relative magnitude of the impact of rehearsals upon these psychological and physiological metrics.

Subjective wellbeing scores were a significant predictor of the increase in arousal – an objective measure, over which the participants have no control (i.e. the bigger the boost in how participants felt, the greater the increase in their arousal readings). The additional analyses indicated that these changes in EDA values were being driven by changes in arousal, due to the fact that the significant changes in EDA held true even when movement, individual differences and choir session number were accounted for.

One common criticism of assessing the impact of participatory creative activities in people living with a dementia is that increases in wellbeing are in fact driven by the nature of participatory activities (e.g. increased opportunity for social interaction and leaving the home environment) rather than the activity itself. While we recognize our experimental design did not include a control group to explore this explicitly, our baseline and post-session intervals recorded physiological responses while participants were out and about with friends, family and relatives, with plenty of opportunity for social interaction. The fact that levels of arousal were higher during the choral rehearsals compared with baseline and post-session readings indicates that this was being driven by the music and the choir. There were no significant differences in either wellbeing or physiological ratings across weeks. Taken together with the

stronger evidence for changes in scores pre- versus post-rehearsals, this seems to support the primacy of in-the-moment over longer term benefits upon wellbeing for this particular choir group.

We recognize the potential of skin conductance to demonstrate how people living with a more advanced stage of dementia who are nonverbal may be responding to a choral program. Further replication studies are needed in populations with early-to-mid stages of dementia before strong evidence-based claims can be made for the overall value of choral singing for wellbeing in people living with a dementia, or to generalize these findings to people living with an advanced dementia.

Naturalistic studies of participatory choral activities involving people with a dementia give rise to multiple complexities. such as intra-group differences in: members' clinical phenotypes, theoretical music experience and preferences, and social status, as well as the recognized potential influences of age, gender and socioeconomic status [11]. The advantage of the current study is the use of multiple measures, which improves confidence in interpreting individual responses (e.g. a relationship between EDA and subjective wellbeing). Nonetheless, we must remain cautious when using a handful of physiological measures to infer changes in complex psychological constructs within naturalistic settings [12]. Combining methodologies (e.g. Empatica, mobile electroencephalography (EEG) and pupillometry data) which each focus on measuring different aspects of a participatory activity, and integrating these findings, may provide a more nuanced insight into the extent to which we can reliably infer the neural and cognitive mechanisms at play when people living with different presentations of dementia participate in music activities.

5. ACKNOWLEDGMENTS

We are indebted to members of the choir, and their supporters, Vicky McClure, and the team at Curve Media led by Peter Coventry. This work is part of the Created Out of Mind research program, funded as 'Created Out of Mind: Shaping Perceptions of Dementias, Grant Ref: 200783/Z/16/Z, by Wellcome as a part of the Hub Award. (Principal Investigator S. Crutch; Core Group: P. Ball, C. Evans, N. Fox, C. Murphy, F. Walsh, J. West, G. Windle, P. Camic).

REFERENCES

[1] D. Petrovsky, P. Z. Cacchione, and M. George, "Review of the effect of music interventions on symptoms of anxiety and depression in older adults with mild dementia," International Psychogeriatrics, vol. 27, pp. 1661-1670, April 2015

[2] T. Särkämö, M. Tervaniemi, S. Laitinen, A. Numminen, M. Kurki, J. K. Johnson, and P. Rantanen, "Cognitive, emotional, and social benefits of regular musical activities in early dementia: Randomized controlled study," Gerontologist, vol. 54, pp. 634–650, August 2014

[3] J. Johnson, A. Culverwell, S. Hulbert, M. Robertson, and P. M. Camic, P.M, "Museum activities in dementia care: Using visual analogue scales to measure subjective wellbeing," Dementia, vol. 16, pp. 591-610, July 2017

[4] P. M. Camic, S. J. Crutch, C. Murphy, N. C. Firth, E. Harding, C. R. Harrison, S. Howard, S. Strohmaier, J. Van Leeuwen, J. West, G. Windle, S. Wray, and H. Zeilig, "Conceptualising artistic creativity in the dementias: Interdisciplinary approaches to research and practice," Frontiers in Psychology, vol. 9, pp. 1-20. October 2018

[5] G. E. Thomas, S. J. Crutch, and P. M. Camic, "Measuring physiological responses to the arts in people with a dementia," International Journal of Psychophysiology, vol. 123, pp. 64–73, January 2018

[6] H. D. Critchley, "Review: Electrodermal Responses: What Happens in the Brain," The Neuroscientist, vol. 8, pp. 132–142, April 2002

[7] P Bourne, P. M. Camic, S. J. Crutch, S. Hulbert, N. Firth, and E. Harding, "Using psychological and physiological measures in arts-based activities in a community sample of people with a dementia and their caregivers: A feasibility study and pilot study," Journal of Ageing Studies & Therapies, vol. 1, pp. 1-11, April 2019

[8] https://www.bbc.co.uk/mediacentre/proginfo/2019/18/our-dementia-choir-with-vicky-mclure

[9] P. M. Camic, S. Hulbert, and J. Kimmel, "Museum object handling: A health promoting community-based activity for dementia care," Journal of Health Psychology. Advanced access, January 2017

[10] P. M. Camic, C. M. Williams, and F. Meeten, "Does a 'singing together group' improve the quality of life of people with a dementia and their carers? A pilot evaluation study," Dementia: The International Journal for Social Research and Practice, vol. 12, pp. 152-171, October 2011

[11] D. Fancourt, A. Ockelford, and A. Belai, "The psychoneuroimmunological effects of music: A systematic review and a new model," Brain, Behavior, and Immunity, vol. 36, pp. 15–26. February 2014

[12] E. Harding, M. Sullivan, and S. J. Crutch, "Engaged or exasperated? Interpreting physiological data in dementia research," presented at the Powerful Partners: Advancing Dementia Care through the Arts and Sciences Conference, Royal Society for Public Health, London, 24 November 2017. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15-64.

Ageing reduces performance asymmetry between the hands in force production and manual dexterity

Qun Fang [a], Chris Aiken [b], Arend W. A. Van Gemmert [c], Jian XU a Zhujun PAN [a*]

[a] Mississippi State University (Kinesiology) - 241 McCarthy Gymnasium
39762, Mississippi State, MS, United States of America

[b] New Mexico State University (Kinesiology and Dance) - MSC 3M
88003, Las Cruces, New Mexico, United States of America

[c] Louisiana State University (Kinesiology) - 112 HP Long Fieldhouse
70803, Baton Rouge, LA, United States of America
qf37@msstate.edu, aiken@nmsu.edu, gemmert@lsu.edu, jx102@msstate.edu, zp147@msstate.edu

1. INTRODUCTION

Aging involves an overall degeneration in the nervous, musculoskeletal, and sensory systems which affects movement patterns. Cabeza (2002) identified age-related reductions of hemispheric lateralization for cognitive processes on which he developed his model of hemispheric asymmetry reduction in older adults, or HAROLD. Following research attempting to extend the cognitive-perceptual model to motor behavior, one study examined movement patterns with respect to multidirectional reaching tasks (Przybyla, Haaland, Bagesteiro, & Sainburg, 2011). This study found robust asymmetries between the left and right arm in young adults, but older adults showed more symmetrical movement patterns, suggesting reduced motor asymmetry as result of the aging process. Although conclusive evidence is lacking for the connection between motor asymmetry and hemispheric lateralization, findings in current research provides support for the suggestion that symmetric neural recruitment associated with aging may contribute to an age-related reduction in motor asymmetry (Przybyla et al., 2011).

In addition to these age-related changes in neural functions, degeneration of physiological structures may also affect the expression of motor asymmetry (Francis & Spirduso, 2000). Decreased maximum grip force, due to the loss of muscle mass, is associated with a selective effect on the right hand (in this study the dominant hand as well) substantially reducing the right-hand advantage identified in young adults (Teixeira, 2008). Therefore, manual performance tends to become more symmetric as people age.

The current study used two manual function assessments; the Jamar hand function and Purdue Pegboard tests. These assessments enabled us to focus on force production and manual dexterity, respectively. The primary purpose of our study was to add knowledge about age-related impacts on performance-based motor asymmetry. According to Przybyla and colleagues' findings (2011), which are consistent with HAROLD model, reduced asymmetric movement performance can be attributed to compensatory effects associated with improved functional performance of the non-dominant hand. However, other research suggests that the reason for a symmetric movement patterns in aging occur due to a significant drop in dominant hand function (Francis & Spirduso, 2000; Teixeira, 2008). It is clear that more research is needed to provide conclusive evidence supporting either or possibly partly both suggestions explaining age-related reductions in motor asymmetry.

2. METHODS

Participants

Forty-one young adults (21 ± 3 years, 16 males) and 25 older adults (77 ± 3 years, 1 male) participated in the study. The inclusion criteria for older participants were: 60 years of age or older, right handed, no or minimal neurological deficits and/or cognitive impairments (Mini-Mental State Examination score > 21), no restrictive cardiovascular
and/or respiratory ailments, and/or no recent surgeries. Young participants were right-hand dominant university students who reported no recent injury and/or disease that might impair performance in the study. Informed consent was obtained from all participants before the study started. The project was approved by the Institutional Review Board of the Mississippi State University.

Testing protocol and outcomes

Participants completed the Edinburgh Handedness Inventory (Oldfield, 1971) at the beginning of the study and then underwent two hand function assessments; i.e., the Jamar hand function test and the Purdue Pegboard test. In the Jamar hand function test, grip strength was first measured, followed by a tip pinch, a palmar (three-jaw chuck) pinch, and a lateral pinch. Participants performed each of the tests in seated position, with the shoulder adducted, elbow flexed to ninety degrees, and forearm and wrist in neutral positions.
The Purdue Pegboard test (Lafayette Instrument Model #32020) measures manual dexterity. The test includes four subtests which require participants to use solely their right hand, left hand, and two subtests using both hands. Performance was scored according to the number of pegs correctly placed in 30 seconds. Each subtest included three trials.

Statistical analysis

A 2 (age-group: young and older) × 2 (hand: left and right) mixed-design ANOVA was conducted to examine age-related effects on manual performance in five of the unimanual tests, including Jamar grip strength, tip pinch force, palmar pinch force, lateral pinch force, and Purdue Pegboard test performed by a single hand. Age-group was entered as a between-group factor while hand was entered as a within-group factor. The dependent variables for both statistical analyses were the score of the hand function test for grip strength, pinch force, and manual dexterity.
We also examined the Pearson's correlation between the left and right hand when older and young adults were performing each task, and we attempted using a Pearson's correlation to identify potential associations among the tasks. As the Jamar hand function test and Purdue Pegboard test measures different aspects of manual function, i.e., force and dexterity respectively, correlations between the two tests may provide additional information on age-related changes in movement patterns. All data analyses were conducted using IBM SPSS 24.

3. RESULTS

ANOVA on manual asymmetry
For grip strength, main effects of age, $F(1, 64) = 38.52$, $p < .001$, $\eta^2 = .37$, and hand, $F(1, 64) = 63.27$, $p < .001$, $\eta^2 = .50$, were found to be significant. Young adults ($M = 38.81$, $SE = 1.85$) showed stronger grip force than older adults ($M = 20.14$, $SE = 2.37$). Participants also showed more force production with the right hand ($M = 30.90$, $SE = 1.49$) than the left hand ($M =$

28.09, SE = 1.54). The age by hand interaction was also significant, $F(1, 64) = 5.60$, $p = .021$, $\eta^2 = .08$. Further investigation suggested a more profound decrease in grip strength of the right hand leading to more symmetric function of the hands in older adults (Fig. 1A).

For tip pinch force, there was no significant main effect of age, $F(1, 64) = 1.91$, $p = .17$, $\eta^2 = .03$, or hand, $F(1, 64) = 1.52$, $p = .22$, $\eta^2 = .02$. A comparison between the young (M = 5.54, SE = .24) and older group (M = 4.99, SE = .31) suggested a similar test performance. Also, no significant right-hand advantage was identified although participants produced greater strength with the right hand (M = 5.33, SE = .21) than the left hand (M = 5.20, SE = .20). The interaction of age by hand proved to be significant, $F(1, 64) = 7.04$, $p = .01$, $\eta^2 = .10$, which was assumed to be caused by more profound reduction of right hand strength (Fig. 1B).

For palmar pinch force, main effects of age, $F(1, 64) = 57.87$, $p < .001$, $\eta^2 = .48$, and hand, $F(1, 64) = 7.71$, $p = .007$, $\eta^2 = .11$, were significant, as well as the interaction between age and hand, $F(1, 64) = 8.40$, $p = .005$, $\eta^2 = .12$, which could be attributed to greater drop in right hand strength during aging (Fig. 1C).

For lateral pinch force, the analysis indicated significant main effects of age, $F(1, 64) = 32.04$, $p < .001$, $\eta^2 = .33$, and hand, $F(1, 64) = 14.50$, $p < .001$, $\eta^2 = .19$. The right hand (M = 7.49, SE = .25) produced greater force than the left hand (M = 7.04, SE = .24), but the hand difference decreased in the older group compared to the young group.

Such an age-related change may contribute to the significant interaction effect found for age by hand, $F(1, 64) = 5.13$, $p = .027$, $\eta^2 = .07$ (Fig. 1D).

For the manual dexterity test, a significant main effect of age was identified, $F(1, 64) = 84.12$, $p < .001$, $\eta^2 = .57$, suggesting that young adults (M = 13.32, SE = .31) performed better than older adults (M = 8.75, SE = .39). The main effect of hand proved also significant, $F(1, 64) = 17.11$, $p < .001$, $\eta^2 = .21$, which indicated higher scores for manual dexterity of the right hand (M = 11.43, SE = .27) when compared to the left hand (M = 10.64, SE = .27). There was also a significant interaction effect between hand and age, $F(1, 64) = 9.52$, $p = .003$, $\eta^2 = .13$. This finding suggests that manual dexterity becomes more symmetric as people age probably due to greater age-related functional reductions of the right hand (Fig. 1E).

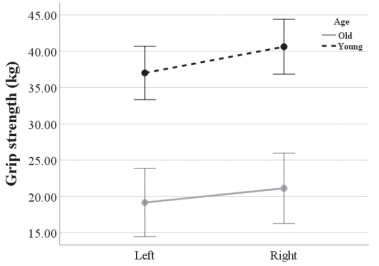


Fig. 1A

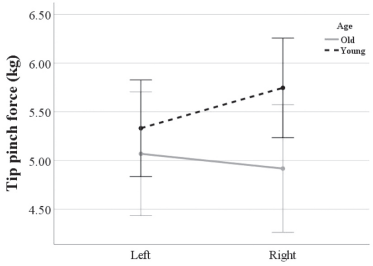


Fig. 1B

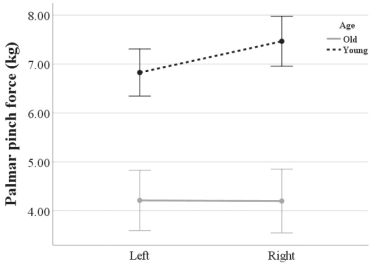


Fig. 1C

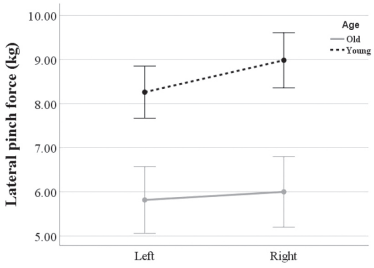


Fig. 1D

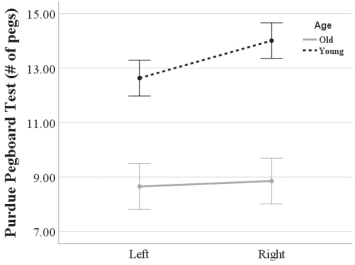


Fig. 1E

Correlation between the tests

A noticeable relationship existed between the Jamar hand function test for grip strength and the Purdue Pegboard test for manual dexterity. Significant correlations between the two tests were only found in the older group (Table 1). Further examination indicated that the Purdue Pegboard test scores of the left hand and both hands together correlated significantly with grip strength.

Purdue Pegboard scores	Jamal Grip Strength	
	Young	Old
Right Hand	-0.17	0.37
Left Hand	0.02	0.53**
Bimanual Test	-0.05	0.41*
Assembly Test	-0.02	0.40*

*$p < .05$, **$p < .01$

Table 1. Correlation coefficients (r) between the scores on the two tests.

Correlation between left and right hand for each test

The analysis indicated a significant correlation between the left and right hand for all sub tests (Table 2). It is worth noting that between-hand correlations of the scores for the young group is greater than the correlations for the older group. Therefore, we used Fisher Z-transformations to examine the potential differences in the correlation coefficients (Pearson's r) of the two groups. The result indicated that young adults showed significantly higher between-hand correlations than older adults in all the hand strength tests, but not in the manual dexterity test (Table 3).

Test	Young	Old
Grip Strength	0.98*	0.92*
Tip Pinch	0.93*	0.74*
Palmar Pinch	0.90*	0.66*
Lateral Pinch	0.93*	0.71*
Purdue Pegboard	0.73*	0.76*

*$p < .001$

Table 2. Correlation coefficient (r) between left and right for each task

Test	Z_{Young}	Z_{Older}	$Z_{Young-Older}$	*p*-value
Grip Strength	2.38	1.58	2.98	0.003
Tip Pinch	1.66	0.94	2.68	0.007
Palmar Pinch	.147	0.79	2.51	0.012
Lateral Pinch	1.67	0.89	2.91	0.004
Purdue Pegboard	0.93	0.98	-0.22	0.826

Manual preference

Table 3. Fisher Z-transformation based on correlation coefficients between young and older group

Analysis of manual preference was conducted based on the handedness score of the Edinburgh Inventory. One-way ANOVA suggested significant difference in self-reported manual preference between young and older group, $F(1, 64) = 5.60$, $p = .021$. Older adults reported stronger preferences for the right hand ($M = 94.40$, $SD = 12.61$) than their younger counterparts ($M = 80.98$, $SD = 26.53$).

4. DISCUSSION

The primary purpose of the current study was to examine potential changes in manual performance associated with aging. Different aspects of movement performance possibly resulting in often-observed differences in movement patterns were identified by comparing young adults with older adults performing the Jamar hand function and the Purdue Pegboard test. Assessments on grip strength, pinch force, and manual dexterity indicated reduced manual asymmetry in older adults. Further examination on these changes in movement characteristics suggested that reduced manual asymmetry in movement performance can be attributed to a greater degeneration of right hand functioning which eliminates the right hand's advantage observed in young adults (Francis & Spirduso, 2000). The finding appears to be inconsistent with the HAROLD model that predicts improved performance in the non-dominant hemisphere to compensate for a decreased function of the dominant hemisphere, thus suggesting that the non-dominant hand would compensate for a decrease in function of the dominant hand. Our findings provided complementary evidence to Teixeira's research (2008) in which an age by hand interaction for maximum grip strength was only found be to marginally significant ($p = .056$). In the current study, all tests for hand strength, including grip strength and pinch force, showed a significant interaction effect between age and hand. Therefore, age-related effects on motor asymmetry seem to be task specific necessitating more future research on this topic.

A significant relationship between grip strength and manual dexterity was identified in older adults, but not in young adults (Table 1). This finding suggests that older adults tend to activate larger muscles even when performing motor tasks that only require smaller muscles. However, more evidence-based support measuring muscle activity is necessary to verify this hypothesis. A closer examination of the correlations revealed different movement characteristics associations between the left and right hand in older adults which possibly caused age-related changes of movement performance patterns. No significant correlation between grip strength and manual dexterity was identified in older adults' right-hand performance. In contrast, left hand manual dexterity, bimanual, and assembly tasks were significantly correlated (Table 1). The results might imply that changes in grip strength of the left hand affect performance of bimanual and assembly test scores. If so, despite right-hand function declines due to aging the movement patterns remain constant. Thus, manual performance becomes less asymmetric in older adults, while movement patterns performed with the left and right hand remain distinct.

It is worth noticing is that, even though correlations (r) were significant in both age groups, young adults in comparison to older adults showed stronger between-hand correlations in each task (Table 2 and 3). The latter finding suggests current theories in motor behavior are incomplete and thus new theories need to be developed or we need to adapt existing ones.

REFERENCES

[1] Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: the HAROLD model. *Psychology and aging, 17*(1), 85.

[2] Francis, L., & Spirduso, M. (2000). Age differences in the expression of manual asymmetry. *Experimental aging research, 26*(2), 169-180.

[3] Mathiowetz, V., Kashman, N., Volland, G., Weber, K., Dowe, M., & Rogers, S. (1985). Grip and pinch strength: normative data for adults. *Arch Phys Med Rehabil, 66*(2), 69-74.

[4] Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

[5] Przybyla, A., Haaland, K. Y., Bagesteiro, L. B., & Sainburg, R. L. (2011). Motor asymmetry reduction in older adults. *Neuroscience letters, 489*(2), 99-104.

[6] Teixeira, L. A. (2008). Categories of manual asymmetry and their variation with advancing age. *Cortex, 44*(6), 707-716.

The effect of heat stress on the performance of a graphomotor choice-reaction time task

Kevin A. Becker, Chris A. Aiken, Cheng-Ju Hung, Arend W.A. Van Gemmert

Abstract - The relationship between hot environmental temperatures and reaction time is not entirely clear. Simple reaction time is reportedly shorter in heated conditions (Schlader et al., 2015), while studies examining choice-reaction time tasks have demonstrated either a length-ening of reaction time (Hancock & Dirkin, 1982), or no differences (Leibowitz et al., 1972). Recently, Aiken et al. (2016) demonstrated that for a graphical aiming task, choice-reaction time was longer when the environmental temperature was high for more difficult task varia-tions (i.e., smaller targets at a greater distance), but not easier task variations. The purpose of this study was to extend upon these findings and fully explore the stress effects of hot envi-ronments upon performance by manipulating the difficulty of a pointing task, i.e., manipulat-ing the target size and target distance independently in a two-experiment approach. In both experiments, participants completed 108 trials of a choice-reaction time task with the move-ment accuracy constraint varied between 0.25 and 0.80cm. The only difference between the two experiments was that in the first experiment (N=26) the movement distance was 12 cm while it was 5 cm in the second experiment (N=24). For each experiment, separate 2 (Heat) x 2 (Target Size) ANOVAs were conducted for each dependent variable (reaction time, move-ment time, initial direction error at peak velocity, and axial pen pressure). Results revealed that during Experiment 1, reaction time was significantly slower in the heated condition ($p = .048$). Heat did not significantly affect any other dependent variable (p's $> .05$). In Experiment 2, the main effect of heat failed to reach significance for reaction time ($p > .05$). A significant interaction between heat and target size was detected ($p = .036$), but pairwise comparisons failed to reveal any significant differences between the different heat and target size combi-nations (p's $> .05$). The results of these two experiments show support for the suggestion that heat results in stress, which previously has been shown to have its impact mediated by task difficulty. Furthermore, these experiments also suggest that the impact of heat-related stress on choice-reaction time is more affected by target distance (i.e., the magnitude of the move-ment) than target size (i.e., the accuracy constraint of the movement).
Keywords: Neuromotor noise, physical stressors, drawing task, heat stress.

1. INTRODUCTION

Research examining the impact of heat-related stress on reaction time has produced varied results. The performance of a simple reaction time task has been shown to be faster when heat is used as a physical stressor (Schlader et al., 2015), while choice-reaction time has been shown to become slower (Hancock & Dirkin, 1982). Other research has shown no effects of heat on reaction time performance (Leibowitz, Abernethy, Buskirk, Bar-Or, & Hennessy, 1972). More recently, Aiken, Becker, Lee, Post & Van Gemmert (2016) tested choice-reaction time using a paradigm that required participants to draw a line from a start target to one of nine randomly selected end targets. Their results showed that reaction time was slower in a hot environment (i.e., heat stress), but only for the more difficult task variations. Difficulty in the task varied in both accuracy demands (i.e., target size) and movement magnitude (i.e., distance between targets), making it unclear whether both or just one of the two factors

influenced the observed effect.

The Neuromotor Noise Perspective (Van Gemmert & Van Galen, 1997) suggests that when an individual encounters stress, there is a resultant increase in noise in the motor system. For simple motor skills, it is possible that the system reduces the noise by increasing limb stiffness (Van Gemmert & Van Galen, 1997). For a more difficult task, the noise in the system may be further increased and thus reduces the signal to noise ratio. This would create a scenario where the biomechanical filter is unable to reduce the noise in the system without performance deterioration due to a signal to noise ratio too low for meeting task requirements. When this biomechanical filter is not sufficient the system increases the processing times since signal increases while noise decays over time. Thus from this perspective, we predict that when a task requires fast responses with a difficulty that is relatively high, performance will suffer when a stressor like heat is added.

The purpose of this study, therefore, was to determine how heat stress (i.e., a hot environment) affects the performance of a choice-reaction time task with various target sizes and distances. This study used a two-experiment approach to extend the work of Aiken et al. (2016) by determining whether increased accuracy demands, movement magnitudes, or both factors influence the presence of effects due to stress as result of heat. We hypothesized that difficulty would influence reaction time, movement time, initial direction error at peak velocity, and axial pen pressure. However, due to the exploratory nature of the study we made no predictions about which specific conditions would differ in each experiment.

2. METHODS

A. Participants, Task, and Setting

In Experiment 1, 26 volunteers between the ages of 18 and 45 completed the study, and in Experiment 2, 24 volunteers in the same age range completed the study. With the exception of the target distance, the two experiments were identical. All 50 participants had normal or corrected to normal vision and full use of both upper extremities. Additionally, they had no neurological conditions, which could impact motor control, and they had no contraindications from a medical professional to experience environments exceeding 42° C (~107° F).

The motor task used in the study was a choice-reaction time task that was completed on a digitizing tablet (Wacom, Intuos Pro, 15" x 10"). The task involved drawing a line from a starting target to one of nine equally spaced end targets that changed colors at random. A color change of a target indicated to the participant that one had to draw to that specific target (see Figure 1). Participants completed the task with targets placed either 12cm from the start target (Exp. 1), or 5cm from the start target (Exp. 2). In both experiments, half of the trials required participants to end their drawing in a target of .25cm in diameter, while the other half of the trials required them to end their drawing in a target of .80cm in diameter. All participants sat in a chair with the tablet placed 7.5cm from the edge of a table directly in front of the chair. Each participant performed the two variations of the task in an enclosed heat chamber kept at

K. A. Becker is with the School of Health Promotion and Kinesiology at Texas Woman's University, Denton, TX 76204 USA (940-898-2592; e-mail: kbecker1@twu.edu).
C. A. Aiken is with the Department of Kinesiology and Dance at New Mexico State University, Las Cruces, NM 88003 USA (e-mail: aiken@nmsu.edu).
C. J. Hung is with the School of Health Promotion and Kinesiology at Texas Woman's University, Denton, TX 76204 USA (e-mail: chung@twu.edu).
A. W. A. Van Gemmert is with the School of Kinesiology at Louisiana State University, Baton Rouge, 70803 USA (e-mail: gemmert@lsu.edu).

a temperature of 42° C during the heat condition and at 22° C during the control condition. Participants were subjected to the heat and control conditions on separate days.

