ISPA 2019

Proceedings of the 11th International Symposium on

Image and Signal Processing and Analysis

S. Lončarić, R. Bregović, M. Carli, M. Subašić (Eds.)

Dubrovnik, Croatia, September 23-25, 2019

EURASIP

Technical Co-Sponsor
IEEE
Signal Processing Society

IEEE
CROATIA SECTION

USB
version

ISPA 2019

11th International Symposium on Image and Signal Processing and Analysis

Dubrovnik, Croatia, September 23-25, 2019

S. Lončarić, R. Bregović, M. Carli, M. Subašić (Eds.)

Organizing Institution

University of Zagreb
Faculty of Electrical Engineering and Computing
Croatia

Auspices

Croatian Academy of Sciences and Arts

Technical Co-Sponsors

European Association for Signal Processing (EURASIP)
IEEE Signal Processing Society
IEEE Croatia Section
IEEE Croatia Section Signal Processing Chapter

Editors

Sven Lončarić (sven.loncaric@fer.hr)
University of Zagreb
Unska 3, HR-10000, Croatia

Robert Bregović (robert.bregovic@tuni.fi)
Tampere University
33101 Tampere, Finland

Marco Carli (marco.carliuniroma3.it)
Roma TRE University
via Vito Volterra, 60 - Corpo B, 00146 Roma, Italy

Marko Subašić (marko.subasic@fer.hr)
University of Zagreb
Unska 3, HR-10000, Croatia

Publisher

University of Zagreb, Croatia

Production and Publishing: Tomislav Petković

Cover Photographs: Ivana Čuljak

2019 11th International Symposium on Image and Signal Processing and Analysis

IEEE Catalog Number CFP19504-ART (online)
IEEE Catalog Number CFP19504-USB (USB)

ISBN 978-1-7281-3140-5 (online)
ISBN 978-1-7281-3139-9 (USB)

ISSN 1849-2266 (online)

Technical support for USB is available from:

ISPA Conference Secretariat
University of Zagreb
Faculty of Electrical Engineering and Computing
Unska 3, HR-10000 Zagreb, Croatia
tel: +385 1 6129 911
fax: +385 1 6129 652

Preface

The 2019 edition of the International Symposium on Image and Signal Processing and Analysis (ISPA 2019) is the eleventh in the series of biennial research meetings and it follows the successful ISPA 2017 meeting held in Ljubljana, Slovenia. ISPA 2019 is held in Dubrovnik, Croatia, which is located on the Adriatic Sea coast in the extreme south of Dalmatia.

The very favourable geographical position of Dubrovnik made its development based on maritime and merchant activities very successful through its history. George Bernard Shaw was enchanted by this beautiful city, about which he said "*those who seek paradise on Earth should come to Dubrovnik and see Dubrovnik*", as well as, famously, describing it as "*the pearl of the Adriatic*". Dubrovnik, with its amazing Old Town, became a UNESCO World Heritage site in 1979.

The ISPA 2019 venue is the Centre for Advanced Academic Studies in Dubrovnik, which was founded by the University of Zagreb as a public academic institution for international scientific programmes.

We received 103 submissions presenting research and applications of signal and image processing. The program committee selected 65 submissions for presentation at ISPA 2019. The papers were organized into 13 oral sessions that were scheduled into two parallel program tracks. The ISPA 2019 program also includes three special sessions and one workshop. The special sessions were organized and are chaired by distinguished researchers from both academia and industry and present the state of the art on specific aspects of image and signal processing. The topics of the special sessions are:

- "methods and applications of time-frequency signal analysis",
- "signal processing and machine learning for finance", and
- "immersive visualisation for safety-critical applications".

The workshop is on color vision and includes a illumination estimation challange which attracted a total of nine submissions.

The goal of the symposium is to foster exchange of information between researchers and serve as a meeting point to establish further contacts for future research collaborations. We wish all ISPA participants a successful meeting and an enjoyable stay in Dubrovnik. We also hope to see you at many future ISPA symposia.

September 2019

Robert Bregović, Marco Carli and Marko Subašić

Program Co-Chairs

Sven Lončarić

General Chair

Acknowledgements

The 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA) is the result of the dedicated efforts of many volunteers from all over the word. All papers included in these proceedings are results of years of research made by scientists coming from many different countries. Without their contribution and commitment this Symposium would not have been possible. The Program Committee members and other scientists spent many hours reviewing submitted papers and working on the preparation of the program. Their expertise and constant advice have been instrumental for ensuring the scientific quality of the Symposium. Managing the electronic submissions of the papers, the preparation of the abstract booklet and of the proceedings also required substantial effort and dedication that must be acknowledged. The Local Organizing Committee members did an excellent job to guarantee a successful outcome of the Symposium and a pleasant stay in Dubrovnik for all the participants. All this continuous effort has been admirable.

We are grateful to the Technical sponsors, who helped us in granting the high scientific quality of the presentations, to the Sponsors that financially supported this symposium, and to the patronage institutions that encouraged our work. The professional service provided by the Conference Secretariat was always prompt and reliable.