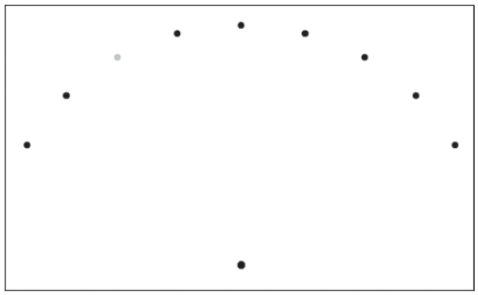


Figure 1. Visual Representation of the Motor Task (start target at bottom, nine end targets at top)

B. Procedures

Upon arrival to the laboratory, participants completed an inclusion criteria questionnaire, provided written informed consent, and consumed 8oz of water to minimize the effects of dehydration on performance. The participant was then seated in the chamber for the assigned thermal condition (heat or control) for a period of five minutes. A research assistant first gave a verbal description of the task. After the 5 minute accommodation and verbal explanation, participants were allowed to manipulate the stylus on the tablet, and view a still photo representation of the task. They were informed that their goal was to move as quickly and accurately as possible from the start target to the one of nine end targets changing its color. The experiment consisted of 108 trials in the hot environment, i.e., heat stress (HEAT) condition, and the control (CON) condition. Within each condition, trial order was randomized, and 6 trials were performed for each of the 18 targets (9 small targets and 9 large targets). The distance between the start target and end targets was the same within each experiment, i.e., 12 cm for experiment 1 and 5 cm for experiment 2. The conditions (HEAT and CON) were counterbalanced across participants with approximately 24 hours between the two conditions.

C. Data Analysis

Data from the two experiments were analyzed separately. Reaction time was defined as the time between the stimulus presentation and the time when 5% of peak velocity was achieved. Movement time was defined as the time interval from
when movement velocity exceeded 5% of peak velocity until the time movement velocity dropped below 5% of peak velocity. Initial direction error at peak velocity was defined as the angular deviation between a vector representing a line directly between the start and end targets, and a vector representing a line between the start target and the position at peak velocity. Axial pen pressure was a direct measure of pressure derived from the digitizing tablet. In each experiment, a 2 (HEAT vs CON) x 2 (small vs large target size) ANOVA was conducted for each dependent variable. Dependent variables were reaction time (ms), movement time (ms), initial direction error at peak velocity (deg), and axial pen pressure (N). For all analyses, alpha was set at .05.

3. RESULTS

A. Experiment 1

Means and standard deviations for each dependent variable across conditions are presented

in Table 1. When targets were presented at a longer distance (12cm), the main effect of heat stress was significant ($p = .048$), with reaction times being slower in the heat stress conditions than the control conditions. The main effect of target size was also significant ($p < .001$), with reaction time being slower with the small targets than with the large targets. The interaction between heat and target size was not significant ($p > .05$) suggesting that the heat stress effect did not differ due to target size. Movement time was shorter with large targets than with small targets ($p < .001$), but was not impacted by heat ($p > .05$). Initial direction error at peak velocity and axial pen pressure did not differ due to either heat or target size (p's > .05).

	Small Target Size		Large Target Size	
	Heat Condition	**Control Condition**	**Heat Condition**	**Control Condition**
Reaction Time (ms)	629.57 (102.14)	596.75 (59.81)	536.26 (85.32)	513.94 (48.27)
Movement Time (ms)	1619.37 (569.90)	1593.30 (558.77)	1289.01 (484.08)	1269.85 (436.75)
Initial Direction Error at Peak Velocity (deg)	5.88 (3.01)	5.73 (2.63)	6.21 (4.30)	4.98 (1.99)
Axial Pen Pressure	525.89 (133.84)	553.37 (149.43)	530.66 (136.12)	559.91 (140.81)

Table 1. Performance measures under different conditions in Experiment 1

B. Experiment 2

Means and standard deviations for each dependent variable across conditions are presented in Table 2. When targets were presented at a shorter distance (5cm), reaction time was slower for the small targets than the large targets ($p < .001$), but the main effect of heat failed to reach significance ($p > .05$). The interaction between heat and target size was significant ($p = .036$). With the small targets, mean reaction times tended to be faster in the control condition than the heat condition and with the large targets the opposite pattern emerged. However, pairwise comparisons failed to reveal significant differences for the different heat by target size combinations. Movement time and initial direction error at peak velocity were both smaller with large targets than with small targets (p's = .001, .018), but were not impacted by heat (p's > .05). Axial pen pressure did not differ due to either heat or target size (p's > .05).

	Small Target Size		Large Target Size	
	Heat Condition	**Control Condition**	**Heat Condition**	**Control Condition**
Reaction Time (ms)	541.19 (78.56)	519.69 (58.77)	478.19 (55.36)	485.60 (45.38)
Movement Time (ms)	1164.35 (383.77)	1181.40 (387.51)	910.35 (354.31)	938.84 (334.49)
Initial Direction Error at Peak Velocity (deg)	5.71 (1.34)	5.60 (1.65)	5.53 (1.57)	4.76 (1.22)
Axial Pen Pressure	478.29 (173.21)	533.65 (183.49)	480.27 (167.86)	533.88 (187.33)

Table 2. Performance measures under different conditions in Experiment 2

4. DISCUSSION

Previous research considering the impact of heat-related stress on reaction time has produced varied results. Aiken et al. (2016) demonstrated that task difficulty influences whether stress as result of heat impacts reaction time, but it was unclear whether heat-related stress effects

were affected by either the accuracy demands, the movement magnitude, or both. The purpose of this study was to determine which of these factors, i.e., movement magnitude, accuracy, or both, influenced the effect of stress due to heat on reaction time. Furthermore, this study aimed to determine whether movement time, initial direction error at peak velocity, and/or axial pen pressure were differentially affected due to heat-related stress.

The results of the current study showed that heat-related stress affected reaction time in the first experiment when targets of differing size were presented at longer distances (12cm). In this experiment, heat-related stress slowed reaction time regardless of target size; while in the second experiment with targets presented at shorter distances (5cm) reaction time did not significantly change due to heat-related stress. Taken together, these results suggest that movement magnitude seems to be the task difficulty factor influencing delays in reaction time in hot environments. Based on the Neuromotor Noise Perspective, it appears when planning a longer duration movement participants may delay initiation of the movement in order to reduce neuromotor noise in the motor system.

The results regarding movement time, initial direction error at peak velocity, and axial pen pressure did not fit predictions made by the Neuromotor Noise Perspective. We expected that at least axial pen pressure would increase when encountering heat-related stress as a means of filtering out neuromotor noise. Our results showed no differences, and these findings were consistent with the results of Aiken et al. (2016). Future work should consider using tasks with more complex motor responses to determine whether complexity increases lead to the need to filter neuromotor noise with strategies beyond delaying initiation of the motor response.

5. ACKNOWLEDGMENTS

This research was supported in part by the Texas Woman's University Research Enhancement Program.

REFERENCES

[1] Aiken, C.A., Becker, K.A., Lee, A., Post, P.G., Van Gemmert, A.W.A. (2016). Performance of a choice-reaction time task is not affected by physical stress in the form of high ambient temperature. *Journal of Sport and Exercise Psychology, 38*, S40.

[2] Hancock, P. A., & Dirkin, G. R. (1982). Central and peripheral visual choice-reaction time under conditions of induced cortical hyperthermia. *Perceptual and Motor Skills, 54*(2), 395-402.

[3] Leibowitz, H. W., Abernethy, C. N., Buskirk, E. R., Bar-Or, O., & Hennessy, R. T. (1972). The effect of heat stress on reaction time to centrally and peripherally presented stimuli. *Human Factors, 14*(2), 155-160.

[4] Schlader, Z. J., Gagnon, D., Adams, A., Rivas, E., Cullum, C. M., & Crandall, C. G. (2015). Cognitive and perceptual responses during passive heat stress in younger and older adults. *American Journal of Physiology-Heart and Circulatory Physiology, 308*(10), R847-R854.

[5] Van Gemmert, A. W., & Van Galen, G. P. (1997). Stress, neuromotor noise, and human performance: A theoretical perspective. *Journal of Experimental Psychology: Human Perception and Performance, 23*(5), 1299.

Chapter 5

Historical document processing

Towards Automated Reading of Historical Vietnamese Steles

Anna Scius-Bertrand, Jeremie Bosom, Philippe Papin, Marc Bui

Ecole Pratique des Hautes Etudes - PSL - 4-14 rue Ferrus 75014 Paris
fanna.scius-bertrand, jeremie.bosom, philippe.papin, marc.buig@ephe.sorbonne.fr

Abstract - Stone engravings in historical Vietnamese steles are a treasure for historians studying the life of villagers. The Vietnamica project aims to discover this history by analyzing a unique collection of about 40,000 digital images of steles. As a support for the historians, automated reading is needed to nd search terms in the images. In this paper, we elaborate an important first step towards this goal: the extraction of text columns from page images. We identify challenges in the dataset and present preliminary results for semi-automatic annotation.
Keywords: Vietnamese Steles, Document Analysis, Text Segmentation, Ground Truth Creation.

1. INTRODUCTION

"Which historian is not interested in the price of chicken in northern Vietnam between 1610 and 1780?"
(Papin, 2003). This question gives rise to the thought of how to know the history of the villagers, knowing that the history of Vietnam was written by the imperial court and the clergy. The answer can be found in the Vietnamese inscriptions engraved in stone on steles that have only recently become accessible. They were used to make written reports in order to be seen by all and to resist bad weather and human destruction. These are texts written for villagers, by villagers. They provide information on the cultural, social, and religious life of the people at the edge of the rice fields.
The data from the scattered steles are worthless because it is impossible to generalize them. To process the inscriptions, the use of quantitative history is a necessity. The unique and unprecedented corpus of about 40,000 digital images of steles oers exactly this opportunity. It is the aim of the Vietnamica project, which recently received an ERC grant. To enable quantitative history, one of the objectives of the project is to carry out an automated reading of the ancient Vietnamese script, such that search terms can be found in the digital images.
The Vietnamese script under consideration consists of Chu N^om characters that are written in vertical columns from top to bottom (see Figure 1a). A first step towards automated reading consists in layout analysis to nd the main text region, followed by column segmentation. The current state of the art to perform these steps is based on machine learning techniques, for example using deep neural networks (Wick & Puppe, 2018; Fink et al., 2018). They require a considerable number of human-annotated learning samples, also called *ground truth*, in order to train models that can extract text columns with high precision afterwards. The resulting column images can then be further processed by keyword spot-ting or handwriting recognition systems, e.g. (Sudholt & Fink, 2016; Wu et al., n.d.), for detecting search terms entered by historians.
In this paper, we provide a preliminary study on the automatic extraction of text columns. First, we identify challenges for layout analysis and column segmentation in the dataset. Afterwards, we present results for semi-automatic ground truth creation using a method originally introduced for Latin script (Fischer et al., 2010) and analyze error cases. Finally, we conclude the paper with an outlook to future work.

2. DATASET OF HISTORICAL VIETNAMESE STELES

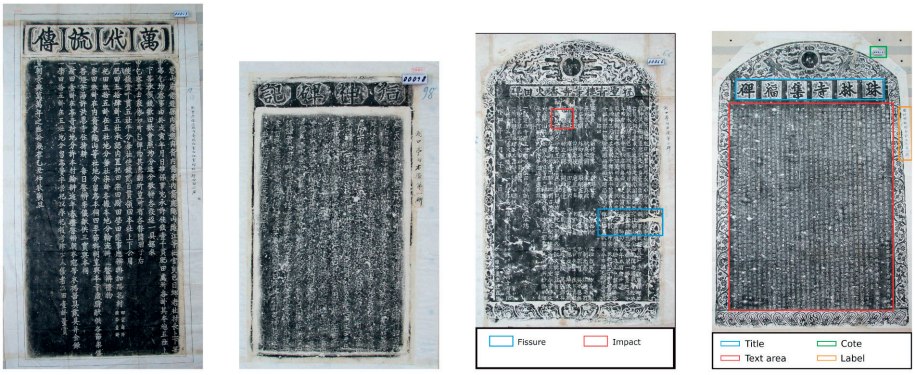
Overall, around 40 000 images of steles exist. They are distributed in two corpora of equal size, one held by the Ecole Fran110caise d'Extr^eme-Orient (EFEO) collected between 1910 and 1954 and the other one held by the Han-N^om Institute collected since 1995 (Papin et al., 2007-2012, 2005-2013). The steles cover a time period of ve centuries. They are mainly dated from the 17th, 18th and 19th centuries. 80% of the corpus originate from northern Vietnam. For creating copies, a traditional stamping procedure is rst applied by sticking a paper to the steles using banana juice and then applying ink with a roller, such that the engraved characters stand out in white from the black background. The stamps are then photographed with a digital camera. This procedure creates images that are better readable when compared with taking photographs of the steles directly. Example images are shown in Figures 1 and 2.

2.1 Evaluation Dataset

To perform a preliminary analysis of the collection, we have randomly selected a subset of 100 images from the corpus and manually inspected challenges that arise for developing automated reading systems.

In general, degradation of the writing support poses a challenge that is common for many types of historical documents. In our case we have noted three types of degradation, namely ssures, impacts, and erosion of the stone (Figures 1b and 1c). On our evaluation dataset, we have observed that 82% of the images show at least one type of degradation, resulting in partially or completely missing characters.

Some of the images are in perfect condition (Figure 1a).



(a) No degradation. (b) Erosion. (c) Fissure and impact. (d) 4 layout elements.

Fig. 1. Historical Vietnamese steles.

2.2 Challenges for Layout Analysis

We aim to retrieve textual information from four image areas, i.e. the main text area, title, border, and reference label (Figure 1d). However, there is a considerable heterogeneity in layout, which is challenging for automatic layout analysis. On our evaluation dataset, we have identied three types of variations regarding the frame, the title, and the reference label, as summarized in Table 1.

Concerning the frame, most of the steles have a frame with motifs around the main text area, the title, and below the title (Figure 1d). Yet 13% of the images have no frame (Figure 1a), 1% have a frame around the main text area but not around the title, and 12% have a solid color frame.

Regarding the title, there are variations in its color that may mislead an automatic layout analysis system. For the majority of cases (51%) the title is in black and the text in white, 40% of the titles are white, same as the text, and 2% of the titles have both black and white characters. 7% of the steles have no title at all (Figure 2b). The main text areas have mostly white characters. However, in some steles we have observed black writing surrounded by white embellishments (Figure 2a).
Finally, there are variations in the location of reference label. Most of the reference labels are outside the frame on the right but 4% are on the left and 4% are within the frame.

Frame			Title				Reference Label	
No frame	Only around text	Solid color	Black	White	B&W	No title	Left	Within frame
13%	1%	12%	51%	40%	2%	7%	4%	4%

Table 1. Challenges for layout analysis.

2.3 Challenges for Column Segmentation

After extracting the main text area, column segmentation is needed as an important pre-processing step of automated reading. We have identied three types of challenges on our evaluation dataset regarding character size, column structure, and number of text zones, as summarized in Table 2.
Just over half of the steles (59%) have almost uniform character sizes. For the rest of the images, we have observed up to three dierent character sizes.
Regarding the column structure, we focus on four criteria, i.e. spacing, discontinuity, subdivision and alignment. We consider a space between two columns of less than a fourth of the character width as small.
This is the case for 32% of the steles. More than half of the steles (52%) have one or more discontinuous columns (Figure 2c) with a gap of at least four characters. 18% of the steles have sub-columns, i.e. at least two columns following a single one (Figure 2d). Finally, 18% of the steles have columns that are not aligned at the top.
As shown in (Figure 2c), a third of the steles have at least two dierent text zones within the main text area. A text zone is considered to be separate from another if it is separated by an empty space greater than a column.

(a) Different text color. (b) No title. (c) Different text zones. (d) Subcolumns.

Fig. 2. Challenges for layout analysis and column segmentation.

Characters	Columns				Text zones
Different sizes	Small spacing	Discontinuous	Subdivided	Not aligned	More than one
41%	32%	52%	18%	18%	33%

Table 2. Challenges for column segmentation.

3. GROUND TRUTH CREATION

As a prerequisite for automatic layout analysis and column segmentation using machine learning, human-annotated learning samples are required. Such ground truth is needed for training as well as evaluation of learning-based methods. In our case, the ground truth consists of bounding polygons around the four layout elements as well as the individual text columns.

To avoid a fully manual annotation, we follow a semi-automatic procedure for ground truth creation originally proposed for Latin manuscripts (Fischer et al., 2010). The method was chosen because it does not require any parameter ne-tuning and has already been successfully applied to a variety of historical manuscripts.

The procedure consists of several consecutive steps. The rst step is manual and consists in using the GIMP software to select the main text area using polygon paths and save it in scalable vector graphics (SVG) format. Only one text area is selected, even if it contains several distinct text zones.

The second step is automatic and aims to segment the text columns. As the original method is intended to segment black horizontal text lines, we rst rotate all images clockwise by 90 degree and invert the colors. Next, the text foreground is enhanced by means of a Dierence of Gaussians (DoG) before the image is binarized using a global threshold. We have employed the radii $\sigma 1$=10:0 and $\sigma 2$=0:5 for DoG and the threshold T=200 for binarization. The columns are then segmented in the black and white image using a dynamic programming method.

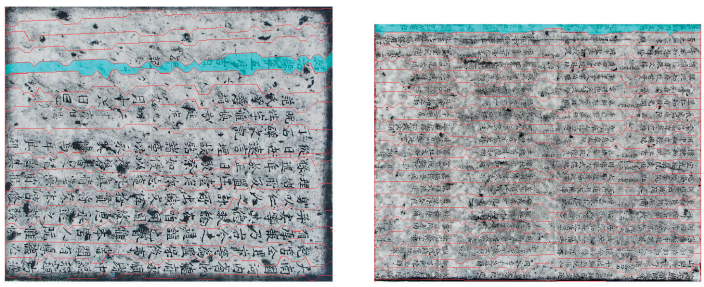
The third and last step is manual and consists in using a graphical user interface (GUI) based on Java to correct the column segmentation.

4. EXPERIMENTAL EVALUATION

The semi-automatic ground truth creation for column segmentation took about 2 minutes per stele for a user familiar with the software. Table 3 summarizes the results of the dynamic programming algorithm used for column segmentation. Exemplary results are shown in Figure 3.

Performance			Error cases			
Recall	Precision	F-score	Merge	Insertion	Incomplete	Missing
79,05%	81,10%	80,08%	59,93%	28,77%	7,88%	3,42%

Table 3. Experimental evaluation of the column segmentation algorithm.



(a) Four types of errors. (b) No error.

Fig. 3. Column segmentation.

On our evaluation dataset, the recall (percentage of retrieved columns) is 79,05%, the precision (percentage of correct columns) is 81,10%, and the F-score is 80,08%. A column with less than ten percent loss of text foreground is considered correct. Otherwise, the column is incomplete (7,88% of all errors).
Other common mistakes include merging two or more columns (59,93%), which may be due to small spacing, inserting an inexciting column (28,77%), which can be triggered by damages in the stone that appear like characters, and missing a column (3,42%), which tends to occur at the periphery of the text area. An error example is provided in Figure 3a, which contains three types of errors: merge, insertion and incomplete. Figure 3b shows an example without any error, which is the case for 21% of the images in our evaluation dataset.

5. CONCLUSIONS AND FUTURE WORK

Despite the dicult challenges for layout analysis and column segmentation for historical Vietnamese steles outlined in this paper, the semi-automatic method for ground truth creation has proven ecient and, surprisingly, quite accurate. Nevertheless, it leaves room for improvements.
In the next step, we intend to use the human annotations for training learning-based methods for text column extraction, which are expected to further improve the results. We will focus, in particular, on methods based on convolutional neural networks and adapt them to the Vietnamese script.
There are several lines of future research to pursue. First, we will focus on keyword spotting and hand- writing recognition methods for the Chu N^om characters. Afterwards, we aim to improve the recognition systems by means of transfer learning as well as data augmentation with synthetic handwriting.

REFERENCES

[1] Fink, M., Layer, T., Mackenbrock, G., & Sprinzl, M. (2018). Baseline detection in historical documents using convolutional U-Nets. In *Proc. 13th Int. Workshop on Document Analysis Systems (DAS)* (pp. 37-42).

[2] Fischer, A., Indermühle, E., Bunke, H., Viehhauser, G., & Stolz, M. (2010). Ground truth creation for handwriting recognition in historical documents. In *Proc. 9th Int. Workshop on Document Analysis Systems (DAS)* (pp. 3-10).

[3] Papin, P. (2003). Aper114cu sur le programme \publication de l'inventaire et du corpus complet des inscriptions sur steles du vi^et-nam". *Bulletin de l' École fran114caise d'Extrême-Orient, 90* (1), 465-472.

[4] Papin, P., Manh, T. K., & Nguyên, N. V. (2005-2013). *Corpus des inscriptions anciennes du vietnam*. EPHE, EFEO, Institut Han-Nôm.

[5] Papin, P., Manh, T. K., & Nguyên, N. V. (2007-2012). *Catalogue des inscriptions du vi^et-nam*. EPHE, EFEO, Institut Han-Nôm.

[6] Sudholt, S., & Fink, G. A. (2016). PHOCNet: A deep convolutional neural network for word spotting in handwritten documents. In *Proc. 15th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR)* (pp. 277-282).

[7] Wick, C., & Puppe, F. (2018). Fully convolutional neural networks for page segmentation of historical document images. In *Proc. 13th Int. Workshop on Document Analysis Systems (DAS)* (pp. 287-292).

[8] Wu, Y.-C., Yin, F., Chen, Z., & Liu, C.-L. (n.d.). Handwritten Chinese text recognition using separable multidimensional recurrent neural network. In *Proc. 14th Int. Conf. on Document Analysis and Recognition (ICDAR).*

Chapter 6

Biometric and Forensic Applications

Synthetic Generation of Online Signatures using a Deep Generative Model

Paul Maergner [a], Taha Sükrü Karabacakoglu [a], Kaspar Riesen [b], Rolf Ingold [a], Andreas Fischer [a,c]

[a]Department of Informatics, University of Fribourg
Boulevard de Perolles 90, 1700 Fribourg, Switzerland
{paul.maergner, rolf.ingold, andreas.fischer}@unifr.ch

[b]Institute for Informations Systems, University of Applied Sciences and Arts
Northwestern Switzerland, Riggenbachstrasse 16, 4600 Olten, Switzerland
kaspar.riesen@fhnw.ch

[c]Institute of Complex Systems, University of Applied Sciences and Arts
Western Switzerland, Boulevard de Perolles 80, 1700 Fribourg, Switzerland

Abstract - Handwritten signatures are a common form of biometric authentication. However, the verication of handwritten signatures is a challenging task since it has to rely on only a few genuine samples. Recently, deep generative neural networks have been introduced that can generate handwriting based on handwritten samples. In this work, we apply these networks to handwritten signatures to generate synthetic signatures to aid the verication process. Using a verication system based on dynamic time warping, our proposed approach leads to signicant improvements in verication accuracy on the MCYT-100 benchmark dataset.
Keywords: Online Signature Verication, Conditional Variational Recurrent Neural Network.

1. INTRODUCTION

Handwritten signatures are one of the most accepted forms of biometric authentication, both legally and culturally. With this popularity comes the need for a robust way to verify the authenticity of signatures.

Indeed, signature verication has been an active research area for many decades (Plamondon & Lorette, 1989; Hafemann et al., 2017; Diaz et al., 2019). Two types of signature verication are commonly distinguished: online signature verication, which relies on the pen-tip trajectory recorded over time, and oine signature verication, which operates only on the static image of the signature. The present paper is focused on the online scenario.

Signature verication remains a challenging task due to two main factors. The intrapersonal variability of each writer paired with the usually small amount of genuine samples available. Two signatures of the same writer are never completely identical due to the strong in
uence of behavioral and social conditions (Diaz et al., 2018). At the same time, it is often unpractical to collect the number of signature samples needed to cover this intrapersonal variability to its full extent. This discrepancy is unfortunate, especially since it is known that the verication performance increases signicantly the more genuine samples are available for training.

To overcome this issue, several works propose the generation of additional synthetic samples based on the available genuine samples, including (Munich & Perona, 2003; Rabasse et al., 2008; Galbally et al., 2009, 2012a, 2012b; Diaz et al., 2014, 2018). The work proposed by (Galbally et al., 2009) tackled this task by adding some noise directly to trajectory points, pressure signal, and duration of the signatures.

More recently, (Diaz et al., 2018) proposed to generate additional reference signatures

from a single genuine sample by leveraging the Kinematic Theory of rapid human movement. They introduced two distortions for generating duplicated samples. A stroke-wise distortion by adding Gaussian noise to the sigma-lognormal parameters and a target-wise distortion by adding sinusoidal noise to the virtual targets of the sigma-lognormal model. All previous works have shown that the addition of synthetic signatures is able to improve verication performance.

In this paper, we explore the generation of synthetic signatures using the conditional variational recurrent neural network (C-VRNN) architecture introduced (Aksan et al., 2018). C-VRNN models can produce realistic handwriting for dierent writers after being trained with real handwriting samples.We follow this approach to synthesize realistic handwritten signatures for dierent signers. To evaluate the eectiveness of the proposed method, we use a signature verication system based on dynamic time warping (DTW). The evaluation is performed on the publicly available MCYT-100 benchmark dataset.

In the remainder, we introduce the signature generator based on C-VRNN, describe the signature verification system based on DTW, present experimental results, and nally draw some conclusions.

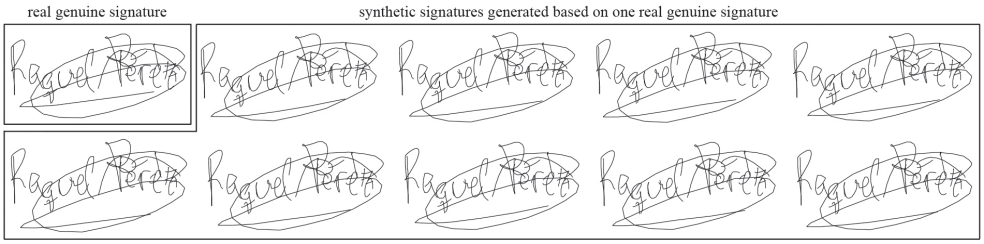


Fig. 1. Example of a real genuine signature from the MCYT-100 dataset and nine synthetic signatures. The variations can best be seen in the ourish of the signature.

2. SIGNATURE GENERATOR

We are using the conditional variational recurrent neural network (C-VRNN) introduced by (Aksan et al., 2018) to generate new signatures. As the name suggests, C-VRNN derives some aspects from recurrent neural networks (RNN), which are capable of training with variable length input data (Goodfellow et al., 2016), and variational recurrent neural networks (VRNN), which include latent random variables (Chung et al., 2015). C-VRNN augments the model with the capability to condition the output on the user, i.e. the signer in our application. For a detailed description of C-VRNN, see (Aksan et al., 2018).

We are using the Tensorflow implementation of C-VRNN published by (Aksan et al., 2018). The input and latent layers are long-short-term memory cells (LSTM) with 512 units. The output layer is a 1-layer feed-forward network with 512 units and rectied linear unit (ReLU) activation function. The latent variables are represented by 32-dimensional isotropic Gaussian distributions. The network is trained using a learning rate of 0:001 and a mini-batch size of 32 as well as dropout.

To train the C-VRNN as a signature generator, we create a training set T containing n reference signatures for each user in the dataset. For training, each signature $t = (f_1, f_2, ...)$ is represented by a sequence of feature vectors $f_i = (x, y, p, u, b, e)$, where x, y are relative pen movements, p is marking a pen-up event, u is the user ID, and b and e mark the begin and end of the signature.

Once the network is trained, it can produce additional genuine signatures for specic users. Fig. 1 shows a plot of a real genuine signature in the top left corner and nine synthetic signatures generated from it.
These synthetic signatures can then be used in the signature verication system described in the next section.

3. SIGNATURE VERICATION SYSTEM

For signature verication, we represent each signature using a sequence (v_1, v_2, ...) of six features vi =$(x, y, \dot{x}, \dot{y}, \ddot{x}, \ddot{y})$, where (x, y) are the pen coordinates, $(\dot{x}, \dot{y})$ is the velocity, and $(\ddot{x}, \ddot{y})$ is the acceleration.
We compute the velocity and the acceleration with respect to two sampling points before and after using second order regression (Fierrez et al., 2007). To compare two signatures r and t, the respective signature sequences are compared using dynamic time warping (DTW) (Berndt & Cliord, 1994; Kholmatov & Yanikoglu, 2005). DTW oers a way to compare two sequences of temporal data, which may be dierent in length. In order to speed up the computation, a Sakoe-Chiba band (Sakoe & Chiba, 1978) is used to limit the deviation from the diagonal alignment to 10% of the length of the reference signature.
The final verication score is dened by the dissimilarity score $d_R(t)$ of the test signature t to the reference signatures r_1,..., $r_n \in R$ of the claimed user. Formally,

$$d_R(t) = \frac{\min_{r \in R}(\mathrm{dtw}(r, t))}{\mu_R}, \tag{1}$$

where μ_R is the average dissimilarity of the reference graphs to each other (Fischer et al., 2015).

$$\mu_R = \frac{1}{|R|} \sum_{r \in R} \min_{s \in R \setminus r} dtw(r, s) \tag{2}$$

A signature t is accepted as genuine, if the dissimilarity score $d_R(t)$ is below a certain threshold, otherwise, the signature is rejected as a forgery.

Method	# of references	EER (in %)
Proposed system	**1**	**10.32**
SW Duplicates + DTW (Diaz et al., 2018)	1	13.56
Cost Matrix features + DTW (Sharma & Sundaram, 2018)	5	2.76
Matcher + DTW (Tang et al., 2018)	5	3.16
$\sum \Lambda$ + DTW (Fischer & Plamondon, 2017)	5	3.56
Histogram + Manhattan (Sae-Bae & Memon, 2014)	5	4.02
Proposed system	**5**	**4.44**
Neuro-fuzzy system (Cpałka et al., 2016)	5	4.88
Wavelet transform (Nanni & Lumini, 2008)	5	5.20

Table 1. Comparison of the results of our proposed systems with published results on the MCYT-100 dataset. Results are grouped by the number of reference signatures per user, then sorted by the achieved EER on skilled forgeries.

4. EXPERIMENTAL EVALUATION

We evaluate the proposed system on the publicly available MCYT-100 signature corpus (Ortega-Garcia et al., 2003). The dataset consists of signatures from 100 individual users. For each user, there are 25 genuine signatures and 25 skilled forgeries. The signatures have been captured using a pen tablet device. All signatures from the dataset are used in our evaluation. The first n genuine signatures of each user are employed to train the signature generator and

as reference signatures for the signature verication. The remaining genuine signatures are used for the evaluation of the signature verication system. In this evaluation, we have chosen $n \in \{1, 5\}$, referring to them as R1 and R5, respectively, i.e. we train the C-VRNN with 100 signatures (R1) or 500 signature (R5).

The verication performance of the systems is measured using the equal error rate (EER), which is the point in the receiver operating characteristic (ROC) curve where the false acceptance rate is equal to the false rejection rate. When using our verication system without additional synthetic signatures, we get an EER result of 19.64% (R1) and 4.48% (R5). For the R1 system, we generated nine synthetic signatures for each user based on the real reference signature. We refer to these synthetic signatures as **9S**. For the R5 system, we generated ve synthetic signatures for each user, one based on each of the ve real reference signatures. We refer to these synthetic signatures as **5S**.

Surprisingly, we found that using only the synthetic signatures for validation yields better results (9S:10.32%, 5S: 4.44%) than using both reference signatures and synthetic signatures (R1+9S: 11.28%, R5+5S: 12.44%). Especially for the R5 system, the dierence is signicant. We suspect that the normalization value (Eq. 2) is negatively aected by the combination of real and synthetic signatures due to their high similarity. In the following, we are using only the synthetic signatures as reference samples for verication.

Table 1 shows the results of our proposed systems in comparison with published results on the MCYT 100 dataset. When using ve reference signatures per user (R5), our proposed approach achieves results that are comparable to the state of the art. For the dicult case of having only one genuine signature available per user (R1), our proposed synthetization method outperforms the current state of the art reported in (Diaz et al., 2018) on this dataset.

5. CONCLUSIONS AND OUTLOOK

In this paper, we have introduced a signature generator based on conditional variational recurrent neural networks (C-VRNN). To evaluate the signature verication performance, we use a signature verication system based on dynamic time warping (DTW). On the publicly available MCYT-100 benchmark dataset, our proposed approach achieves strong results when compared with the current state of the art, especially when relying on only one genuine reference signature.

We see several future lines of research. First, the eect of dierent numbers of synthetic signatures should be investigated in greater detail. Additionally, the quality of synthetic signatures might benet from a dierent network architecture, e.g. bi-directional VRNN models. A quality control mechanism would be interesting for deciding which synthetic signatures should be added as references. Furthermore, an HMM or other feature-based signature verication methods might lead to better results than the presented DTW-based approach. Finally, the approach should be evaluated on more benchmark dataset to determine the robustness of the approach.

Acknowledgment

This work has been supported by the Swiss National Science Foundation project 200021 162852.

REFERENCES

[1] Aksan, E., Pece, F., & Hilliges, O. (2018). DeepWriting: Making Digital Ink Editable via Deep Generative Modeling. In *Proc. sigchi conf. on human factors in computing systems* (pp. 1-14).

[2] Berndt, D. J., & Cliord, J. (1994). Using dynamic time warping to nd patterns in time series. In *Proc. 3rd int. conf. on knowledge discovery and data mining* (pp. 359-370).

[3] Chung, J., Kastner, K., Dinh, L., Goel, K., Courville, A. C., & Bengio, Y. (2015). A recurrent latent variable model for sequential data. In *Proc. 28th int. conf. on neural information processing systems* (pp. 2980-2988).

[4] Cpa lka, K., Zalasinski, M., & Rutkowski, L. (2016). A new algorithm for identity verication based on the analysis of a handwritten dynamic signature. *Applied Soft Computing, 43*, 47-56.

[5] Diaz, M., Ferrer, M. A., Impedovo, D., Malik, M. I., Pirlo, G., & Plamondon, R. (2019). A perspective analysis of handwritten signature technology. *ACM Computing Surveys, 51* (6), 1-39.

[6] Diaz, M., Ferrer, M. A., & Morales, A. (2014, Sep.). Cognitive inspired model to generate duplicated static signature images. In *Proc. 14th int. conf. on frontiers in handwriting recognition* (pp. 61-66).

[7] Diaz, M., Fischer, A., Ferrer, M. A., & Plamondon, R. (2018). Dynamic signature verication system based on one real signature. *IEEE Transactions on Cybernetics, 48* (1), 228-239.

[8] Fierrez, J., Ortega-Garcia, J., Ramos-Castro, D., & Gonzalez-Rodrguez, J. (2007). HMM-based on-line signature verication: Feature extraction and signature modeling. *Pattern Recognition Letters, 28* , 2325-2334.

[9] Fischer, A., Diaz, M., Plamondon, R., & Ferrer, M. A. (2015). Robust score normalization for DTW-based on-line signature verication. In *Proc. 13th int. conf. on document analysis and recognition* (pp. 241-245).

[10] Fischer, A., & Plamondon, R. (2017). Signature verication based on the kinematic theory of rapid human movements. *IEEE Transactions on Human-Machine Systems*, 47 (2), 169-180.

[11] Galbally, J., Fierrez, J., Martinez-Diaz, M., & Ortega-Garcia, J. (2009, July). Improving the enrollment in dynamic signature vercation with synthetic samples. In *Proc. 10th int. conf. on document analysis and recognition* (pp. 1295-1299).

[12] Galbally, J., Plamondon, R., Fierrez, J., & Ortega-Garcia, J. (2012a). Synthetic on-line signature generation. Part II: Experimental validation. *Pattern Recognition, 45* (7), 2622-2632.

[13] Galbally, J., Plamondon, R., Fierrez, J., & Ortega-Garcia, J. (2012b). Synthetic on-line signature generation. Part I: Methodology and algorithms. *Pattern Recognition, 45* (7), 2610-2621.

[14] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. (http://www.deeplearningbook.org)

[15] Hafemann, L. G., Sabourin, R., & Oliveira, L. S. (2017). Oine handwritten signature verication - literature review. In *Proc. int. conf. on image processing theory, tools and applications* (pp. 1-8).

[16] Kholmatov, A., & Yanikoglu, B. A. (2005). Identity authentication using improved online signature verication method. *Pattern Recognition Letters, 26* , 2400-2408.

[17] Munich, M. E., & Perona, P. (2003). Visual identication by signature tracking. *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, 25 (2), 200-217.
[18] Nanni, L., & Lumini, A. (2008). A novel local on-line signature verication system. *Pattern Recognition Letters, 29* (5), 559-568.
[19] Ortega-Garcia, J., Fierrez-Aguilar, J., Simon, D., Gonzalez, J., & et al. (2003). MCYT baseline corpus: a bimodal biometric database. *IEEE Proceedings Vision, Image and Signal Processing, Special Issue on Biometrics on the Internet, 150* (6), 395-401.
[20] Plamondon, R., & Lorette, G. (1989). Automatic signature verication and writer identication - the state of the art. *Pattern Recognition, 22* (2), 107-131.
[21] Rabasse, C., Guest, R. M., & Fairhurst, M. C. (2008). A new method for the synthesis of signature data with natural variability. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 38* (3), 691-699.
[22] Sae-Bae, N., & Memon, N. (2014). Online signature verication on mobile devices. *IEEE Transactions on Information Forensics and Security, 9* (6), 933-947.
[23] Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 26* (1), 43-49.
[24] Sharma, A., & Sundaram, S. (2018). On the exploration of information from the dtw cost matrix for online signature verication. *IEEE Transactions on Cybernetics, 48* (2), 611-624.
[25] Tang, L., Kang, W., & Fang, Y. (2018). Information divergence-based matching strategy for online signature verication. *IEEE Transactions on Information Forensics and Security, 13* (4), 861-873.

Forensic examination of dynamic signatures: Multivariate signature data and evidence evaluation

Jacques Linden, Silvia Bozza, Raymond Marquis, Franco Taroni

Abstract - This study investigates a forensic 'signature authentication' system, using the Bayesian approach and Case Assessment and Interpretation (CAI) recommendations. A Bayesian model operating on dynamic signatures collected from three individuals was developed. The objective of our study was twofold; First, we aimed to develop a helpful tool for forensic examiners in discriminating between genuine and 'forged' signatures. Second, we wanted to provide information on the amount of signatures necessary to produce reliable inference. The proposed model was tested in practical operational conditions, where few reference signatures are available for comparison. Accuracy and repeatability, measuring the overall reliability and performance were measured throughout repeated trials. The system had above 90% accuracy rate even when only 5 reference signatures are used as control material. Further, the variance of the accuracy suggests that 15 signatures is a reasonable reference set size leading to reproducible results. Performance, critical issues and limitations are discussed.
Keywords: Forensic Science, Handwritten Signatures, On-line Signatures, Forgery, Likelihood ratio, Bayesian statistics

I. INTRODUCTION

This research started in response to an increasing interest in dynamic signatures, a form of biometric authentication. Our main concerns were the lack of forensic methodology and tools to conduct dynamic signature examination and assess the evidence's value [1] to discriminate genuine from forged signatures. At large, forensic science presents a solid methodological framework [2, 3] and valuable principles for communicating and expressing evidential value in a court of law [4-8] using the Bayesian paradigm. The Bayes factor (BF) is a rigorous concept providing a measure of probabilistic and quantified support to discriminate among competing propositions [8-11]. The Bayes Factor is often called 'likelihood ratio' or 'LR' in forensic context. The LR is actually a type of Bayes Factor. The 'likelihood ratio approach' has started to gain strong support in the forensic community [8-11]. This research aims to deliver a readily implementable model based on this approach for signature examination, evaluation and presentation in forensic science contexts.
First, we aim to develop a helpful tool for forensic examiners in discriminating between genuine and 'forged' signatures. Second, we want to provide information on the amount of signatures necessary to produce reliable inference. The proposed model was tested in practical operational conditions, where few reference signatures are available for comparison.

2. METHODS

A. Method

In most signature cases, the relevant question is about the signature being either a genuine signature or an insincere one. Insincerity can be seen as intent to deceive, which manifest as either forgery by another person, or by disguising their own signature. For the sake of simplicity, let us consider a scenario where no reason to disguise the signature exists and no

particular contextual information concerning position, posture or health conditions is available. A pair of 'default' propositions can be used, namely H_1: the signature was made by the presumed source and H_2: the signature was made by an alternative source. Evidence strength is expressed through a Bayes Factor (1) as

$$BF = f(y \mid x, H1) * [f(y \mid x, H2)]^{-1} \quad (1)$$

where y and x represent the measurements on the questioned signatures and on the reference signatures, respectively. The underlying statistical model is adapted from the one proposed by Bozza et al. [12] for handwriting comparison, where a two level model was proposed to deal with the within-writer and the between-writers variability. Bozza et al. [12] assumed independence between questioned and reference handwriting under the hypothesis in the denominator. This assumption is however not reliable in the questioned signature context. Forgery is type of mimicry [13] and undeniably depends on a 'target' signature. The proposed approach briefly sketched in (1) allows taking into account this critical issue.

The proposed model must provide accurate and reproducible evidential values in order to be accepted in a court of law. The performance, as defined in this study, is composed of the accuracy and the reproducibility. The concept of accuracy in the Bayesian framework (2) is here defined as the percentage of BFs supporting the proposition the questioned data came from. The reproducibility is represented through the variance of the accuracy throughout the repetitions and using different reference sets. The 'errors' are quantified through the rate of misleading evidence (RME). The evidence can be misleading in favor of any of the propositions of interest. Here $RME(H_1)$ expresses the proportion of forged signatures the model supports as being genuine, with $RME(H_2)$ being the opposite. The RME is the sum of both rates. By minimizing the criterion defined in (3), we obtain the best performing model. Performance is evaluated for the selected feature sets, throughout simulated cases. All of these simulations are random trials. Cases were generated using both reference signatures and forgeries of the same signature, all with known-source data. Questioned signatures are randomly drawn from both the genuine and forgery databases. The reference signatures are randomly drawn from the reference signature database. Two other databases served as background information and contained non-case-related genuine signatures and simulated signatures.

$$\text{Accuracy} = \#\text{Correct BF} / \#\text{BF} \quad (2)$$

$$\text{PerfCrit} = \text{RME} * \sigma 2_{\text{Accuracy}} \quad (3)$$

Features used in this study were both dynamic and static, but they were exclusively global (or 'parameter') features. According to Richiardi et al. [14] a global feature is: '*where one feature is extracted for a whole signature, based on all sample points in the input domain*'. A total of 60 global features covering timing, pressure, distance and velocity information was available. The preliminary feature selection was conducted by assessing the performance of every possible trivariate (multivariate combination of three variables) feature set with (3). The best performing feature set per signature (when trained with 100 reference signatures) was select-

J. Linden is with the School of Criminal Justice, University of Lausanne, Lausanne, CH 1015 Switzerland (corresponding author: +41 21 692 46 54; fax: +41 21 692 46 05; e-mail: Jacques.Linden@unil.ch).
S. Bozza is an Associate Professor at the Ca' Foscari University of Venice, Dorsoduro 3246, 30123 Venice. She is also teaching at the School of Criminal Justice, at the University of Lausanne, Lausanne, CH 1015 Switzerland (e-mail: Silvia.Bozza@unil.ch).
R. Marquis, Ph.D, is with the School of Criminal Justice, University of Lausanne, Lausanne, CH 1015 Switzerland (e-mail: Raymond.Marquis@unil.ch).
F. Taroni is a Full Professor at the School of Criminal Justice, University of Lausanne, Lausanne, CH 1015 Switzerland (e-mail: Franco.Taroni@unil.ch).

ed for the actual experiment. The selected feature sets are summarized in Table I. Feature sets differed between the signatures, especially between signature 3 and signatures 1 and 2. The only common feature type in all three sets was timing information. Signature 3 differed most in feature selection, possibly because the signer is left-handed. The laterality of the signer may affect the discriminative power of the writing angles. The selected feature sets were then used in the main experiment. In this experiment, multiple trials were carried out while varying the number of reference signatures available to the model. For each selected feature set 10'000 trials per experimental condition were carried out. These trials were used to investigate the model's accuracy and reproducibility in a diverse set of conditions.

	Feature set	Feature Types	Description
Signature 1	Height-Totaltime-dt1_bar	Spatial, Time, Velocity	Height : Vertical distance from upmost to lowest pixel Totaltime : Time to complete signature dt1_bar : Mean tangential velocity
Signature 2	Width-Downtime-dt1_bar	Spatial, Time, Velocity	Width : Horizontal distance from leftmost to rightmost pixel Downtime : Time spent with pen touching the tablet dt1_bar : Mean tangential velocity
Signature 3	Totaltime-TVD_var-dp2_max	Time, Angle, Pressure	Totaltime : Time to complete signature TVD_var : Variance of the velocity's angle (to horizontal axis) dp2_max : Maximum of 2nd temporal differentiate of Pressure

Table 1. Feature sets. a. Summary of the feature selection (after a preliminary trial).

B. Signature Data

For the non-case-related signatures, referred to as background data, people were asked to sign twenty times on a Wacom DTU 1141 signature tablet. Twenty-three people participated in the acquisition. The population contained a total of 460 genuine signatures. For the reference signatures, three individuals were asked to sign regularly, in the same conditions as for the background data. They signed approximately once every three weeks over the length of one year. Two hundred signatures per signer were selected from this pool of genuine signatures. They served as potential materials to be drawn as "reference" signatures in the simulated cases. Forgers were volunteers, recruited from the University of Lausanne's forensic science department. All of the forgers participated in a "contest" with a reward for the best forger. All of the forgers had a briefing on dynamic signatures and data privacy before participating. No instructions were given regarding forgery strategy. All forgers chose freehand simulations, except for one person who preferred a dynamic tracing. Forgers were presented with six specimen of one of the reference signatures. They were free to train on both paper and tablet for 15 minutes prior to forging. They could keep the references in sight during the process. They could also choose to reject their forgeries and start anew, until they produced a total of 10 forgeries. The collected data for every one of the three reference signatures is summarized in Table 2.

	Signature 1	Signature 2	Signature 3
# Genuine	200	200	200
# Forgeries	280	400	160
# Forgers	28	40	16

Table 2. Signature datasets. a. Summary of case-related data used to for randomly drawing the trial materials.

A Wacom DTU 1141 signature acquisition tablet was used for the data acquisition. The sampling rate of the tablet is 200 Hz, with a coordinate resolution of 2540 lines per inch (lpi) and 1024 levels of pen pressure measured on the pen axis. Wacom drivers and software were used for data acquisition. The signature data was imported into the R Statistical Software package for data treatment, feature extraction, comparison and probabilistic evaluation.

3. RESULTS

The accuracy of the selected features sets for the three studied signatures is represented on Figure 1. The reproducibility of the models is represented by the accuracy's variance for each condition, visible on Figure 2. The lower the variance of the accuracy, the more reproducible the model is when repeated with a different sample. Accuracy is above 90% for every one of the signatures, even when only 5 reference signatures are available. First, a steep improvement in accuracy occurs when more reference signatures are available. This addition does affect both accuracy and reproducibility, as can be seen in Figure 2. The variability of accuracy between trials lowers quickly with an increasing number of reference signatures. As the sample size grows, the sensitivity to sampling effects decreases and the modeling of the signatures variation improves. The model then attains a maximum for both accuracy and reproducibility. This particular zone exists with reference sets between 15 and 40 reference signatures. When using more sizeable reference material sets, accuracy and reproducibility decrease for two of the three signatures. We expected reproducibility to increase with the number of reference signature, and is acceptable around sets of 15 reference signatures. Here, the decrease in both accuracy and reproducibility may be explained by a combination of two reasons. First, the reference data contains outliers, which produce misleading BFs when included. When the size of the reference sets increases, the probability of incorporating outliers increases. The second reason is due to the data quantity and the model. The chosen model is symmetrical and unimodal. The more data is available, the more the model will focus on the 'main class' of reference signatures, therefore centering on those 'normal' signatures. The outliers become less probable under the distribution, as it shifts away from them. This may cause an increase in misleading evidence (RME). It was noted that the decrease in accuracy is especially apparent in signature 2, the shortest and simplest of the three signatures.

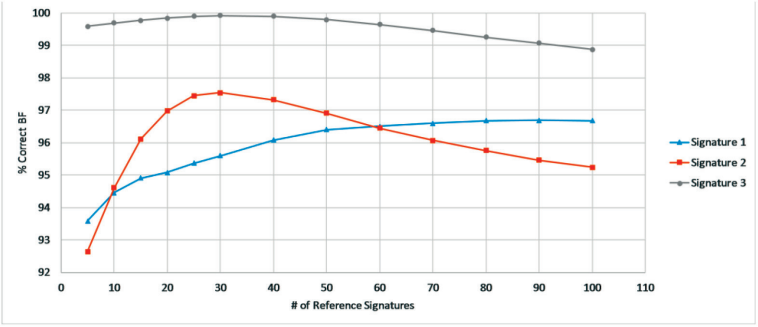


Figure 1. Accuracy for the 'best' performing feature set per signature.

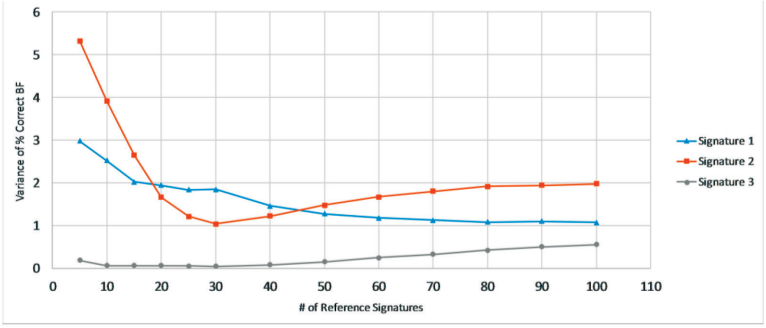


Figure 2. Reproducibility for the 'best' performing feature set per signature.The lower the value of variance was, the higher the reproducibility for the experimental condition.

Dynamic features may be subject to more substantial variability than static features, especially in short and simple signatures. As a possible explanation, contemporaneity of questioned and reference signatures may impact performance more strongly than for static features. The same goes for writing conditions, including posture, writing surface and stability, although we did not test this. The examiner should take care when using dynamic features in short signatures, until their specific variation and the impact of extrinsic and intrinsic factors is better known. As a result using an undersized data sets may give a false impression of robustness and strength when presenting the evidence, because accuracy and reproducibility cannot be guaranteed for dynamic features with small reference signature samples. Overall, most of the selected features behave well with as few as 15 reference signatures. Below this number, performances might still be acceptable, but especially the reproducibility of results suffers. Results may then strongly depend on the available reference set, forcing the examiner to carefully select the included signatures.

4. DISCUSSION

The proposed model aims to develop a methodology in the examination of dynamic signatures and provides a promising tool for forensic examiners. Although other methodological approaches exist in the literature [15, 16] to address the thematic under study, the proposed model is the only one to use a forensic approach to the evaluation of evidence, through the transparent quantification of performance and communication of results. Further, it can be applied to both dynamic and static signatures, on any continuous global feature set. It is versatile and provides high accuracy and reproducibility. Performance for the three tested signatures is high. Most of the generated BFs correctly support the correct hypothesis, even when few reference materials are available. Accuracy of the models is above 90% and reproducibility high, even though 'only' global features are used. Although feature selection has a strong impact on the results, the proposed model is itself logically coherent and provides justifiable evidential value statements. The importance of reproducibility was highlighted through the low sample number scenarios, especially when providing estimates of the rate of misleading evidence. As transparency on "error rates" is an essential part of the evidence requirements according to the Federal Rules of Evidence (FRE702) and Daubert criteria, one should only provide results for the selected feature set, as well as avoid estimates based on small samples. Reference materials play a large role in the reproducibility and accuracy of the model.
The proposed model performed well in (forensic) operational conditions, respects forensic science principles and recent evidence interpretation recommendations. When using global features, using around 15 reference signatures is recommended to guarantee reproducible results. Global features such as the ones used in this study may complement the features used by forensic signature examiners. The model is also applicable to any type of features and not limited to dynamic signatures. Additionally, the statistical approach and the use of empirical data provide additional support for the examiner's conclusions.

5. ACKNOWLEDGMENTS

The authors would like to thank Wacom Europe GmbH for providing the soft- and hardware necessary for the study. The authors also thank all the volunteers that have provided genuine and forged signatures.

REFERENCES

[1] J. Linden, R. Marquis, S. Bozza, F. Taroni, Dynamic signatures: A review of dynamic feature variation and forensic methodology, Forensic Sci Int 291 (2018) 216-229.

[2] B. Found, C. Bird, The Modular Forensic Handwriting Method, Journal of Forensic Document Examination 26(1) (2016) 7-83.

[3] R. Huber, A. Headrick, Handwriting Identification: Facts and Fundamentals, CRC Press, Boca Raton, 1999.

[4] C.G.G. Aitken, F. Taroni, Statistics and the Evaluation of Evidence for Forensic Scientists, Second Edition, John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 2004.

[5] B. Robertson, G.A. Vignaux, C.E.H. Berger, Interpreting Evidence: Evaluating Forensic Science in the Courtroom, John Wiley & Sons, Ltd2016.

[6] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, A model for case assessment and interpretation, Science & Justice 38(3) (1998) 151-156.

[7] I.W. Evett, Towards a uniform framework for reporting opinions in forensic science casework, Science & Justice 38(3) (1998) 198-202.

[8] ENFSI, ENFSI Guideline for evaluative reporting in forensic science, Strengthening the Evaluation of Forensic Results across Europe (STEOFRAE), European Network of Forensic Science Institutes, 2015.

[9] ENFSI, Best Practice Manual for the Forensic Examination of Handwriting Version 1, European Network of Forensic Science Institutes, 2015.

[10] C. Aitken, P. Roberts, G. Jackson, Fundamentals of Probability and Statistical Evidence in Criminal Proceedings, in: C. Aitken (Ed.) Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, Royal Statistical Society, http://www.rss.org.uk/statsandlaw, 2010, p. 122.

[11] Association of Forensic Science Providers, Standards for the formulation of evaluative forensic science expert opinion, Science & Justice 49(3) (2009) 161-164.

[12] S. Bozza, F. Taroni, R. Marquis, M. Schmittbuhl, Probabilistic evaluation of handwriting evidence likelihood ratio for authorship, Applied Statistics. 57(3) (2008) 329–341.

[13] J. Galbally, M. Gomez-Barrero, A. Ross, Accuracy evaluation of handwritten signature verification: Rethinking the random-skilled forgeries dichotomy, 2017.

[14] J. Richiardi, H. Ketabdar, A. Drygajlo, Local and global feature selection for on-line signature verification, in: B. Werner (Ed.) Eighth International Conference on Document Analysis and Recognition, IEEE Computer Society, Seoul, Korea, 2005, pp. 625-629.

[15] M.I. Malik, Automatic Signature Verification Bridging the Gap between Existing Pattern Recognition Methods and Forensic Science, Department of Computer Science, Technische Universität Kaiserslautern, 2015.

[16] A. Balestrucci, D. Impedovo, G. Pirlo, Processing of Handwritten Online Signatures: An Overview and Future Trends, in: B.L.D. Bezerra, C. Zanchettin, A.H. Toselli, G. Pirlo (Eds.), Handwriting: Recognition, Development and Analysis, Nova Science Publishers, Inc., Hauppauge NY, 2017, pp. 363-386.

Adaptive learning rates schedules: performance evaluation

Gianmarco Ragognetti, Antonio Parziale, Rosa Senatore, Angelo Marcelli
Department of Information and Electrical Engineering and Applied Mathematics
University of Salerno - Fisciano (SA), Italy

Abstract - In the framework of reinforcement learning, we compare the performance of the most common adaptive learning rate schedules in explaining the behavior of human subjects while learning the sequences of actions to be executed in response to sequences of stimuli for achieving a reward. For the purpose, we find the best fitting between the linear time decay, the exponential time decay and the reward based adaptive learning schedules with the data collected in a set of experiments with human subjects, and estimate the hyper-parameters of the schedules. The experimental results show that the reward based schedule outperforms the other ones, but that its performance decreases as the difficulty of the task increases.
Keywords: reinforcement learning, machine learning, adaptive learning rate, performance evaluation

1. INTRODUCTION

Among the models of learning employed in machine learning, reinforcement learning models hail from the behaviorist psychology domain: these models allow a software agent to explore an environment and learn by positive outcomes, *rewards*, or negative outcomes, *punishments*. The idea behind reinforcement learning (RL) is that an agent will learn from the environment by interacting with it and receiving rewards/punishments for performing actions. While learning process goes on, those actions leading to more consistent rewards will become more frequent. That is why in reinforcement learning, the best behavior is the one that maximizes the expected cumulative reward.
The most widely adopted architecture of a RL agent is the Actor-Critic one, depicted in figure 1: the *Actor*, given the sensory stimuli it receives in input, chooses the action to perform, while the *Critic*, which receives the reinforcing (*rewards*) or punishing stimuli (*punishments*) from the environment, favors the linking of stimuli to the actions leading to the maximum reward. In the literature there have been proposed many different implementations of the Actor-Critic architecture, depending on the policy adopted by the actor to select the actions given the stimuli, by the function adopted by the Critic to represent the reward/punishment feedback and the intertwining between them (Kaelbling at al, 1996), (Witten et al, 1977), (Barto et al 1983).
In this context, we addressed the problem of comparing the performance of three different models proposed in the literature for dynamically adjusting the learning rate adopted by the RL algorithms when the learning task aims at selecting a sequence of actions in response to a sequence of stimuli, and the reward/punishment is not provided for each action, but only when the whole sequence of actions is executed. For the purpose, we have designed a learning task, recorded the performance of 17 human subjects recruited for the experiments and eventually fit the data with the three most widely adopted models of adaptive learning rates, namely linear time decay, exponential time decay (Powell et al, 2006) and reward based decay (Gershman, 2015). Eventually, we compare the behavior of the agent incorporating the best adaptive learning rate schedule with the human one.

In the remaining of the paper, In Section 2 we describe the learning experiment performed by humans and the data we collected for estimating the learning curve, while in Section 3 we describe the implementation of the artificial neural network model we use comparing the performance of the adaptive learning schedules. In Section 4 we present the experimental results, and in the Conclusion we summarize the results and outline the future investigations they suggest.

2. LEARNING EXPERIMENT

Experiments involved the learning of the association between sequences of visual stimuli and sequence of actions, where reward/punishments were provided at the end of the sequence. Visual stimuli were displayed on monitor, and consisted in circles of different colors. The action consisted in pressing one of two buttons of a cloche. Reward and punishment consisted in a visual and acoustic stimulus, consisting, respectively of two green/red vertical bars prompted on both sides of the screen and a cheering/noisy sound. In case none of the buttons were pressed in a time span of 5 seconds from the presentation of the visual stimuli the response was considered wrong and punishment provided as feedback. Figure 1 illustrates the visual stimuli and the cloche for actuating the motor commands.

We have designed the following learning task: a visual stimulus is presented and the learner selects one of the possible actions, receiving as feedback the next visual stimulus, depending on the selected action. At the end of a sequence of visual stimuli, a reward/punishment is provided as feedback. The purpose of the learning it selecting the sequence of the actions that leads to the reward. In order to investigate to which extent the complexity of the learning task affect the performance, we have implemented three tasks, varying the number of visual stimuli, the length of the sequence and the number of sequence to learn:

- *Task 1*: 4 visual stimuli, grouped in two sets of 2 visual stimuli, and 2 steps long sequences; in the first step one of the stimulus of the first set is randomly presented, while in the second step the visual stimulus is selected among the second set. Among the possible 4 sequences of stimuli, 2 of them, one for each of the visual stimulus presented in the first step lead, to rewards.
- *Task 2*: 6 visual stimuli, grouped in two sets, consisting of 4 and 2 visual stimuli respectively, and 2 steps long sequences; in the first step one of the stimulus of the first set is randomly presented, while in the second step the visual stimulus is selected among the second set. Among the possible 8 sequences of stimuli, 4 of them, one for each of the visual stimulus presented in the first step, lead to rewards.
- *Task 3*: 6 visual stimuli, grouped in three sets of 2 visual stimuli, and 3 steps long sequences; in the first step one of the stimulus of the first set is randomly presented, while in the second and in the third step the visual stimulus is selected among the second and third set respectively. Among the possible 8 sequences of stimuli, 2 of them, one for each of the visual stimulus presented in the first step, lead to rewards.

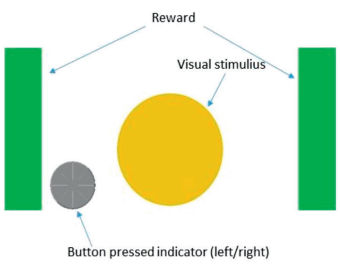


Figure 1. The experimental setting: the subject has to press a button on the cloche after being presented with a coloured circle. At the end of the sequence a reward (green bars and cheery sound) or a punishment (red bars and noisy sound) is provided to the subject. The task consists in learning to match each stimuli with the right button to receive the reward.

In every experiment, each sequence starting with a different visual stimulus was shown 20 times (epochs), leading to 40 sequence presentation (trials), resulting in 80 stimuli presentations for task 1 and 3, and 80 trials and 160 stimuli presentations for task 2.
The experiments were carried out by 17 volunteers, mainly master students enrolled in our departments. Their age is in the range 22-24, with a mean value of 22.7 and a standard deviation of 1.3. All the subjects have normal or correct to normal vision, do not report motor impairments and signed an informed consensus form.
For each subject, and for each sequence of visual stimuli, we recorded whether the selected actions received a reward or a punishment. For each trial, we derived a *score vector* made of as many entries as the number of trials, and the value of *i-th* entry is the sum of the reward/punishment achieved after *i* trials. To obtain the learning curve of the experiment, i.e. the curve representing the learning performance during the experiment, we use as measure of performance the *normalized score vector*, which is obtained by normalize the value of the *i-th* entry with respect to *i*. Thus, for each task we have 17 learning curves. Eventually for characterizing the collective behavior of the subjects performing the task, we compute the collective learning curve by averaging the performance of the subjects at every epoch.

3. PERFORMANCE EVALUATION

In machine learning, the most widely adopted architecture of a RL agent is the Actor-Critic one, depicted in figure 1: the Actor, given the sensory stimuli it receives in input, chooses the action to perform, while the Critic, which receives the reinforcing (rewards) or punishing stimuli (punishments) from the environment, favors the linking of stimuli to the actions leading to the maximum reward.

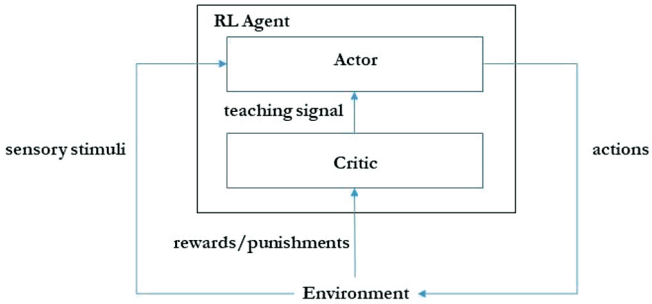


Figure 2. Architecture of a Reinforcement Learning (RL) agent. The agent interacts with the surrounding environment performing actions and receiving sensory stimuli and rewards or punishments according to the action performed. The agent is modeled within the Actor–Critic framework, in which the Actor selects the Action to perform and Critic provides a teaching signal to the Actor on the basis of the rewards and punishment provided from the environment.

In the literature there have been proposed many different implementations of the Actor-Critic architecture, depending on the policy adopted by the actor to select the actions given the stimuli, by the function adopted by the Critic to represent the reward/punishment feedback and the intertwining between them (Kaelbling et al, 1996), (Witten et al, 1977), (Barto et al 1983). In particular, we adopted the learning rule suggested by Barto (Barto et al 1983), as it reflects more closely the underlying architecture of the neural correlates, and taking into accounts that the rewards/punishment are provided only at the end of the sequence, the learning rule becomes:

$$w_{ij}(t+1)= w_{ij}(t) + \eta(t)*(r_t+1-w_{ij}(t)) \quad (1)$$

where η(t) is computed for the linear time decay, the exponential time decay and the reward based schedules as it follows (Powell et al,2006), (Gershman, 2015):

1. *linear time decay* $\eta(t+1)= \eta(t)/(1+ \alpha*t)$ (2)
2. *exponential time decay* $\eta(t+1)= \eta(t)*\exp(-\alpha*i)$ (3)
3. *reward based* $\eta(t+1)= \eta(t)+ \alpha*(1-r_{t+1}-\eta(t))$ (4)

where α is a parameter model.

Given the tasks, the implementation of eq. (1) incorporating the schedules (2-4) resulted in three artificial neural network having 6, 12 and 14 input nodes for Task 1, Task 2 and Task 3, respectively. These implementations were then fitted with the experimental data by minimizing the root mean squared error (RMSE) between the collective learning curve of the subjects with each model. The best value for the parameter α in the equations (2-4) was obtained brute force by computing the best fitting for values of α within the range 0.001 to 0.999, with 0.001 step size.

4. EXPERIMENTAL RESULTS

The results obtained by fitting the experimental data with the three implementations are shown in Table 1, which reports for each task and for each learning schedule, the RMSE and the values of α corresponding to the best fit. The data show that for Task 1 all the schedules exhibit their best performance in correspondence of the same value of α, but only the reward based rule achieves the best performance with the same value of α for all the tasks, while the other schedules required larger value as the complexity of the task increases.

Figure 3 illustrates the results of the fitting by showing, for each task, the learning curve derived from the data, i.e. the human one, and the one exhibited by the three models, making even more evident that the reward based rule fits almost perfectly the data. They also show that, as task difficulty increased, the performance of the rules based on time decay worsens, and this is in according with recent findings reported in the literature (Gershman, 2015), according to which learning rate adapts to the distribution of rewards and less on other features of the reward structure.

	Reward based		*Linear time decay*		*Exponential time decay*	
	α	RMSE	α	RMSE	α	RMSE
Task 1	0.010	0.003	0.010	0.047	0.010	0.049
Task 2	0.010	0.017	0.129	0.200	0.129	0.207
Task 3	0.010	0.013	0.027	0.207	0.024	0.214

Table 1. Model evaluation. For each row, the data report the RMSE corresponding to the best fitting for each of the adaptive learning rate rule and the corresponding value of α.

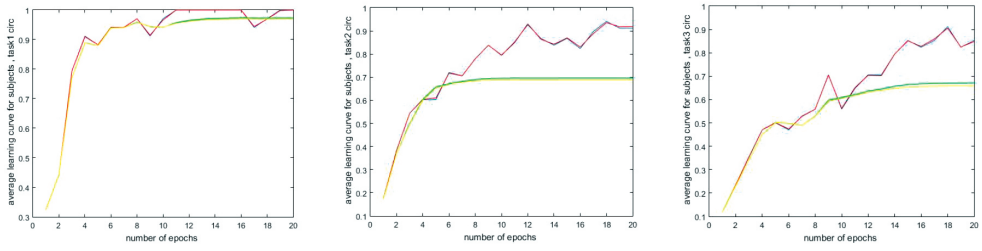


Figure 3. Model validation. The learning curves corresponding to the experimental data (blue), and the best fitting learning curves exhibited by the reward based (red), linear decay (green) and exponential decay (yellow) rules for learning rate adaptation. The panels, from left to right, report the result for Task 1, Task 2 and Task 3, respectively.

5. CONCLUSIONS

We have addressed the problem of comparing the performance of the most widely adapted adaptive learning rate schedules. The performance where evaluated by incorporating the schedules into a fully connected artificial neural network, and comparing its learning curve with that exhibited by 17 subject performing 3 different learning tasks.
The experimental results show that the implementation adopting the reward based schedule provides the best fitting among the three, with an RMSE which is one order of magnitude smaller that the ones exhibited by the other. Even more interesting, the RMSE values for the best model are very small, denoting an almost perfect fit between the data and the model.
They also show that reward based adaptive learning rate schedule exhibits performance comparable to human in two of the learning tasks. As it happens for the most complex task among the considered one, we speculate that it may depend on some hidden information humans bring in while approaching more complex learning tasks. This aspect is the focus of our current and future investigations.

REFERENCES

[1] Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). Reinforcement learning: A Survey. *Journal of Artificial Intelligence Research*, 4:237-285.

[2] Witten, I. H. (1977). An adaptive optimal controller for discrete-time Markov environments. *Information and Control*, 34:286-295

[3] Barto, A. G., Sutton R. S., and Anderson C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems." *IEEE transactions on systems, man, and cybernetics* 5: 834-846.

[4] Abraham, G.P., and Powell, W.B (2006). Adaptive step size for recursive estimation with applications in approximate dynamic programming, *Machine earning*, 65(1): 167-198

[5] Gershman, Samuel J. "Do learning rates adapt to the distribution of rewards?." *Psychonomic bulletin & review* 22.5 (2015): 1320-1327.

Biometric Signature:
Analysis in forensic field of forged tracing over the original

Massimo Baraldi[a], Silvia Benini[a], Tatiana Zucco[a]

[a]Researchers of Centro Ricerche sulla Scrittura
Via Antonio Gramsci, 17, 61040 Monte Porzio PU - Italy
(massimobaraldi@grafologi-forensi.it; silviabenini@beninistudio.com; tatiana_zucco@yahoo.com)

Abstract - The biometric signature is a tool used for an electronic document subscription with legal value. By European law the biometric signature has the same legal validity and judicial evidence as signature on paper. In the same way, in fact,
biometric signature may be disowned and become object of expertise in a court. The biometric signature doesn't have the same characteristics as the signature on paper. For this reason the investigation methodology used for paper signature isn't useful to solve forensic questions coming from a biometric signature. In this study we have considered one of the most classic systems of signature falsification: the forged tracing over the original (forged trace), related to tablets. Our hypothesis is that reproduction by tracing could be one of the most frequent signature falsification methods, because it is very easy to achieve.
We have set ourselves the goal of identifying the biometric indices that define the forged trace with respect to those spontaneous of typical signatures. In the forensic field it is essential to refer to objective evaluation indexes to ascert whether a contested signature is really a forgery or is one of the variants of the model of spontaneous signature of the subject. A subsequent extension to our research can be carried out by analyzing the same signature sampling with the aim of identifying the author of the forgery.
Keyword: biometric signature, tracing, forged tracing over the original (forged trace), forged signature, forensic investigation, false verification on tablets.

1. INTRODUCTION

In the presence of new computer technologies also used in the forensic field, we wanted to verify if, and in what way, information derived from biometric data of a signature on a tablet can be considered predictive of a falsification. The signature, as a recognition system, is considered a behavior. Y. Bay et al. (2017) argues that "physical biometrics including physical part of the human body such as, fingerprint, retina, vein, palm, etc. Signature reveals some information about the signer and involves data derived from an action therefore and in contrast to physical biometrics, signature recognized as strong behavioral biometrics". The writing act is influenced by both internal and external factors (M. Diaz, M. A. Ferrer, D.
Impedovo & G. Pirlo, 2018). Likewise, the biometric signature is also an expression of dynamic characteristics that can be identified in the "pressure exerted, tilts, position or velocity of the stylus. All this signals provide not only information of the signature but also information about the act of signing, which is considered more related to the specific user (A.Mendaza-Ormaza -2009) ". In the scientific field, several researches have already been carried out in terms of recognition of signature and degree of validity inherent dynamic recognition compared to others of a static type (Y. Bay, M. Erbilek, A. Fosuah Gyasi and E. Celebi Cyprus, 2017), (Aini Najwa Azmi, Dewi Nasien & Fakhrul Syakirin Omar, 2016), (A. S. Syed Navaz & K. Durairaj, 2016), (František Hortai, 2017). Particularly in an interesting research carried

out on the simulation of cursive writing "all statistical analyses unequivocally showed that simulated handwriting results in longer reaction times, is produced at a slower rate, and is generated by more frequent but smaller force pulses. The third conclusion is that the stiffness of the writing limb is greater when imitating another person's style" (G. P. Van Galen & A.W.A. Van Gemmert, 1996). The consideration data about the analysis of signing on tablets concern, in particular, the presence of constants with respect to the variants. Various authors have been interested in the study of the biometric signature. M. Soltane et Al (2010) stated that "BA is becoming an important alternative to traditional authentication methods such as keys ("something one has", i.e., by possession) or PIN numbers ("something one knows", i.e., by knowledge) because it is essentially who one is, i.e., by biometric information". Another research on speed seems to confirm that with the increase of this the precision of execution decreases (N. Dounskaia · A.W.A. Van Gemmert · G.E. Stelmach, 2000) and that "a writer cannot reproduce a writing pattern exactly" (H.L. Teulings,1996). As it happens for the tracing of the signatures with ink, "retraced signatures involve the nearly exact replication of a signature model. This can be achieved by tracing over the genuine signature in such a way where an indentation or image of the signature is copied onto a surface under the signature model" (H. H. Harralson, 2013).
In 2011 the margin of error of recognition seemed still imprecise: "In online signature verification systems, additional features such as pen pressure, pen speed and pen tilt angle have made the process of forging online signatures more difficult. Equal error rate of available online signature verification systems lies between 1 to 10%." (S. Tariq, S. Sarwar & W. Hussain, 2011). To date, research is being expanded with the aim of making recognition and therefore the instrument of biometric, more secure and reliable signature. Our research intends to extend this topic, with specific regard to the falsification of the signature through the tracing of the track. The study was set up to compare spontaneous signatures and signed signatures, identifying the categories of writing that highlight the most evidence.

2. METHODS

Subjects

A sample of 150 individuals (79 females and 71 males) who have no skills on writing analysis aged between 20 and 70 years, right-handed, with a minimum diploma qualification, was chosen. The participants were given essential performance indications without harsh conditioning, but taking care that everyone assumed a comfortable and natural posture.

Instrumentation

A Wacom DTU1141 Tablet and the Movalyzer Neuroscript Software for the acquisition of signatures and data analysis were used for the experiment.

Conditions

The conditions applied are Natural Writing (NW) and Retraced Writing (RW).

Experimentation course

The neutral phrase "firma grafometrica" has been used, because it contains a sufficient variability of literal forms and spatial development. A person unrelated to the work group, which we called "master", wrote 10 times "firma grafometrica" using the tablet. We have recorded the biometric data of the 10 master writings and among these we have chosen a random one

to be used as a tracing matrix (RM from RETRACED MASTER). To guarantee the same performance conditions, 150 copies of the RM render were used on paper with a weight of 80 g/m^2 and each participant used an intact (untouched) copy. The 150 sample subjects were asked to write 20 times the neutral phrase "firma grafometrica" on the tablet in NW condition. In this way the biometric data of each subject, identifying the subjective parameters of each one, were recorded. In this way the biometric data of each subject, identifying the subjective parameters of each one, were recorded. Subsequently, each subject has traced the signature of the RM rendering 10 times by placing the same sheet on the tablet and holding it firmly with the left hand, as per RW condition. Of the 20 spontaneous signatures and 10 traces, all trials and the entire graphometric writing data package including on-air movements were taken into consideration. For STROKE / SEGMENT we mean "at velocity zero crossing" as foreseen among the options of the Movalyzer program.

3. RESULTS

The following indexes analyzed by the software were considered1: Segment (SGT), Start Time (ST), Duration (DUR), Start Vertical Position (SVP), Vertical Size (VS), Peak Vertical Velocity (PVV), Peak Vertical Accelleration (PVA), Start Orizzontal Position (SOP), Orizzontal Size (OS), Straightness Eerror (SE), Slant (SLA), Loop Surface (LS), Relative Initial Slant (RIS), Relative Time To Peak Vertical Velocity (RTTPVV), Relative Pen Down Duration (RPDD), Absolute Size (AS), Average Absolute Velocity (AAV), Road Length (RL), Absolute Y Jerk (AYJ), Average Normalized Y Jerk Per Trial (ANYJPT), Absolute Jerk (AJ), Normalized Jerk (NJ), Average Normalized Jerk Per Trial (ANJPT), Score (SCO), Average Pen Pressure (APP), Numbers Of Strokes (NOS). The summarized data available in Movalyzer were exported to an Excel spreadsheet and, by query; the statistical averages for each index were counted. The classification of the data follows two perspectives of analysis: one for Subject / Condition (SC) and the other for Subject / Condition / Trial (SCT). In this way it was first outlined the differences between NW and RW and then go into the details of individual trials to see if the result of the SC behavior also occurs in the one to one comparison in SCT. In the SC perspective the mean values of the RW condition were compared with those of the NW condition and the results of the subjects were grouped into: major (>), minor (<), equal (=), expressing them as a percentage (Table 1). In the SCT (Fig. 1) perspective it is compared the mean values of each single trial of the RW condition with those of the NW condition and we grouped the results of the subjects in: Always Highter (AH), Always Less (AL), Variable (VAR) expressing them in percentage (Table 1).

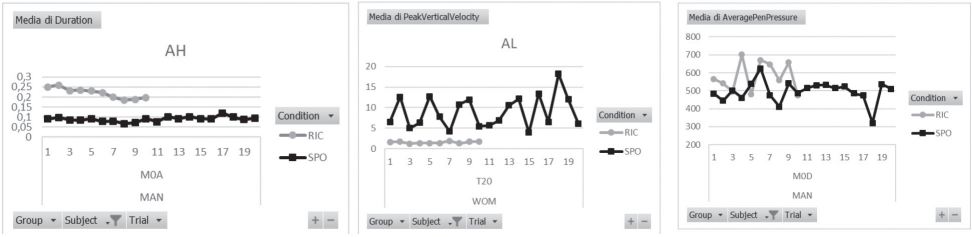


Fig. 1.

[1]Neuroscript Movalyzer Help - Processing, charting ad analysis - Viewing trials.

The percentage of distribution of the examined subjects is indicated for each index considered, following the comparison between RW and NW, according to the SC and SCT analysis perspectives. In summary, the indices observed by the SC perspective are the following:

- 14 indices group more than 90% of the population as a whole or in a group < Of these 14 indices, 8 affect - 99-100% of the population.
- the SEG index gave an incongruous result with a result of 67% >, 31% <, 1% =.

Observed from the SCT perspective are the following:

- SEG confirms inconsistent outcome resulting in 18% AH, 1% AL, 81% VAR;
- 20 indexes show the whole VAR more than 50% with a maximum value of 97% for RTTPVV;
- 8 indexes show a VAR grouping of less than 50% with a minimum value of 21% for AAV (which in any case represents 31 subjects out of 150);

Indexes	SC				SCT		
	>	<	=		AH	AL	VAR
SGT	67%	31%	1%		18%	1%	81%
ST	99%	1%			53%		47%
DUR	100 %				73%		27%
SVP	47%	53%			11%	14%	75%
VS	31%	69%			2%	17%	81%
PVV	1%	99%			2%	67%	33%
PVA	1%	99%				69%	31%
SOP	90%	10%			40 %		60%
OS	36%	64%			1%	11%	89%
SE	67%	33%			3%		97%
SLA	49%	51%			2%	4%	94%
LS	19%	81%			1%	12%	87%
RIS	94%	6%			18%		82%
RTTPVV	20 %	80%				3%	97%
RPDD	87%	13%			13%	1%	86%
AS	25%	75%			1%	20%	79%
AAV		100%				79%	21%
RL	23%	77%			1%	21%	78%
AYJ	1%	99%				60%	40%
NYJ	97%	3%			43%		57%
ANYJPT	97%	3%			45%		55%
AJ	1%	99%				58%	42%
NJ	98%	2%			43%		57%
ANJPT	97%	3%			44%		56%
SCO	32%	68%			1%	4%	95%
NOPOAP	99%	1%			73%		27%
APP	79%	21%			17%	1%	83%
NOS	70%	30%			18%	1%	81%

Table 1. (percentage of subject)

4. DISCUSSION

In the evaluation it was not possible to identify indices useful for a legal standpoint in order to recognize with certainty a spontaneous signature from a forged one. Initially, some data aggregated in the SC perspective indicated an objective significance that was however denied by the SCT analysis perspective. The only exception is the net difference in RW with respect to NW of the total duration of each individual signature. It has also emerged that, analyzing the SCT data, in 15 of the 28 indexes the analyzed population is enclosed in only two of the AH-AL-VAR options and there is never the concurrence of the opposing options AH-AL but always or AH-VAR or AL-VAR. These 15 indices describe that in the tracing phase there is an increase (for ST, DUR, SOP, SE, RIS, NYJ, ANYJPT, NJ, ANJPT, NOPOAP) or a decrease

(PVA, RTTPVV, AAV, AYJ, AJ) of the average value investigated. About the remaining 13 indices (SGT, SVP, VS, PVV, OS, SLA, LS, RPDD, AS, RL, SCO, APP, NOS) the contrast is present in AH-AL even if in some cases for 1 subject only. This makes them unsuitable for the detection of forged trace in the forensic field. Regarding to the presence of the strong VAR component, it should be emphasized that it emerges predominantly in 22 indexes, 4 of which report over 90%, while only 7 indexes out of 29 show a variability of less than 50% and for no index falls below 27%. This evidence of preponderance of the variability component can be hypothesized due to:

- Genuine spontaneous individual variability and / or minimum differentiation of behavior between RW and NW trials as shown in Figure 2. Indexes characterized by this type of variability therefore appear to be unsuitable to reveal that it is tracing (Fig.2);
- variability due to the loss of attention in implementing the forged trace and showing the true nature of the counterfeit as shown (Fig. 3). In this case the index would be significant and could reveal the author.

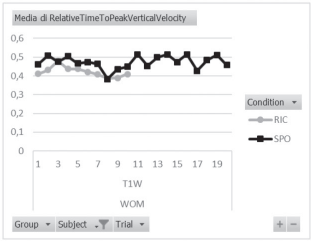


Fig. 2.

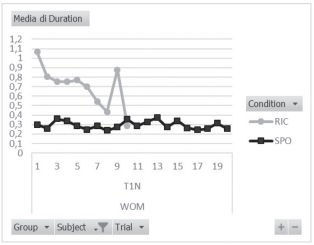


Fig. 3.

These hypotheses could be explored with further research, aimed at investigating the relationship and the nature of variability in the 16 indexes in which there is the absence of the coexistence of the AH-AL dichotomy. A subsequent expansion could also concern the study-comparison of the indexes of the signatures recalculated with the indexes of the master signature, aimed at identifying the forgery.

REFERENCES

[1] M. Diaz, M.A. Ferrer, D. Impedovo & G. Pirlo, 2018, *A Perspective Analysis of Handwritten Signature Technology*.

[2] C. De Stefano, M. Garruto and A. Marcelli, 20 *A Saliency-Based Multiscale Method for On-Line Cursive Handwriting Shape Description 04*.

[3] C. De Stefano, G. Guadagno and A. Marcelli, 2004, *A saliency-based segmentation method for online cursive handwriting*.

[4] S. Habin and F. Javed Zareen, 2015, *Biometric signature verification*.

[5] S James, S Pal, 2000, *Biometrics-Signature Verification Techniques*.

[6] Aini Najwa Azmi1 & Dewi Nasien1 & Fakhrul Syakirin Omar1, 2016, *Biometric signature verification system based on freeman chain code and k-nearest neighbor*.

[7] S. Tariq, S. Sarwar & W. Hussain, 2011, *Classification of Features into Strong and Weak Features for an Intelligent Online Signature Verification System*.

[8] H. H. Harralson, 2013, *Developments in Handwriting and signature identification in the digital age*.

[9] M.Diaz, A. Morales ,M. A. Ferrer, 2014, *Emerging issues for static handwritten signature biometrics*.

[10] Fenton et al, 2006, *Evaluation of Features and Normalization Techniques for Signature Verification Using Dynamic Timewarping*.

[11] Raul Sanchez-Reillo, (member, ieee), Judith Liu-Jimenez, and Ramon Blanco-Gonzalo, 2018, *Forensic Validation of Biometrics Using Dynamic Handwritten Signatures*.

[12] H.L. Teulings,1996, *Handwriting Movement Control*.

[13] R. Sanchez-Reillo, H.C. Quiros-Sandoval, I. Goicoechea-Telleria, W. Ponce-Hernandez, 2017, *Improving Presentation Attack Detection in Dynamic Handwritten Signature Biometrics*.

[14] N. Dounskaia · A.W.A. Van Gemmert · G.E. Stelmach, 2000, *Interjoint coordination during handwriting-like movements*.

[15] H-L. Teulings & L.R.B. Schomaker, 1993, *Invariant properties between stroke features in handwriting*.

[16] A.Mendaza-Ormaza, Oscar Miguel-Hurtado, Ivan Rubio-Polo, Raul Alonso-Moreno, 2009, *On-line Signature Biometrics using Support Vector Machine*.

[17] Jawad-ur-Rehman Chughtai, Dr. Shehzad Khalid, Dr. Imran Siddiqi, 2014, *Online Signature Verification: A Review*.

[18] František Hortai, 2017, *Possibilities Of Dynamic Biometrics For Authentication And The Circumstances For Using Dynamic Biometric Signature-*

[19] A. S. Syed Navaz1, K. Durairaj2, 2013, *Signature Authentication Using Biometric Methods*.

[20] Deepali H. Shah1, Dr. Tejas V. Shah2, 2015, *Signature Recognition and Verification: The Most Acceptable Biometrics for Security*.

[21] M. Soltane, N. Doghmane and N. Guersi, 2010, State of the Art: *Signature Biometrics Verification*.

[22] Y. Bay, M. Erbilek, A. Fosuah Gyasi and E. Celebi Cyprus, 2017, *The impact of blind and visual Signing on signature Biometrics*.

[23] G. P. Van Galen & A.W.A. Van Gemmert, 1996), *Kinematic and dynamic features of forging [fn2] another person's handwriting*.

Chapter 7

Special Session:
Lognormality Principle and its Applications in Graphonomics

Lognormality of Velocity Proles in Rapid Robotic Arm Movements

Moises Diaz [a,b], Jose J. Quintana[b], Miguel A. Ferrer[b], Cristina Carmona-Duarte[b], Adam Wolniakowski[c], Konstantsin Miatliu[c]

[a]Universidad del Atlantico Medio, Las Palmas de Gran Canaria, 35017, SPAIN

[b]Instituto Universitario para el Desarrollo Tecnologico y la Innovación en Comunicaciones. Universidad de Las Palmas de Gran Canaria, 35017, SPAIN

[c]Faculty of Mechanical Engineering, Bialystok University of Technology, POLAND
moises.diaz@atlanticomedio.es, {josejuan.quintana, miguelangel.ferrer}@ulpgc.es, ccarmona@idetic.eu, {a.wolniakowski, k.miatliuk}@pb.edu.pl

Abstract - Robotic arms commonly execute movements at a constant speed, but a sequence of short and rapid movements require the robotic arm to speed up and slow down continuously. This paper is aimed to assesses if these movements can be approximated by the Sigma-Lognormal model.

The parameters of the Sigma-Lognormal model have been calculated with two dierent methods: ScriptStudio and iDelog. In both cases, the SNR is greater than 15 dB assessing the lognormality of the rapid robotic movements. This is highlighted by the fact that the solutions reached by each method are statistically dierent.

1. INTRODUCTION

Robots are having exponential growth in recent years. Typically, they have been restricted to industrial environments, especially because they were costly. However, lately, due to their price reduction, they are starting to be used in new areas and applications.

The robots can be categorized according to their use as industrial or service robots. The rst category is used in factories like in the assembly lines of automobiles. The latter category is exponentially growing because of the wide variety of applications outside the industry such as the quadruped robot Spot and the biped Atlas of Boston Dynamics, the surgeon robots such as the Da Vinci of Intuitive Surgical, delivery systems at home by drones, and a long etcetera. Another area of robots is that of humanoid robots. There are several elds of research related to robotic partners that interact with children to increase their reading comprehension Yadollahi et al. (2018), help them to write Hood et al. (2015) or to better understand their disease Lewis et al. (2015) and so on.

Coming back to industrial robots, they are very precise devices both in the positioning and orientation of their links. Their trajectories are commonly described in either straight or circular lines between target points. These trajectories are carried out at a constant or variable speed depending on the robot task context, and they may have the option of linking consecutive targets points. Even though they are very precise to reach a point, it can be congured to make the robot does not pass exactly by the intermediate point. So, the robot would round that corners. Because the robot keeps the speed almost constant on each trajectory, its velocity patterns are very dierent from a human.

In order to improve the human-machine interaction, it is recommended that robots move to describe the most human possible patterns. In literature, there are several models that can approximate rapid human movement patterns. One of them is the kinematic theory of rapid

movements. This theory and its Sigma-Lognormal model decomposes a rapid movement into a vector sum of elementary movements that model the neuromuscular impulses of said movement. As an example, in Berio et al. (2016) authors congure a lognormal grati movement to a compliant Baxter robot, by taking advantage of the torque control capabilities of the robot with low gain feedback control. The visual feedback suggested that a proper style in the grati's trajectory executed by the robot was obtained.

In this work, we have programmed an industrial robot ABB IRB120 to execute fast and short consecutive movements. It is expected that the robot will speed up and speed down between target points owing to the speed required is high and the distance between control points short. So the velocity is expected to be bell-shaped. To evaluate the exact shape of the velocity, it is placed a pen capable of communicating with a Wacom Intuos Pro tablet at the end of the IRB 120 robot and the movement is carried out in the plane of the Wacom Intuos Pro. The movement has been registered and analytically analyzed with ScriptStudio and iDeLog, achieving encouraging performances in the lognormality of the velocity profiles.

This paper is organized as follows. Section 2 describes ScriptStudio and iDeLog methods to work out the parameters of the Sigma-Lognormal model. Section 3 describes the experiment setup and materials, experimental results are given in Section 4. The paper nishes with the conclusions.

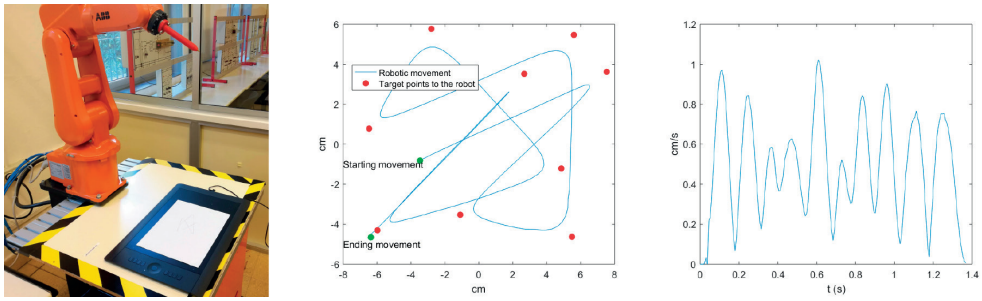


Fig. 1. Experimental setup (left), an example of a sequence of movements carried out by the robot (centre); velocity prole of the robotic movement (right).

2. Sigma-Lognormal model implementations for rapid movements: ScriptStudio and iDeLog

ScriptStudio framework O'Reilly and Plamondon (2009) has been widely used by the research community and dierent laboratories around the world. Its simplicity of use has allowed many advances in dierent elds like e-health, e-security or e-learning applications Plamondon et al. (2018). This framework recon- structs a velocity prole of a rapid movement by a vectorial sum of lognormal functions. The accuracy of the reconstruction is measured in terms of Signal-to-Noise Ratio between the observed and reconstructed velocity proles, i.e. SNR_v. The bell-shaped velocity prole is modelled with a sum of weighted lognor- mal functions displaced to each peak of the signal. It represents an elementary movement or stroke. In a second step, ScriptStudio tries to improve the SNR_v by adding small lognormals to the main ones.

A novel framework has been recently published, which also works out a movement by means of Sigma-Lognormal model. It is known as iDeLog, Ferrer et al. (2018), and models the velocity and the trajectory at the same time. In this case, once the velocity prole has been approached by a sum of weighted lognormals displaced to each peak, iDeLog iterates the six parameters of each lognormal as well as the position of the virtual target points in order to improve the velocity and trajectory signal-to-noise ratio, i.e. SNR_v and SNR_t, simultaneously.

Certainly, the signal acquisition hardware introduces noise. Script Studio overdraws this issue by smoothing the signal. As smoothing may erase some important characteristics of the movement in some applications, iDeLog is able to model the observed movement with and without smoothing.
It is worth pointing out that both ScriptStudio and iDeLog use circular traces to link the virtual target points. As such, a circular action plan is created. However, in Berio and Leymarie (2015) is suggested that such an action plan could be described by other patterns like Euler curves Berio and Leymarie (2015) for better performances.

3. MATERIALS AND EXPERIMENT SET UP

We have used the ABB IRB120 robot in our experiments. This robot is an anthropomorphic robotic arm widely used in both industry and academy applications. It has six degrees of freedom, which allows to position and orient a tool placed at the arm's end at our convenience. To develop and transfer the code to the robot, RobotStudio software was used.
In this work, the tool is a ballpoint pen capable of communicating with a Wacom Intuos Pro A4 tablet. It allows capturing about 120 points per second at a resolution of 5080 dots per inch. In order to attach the ballpoint pen to the end of the robot, a tool was designed and built using a 3D printer. A picture of the experimental setup is shown in Figure 1, left.
We designed 100 movements, which had between 5 and 13 target points, randomly selected on a 1712 cm surface and connecting them by means of straight lines. An example is shown in Figure 1 (centre and right).
On the other hand, we set up dierent speeds in the robot to reach the target points in a trial and error way. However, it was observed that the lower the speed, the less bell-shaped the velocity prole in the movements. This can be explained by the fact that robots of this type are commonly used in industrial applications where all individual movements are usually required at a constant speed. In order to avoid such constant patterns in the velocity proles of the sequence of movements, we dene the movements at the maximum robotic speed. To round the target points, the tolerance of passage through the intermediate points was set to two centimeters.
With this conguration, the 100 movements were plotted by the robot, with a duration of 1.17 s. in average terms. Obviously, the more target points, the longer it took.

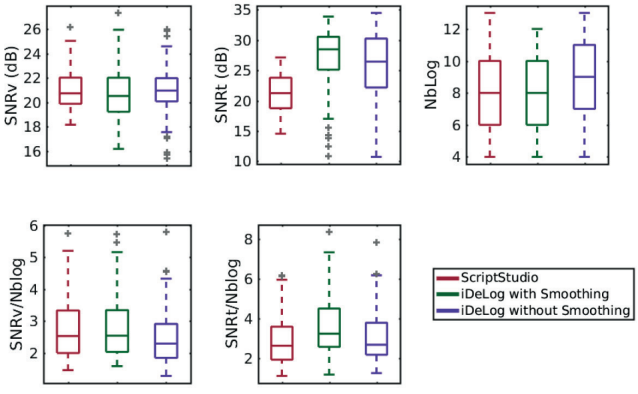


Fig. 2. Reconstruction performances with ScriptStudio, iDelog with and without smoothing the observed trajectories.

4. RESULTS AND DISCUSSION

We segmented the velocity proles with Script Studio and iDelog with and without smoothing the observed pen trajectories. It is worth pointing out that some connection hardware errors between tablet and computer caused some noise in the x-y trajectories. While ScriptStudio and iDeLog with smoothing corrected these mistakes by ltering out the observed trajectories, iDeLog without smoothing does not.

In a boxplot representation, Figure 2 shows the performance results obtained with the three methods. In the SNR_v we can observe that all methods report similar results in mean terms and over 20 dB. Examining the ScriptStudio boxplot whiskers in SNR_t, all of the reconstructions were above 18 dB.

Although slightly better performances were obtained with both iDeLog modalities all SNR_t were above 15 dB. In the case of the number of lognormals, *NbLog*, the glitches that ScriptStudio makes for improving the SNR_v increments *NbLog* in some movements, compared to iDeLog with smoothing. Also, higher *NbLog* in iDeLog without smoothing is justied since hardware communications introduce additional peaks in the bell-shape velocity prole. One way to observe the performance of the reconstruction is examining both SNR_v=*NbLog* and SNR_t=*NbLog*. The higher its value the better the reconstructions, because less lognormals were necessary to model the movement. While the minimum whiskers are the same across the three methods, the best mean is achieved with iDeLog with smoothing.

The three analysis demonstrate that the robotic movements satisfy the lognormality principles. This statement could be statistically signicative if the three Sigma-Lognormal model implementations were statistically dierent. Firstly, a Jarque-Bera test at 5% was made. The null hypothesis of this test assesses if a distribution is normally distributed by working out the upper tail probability of the chisquare distribution dened by the skewness and kurtosis of raw data. Apart from SNR_t in ScriptStudio and iDeLog without smoothing as well as the number of lognormals for the three methods ($p > 0{:}05$), all of them rejected the null hypothesis in all cases at 5% of signicance.

For this reason, two parameters and two non-parameter tests were used in the analysis. To assess the statistical dierence between the three methods, we used the Kruskal-Wallis non-parametric test and the one-way ANOVA, which evaluate whether all methods are statistically similar. To assess the statistical similarity between a pair of methods, the non-parametric Mann-Whitney U-test and the parametric t-test were used.

In order to evaluate how strong is the statistical signicance of the studied methods, the p-value is evaluated as follows:

- p - *value* < 0:01 (*)
- p - *value* < 0:05 (**)
- p - *value* > 0:05 ( )

Experimental results are given in Table 1 by highlighting in bold the cases where the null hypothesis is rejected. When we compare all three methods, we can observe dierences in all parameters which indicates that all three methods are statistically different, except for SNR_v. More similarities are found across ScriptStudio and iDeLog with smoothing for the SNR_v, *NbLog* and SNR_v/*NbLog*, because both methods identify the velocity bells in a similar way. However, dierences are seen in the remaining parameters, especially in SNR_t ($p = 2{:}02_e$ - 15). It justies that the reconstructed trajectory-based parameters are not statistically comparable in this problem. On the other hand, dissimilarities are observing between ScriptStudio and iDeLog without smoothing. In addition to the trajectory reconstruction metrics, the number of lognormals are not statistically similar between them.

Method	SNR_v	SNR_t	$NbLog$	$SNR_v/NbLog$	$SNR_t/NbLog$
SS- iDeLog^{Y}- iDeLog^{N}	0.2376	**2.66e-18(*)**	**0.0046(*)**	**0.01609(**)**	**0.03694 (**)**
SS-iDeLog^{Y}	0.2066	**2.02e-15(*)**	0.7403	0.9235	**0.01227 (**)**
SS- iDeLog^{N}	0.6159	**7.54e-12(*)**	**0.0086(*)**	**0.01024(**)**	0.21
iDeLog^{Y}- iDeLog^{N}	0.1166	0.08064	**0.0028(*)**	**0.01643(**)**	**0.001693(*)**

Table 1. p - values for parameters and non-parameters tests. In bold the p-values that reject the null hypothesis. SS, iDeLogY, iDeLogN refer to ScriptStudio, iDeLog with smoothing and iDelog without smoothing.

It could be explained by the fact that more peaks are detected with iDeLog without smoothing and, therefore, more lognormals are estimated, as it is seen in Figure 1. Instead, the trajectory ratio seems to bring a high similarity in both data distributions. Finally, when both iDeLog methods are compared, they accept the null hypothesis for the signal-to-noise-ratio metrics. As it was expected, they are not statistically similar for the number of lognormals, and, therefore, for the velocity and trajectory ratios.

Overall, we can observe that quite signicant similarities were observed in the SNR_v across the studied methods. Conversely, the study suggests that SNR_t has quite higher dierences when ScriptStudio is compared to iDeLog. It can be explained since the reconstructed trajectory is slightly dierent with both methods.

5. CONCLUSIONS

In this paper, we studied whether rapid movements made by a robotic arm could be modelled by a vec- torial sum of lognormals. For this aim, the Sigma-Lognormal model is used to approach the robotic arm trajectories. Specically, two methods are studied in the papers: ScriptStudio and iDeLog. Promising performances were obtained with both methods when the robots write on a digital tablet with rapid movements. Additionally, a statistical analysis is performed in order to highlight the statistical independence of the three methods, i.e. ScriptStudio, iDelog with and without smoothing, to obtain lognormal movements.

Our next goal is to compare the real target points of the movements with the target points estimated by Sigma-Lognormal methods. Additionally, we are also exploring human-like velocity patterns when robots move at lower speeds. Human-like movement in machines can improve their interaction with humans. In the meantime, our results encourage us to continue bridging the gap between robots and humans.

REFERENCES

[1] Berio, D., Calinon, S., and Leymarie, F. F. (2016). Learning dynamic graffiti strokes with a compliant robot. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 3981-3986.

[2] Berio, D. and Leymarie, F. F. (2015). Computational models for the analysis and synthesis of graffiti tag strokes. In *Proceedings of the workshop on Computational Aesthetics*, pages 35-47. Eurographics Association.

[3] Dynamics, B. Boston dynamics is changing your idea of what robots can do. www.bostondynamics.com. 2019 [Accessed: February 26, 2019].

[4] Ferrer, M. A., Diaz, M., Carmona, C. A., and Plamondon, R. (2018). iDeLog: Iterative dual spatial and kinematic extraction of sigma-lognormal parameters. *IEEE transactions on pattern analysis and machine intelligence*.

[5] Hood, D., Lemaignan, S., and Dillenbourg, P. (2015). When children teach a robot to write: An autonomous teachable humanoid which uses simulated handwriting. In *Proceedings of the 10th Annual ACM/IEEE Int. Conf. on Human-Robot Interaction*, pages 83-90. ACM.

[6] Intuitive. Intuitive, robotic assisted systems, da vinci robot. www.intuitive.com/en/products-and-services/da-vinci/systems. 2019 [Accessed: February 26, 2019].

[7] Lewis, M., Oleari, E., Pozzi, C., and Canamero, L. (2015). An embodied ai approach to individual dierences: supporting self-ecacy in diabetic children with an autonomous robot. In *Int. Conf. on Social Robotics*, pages 401-410. Springer.

[8] O'Reilly, C. and Plamondon, R. (2009). Development of a sigma-lognormal representation for on-line signatures. *Pattern Recognition*, 42(12):3324-3337.

[9] Plamondon, R., Pirlo, G., Anquetil, E., Remi, C., Teulings, H.-L., and Nakagawa, M. (2018). Personal digital bodyguards for e-security, e-learning and e-health: A prospective survey. *Pattern Recognition*, 81:633-659.

[10] Yadollahi, E., Johal, W., Paiva, A., and Dillenbourg, P. (2018). When deictic gestures in a robot can harm child-robot collaboration. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, pages 195-206. ACM.

Looking at the principle of motor equivalence through the "lognormal-glasses"

Antonio Parziale, Rosa Senatore, Angelo Marcelli

Abstract - According to the kinematic theory of human rapid movements, handwriting can be represented by a temporal sequence of strokes, each of which encodes the muscular commands to perform an elementary movement, i.e. a movement aimed at bringing one articular joint from its current position to the desired one. Motor theory of human movement, on the other hand, have hypothesized that skilled, aka fluent handwriting is achieved by executing a motor program that is stored in form of both a sequence of target points, representing the desired positions to be reached by executing each stroke, and the commands to properly recruiting and configuring the muscles for actuating the movements. We investigate to which extent the two representations can be mapped, and particularly how the target points encoded in the motor program can be located in the writing area. For achieving this aim, we extracted the motor program from multiple executions of the same handwritten pattern under the further constraint of minimizing the distances among the motor program's target points and the patterns' target points.
We compared the velocity profiles obtained from such a motor program representation with the actual movements to assess to which extent the execution of the extracted motor program can account for the variability among multiple executions, possibly providing an insight on the neural correlates behind the principle of motor equivalence.
Keywords: Motor Equivalence, Handwriting Representation, Motor Program extraction.

1. INTRODUCTION

Studies have shown that writing movements learned through the dominant hand could be repeated using different body parts, such as non-dominant hand, the mouth and foot, even if the subject had essentially no previous experience writing with any of this body parts [1]. The writing style is the same independently of the body part producing the movement even if there are differences among muscles recruited for executing the movement. This phenomenon is referred to as "motor equivalence" and it is of theoretical importance, because "*it suggests that actions are encoded in the central nervous system in terms that are more abstract than commands to specific muscles*" [2]. This abstract representation of the action is known as motor program, which has been defined as "*a central representation of a sequence of motor actions*" [3].

Motor learning and execution studies suggest that the implementation of a movement through the interaction between the central nervous system and the musculoskeletal system can be interpreted as the realization of a motor program stored in the brain [4]. The idea is that the practice of a certain movement over time allows to create a compact representation of a complex movement that, in the final stages of learning, is stored as a succession of elementary motor commands describing the motor program. Accordingly, after a movement has been learned, the variability observed in repeated executions may be ascribed to the neuromuscular system executing the movement. Further investigations into movement execution have shown that the actual movements result from the interaction between the central nervous

system, the spinal cord, the muscles and the proprioceptive receptors [5, 6]. In a nutshell, to initiate the movement, the central nervous system sends commands to recruit the muscles and to set the forces they have to exert on the bones they are connected to, while, during execution, the spinal cord modulates such command depending on the information received by the proprioceptive receptors in order to keep the execution as close as possible to the learned one. Therefore, those modulations introduce corrective movements that cannot be considered as the result of commands stored in the motor plan. The resulting sequence of strokes represents the *action plan* for that execution of the motor program.

The Kinematic Theory of Rapid Human Movements [7] suggests that complex human movements, such as handwriting, can be represented as the time superimposition of strokes, each of which results from a command generated by the central nervous system for reaching a target point and exhibits a lognormal velocity profile [8]. In the framework of this theory, a *stroke* is the ideal output of a neuromuscular system, an *action plan* is a sequence of strokes, while a *virtual target point* is defined as the end point of a stroke when it is executed in isolation. The effect of time superimposition of strokes is that virtual target points are never reached and inferring their position in the space, by analyzing the executed pattern, is not trivial. Different algorithms have been proposed for extracting the action plan from a handwritten pattern, i.e. to identify within a handwriting movement its strokes and estimate the values of the parameter describing each of them [9, 10, 11]. Because variations in the writing conditions and psychophysical state of a subject influence the execution of a complex movement, we can observe differences in the action plans extracted from different executions of the same trajectory. Differences can be observed in the number of extracted strokes, in the parameter values used for representing strokes and in the x-y position of virtual target points.

The aim of this paper is to extract from the *action plan* the *motor program* learned by the writer, which, by definition, is independent from the variability affecting different executions of the same pattern. Therefore, the desired representation should be stable with respect to the number of elementary movements that constitute the description of each handwritten sample and to the position of the virtual target points in the space.

The remaining of the paper is structured as follow: Section II describes the method proposed for inferring the motor program from multiple execution of the same pattern and explains how data were collected, Section III shows the application of the method to the pattern produced by a subject, eventually Section IV discusses further investigation of this preliminary work.

2. METHODS

A. Data collection

In a previous work [11], we asked 22 subjects, whose age ranges in the interval 18-30 years,

A. Parziale is with the Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno, 84084 Fisciano (SA), ITALY (phone: +39- 089-96-4177; e-mail: anparziale@unisa.it).
R. Senatore, is with the Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno, 84084 Fisciano (SA), ITALY (e-mail: rsenatore@unisa.it).
A. Marcelli is with the Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno, 84084 Fisciano (SA), ITALY (e-mail: amarcelli@unisa.it).

to reproduce 10 times the handwriting pattern "*llll*". Such a pattern, or very similar ones, have been used in many experiments on handwriting generation modelling because it is reasonable to believe that its motor program has been learned by the subjects involved in the experiments, independently of their individual characteristics, being at the same time fairly complex to allow a quantitative evaluation of the estimated motor program parameters.
The handwriting samples were collected by using an ink-and-paper WACOM Intuos 2 digitizing tablet with a 100 Hz sampling rate to record the handwriting movements. We adopted an ink-and-paper digitizing tablet as we aimed at extracting the strokes corresponding to the motor program of the subject, that most likely were learned under the same condition, so as to avoid as much as possible the influence of the spinal cord during the movement execution due to unexpected proprioceptive feedbacks that may arise in different writing condition, as it would have been the case by using a stylus-and-screen digitizing tablet.

B. Motor program Extraction and Representation

We adopted the algorithm *MPE* for segmenting handwriting movements into the sequence of strokes that most likely correspond to the actual commands issued by the central nervous system [8]. *MPE* incorporates heuristic criteria to filter out movements generated by spinal networks for keeping the trajectory under execution closest to intended one. The advantage of *MPE* is that it provides the same number of strokes for the large majority of, if not for all, the repetitions of the pattern drawn by subjects [11]. It is expected that, under normal conditions, the pattern "*llll*" is segmented in 8 strokes and *MPE* extracted, on average, 7.98 ± 0.92 strokes for each pattern. The Sigma-Lognormal model [9, 12] was adopted for describing each stroke in term of the six parameters $\{t_0,D,\mu,\sigma,\theta_s,\theta_e\}$.
Given K_j patterns drawn by a subject i and segmented in 8 strokes by *MPE*, we extracted virtual target points from each of them. Virtual target points were divided in 8 sets, with the *j-th* set made by the end points belonging to the *j-th* stoke of each pattern. The convex hull and centroid of each set of points were computed. We suggest that the centroid of a set represents the real target point stored in the motor program learned by the writer while the convex hull area is a graphical representation of the variation between the learned stroke and the one produced at each repetition of the task.
We were interested in identifying the sequence of strokes reaching the 8 centroids because we hypothesize this sequence is the *motor program* we are looking for and it is a generalization of the ones extracted from single handwritten patterns. Because of an infinitive number of strokes can be generated between two consecutive centroids, some criteria were introduced for reducing the search space and finding the sequence of strokes representing the motor program, which was described by 48 parameters (6 parameters for each stroke). In particular, in order to find the best values of the 48 parameters, Differential Evolution [13], an evolutionary algorithm, was run with the aim of minimizing:

- the sum of *RMSE*s computed between the execution of the motor program and each of the K_j - patterns;
- the sum of $(SNR_S)^{-1}$ computed between the execution of the motor program and each of the K_j - patterns;
- the sum of distances between the virtual target points of the motor program and the centroids.

Differential Evolution was run 8 times, so that during the *j-th* run only the 6 parameters of the *j-th* stroke were optimized. In particular, strokes were optimized according to the order of execution so that during the *j-th* run the parameters of strokes till the *(j-1)-th* were kept con-

stant to their best values while the parameters of the strokes following the *j-th* were not taken into account because they were not yet computed. Figure 1 shows the 8 convex hulls with their centroids and the motor program identified by the proposed algorithm.

3. RESULTS

The method has been applied to all the patterns collected in [8] while here the results obtained on samples drawn by subject S01 are shown. *MPE* extracted exactly 8 strokes for each of the patterns and the corresponding virtual target points are depicted in Figure 1. The convex hull with the smallest area is the one related to the first stroke (in magenta), while the one with the greatest area is the one related to the second stroke (in brown). The inferred motor program is represented in black in Figure 1 and its target points are, as desired, the centroids of the convex hulls. The trajectories and the velocity profiles of the 9 patterns drawn by S01 are shown in blue in Figure 2 and 3, respectively. The trajectory and the velocity profile corresponding to the execution of the inferred motor program are depicted in red in Figure 2 and 3, respectively. It is interesting to note as the motor program show the same skew of the patterns while there are differences between the motor program and patterns, especially with those executed too fast or too slow respect the others.

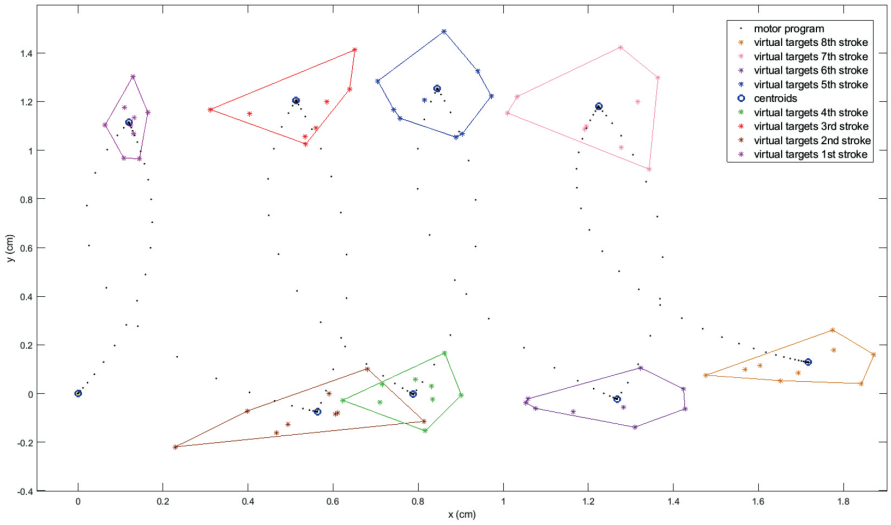


Figure 1. Virtual target points extracted from 9 samples written by subject S01 and plotted in the same x-y plane. For each stroke, the convex hull containing the virtual targets is depicted. The motor program identified by applying the method described in this paper is in back.

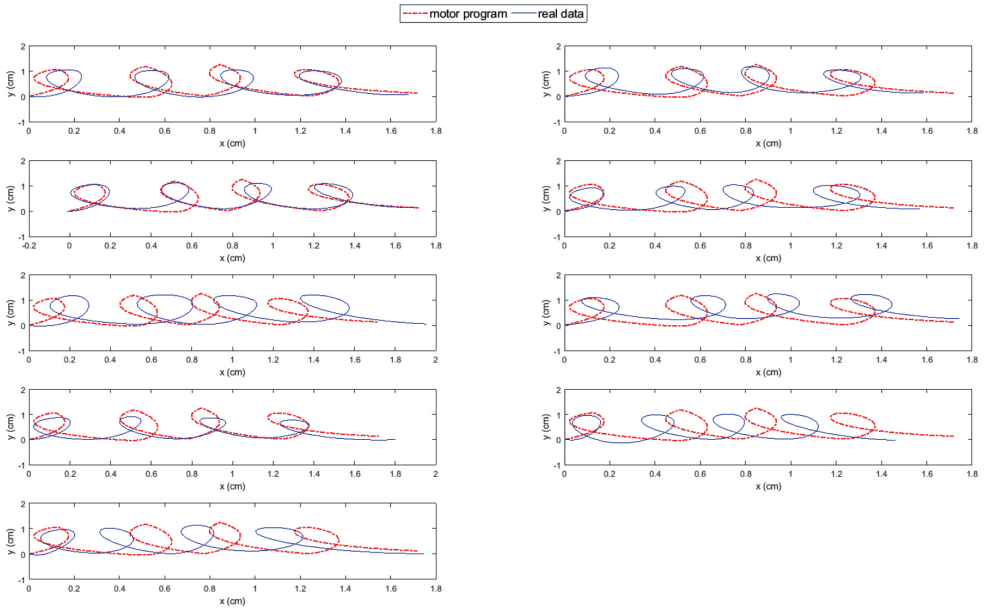


Figure 2: The trajectories written by subject S01 are depicted in blue while in red is represented the trajectory obtained by executing the motor program shown in Figure 1. For S01 only 9 samples are available instead of 10, due to a problem in the collection of the last sample.

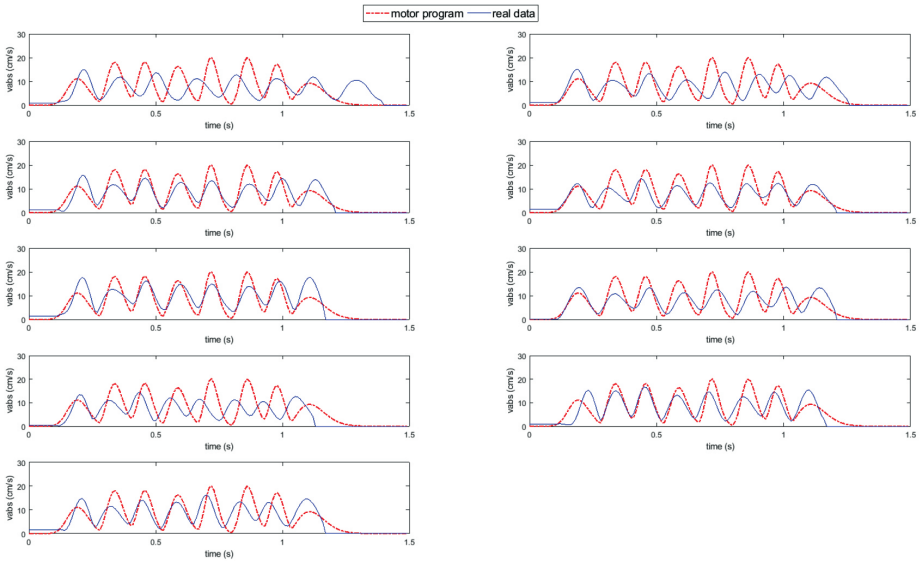


Figure 3: The velocity profiles of the 9 patterns written by subject S01 and of the execution of the motor program are depicted in blue and red, respectively.

4. DISCUSSION

The paper presents the preliminary results of a method for inferring the motor program from the action plans adopted for its execution. Figure 2 and 3 help to get an idea of the variability introduced by the writer in repeated execution of a learned motor program, even if it is a "simple" pattern as "*llll*". *MPE* guarantees stability in the number of strokes extracted from the repetitions.

Nevertheless, Figure 1 shows variability in the x-y position of the virtual target points of each stroke. It is interesting to note that the convex hull corresponding to the 2nd stroke has an area greater than all the others and it can be in part explained by observing the patterns executed by the subject. In fact, the second movement is the one that shows the greatest variability among the patterns. Figure 3 shows that patterns' velocity profiles and the velocity profile of the motor program execution are globally similar but there are differences in their temporal executions. These results are in line with the idea that a motor program represents the invariant features of a class of movements while specific movement parameters are varied during each execution to meet specific demands [3]. In further studies, we will investigate on other criteria for inferring the motor programs and on methods to quantify the amount of similarity between a motor program and an execution paying particular attention to the variability of stroke parameters.

5. ACKNOWLEDGMENTS

The work reported in this paper was partially funded by the "Bando PRIN 2015 - Progetto HAND" under Grant H96J16000820001 from the Italian Ministero dell'Istruzione, dell'Università e della Ricerca.

REFERENCES

[1] M. H. Raibert, *Motor control and learning by the state space model*. Cambridge: Artificial Intelligence Laboratory, MIT, 1977.

[2] Alan M. Wing, *Motor control: Mechanisms of motor equivalence in handwriting*, Current Biology, Volume 10, Issue 6, 2000, pp. 245-248,

[3] J. J. Summers, J. G. Anson. *Current status of the motor program: Revisited*. Human Movement Science, Volume 28, Issue 5, 2009, pp. 566-577.

[4] A. Marcelli, R. Senatore, *A Neural Scheme for Procedural Motor Learning of Handwriting*, 2012, Proc. ICFHR 2012, New York IEEE, pp.659-664

[5] A. Parziale, *A neurocomputational model of reaching movements*, PhD Thesis, Salerno, Italy: University of Salerno, 2016

[6] A. Parziale, J. Festa, and A. Marcelli. *A neurocomputational model of spinal circuitry for controlling the execution of arm voluntary movements*, 17th Biennial Conference of the International Graphonomics Society. 2015.

[7] R. Plamondon, *A kinematic theory of rapid human movements*. Part I., Biological Cybernetics, 1995, vol. 72, pp. 295-307.

[8] R. Plamondon, W. Guerfali, *The generation of handwiritng with delta-lognormal synergies*, Biological Cybernetics, 1998, vol. 78, pp. 119-132

[9] C. O'Reilly, R. Plamondon, *Development of a Sigma–Lognormal representation for on-line signatures*, In Pattern Recognition, Volume 42, Issue 12, 2009, pp. 3324- 3337.

[10] M. A. A. Ferrer, M. Diaz, C. A. Carmona and R. Plamondon, *iDeLog: Iterative Dual Spatial and Kinematic Extraction of Sigma-Lognormal Parameters*, in IEEE Transactions on Pattern Analysis and Machine Intelligence.

[11] R. Parisi, A. Parziale, A. Marcelli. *Some observations on lognormality and motor control in handwriting*, Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence Montreal CENPARMI, Concordia University, 2018, pag.732-737

[12] R. Plamondon, M. Djioua, *A multi-level representation paradigm for handwriting stroke generation*, Hum. Mo. S., 2006, vol. 25, issue 4, pp. 586-607.

[13] R. Storn, K. Price, *Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces*, Journal of Global Optimization (1997), vol. 11, pp. 341-359

Graphomotor Evolution in the Handwriting of Bengali Children Through Sigma-Lognormal Based-Parameters: A Preliminary Study

Moises Diaz [a,b], Miguela. Ferrer[b], Richard Guest[c], Umapada Pal[d]

[a]Universidad del Atlantico Medio, Las Palmas de Gran Canaria, 35017, SPAIN

[b]Instituto Universitario para el Desarrollo Tecnologico y la Innovacion en Comunicaciones. Universidad de Las Palmas de Gran Canaria, 35017, SPAIN

[c]School of Engineering and Digital Arts, University of Kent, Canterbury, CT2 7NT, UK.

[d]Computer Vision and Patter Recognition Unit, Indian Statistical Institute, Kolkata 700108, INDIA
moises.diaz@atlanticomedio.es, miguelangel.ferrer@ulpgc.es,
R.M.Guest@kent.ac.uk, umapada@isical.ac.in

Abstract - Handwriting has a natural evolution throughout child development. It is expected that the higher their educational level, the more fluent their handwriting. This paper is the first study in graphomotor evolution in Bengali children. This study is focused on the handwriting of the first five letters learnt in Bengali (Bangla) schools in Kolkata, India. We selected between three and six participants within three dierent classrooms and applied the Sigma-Lognormal model to their handwriting by using the iDeLog method. We present conventional and novel parameters extracted from the lognormality analysis of their handwriting. An assessment of extracted parameters is carried out with meaningful statistical dierences observed within their handwriting evolution. Our results adds to the growing list of evidence supporting the lognormality principle of graphomotor evolution in children.

1. INTRODUCTION

Among the dierent modelling-techniques associated with rapid movement, handwriting generation in children has already been analytically represented using the Kinematic Theory of rapid human movements and its associated Sigma-Lognormal model. In Duval et al. (2015) three groups of kindergarten children between 3 and 6 years old executed simple and oblique traces, with 10 lognormals approximated in average terms. Such trajectories were reconstructed with a sum of lognormals through the Sigma-Lognormal model. The study showed that lognormality patterns in velocity proles were related to the age of the children with signal-to-noise-ratio in velocity SNR_v, number of lognormals *NbLog* and the ratio between them SNR_v=*NbLog* statistically assessed. Similar parameters were discussing in Laniel et al. (2019) to determine the ne motor control in healthy and attention decit/hyperactivity disorder children. Other similar patterns and characters were executed by children in Plamondon et al. (2013). The authors discussed a combination of classical and Sigma-lognormal features to assess the handwriting evolution of three groups of children aged between 3 and 6 years old. Apart from observing successful writing reconstruction performances related to the rapidity, uidity and regularity in childrens handwriting, the authors statistically determined which parameter can better identify children in a classroom. Additionally, iterative tools and applications for learning to write are discussed in Plamondon et al. (2018) on the basis of Sigma-Lognormal model.

In this paper, our contribution is twofold. Firstly, we present novel and classical Sigma-Lognormal parameters to study handwriting fluidity. Some features can be extracted directly

from the Sigma-Lognormal decomposition, however, others are related to geometrical relationships with trajectory, virtual target points, and the relationship between lognormals that represents elementary neuromuscular movements.
For these purposes, iDeLog (Ferrer et al. (2018)) was deemed to be a suitable option for extracting such parameters. Secondly, a preliminary study of these parameters within Bengali handwriting is provided.
Children from dierent classrooms enrolled in a school based in Kolkata, India wrote the rst learnt letters in this script. Results showing the childrens' graphomotor evolution with Sigma-Lognormal parameters are obtained and discussed from statistical analysis.
The paper is organized as follows: Section 2 makes a short description of the iDeLog method for extracting the Sigma-Lognormal parameters; Section 3 introduces the experiments and the study whereas Section 4 is devoted to the data analysis results. Final remarks are given in Section 5.

2. iDeLog method for Sigma-Lognormal model: a short review

A novel framework to implement the Kinematic Theory of rapid movement and its Sigma-Lognormal model is iDeLog (Ferrer et al. (2018)), which has been developed in MATLAB. This method decomposes a spatiotemporal trajectory as a sum of temporally overlapped circumferences. It also divides the trajectory into strokes, where j is the index of the stroke. Then, iDeLog estimates the virtual target points tp_{j1} and tp_j , the starting and ending angle of the circumference θ_{sj} and θ_{ej} , and the lognormal parameters $t0_j$,μ_j , σ_j and D_j for each stroke, j. After this rst solution, the algorithm iterates moving the target point to improve the representation of the spatiotemporal trajectory by the Sigma-Lognormal transform.

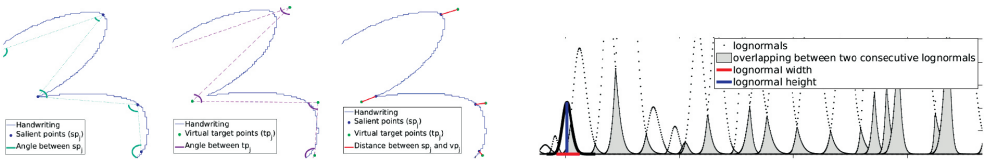


Fig. 1. The novel parameters in both handwriting velocity and trajectory. Left: parameters p_4; p_5; p_6. Right: parameters p_7; p_8; p_9.

The values μ_j and σ_j are obtained by means of minimizing the velocity prole and the estimated lognormals by means of a Levenberg-Marquardt algorithm (LMA). The parameter $t0_j$ is obtained as $t0_j = t_{min,j} - 0{:}5$ where $t_{min,j}$ is the velocity minima in which the stroke starts. Moreover, the iDeLog method estimates the virtual targets points tp_j from the salient point sp_j. The closer is the angle of the vertex spj , calculated with sp_{j-1} and sp_{j+1}, the further is tp_j from sp_j. iDeLog accomplishes a renement algorithm to optimize the position of the virtual target points tp_j to improve both the reconstructed trajectory and velocity prole. The improvement is carried out by means of an iterative Least Mean Square (LMS) algorithm applied stroke by stroke in the same order than the original movement.

2.1 Sigma-Lognormal parameters for handwriting evolution

The handwriting decomposition with Sigma-Lognormal model is an appropriate technique to study the evolution of handwriting in children (Duval et al. (2015); Plamondon et al. (2013)). Beyond the six parameters to dene each of lognormal function in the handwriting, it is possible to calculate extended parameters to further explore the physical relationships with handwriting evolution. In this paper we have studied the following additional nine parameters:

- p_1 is the variance distribution of the lognormal functions, σ_j.
- p_2 is the mean distribution of the lognormal functions, μ_j.
- p_3 denotes the amplitude of the elementary movement,D_j , which is a geometrical relationship from circular trajectories that dene the writing action plan.
- p_4 refers to the distance between salient points and its corresponding virtual target point. The larger the distance, the more fluent the handwriting
- p_5 is the angle between salient points (Ferrer et al. (2017)). The more acute the angle, the more fluent the handwriting.
- p_6 contains the angle between virtual target points with an interpretation as for p_5.
- p_7 stores the overlapping area between two consecutive lognormals. The greater the area the more fluent the handwriting as the writer does not pause in every stroke traced.
- p_8 refers to the height or maximum of the peaks of the bell-shaped lognormal functions. The higher their values, the more speed used for tracing the strokes.
- p_9 contains the widths of the individual lognormal functions. The larger their values, the more time was spent for tracing a stroke.

The novel parameters dened above are illustrated in Figure 1.

3. STUDY AND EXPERIMENTS

3.1 Participants and data collection

Fifteen Bengali children from Kolkata participated in this study. Six children were enrolled in the kindergarten 2 (KG2) classroom (aged between 4 and 5), six from Grade 2 (G2) classroom (ages 6 to 7) and, finally three from Grade 5 (G5) classroom (ages 9 to 10). The task was to write twice the first five letters that native Bengali children learn. As they were the rst letters learnt, it is expected that their production was automated and intrinsic. Participants wrote in a worksheet, which included the first five letters printed on the top of the sheet.

To register their on-line handwriting, the worksheet was placed over a Wacom tablet, so the participants could write with an inked pen and receive visual feedback. The Wacom tablet was congured at 5080 dpi and a 135 Hz sampling rate. We analysed the letters from the second writing attempt since children may be more condent with the task and familiar with the work environment.

From the obtained writing sequences, we concatenated the pen-down data of their handwriting, applied a cubic splines resampling at 200 Hz and smoothed by a Chebyshev lter. The resulting concatenated handwriting string for each participant was analysed with iDeLog. It is worth noting that iDeLog is especially convenient for modelling trajectories and velocities of long and complex handwriting (Ferrer et al. (2018)). In total, nine long pieces of handwriting were analysed.

3.2 Statistical study

The aim of this study is to determine whether the proposed Sigma-Lognormal parameters can estimate the natural graphomotor evolution in Bengali children. First of all, a test of normality was applied to the data to conrm whether a parametric or non-parametric hypothesis test should be used. We use the Jarque-Bera test, which calculates the skewness and kurtosis of a distribution. It evaluates the null hypothesis which arms whether a parameter is normally distributed. In our data, the null hypothesis was rejected in all cases ($p < 0{:}01$), so we conclude that statistical parametric test cannot be used to study our data. Therefore, we use the non-parametric Mann-Whitney U-test to study the equal distribution mean of two samples

from a continuous distribution. Moreover, we have used the Kruskal-Wallis non-parametric test to study whether three or more conditions comes from the same distribution.
Thus, in this study we evaluate the following three research questions:

- Q1: *What parameter is more trustworthy to describe the graphomotor evolution of a particular classroom?*
- Q2: *What parameter is more relevant for describing the graphomotor evolution of handwriting in Bengali kindergarten?*
- Q3: *What two classrooms present signicantly more similar graphomotor evolution?*

The first question can help to understand the more stable parameters for characterize the handwriting in a classroom. Second question is towards study the parameters that reports better performances in line with the progress in the school. Last question nds the parameters that can dierentiate more the graphomotor evolution in two classrooms.

4. STATISTICAL DATA ANALYSIS

This analysis is focused on studying the capacity of using the Sigma-Lognormal model in the analysis of handwriting by children as well as a graphomotor evolution analysis with novel related parameters.

4.1 Reconstruction analysis

The rst step was to analyse the quality of the reconstruction. As such, we calculate the mean and standard deviation of the nine Signal-to-noise-ratio values for both velocity and trajectory in dB. We obtained $SNR_v = 17{:}62_{1:84}$ and $SNR_t = 38{:}63_{3:71}$. These values suggest that the reconstructed signals can be suitable for the analysis, since they are above 15dB on average (Djioua and Plamondon (2009)).
Examining the number of lognormals averages, we have $287{:}08_{71:74}$ for classroom KG2, $327{:}25_{38:01}$ for classroom G2 and $133{:}50_{38:34}$ for classroom G5. At the same time, we can see the average in the ratio $SNR_v/NbLog$ as $0{:}07_{0:02}$, $0{:}06_{0:01}$ and $0{:}13_{0:04}$, for classrooms KG2, G2 and G5 respectively. Similarly, in the case of trajectory, we observe $0{:}14_{0:04}$, $0{:}13_{0:02}$ and $0{:}27_{0:05}$. The higher this value, the fewer lognormals required to trace the same letters, as is the case within classroom G5. Although there are only a few participants within each classroom, these results suggest that the oldest classroom group observed in this study need nearly half the elementary neuromuscular movements for approaching the same task than classrooms KG2 and G2. This observation is repeated in the average handwriting duration per classroom where we calculated $40{:}37_{8:73}$ s, $43{:}13_{4:60}$ s and $22{:}40_{5:79}$ s in classrooms KG2, G2 and G5 respectively.
It was observed that there is a disparity of graphomotor evolution within group KG2 according to the standard deviation of number of lognormals and duration.
Furthermore, we have analysed the shimmer and jitter of each concatenated piece of handwriting in each classroom. These features assess the cycle-to-cycle variations between amplitudes and mode of consecutive lognormals respectively (Farrus et al. (2007)). For KG2, G2 and G5 classrooms, the jitter was $0{:}15_{0:01}$, $0{:}14_{0:01}$ and $0{:}17_{0:01}$ respectively. This indicates that children in classroom G5 produced longer and fewer lognormals for the same tasks, as to be expected. The shimmer values were $7{:}39_{0:69}$, $7{:}57_{0:68}$ and $7{:}11_{1:49}$ respectively, which suggest slightly higher amplitude dierences in the youngest children.

4.2 Sigma-Lognormal-based parameter analysis

Our rst question (Q1) seeks to validate the robustness of handwriting parameters within

classrooms. As we have between three and six participants per classroom, the Kruskal-Wallis test was used, where the null hypothesis is that all children have the same distribution at 1% of signicance level. The results in Table 1 shows that all of the studied parameters reject the null hypothesis for the youngest children. However, for children within classroom G2, the angle between target points (p_6) indicates that the population are similar within a particular classroom. The oldest children have three out of nine parameters that indicate statistical uniformity with their handwriting. Overall, it suggests that the oldest the children the more statistically similarities using the proposed Sigma-Lognormal parameters of their handwritings, this indicating uniformity between the handwriting of children of a similar age.

Grouping	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9
Within Classroom KG2	1.73e-12	1.78e-06	9.67e-07	9.74e-25	1.35e-10	4.83e-07	3.06e-15	2.25e-06	3.18e-12
Within Classroom G2	1.96e-36	4.58e-09	4.81e-19	3.51e-51	0.00813	**0.0628**	9.68e-61	7.38e-13	4.98e-34
Within Classroom G5	**0.0727**	**0.271**	5.07e-07	2.59e-08	0.00040	0.00307	5.34e-09	8.99e-09	**0.0908**

Table 1. Non-parametric Kruskal-Wallis test *p - values* at 0:01 of signicance for Q1 for the studied parameters

Grouping	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9
(KG2, G2, G5)	2.02e-40	4.31e-26	1.33e-38	6.97e-77	2.96e-92	9.92e-62	2.36e-45	7.44e-29	1.25e-40
(KG2, G5)	1.46e-14	1.6e-12	3.69e-23	9.67e-39	4.33e-76	2.44e-56	7.54e-20	7.41e-21	5.21e-15
(KG2,G2)	1.76e-13	1.56e-06	1.04e-05	1.76e-13	**0.0646**	**0.774**	1.27e-11	**0.0203**	3.19e-13
(G2,G5)	8.34e-39	2.45e-26	3.74e-38	4.19e-76	3.01e-81	2.84e-51	1.75e-44	1.99e-28	4.46e-39

Table 2. Non-parametric Kruskal-Wallis and Mann-Whitney tests *p - values* at 0:01 of signicance for Q2 and Q3

The second question (Q2) is again assessed with the Kruskal-Wallis test enabling an assessment across the three classrooms. As shown in Table 2, there were signicant dierences between classrooms for all parameters, especially for p5, i.e. the distance between salient points and virtual target points ($p = 2{:}96_e$-92). The order of the parameters about statistical signicance was found in: p_5, p_4,p_6, p_7, p_1, p_9, p_3, p_8, p_2.
Note that a low rank number indicates higher signicant dierences.
The third question (Q3) evaluates whether two groups come from the same distribution. In this way, we have used the Mann-Whitney U-test. We observed in Table 2 that children in classroom KG2 and G2 produce similar outputs for the angle between salient points (p_5), the angle between virtual target points (p_6) and the height of the bell-shaped lognormal peaks (p_8). The major dierences between classrooms are observed in classrooms G2 and G5, since their *p-values* have the lowest results. It suggests a higher improvement in handwriting development which can be associated to a better graphomotor abilities after passing G2.

5. CONCLUSIONS

A series of novel, alongside conventional, Sigma-Lognormal parameters are presented in this paper to assess the graphomotor evolution of children in a Bengali school. We present these parameters and their calculation using the iDeLog method (Ferrer et al. (2018)).We have validated the use of Sigma-Lognormal across the handwriting of three groups of children within a Bengali school with fteen participants. Moreover, statistical data analysis is performed to evaluate the graphomotor evolution of children with these parameters. Our future ideas include extending the number of participants per classroom as well assessing additional Sigma-Lognormal parameters related to the graphomotor evolution. New insights could be also obtained studying letter individually. It is also interesting to analyse pen-ups within the writing sequence since children usually loose attention to the task, especially within those in classroom KG2. Even though our experiments show a range of interesting ndings across a number of parameters such as the angle between salient points or target points, more experiments

would be necessary to conrm robustness across a wider population. Meanwhile, our results adds to the growing list of evidence supporting the lognormality principle of graphomotor evolution in children.

REFERENCES

[1] Djioua, M. and Plamondon, R. (2009). A new algorithm and system for the characterization of handwriting strokes with delta-lognormal parameters. *IEEE Trans. on Pattern Ana. and Machine Int.*, 31(11):2060-2072.

[2] Duval, T. et al. (2015). Combining sigma-lognormal modeling and classical features for analyzing graphomotor performances in kindergarten children. *Human movement science*, 43:183-200.

[3] Farrús, M., Hernando, J., and Ejarque, P. (2007). Jitter and shimmer measurements for speaker recognition. In *Eighth annual conference of the international speech communication association*.

[4] Ferrer, M. A., Diaz, M., Carmona, C. A., and Plamondon, R. (2018). iDeLog: Iterative dual spatial and kinematic extraction of sigma-lognormal parameters. *IEEE Trans. on Pattern Ana. and Machine Int*.

[5] Ferrer, M. A. et al. (2017). Two-steps perceptual important points estimator in 8-connected curves from handwritten signature. In *7th Int. Conf. on Image Processing Theory, Tools and Appl.*, pages 1-5.

[6] Laniel, P., Faci, N., Plamondon, R., Beauchamp, M. H., and Gauthier, B. (2019). Kinematic analysis of fast pen strokes in children with adhd. *Applied Neuropsychology: Child*, pages 1-16.

[7] Plamondon, R., O'Reilly, C., Remi, C., and Duval, T. (2013). The lognormal handwriter: learning, performing, and declining. *Frontiers in psychology*, 4:945.

[8] Plamondon, R., Pirlo, G., Anquetil, E., et al. (2018). Personal digital bodyguards for e-security, e-learning and e-health: A prospective survey. *Pattern Recognition*, 81:633-659.

Pre screening for Central or Peripheral Shoulder Fatigue using the Sigma lognormal Model

A. Laurent, R. Plamondon, M. Begon

Abstract - Fatigue assessment is important to prevent injuries or overtraining in sport for example. Current methods for detecting fatigue cannot discriminate peripheral from central fatigue, often require expensive equipment (*e.g.*, EMG) or are subjective. As part of the different troubles, musculoskeletal disorders at the rotator cuff are among the most difficult to diagnose. The Kinematic Theory of Rapid Human Movement is the most accurate model describing end effector human movements, and it has been exploited to study neuromuscular disorders in children and aged subjects. Data collection requires only a digitizing tablet, but so far it has never been employed to evaluate upper-limb fatigue. This paper deals with assessing shoulder fatigue detection on healthy subjects with the Kinematic Theory, after a series of exertions soliciting the rotator cuff muscles. The results obtained show that it is possible to detect changes in the parameters of the Kinematic Theory and to discriminate a peripheral from a central fatigue.
Keywords: Fine motor control, Sigma-Lognormal model, Kinematic Theory, central and peripheral fatigue, rotator cuff

1. INTRODUCTION

Musculoskeletal disorders (MSDs) are among the most common chronic diseases. The ones affecting shoulders are the third most prevalent, after those affecting the lumbar and cervical regions. Rotator cuff muscle pathologies account for up to 85% of all shoulder injuries treated by clinicians [1]. MSDs are omnipresent in sport, particularly in those requiring overhead movements, with repetition being an important risk factor [2-3]. Neuromuscular fatigue is the precursor to these MSDs and its detection can help reduce the risk of future injuries [4-5]. Fatigue can be decomposed into a central part, which refers to the neural system, and a peripheral part, which involves the muscles. A submaximal effort can lead to central and peripheral modifications [2-6].

The Kinematic Theory of Rapid Human Movement is a model that accurately describes the fine motor control of someone in perfect condition. In case of neuromuscular disorders, motor control is affected and the parameters extracted from this Theory are modified [7-8]. Performing tests to assess someone's neuromotor system through this Theory is now easily applicable using an experimental set-up, consisting of a Wacom Cintiq 13HD tablet [9]. This system effectively captures the coordinates of the movement as a function of time from a stylus tip. Then, the velocity profile is reconstructed, and the Kinematic parameters associated with this gesture are extracted.

In this study, we emphasize our first analysis on fatigue to see if, by drawing fast strokes, it is possible to detect rotator cuff muscle fatigue following an effort, in healthy subjects. This study is preliminary to the analysis of shoulder injuries through the Kinematic Theory. We hypothesized that the transmission of the nerve signal would be less effective after fatigue and that lognormal parameters would reflect this problem.

2. METHODS

A. Participants

Twenty healthy active participants (11 males and 9 females, 2 left-handed and 18 right-handed, age: 23.2±3.2 years; height: 173.9±8.3cm), with no history of shoulder surgery nor upper-limb pain or neuro-musculoskeletal disorder, performed a set of fast strokes on a tablet, followed by a fatigue protocol and then another set of fast strokes. All the tests were completed with the dominant arm. The protocol was approved by the research ethics committee (CER-1819-23 v.3).

B. Experimental part

Each participant had to draw fast strokes on a Wacom Cintiq 13HD tablet with a stylus [9], from a starting point to a specific finish area (Fig. 1a). The bottom of the tablet was placed at a participant's shoulder height (Fig. 1b). A guide sheet was integrated into the tablet and was reversed for left-handed participants. A beep (at 1 kHz for 500 ms) was emitted randomly (from 1 to 10 s) and unpredictably before each stroke, acting as a go signal. Participants were instructed to place the stylus on the starting zone, wait for the beep and then draw the fast strokes one at a time till the finish area. At the end, they had to wait with the stylus still on the tablet for minimum 1 s. The in-house program Sign@medic was used to synchronize the tablet and simulator and was used as interface for the guide sheet. A training period (7 to 10 strokes) to draw strokes on the tablet was given in order to make sure that the participant correctly understand and follow the sequencing of instructions. Afterward 30 valid fast strokes were recorded on the tablet.

For the fatigue protocol, the participant was placed on an isokinetic dynamometer (CONTREX® MJ; Schnaittach, Germany), a machine for assessing the joint function in dynamic at a constant angular velocity. Especially she/he was in a sitting position fastened with a seat belt with an arm elevation of 30° in the scapular plane and the elbow bent at 90°; the upper-arm was aligned with the moving axis of the dynamometer (Fig. 1c) [10]. The range of motion in external/internal rotation was established at 70° and the mechanical arm was set at a constant angular velocity of 90°/s. A period of familiarization with the dynamometer was done before the session so that the participant could get used to a shoulder isokinetic exercise. Then the participant was instructed to perform two maximum voluntary isokinetic contractions in external rotation (ER). The dynamometer moved back the arm passively in internal rotation (continuous passive mode). ER exertion targets the infraspinatus, one of the rotator cuff muscles. The maximum torque of the two contractions was reported and used in a Matlab visual biofeedback that presents the instantaneous relative torque value and a target area of 50±7.5%. This feedback was displayed in front of the subject on a large screen. They were instructed to reach the target area at each ER repetition and to not force during the IR. They could use the small grip on the side of the dynamometer for more comfort and were instructed to not use their back to facilitate the movement. Every minute, they were asked to rate their perceived exertion using Borg CR10 Scale from 0 (no effort) to 10 (the most difficult effort ever made). As in [11], the fatigue protocol was stopped when one of the three conditions was reached:

A. L. is with the Département de génie biomédical, Polytechnique Montréal, Montréal, Canada(e-mail: anais.laurent@polymtl.ca).
R. P. is with the Département de génie électrique, Polytechnique Montréal, Montréal, Canada
(e-mail: rejean.plamondon@polymtl.ca).
M. B. is with the Faculté de médecine, Université de Montréal, Montréal, Canada (e-mail: mickael.begon@umontreal.ca).

(i) the participant could not reach the target area in three consecutive repetitions; (ii) during eight consecutive minutes his/her Borg number reached 8; or (iii) after 30 minutes. Verbal encouragement was given when the person had difficulty in reaching the area. The stopping conditions were not mentioned to the participants. After fatigue, the participant had to draw 30 valid fast strokes as described above.

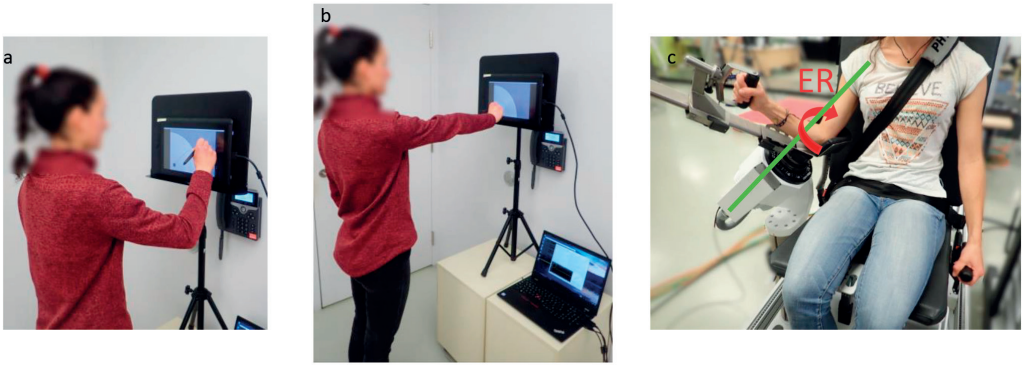


Figure 1. a) Experimental set-up with the tablet. b) Positioning of the tablet. c) Positioning on the isokinetic dynamometer.

C. Data extraction

Kinematic parameters were extracted for each participant's stroke. The reconstruction of the speed profile was done using Script Studio according to the Sigma-Lognormal model [12]. Each stroke was split into its agonist and antagonist components. The agonist component was seen as the largest Lognormal allowing to go in the same direction of the stroke while the antagonist component was seen as the largest Lognormal going in the opposite direction of the stroke. Strokes having only one Lognormal were automatically classified as agonist. Parameters of the Kinematic Theory were extracted from each Lognormal for further analysis [13-14]. They can be separated into three categories. Two of them describe the global state of the neuromotor system of each participant:

- Signal-to-Noise Ratio [SNR in dB]: Measure of the quality of the participant's motor control. In other words, an evaluation of its lognormality. An augmentation of the SNR means that the reconstruction is better and that the velocity profile is nearer to the ideal behavior. The SNR should be at least 15 dB.
- Number of Lognormal used for reconstruction [NbLog]: An augmentation of the NbLog means that the movement is less fluid.

The state of the central motor command of a participant is assessed for each Lornormal with:

- Time when the brain sends a motor command [t_0 in s]: Longer t_0 means that the brain takes more time to send its motor command.
- Amplitude of the input command [D in mm]: Larger D means that the participant makes longer strokes or movements.
- Starting and ending angles of the stroke [θs and θe in degrees]: They are also intrinsic parameters of the action plan. Larger θs and θe means that the participant changes its action plan to compensate for example some articulation or postural troubles.

The peripheral system is assessed for each lognormal with:

- Logtime delay [μ in log(s)]: Longer μ means that the general speed of the neuromuscular

system to respond to a command is slower.
- Logresponse time [σ in log(s)]: Larger σ means that the duration of the movement is longer.

Derived parameters of the Theory (Mode, Median, Delay, Response time, Asymmetry, amplitude of the movement, SNR/NbLog) were also extracted to describe the global system of the participant (refer to [8] for further details).
Another classical parameter, not from the Theory, was also measured:

- Reaction time [RT in s]: Time at which the movement starts [14]. An augmentation of the reaction time means that the time before starting the movement is longer.

D. Statistical analysis

A Wilcoxon signed-rank test (i.e. non-parametric paired t-test) was applied to each subject to assess the changes of parameters due to fatigue. All analyses were conducted using Matlab. The level of statistical significance was set at $p<0.0055$ (0.05/9) to account for multiple comparisons using a Bonferroni correction.

3. RESULTS

The fatigue protocol lasted in average 14.45±9.8 min and the Borg perceived exertion was 7.65±1.64 out of 10. Table 1 presents the number of subjects presenting significant differences after fatigue in their parameters reflecting central system (t_0, D, θs, θe), peripheral system (μ, σ) or both systems. Both central and peripheral systems were affected by fatigue. Agonist components seemed to be more altered by fatigue than the antagonist ones in the central and peripheral systems. In fact, 60% of the participants have shown changes in their central system parameters for the agonist components, as compared to 35% for their antagonist components. Regarding the peripheral system, parameters were significantly modified after fatigue in 40% of the population for the agonist components compared to 15% for the antagonist component. Parameters reflecting the global state of the neuromotor system changed in 40% and 75% of the population respectively for the agonist and agonist components and the reaction time in 45%. For this latter parameter no distinctions between agonist and antagonist components were made as it is calculated from the whole fast stroke.
Overall, all but one participant presented statistical significant differences in some parameters for both agonist and antagonist components.

	Central system	Peripheral system	Both systems	Global state of the neuromotor system	RT
Agonist	12 (60%)	8 (40%)	5 (25%)	8 (40%)	9 (45%)
Antagonist	7 (35%)	3 (15%)	2 (10%)	15 (75%)	

Table 1. Number of participants (percentage) with significant differences for their agonist or antagonist components.

To illustrate how fatigue can affect the Kinematic parameters, Fig. 2a presents the velocity profile of the mean stroke of participant 1 before fatigue (blue) and after fatigue (red). We can observe that the velocity profile after fatigue was wider with a lower amplitude and shifted to the right. Fig. 2b and 2c show respectively that t0 and σ are higher after fatigue for both the agonist and antagonist components, suggesting that the central and peripheral systems of this participant are altered. The time before starting the movement is also higher after fatigue, which means that the information takes more time to be processed by the brain and directed toward the end effector.

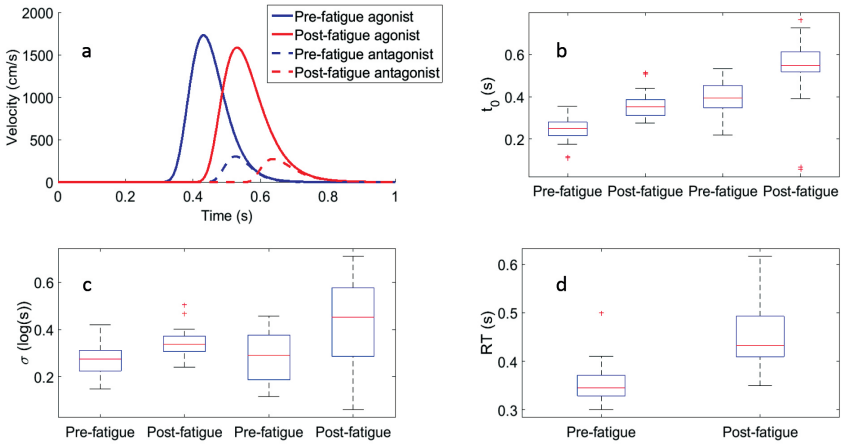


Figure 2. a) Participant's 1 velocity profile pre- and post-fatigue. b to d) Participant's 1 boxplots of t0 (b), σ (c) and RT (d) pre- and post-fatigue. For b and c the first two boxplots are for the agonist components and the last two for the antagonist components. Red crosses represent outliers (> 2.7 standard deviations).

4. DISCUSSION

The purpose of this study was to assess the detection of shoulder fatigue using the Kinematic Theory. We found that Kinematic parameters from the central system (t_0, D, θs, θe), peripheral system (μ, σ) or global system (SNR, NbLog and derived parameters of the Theory) reflected motor control changes. The only participant without any significant difference stopped the fatigue protocol due to the time limit with a Borg CR10 perceived exertion of only 6. It is possible that this participant did not reach his maximum force in the calibration phrase.

With this method, analyses could show that the agonist components seemed to be more impacted by the fatigue than the antagonist components. It could be due to the movement performed, as the participant's arm was maintained elevated and the movement performed similar to ER. The global state of the neuromotor system had more significant differences after fatigue compared to the central or peripheral systems, which means that the global neuromuscular system of the participants was modified by the fatigue. Moreover, results showed that the protocol impacted the central system of some participants and the peripheral system of some other. However, only a few had both a central and peripheral fatigue. As fatigue is multi-causes induced, its expression differs from one individual to another [15]. Some participants had a reduction in their motor command, with a higher t_0 or smaller D, meaning respectively that their brain took more time to send their motor command and that they drew smaller strokes. Some other participants improved the capacities of their central system command after fatigue, with a smaller t_0. This can be explained by the fact that physical exercise can improve cognitive function [19]. Nevertheless, those participants had a higher μ after fatigue, meaning that their system's neuromuscular response was slower. They compensated the impact of the central system (t_0) with a degradation of parameters linked to their peripheral system (μ and associated parameters). Furthermore, participants who drew faster strokes after fatigue (diminution of σ) or had a faster neuromuscular response (diminution of μ) also compensated these ameliorations with degradations in other parameters, for example a modification of their central motor command (θs or θd significantly different). To conclude, motor pattern changes are unique to each one. Many scenarios are possible, but starting from a baseline, it is though possible to customize the tool for each individual. Further tests, such

as EMG, can then be correlated to physiologically determine exactly from which muscle the fatigue comes from [16-17] and to better understand the way a participant will compensate the fatigue. These are very promising results because, to the best of our knowledge, no tools taken separately can discriminate a peripheral from a central fatigue to measure human muscle fatigue. The combination of surface electromyography (EMG) with voluntary contraction and noninvasive transcutaneous electrical stimulation can distinguish central fatigue from peripheral fatigue. But this method remains tough for the participants, whereas the experimental set-up proposed here is really easy to use [18].

From a clinical point of view, this study will serve as basis for other studies regarding fatigue detection in athletes. In this protocol the infraspinatus muscle was targeted for the fatigue. It would be interesting to see how parameters are impacted depending on the type of fatigue, after targeting another muscle of the rotator cuff for the fatigue. In conclusion the Sigma-Log-normal analyses seem appropriate for fatigue pre-screening and later on for shoulder injuries prevention.

Attention still have to be focused on the fact that fatigue impacts all biological levels of the body, from the central command, to the propagation of the signal resulting in movement changes and impairment in motor performance [20-22]. As a result of fact, the Sigma-Log-normal analyses seem appropriate to reflect motor control changes after any kind of fatigue or neural modification. However further studies should be done to assess it. It also has to be highlighted that fatigue is not the only aspect that can impact parameters of the Kinematic Theory [8, 23]. At this level of knowledge of the proposed method, if significant differences are seen on a person, it is imperative for the clinician to make complementary investigations to assess their origins.

5. ACKNOWLEDGMENTS

This research was conducted as part of the TransMedTech Institute's activities and the first author was funded in part by the Canada First Research Excellence Fund. This work was also partly supported by NSERC CANADA Discovery Grant RGPIN-2015-06409 to Réjean Plamondon.

REFERENCES

[1] United States Bone and Joint Initiative, «The Burden of Musculoskeletal Diseases in the United States (BMUS), Third Edition,» 2014. [Online document].Available: https://www.boneandjointburden.org/.

[2] D. D. Ebaugh, P. W. McClure and A. R. Karduna, «Effects of shoulder muscle fatigue caused by repetitive overhead activities on scapulothoracic and glenohumeral kinematics,» *Journal of Electromyography and Kinesiology*, vol. 16, n° 3, pp. 224-235, 2006.

[3] S. Gaudet, J. Tremblay and F. Dal Maso, «Evolution of muscular fatigue in periscapular and rotator cuff muscles during isokinetic shoulder rotations,» *Journal of sports sciences*, vol. 36, n° 18, pp. 2121-2128, 2018.

[4] R. L. Sterner, D. M. Pincivero and S. M. Lephart, «The Effects of Muscular Fatigue on Shoulder Proprioception,» *Clinical Journal of Sport Medecine*, vol. 8, pp. 96-101, 1998.

[5] S. D. Mair, A. V. Seaber, R. R. Glisson and W. E. Garrett, «The Role of Fatigue in Susceptibility to Acute Muscle Strain Injury,» *The American Journal of Sports Medicine*, vol. 24, n° 2, pp. 137-143, 1996.

[6] J. L. Taylor and S. C. Gandevia, «A comparison of central aspects of fatigue in submaxi-

mal and maximal voluntary contractions,» *Journal of Applied Physiology*, 2008.
[7] R. Plamondon, «A kinematic theory of rapid human movements. Part I: Movement representation and generation.,» *Biological cybernetics*, vol. 72, pp. 295-307, 1995.
[8] P. Laniel, N. Faci, R. Plamondon, M. H. Beauchamp and B. Gauthier, «Kinematic Analysis of Fast Pen Strokes in Children with ADHD,» *Applied Neuropsychology*: Child, pp. 1-16, 2018.
[9] N. Faci, S. P. Boyogueno Bidias, R. Plamondon and N. Bergeron, «A New Experimental Set-up to Run Neuromuscular Tests,» In *International Conference on Pattern Recognition and Artificial Intelligence*, Montréal, 2018.
[10] S. Gaudet, J. Tremblay and M. Begon, «Muscle recruitment patterns of the subscapularis, serratus anterior and other shoulder girdle muscles during isokinetic internal and external rotations,» *Journal of Sports Sciences*, vol. 36, n° 9, pp. 985-993, 2017.
[11] C. Yang, J. Bouffard, D. Srinivasan, S. Ghayourmanesh, H. Cantú, M. Begon and J. N. Côté, «Changes in movement variability and task performance during a fatiguing repetitive pointing task,» *Journal of Biomechanics*, vol. 76, pp. 212-219, 2018.
[12] R. Plamondon and M. Djioua, «A multi-level representation paradigm for handwriting stroke generation,» *Human Movement Science*, vol. 25, n° 4-5, pp. 586-607, 2006.
[13] R. Plamondon, C. Feng and A. Woch, «A kinematic theory of rapid human movement. Part IV: a formal mathematical proof and new insights,» *Biological Cybernetics*, vol. 89, n° 2, pp. 126-138, 2003.
[14] C. O'Reilly, R. Plamondon, M. K. Landou and B. Stemmer, «Using kinematic analysis of movement to predict the time occurrence of an evoked potential associated with a motor command,» *European Journal of Neuroscience*, vol. 37, n° 2, pp. 173-180, 2013.
[15] J. Finsterer, «Biomarkers of peripheral muscle fatigue during exercise,» *BMC Musculoskeletal Disorders*, vol. 13, n° 1, p. 218, 2012.
[16] D. D. Ebaugh, P. W. McClure and A. R. Karduna, «Scapulothoracic and Glenohumeral Kinemtics Following an External Rotation Fatigue Protocol,» *Journal of Orthopaedic & Sports Physical Therapy*, vol. 36, n° 8, pp. 557-571, 2019.
[17] M. R. Al-Mulla, F. Sepulveda and M. Colley, «A Review of Non-Invasive Techniques to Detect and Predict Localised Muscle Fatigue,» *Sensors*, vol. 11, n° 4, pp. 3545-3594, 2011.
[18] N. Place, N. A. Maffiuletti, A. Martin and R. Lepers, «Assessment of the reliability of central and peripheral fatigue after sustained maximal voluntary contraction of the quadriceps muscle,» *Muscle & Nerve*, vol. 35, n° 4, pp. 486-495, 2007.
[19] C. H. Hillman, K. I. Erickson and A. F. Kramer, «Be smart, exercise your heart: exercise effects on brain and cognition.,» *Nature Reviews Neuroscience*, vol. 9, n° 1, pp. 58-65, 2008.
[20] R. M. Enoka and D. G. Stuart, «Neurobiology of muscle fatigue,» *Journal of Applied Physiology*, vol. 72, n° 5, pp. 1631-1648, 1992.
[21] N. Cortes, J. Onate and S. Morrison, «Differential effects of fatigue on movement variability,» *Gait & Posture*, vol. 39, n° 3, pp. 888-893, 2014.
[22] S. C. Gandevia, «Spinal and Supraspinal Factors in Human Muscle Fatigue,» *Physiological Reviews*, vol. 81, n° 4, pp. 1725-1789, 2001.
[23] R. Plamondon, C. O'Reilly, C. Rémi and T. Duval, «The lognormal handwriter: learning, performing, and declining,» *Frontiers in Psychology*, vol. 4, n° 945, 2013.

Combining Interval Arithmetic with the Branch and Bound Algorithm for Delta-lognormal Parameter Extraction

Simon Pierre Boyogueno Bidias, Jean-Pierre David, Yvon Savaria, Réjean Plamondon

Abstract - In this paper, an interval arithmetic approach is proposed to compute guaranteed enclosures of the sets of points that bound a multivariate delta lognormal function. The proposed algorithm is used in a branch and bound scheme, to extract parameters of the delta lognormal function from ideal handwriting velocity profiles. Following the Kinematic Theory of rapid human movements, the basic timing properties of the neuromuscular system is represented by the seven parameters which characterize the delta log-normal function. The basic idea behind the proposed algorithm is to use the natural interval extension to sharply bound the range of the delta lognormal function. The proposed new algorithm is described, tested (under various handwriting strokes) and compared with a previous algorithm developed for the same purpose. The numerical results show that the proposed algorithm gives better results in terms of speed and accuracy. This new algorithm, when embedded in a branch and bound scheme is useful for pointing movement analysis like gesture commands and handwriting strokes.

Keywords: Branch and Bound, interval arithmetic, delta lognormal, parameter extraction, Kinematic Theory, rapid Human movements, Handwriting, algorithm.

1. INTRODUCTION

Basic gestures like pointing movements are part of our daily interaction with the world. They can be defined as a coordinated set of actions performed by small muscles usually involving the synchronization of hands and fingers with the eyes. The complex levels of manual dexterity that humans exhibit are demonstrated in tasks controlled by the nervous system, to which they can be attributed [1] [2]. These simple movements can be seen as the basic primitives [3] that are combined to produce more complex movements like handwriting and signatures. These movements were deeply studied by psychologists, neuroscientists, physicists and computer scientists over the last century [4] [5]. Modeling them is the cornerstone to understand the basic principles of human motor control and its disorders [6].

In this perspective, the need to study the elementary properties of a single stroke for realistic analysis becomes key to understand how the motor control system accomplishes complex movements. The study of graphomotor behavior based on the Kinematic Theory of rapid human movements [7], [8], produces quantitative, precise and objective measures of writing [9]. Among all existing handwriting generation models in the literature, the Delta-lognormal model (1), has proven over the years to be one of the most powerful to describe pointing movements because of its ability to reproduce, with minimal error, the velocity profile of a handwritten stroke. This model derived from the Kinematic Theory of rapid human movements, [10], [8], [11] considers a single stroke as a pen tip movement with a velocity profile expressed as in equation (1).

$$\Delta\Lambda(t;p) = D_1\Lambda(t;t_0,\mu_1,\sigma_2^2) - D_2\Lambda(t;t_0,\mu_2,\sigma_2^2) \tag{1}$$

Where

$$\Lambda(t; t_0, \mu, \sigma^2) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}(t-t_0)} \exp^{-\frac{(\ln(t-t_0)-\mu)^2}{2\sigma^2}}, & for \quad t_0 < t, \\ 0, & elsewhere. \end{cases} \quad (2)$$

With $\sigma > 0$; μ ,$t_0 \in \mathbb{R}$
is the lognormal impulse response function. In this theory, a stroke is produced by a synergy of two neuromuscular systems; the agonist, acting in the direction of the movement and the antagonist, acting in the opposite direction. The agonist and antagonist are lognormal convolved with their corresponding input commands, represented in equation (1) by the subscripts 1 and 2 respectively. The parameters in equation (3) are defined as follows:

- t_0: represents the system activation time.
- μ_1, σ_1, μ_2, σ_2 correspond to the timing properties (on a logarithmic scale) of the two neuromuscular commands
- D_1 and D_2 are the input amplitudes of the agonist and antagonist neuromuscular systems respectively.

The study of such strokes relies on the extraction of the parameters that characterize their velocity profile [8] [10], [11]. This problem can be achieved by minimizing the nonlinear least square function f (3), that maximizes the signal to noise ratio *SNR* (4). This will be the subject of our next two sections.

$$f = \int (v_t(t) - \Delta\Lambda(t; p))^2 dt \quad (3)$$

$$SNR = 10 \log\left(\frac{\int v_t^2(t) dt}{f}\right) \quad (4)$$

Although the parameter extractors developed in [12], [13], [14] show good data fitting, none of them can guarantee the optimality of the solution obtained. To address this need, a first branch and bound algorithm was developed in [15]. This algorithm is based on a delta lognormal envelope which bounds all the possible delta lognormal parameters within a subspace, and then computes the error between the signal and its nearest envelope. This method, even though it is promising, suffers from long computation time, very large memory requirements and lower bounding problems to extract a solution. In this paper, we propose a new algorithm, based on interval arithmetic, to solve the boundaries problems of the extractor proposed in [15] and show the effectiveness of the proposed algorithm when combined in a branch and bound scheme to extract model parameters from ideal velocity profile data.
The rest of the paper is organized as follows. Section II reviews some theoretical aspects of interval arithmetic and presents an algorithm for computing verified enclosures of the bound of the objective function. The uses of this algorithm in a branch and bound scheme will also be highlighted. Section III presents the numerical results and performance comparisons with [15]and [13]. Section IV proposes possible future research and summarizes the main contributions of this paper.

Authors are with the Electrical Engineering Department, Polytechnique Montréal, Montréal, H3C 3A7 CANADA, simon-pierre.boyogueno-bidias@polymtl.ca, JPDavid@polymtl.ca, yvon.savaria@polymtl.ca, rejean.plamondon@polymtl.ca

2. ALGORITHM FOR COMPUTING THE UPPER AND LOWER BOUND OF THE DELTA LOGNORMAL FUNCTION.

Some definitions and notations need to be presented to support the proposed algorithm. We denote by $I=[a,b] = \{x \in \mathbb{R}: a \leq x \leq b\}$ the set of real numbers and one dimensional closed intervals. In this paper, an interval will be denoted by a variable in upper case while the set of values contained in the interval will be denoted by a lower case. Thus $X = [\underline{x}, \overline{x}]$ is an interval variable where $\underline{x}$ and $\overline{x}$ represent the left and right end points of X respectively. The width of X is defined and denoted by $(X) = \overline{x} - \underline{x}$ and the midpoint is given by $m(X) = \frac{1}{2}(\overline{x} + \underline{x})$

We denote by $\boldsymbol{X^I} = (X_1, X_2, \ldots \ldots, X_n)$ the set of n dimensional interval vectors or hyper box such that:

$\boldsymbol{X^I} = (X_1, X_2, \ldots. X_n) = \left([\underline{x}_1, \overline{x}_1], [\underline{x}_2, \overline{x}_2], \ldots [\underline{x}_n, \overline{x}_n]\right)$.

The width and the midpoint of an interval vector are defined as:

$w(\boldsymbol{X^I}) = \max_i w(\boldsymbol{X_i^I})$

and

$m(\boldsymbol{X^I}) = (m(X_1), m(X_2), , m(X_n))$.

We call F an inclusion function of $f: \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, the interval extension of f, such that for $x \in X \Rightarrow f(x) \subseteq F(X)$. $f(X)$ is the real range of f on X and $f(X) \subseteq F(X)$.

Now that we have these definitions and notations, it is possible to use interval arithmetic [16], [17] to bound a function defined by a mathematical expression such as the delta-lognormal function.

A. Interval arithmetic:

In this section, we introduce the interval tools needed to derive the bounding operation of equation (3). For more details, we refer the reader to [17], [16].

Theorem (fundamental theorem). Let $F(X_1,X_2,\ldots X_n)$ be the natural interval extension of f $(x_1, x_2,\ldots x_n)$ then $f(X_1, X_2,\ldots X_n) \subseteq (X_1, X_2,\ldots X_n)$.

And for all intervals,

$if\ Y_k \subset X_k\ for\ k = 1, \ldots . n\ ,\ then\ f(Y_1, Y_2, \ldots Y_n) \subseteq F(Y_1, Y_2, \ldots Y_n)$.

Where $f(Y_1, Y_2, \ldots Y_n) = \{ f(x_1, x_2, \ldots x_n): x_k \in Y_k\ for\ k = 1, \ldots, n\}$

This theorem is due to Moore and extended by Hansen, the proof can be found in [17]. It shows how easy it is to bound the range of a function, and make finding the solution to the global optimization problem possible. In the next sub-section, we are going to apply this theorem to the delta lognormal function as a specific sequence of interval arithmetic operations to bound the range of equation (3) within an interval vector.

B. Algorithm for computing the upper and lower bounds.

To compute the bounds of equation (3) we first define the bounding space of its variables as:

$$\boldsymbol{P^I} = \left\{[\underline{d}_1, \overline{d}_1], [\underline{\mu}_1, \overline{\mu}_1], [\underline{\sigma}_1, \overline{\sigma}_1], [\underline{d}_2, \overline{d}_2], [\underline{\mu}_2, \overline{\mu}_2], [\underline{\sigma}_2, \overline{\sigma}_2], [\underline{t}_0, \overline{t}_0]\right\}$$

Then, through the fundamental theorem, we apply the natural interval extension to it. This yields:

$$\begin{aligned} f(\boldsymbol{P^I}) \in F(\boldsymbol{P^I}) &= [\underline{F}(\boldsymbol{P^I}), \overline{F}(\boldsymbol{P^I})] \\ &= \int (v_t(t) - \Delta\Lambda(t; \boldsymbol{P^I}))^2 dt \end{aligned} \qquad (6)$$

As provided by the theorem and the natural interval extension, the result of the right-hand side of equation (10) is an interval $\left[\underline{F}(\boldsymbol{P^I}), \overline{F}(\boldsymbol{P^I})\right]$ that bounds *f*.
To compute that result, we have broken down the last term $\Delta\Lambda(t;\boldsymbol{P^I})$ in the integral of equation (10), into a unique finite sequence of interval arithmetic operations as proposed in [18].

Function_compute_Λ $\left(\left[\underline{\mu}_1,\ \overline{\mu}_1\right], \left[\underline{\sigma}_1,\ \overline{\sigma}_1\right], \left[\underline{t}_0,\ \overline{t}_0\right], t\right)$

1. constant $= \sqrt{2\pi}$;
2. Compute equation (2) using the following list of expressions (';' are used as separators for brevity).

$$T_1 = \left[\underline{\sigma}_1,\ \overline{\sigma}_1\right]^2; \quad T_2 = t - \left[\underline{t}_0,\ \overline{t}_0\right];$$

$$T_3 = ln(T_2); \quad T_4 = T_3 - \left[\underline{\mu}_1,\ \overline{\mu}_1\right];$$

$$T_5 = {T_3}^2; \quad T_6 = \frac{T_5}{T_1}; \quad T_7 = 0{,}5 \times T_6;$$

$$T_8 = -T_7\ ; \quad T_9 = exp(T_8);$$

$$T_{10} = constant \times \left[\underline{\sigma}_1,\ \overline{\sigma}_1\right]; \ T_{11} = T_2 \times T_{10}\ ;$$

$$T_{12} = \frac{T_9}{T_{11}}\ ;$$

3. return T_{12};

Algorithm1: Compute Bound ($\boldsymbol{P_i^I}, v_t(t), t, N$)

1. set the initial variables: $\Delta\Lambda[\ N\ ]$, $F = [\ ]$; $\Lambda_1\ [\ N\ ]$; $\Lambda_2[\ N\ ]$
2. $P = \boldsymbol{P_i^I} = \left\{\left[\underline{d}_1,\ \overline{d}_1\right], \left[\underline{\mu}_1,\ \overline{\mu}_1\right], \left[\underline{\sigma}_1,\ \overline{\sigma}_1\right], \left[\underline{d}_2,\ \overline{d}_2\right], \left[\underline{\mu}_2,\ \overline{\mu}_2\right], \left[\underline{\sigma}_2,\ \overline{\sigma}_2\right], \left[\underline{t}_0,\ \overline{t}_0\right]\right\}$.
3. for i=0 to N-1
4. if $\left(t_i - \left[\underline{t}_0,\ \overline{t}_0\right]\right) > 0$
5. $\Lambda_1\ [\ i\ ]$= **Function_*compute_Λ***$\left(\left[\underline{\mu}_1,\ \overline{\mu}_1\right], \left[\underline{\sigma}_1,\ \overline{\sigma}_1\right], \left[\underline{t}_0,\ \overline{t}_0\right], t_i\right)$;
6. $\Lambda_1\ [\ i\ ]$= **Function_*compute_Λ***$\left(\left[\underline{\mu}_2,\ \overline{\mu}_2\right], \left[\underline{\sigma}_2,\ \overline{\sigma}_2\right], \left[\underline{t}_0,\ \overline{t}_0\right], t_i\right)$;
7. $\Delta\Lambda[\ i\ ] = \Lambda_1\ [\ i\ ] - \Lambda_2\ [\ i\ ]$;
   else
   $\Lambda_1\ [\ i\ ]=0$; $\Lambda_2\ [\ i\ ]=0$; $\Delta\Lambda[\ i\ ]=0$;
8. compute the integral
9. if$\left(t_i - \left[\underline{t}_0,\ \overline{t}_0\right]\right) > 0$
   $((v_t[0] - \Delta\Lambda[0])^2 + (v_t[N-1] - \Delta\Lambda[N-1])^2) \times 0.5$
   For i=1 to N-1
10. If $\left(t_i - \left[\underline{t}_0,\ \overline{t}_0\right]\right) > 0$
    $F \mathrel{+}= (v_t[i] - \Delta\Lambda[i])^2$;
    Else
11. $F \mathrel{+}= 0$;
12. return $\frac{F}{sample\ frequency}$;

Algorithm1 starts by expressing *f* as a code list (see **Function_compute_Λ**) [18] where each arithmetic (±, ×, ÷, etc.) and unary elementary functions are expressed with only one unique variable in their expression. In this way we ensure that, the desirable properties (inclusion isotonicity, etc.) of the fundamental theorem are satisfied. Another advantage of using this method is the sharpness of the bound that results from it. In this algorithm, the integral is

computed numerically (see steps 8-11) as proposed in [19]. The condition in steps 4, 9 and 10 in Algorithm1 served to preserve the inclusion of the objective function (when computing the integral) to lie in the whole domain where it is defined, regardless of the subspace to be evaluated. This is necessary because our objective function (see 3) is piecewise defined.
In summary the inputs for Algorithm1 are a box P, the ideal velocity profile $v_t(t)$, the sample time t, and the sample size N. From step 3 to 8, it computes the delta-lognormal function by calling Function_compute_Λ for each lognormal sample and then computes the integral to return the bounds on F. This is done in steps 8 to 11.

C. Applying algorithm1 in a general branch and bound scheme

The general idea behind the branch and bound technique is to bound sub-boxes of a feasible area in which we perform a search to get the sharpest one. The bounding operations can be carried out by operations such as in Algorithm1 and, if the bounds are not sharp enough, some sub-boxes are divided into smaller one according to a splitting rule. The procedure repeats until the optimal solution is obtained within the required accuracy. For our purpose, the objective function to be minimized is provided by equation (3) and the accuracy is fixed by equation (4) to an SNR = 100dB. To verify that the developed algorithm produces valid results, we replaced in [15] the computation of the lower and upper limits of the lognormal parameters by Algorithm1. We also changed the division rule of a subspace from three to two sub boxes.
The feasible area is restricted to a region of search where we know that the solution can be found. The values of the interval within the bounding box are chosen as wide as possible, to see how the algorithm will perform in finding the solution.

3. TESTING THE ALGORITHM UNDER IDEAL CONDITION

A. Test conditions

To test Algorithm1, we used the database proposed in [13]. This database is composed of 7000 delta-lognormal curves, grouped into categories. These ideal data are derived from human handwriting movements, previously extracted using the IIX algorithm presented in [15]. From this database, we randomly choose 300 ideal delta-lognormal velocity profiles, from Ca0 and Ca2 classes. Each of them is sampled at 200Hz to simulate the data collected with a digitizer [20] and satisfy a signal-to-noise ratio of 25dB with respect to the real (collected from real human movement with a digitizer) velocity profile. The implementation is done in C++ using our custom interval arithmetic library based on [16] and [21]. All computations are performed using rounded interval arithmetic in a 3.60 GHz Intel core (i7) with 32 Go of Ram.

B. Results and discussion

In this experience, a box is considered a solution and stored in the list of results if the maximum width or the quadratic error of the chosen box is less than 10^{-5} . That is the signal to noise ratio of the optimal box must be at least 100dB.
Table I presents the performance comparison in terms of accuracy between the proposed algorithm, [13] and [15]. As can be seen in table I, for the class Ca0, our algorithm performs better than Xzero [13] and produces the same results as in [15]. Table II shows the performance comparison in terms of processing time and number of boxes explored between the proposed algorithm and [15]. In [15], the algorithm is a branch and bound executed in parallel, on a 12 CPU node grid. As can be seen in table II below. This algorithm requires a run

time of about 50s to perform 11 iterations of its main loop while the one we propose requires only 702ms. On the other hand, the number of boxes needed to find the optimal solution can reach a million with their algorithm while ours requires at most 70 thousand boxes to extract the optimal solution.

	Method		
Classes	***Xzero***	***[15]***	***Proposed***
Ca0 in %	95.2	100	100
Ca2 %	100	100	100

Table 1. extraction of the exact value of the parameters for an snr of >100db as ending criterion

	Method	
Performances	***[15]***	***Proposed***
Number of Boxes	Min 800000	Max 70000
Processing time (in second)	50	0.702

Table 2. performance comparison of the number of boxes explored to extract one velocity profile and the processing time required to perform 11 iterations in the main loop

4. CONCLUSION

This paper shows that a good bounding algorithm used in a branch and bound scheme can be used to extract the delta-lognormal function parameters of ideal velocity profile data with a lower computational cost. We have focused on the analysis of individual synthesized movements using the kinematic theory of rapid human movement. The developed algorithm that bounds the delta lognormal function and guarantees the inclusion when used in a branch and bound for global optimization is illustrated and evaluated. The proposed algorithm exploits the natural interval extension and the fundamental theorem of interval arithmetic to compute the bounding operations of the delta lognormal function. We also highlighted the use of that algorithm in a branch and bound scheme and compared the performance of the parameter estimation with two previous algorithms developed for the same purpose. Although the proposed algorithm is promising in its ability to recover the global solution in a less time (in comparison to the previous one), more work still remains to be done. Another step could be to test the algorithm on several classes of velocity profiles as provided by the delta lognormal model as well as on real human movement data.

REFERENCES

[1] G. Pirlo, M. Diaz et M. A. e. a. Ferrer, «Early diagnosis of neurodegenerative diseases by handwritten signature analysis. In: International Conference on Image Analysis and Processing.,» *Springer, Cham*, pp. 290-297, 2015.

[2] P. M. Drotár, R. I. Jiří et e. al., «Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease.,» *Artificial intelligence in Medicine*, vol. 67, pp. 39-46., 2016.

[3] A. Woch et R. Plamondon, «Using the framework of the kinematic theory for the definition of a movement primitive.,» *Motor Control*, vol. 8, n° %14, pp. 547-557, 2004.

[4] M. P. Caligiuri, H.-L. Teulings et C. E. e. a. Dean, «Handwriting movement kinematics for quantifying extrapyramidal side effects in patients treated with atypical antipsychotics.,» *Psychiatry Research*, vol. 177, n° %11-2, pp. 77-83, 2010.

[5] R. Plamondon, G. Pirlo et E. e. a. Anquetil, «Personal digital bodyguards for e-security, e-learning and e-health: A prospective survey.,» *Pattern Recognition*, vol. 81, pp. 633-659, 2018.

[6] M. Lorch, «Written language production disorders: historical and recent perspectives.,» *Current neurology and neuroscience reports*, vol. 13, n° %18, p. 369, 2013.
[7] R. Plamondon, «A kinematic theory of rapid human movements: Part III. Kinetic outcomes,» *Biological Cybernetics*, vol. 78, n° %12, pp. 133-145, 1998.
[8] R. Plamondon, «A kinematic theory of rapid human movements Part II. Movement time and control.,» *Biological Cybernetics*, vol. 74, n° %14, pp. 309-320, 1995.
[9] O. Tucha, L. Tucha et G. e. a. Kaumann, «Training of attention functions in children with attention deficit hyperactivity disorder.,» *ADHD Attention Deficit and Hyperactivity Disorders*, vol. 3, n° %13, pp. 271-283, 2011.
[10] R. Plamondon, «A kinematic theory of rapid human movements,» *Biological Cybernetics*, vol. 72, n° %1 4, pp. 295-307, 1995.
[11] R. Plamondon, C. Feng et A. and Woch, «A kinematic theory of rapid human movement. Part IV: a formal mathematical proof and new insights.,» *Biological Cybernetics*, vol. 89, n° %12, pp. 126-138, 2003.
[12] W. Guerfali Et R. Plamondon, « Signal Processing for the Parameter Extraction of the Delta Lognormal Model (ΔΛ).,» *Research in computer and robot vision*, pp. 217-232, 1995.
[13] M. Djioua et R. and Plamondon, « "A New Algorithm and System for the Characterization of Handwriting Strokes with Delta-Lognormal Parameters,» *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, n° %111, pp. 2060-2072, 2009.
[14] R. Plamondon, X. Li et M. Djioua, «Extraction of delta-lognormal parameters from handwriting strokes,» *Frontiers of Computer Science in China*, vol. 1, n° %11, pp. 113-117, 2007.
[15] C. O'Reilly et R. Plamondon, «A Globally Optimal Estimator for the Delta-Lognormal Modeling of Fast Reaching Movements,» *IEEE Transactions on Systems Man and Cybernetics*, vol. 42, n° %15, pp. 1428-1442, 2012.
[16] R. E. Moore, Interval analysis, NJ: Englewood Cliffs, : Prentice-Hall, 1966.
[17] E. Hansen et G. W. Walster, Global optimization using interval analysis: revised and expanded, CRC Press, 2003.
[18] S. Boyogueno Bidias, J.-P. David, Y. Savaria et R. Plamondon, «On the use of Interval Arithmetic to Bound Delta-Lognormal Rapid Human Movements Models,» chez *C.F. Intelligence edition, International Conference on pattern Recognition and artificial intelligence: Workshop on the Lognormality Principles and its Applications* , Montreal, mai 2018.
[19] A. Fortin, Analyse numérique pour ingénieurs, Presses inter Polytechnique, 2001.
[20] N. Faci, S. Boyogueno Bidias, R. Plamondon et N. Bergeron, «A new experimental Setup to Run neuromuscular tests.,» chez *Internationnal conference on pattern recognition and artificial intelligence: worksop on the lognormality Principle and its applications*, Montreal, mai 2018.
[21] O. Knüppel, « Profil/Bias-a fast interval library.,» *Computing*, vol. 53, n° %13-4, pp. 277-287, 1994.

Authors