Contents

Organizing Committee

Honorary Chair

Sanjit K. Mitra, University of California Santa Barbara, USA

General Chair

Sven Lončarić, University of Zagreb, Croatia

Program Co-Chairs

Robert Bregović, Tampere University of Technology, Finland
Marco Carli, University Roma TRE, Italy
Marko Subašić, University of Zagreb, Croatia

Special Sessions Chair

Federica Battisti, University Roma TRE, Italy

Publicity Chair

Federico Sukno, University Pompeu Fabra, Barcelona, Spain

Publications Chair

Tomislav Petković, University of Zagreb, Croatia

Local Arrangements Chair

Ana Sović Kržić, University of Zagreb, Croatia

IEEE Signal Processing Society Liaisons

Touradj Ebrahimi, Swiss Federal Institute of Technology (EPFL), Switzerland
Weisi Lin, Nanyang Technological University, Singapore
Aleksandra Pižurica, Ghent University, Belgium

Reviewers

Decky Aspandi, Spain
Péter Balázs, Hungary
Sara Baldoni, Italy
Nikola Banić, Croatia
Eduardo Batista, Brazil
Federica Battisti, Italy
Stjepan Begušić, Croatia
Maurice Bellanger, France
Xavier Binefa, Spain
Mirjana Bonković, Croatia
Jelena Božek, Croatia
Robert Bregović, Croatia
Michele Brizzi, Italy
Marco Carli, Italy
Stefania Cecchi, Italy
Mujdat Cetin, USA
Paulo Correia, Portugal
Robert Cupec, Croatia
Miloš Daković, Montenegro
Simon Dobrišek, Slovenia
Milos Doroslovacki, USA
Nayee Dousai, Croatia
Anđela Draganić, Montenegro
Konstantinos Drossos, Finland
Stefan Duffner, France
Hazim Ekenel, Turkey
Pierre-Antoine Eliat, France
Matteo Ferrara, Italy
Camelia Florea, Romaia
Said Gaci, Algeria
Dubravko Gajski, Croatia
George Gregoriou, Cyprus
Mislav Grgić, Croatia
Sonja Grgić, Croatia
Karlo Griparić, Croatia
Balázs Harangi, Hungary
Edouard Ivanjko, Croatia
Jelena Ivanković, Croatia
Zoran Kalafatić, Croatia
Hrvoje Kalinić, Croatia
Ivica Kopriva, Croatia
Karlo Koščević, Croatia
Zvonko Kostanjčar, Croatia
Constantine Kotropoulos, Greece
C.-C. Jay Kuo, USA
Ana Kuzmanić Skelin, Croatia
Maja Lakičević Žarić, Montenegro
Kamal Lamichhane, Italy
Jonatan Lerga, Croatia
Tibor Lukić, Serbia
Lidija Mandić, Croatia
Tea Marasović, Croatia
Margrit Gelautz, Austria
Nenad Markuš, Croatia
Andrzej Materka, Poland
Pramit Mazumdar, Italy
Duje Medak, Croatia
Denis Milošević, Croatia
Benedek Nagy, Hungary
Antal Nagy, Hungary
Maciej Niedzwiecki, Poland
László Nyúl, Hungary
Kalman Palagyi, Hungary
Vladan Papić, Croatia
Roland Perko, Croatia
Tomislav Petković, Croatia
Davor Petrinović, Croatia
Luka Posilović, Croatia
Pavle Prentašić, Croatia
Tomislav Pribanić, Croatia
Edoardo Provenzi, Croatia
Anirudh Puligandla, Croatia
Gianni Ramponi, Italy
Slobodan Ribarić, Croatia
Alessandro Rizzi, Italy
Adria Ruiz, France
Frank Sachse, USA
Fredi Šarić, Croatia
Nicoletta Saulig, Croatia
Sinisa Šegvić, Croatia
Giovanni Sicuranza, Italy
Mile Šikić, Croatia
Ana Sović Kržić, Croatia
Vedran Stipetić, Croatia
Michal Strzelecki, Poland
Viktor Sučić, Croatia
Federico Sukno, Spain
Ioan Tabus, Finland
Azra Tafro, Croatia
Attila Tanacs, Hungary
Claus Vielhauer, Germany
Ivan Volarić, Croatia
Mladen Vučić, Croatia
Jurica Vugrin, Croatia
Nicolas Younan, USA
Ivana Zadro Matovinović, Croatia
Petar Zečević, Croatia
Frank Zoellner, Germany

MMOD-COG: A Database for Multimodal Cognitive Load Classification

Igor Mijić
RealNetworks
Zagreb, Croatia
imijic@realnetworks.com

Marko Šarlija, Davor Petrinović
University of Zagreb, Faculty of Electrical Engineering and Computing
Zagreb, Croatia
marko.sarlija@fer.hr, davor.petrinovic@fer.hr

***Abstract*—This paper presents a dataset for multimodal classification of cognitive load recorded on a sample of students. The cognitive load was induced by way of performing basic arithmetic tasks, while the multimodal aspect of the dataset comes in the form of both speech and physiological responses to those tasks. The goal of the dataset was two-fold: firstly to provide an alternative to existing cognitive load focused datasets, usually based around Stroop tasks or working memory tasks; and secondly to implement the cognitive load tasks in a way that would make the responses appropriate for both speech and physiological response analysis, ultimately making it multimodal. The paper also presents preliminary classification benchmarks, in which SVM classifiers were trained and evaluated solely on either speech or physiological signals and on combinations of the two. The multimodal nature of the classifiers may provide improvements on results on this inherently challenging machine learning problem because it provides more data about both the intra-participant and inter-participant differences in how cognitive load manifests itself in affective responses.**

I. Introduction

The load imposed on one's cognitive system when performing a particular task defines the construct of cognitive load [1]. Most practical uses of cognitive load rely heavily on its proper measurement/estimation, which is why cognitive load estimation is heavily researched in the field [1]. Because of the various manifestations of cognitive load on human biological signals, physiological measures like heart rate variability or pupillary responses are sometimes used to estimate it [2], [3]. However, speech data has recently emerged as a possible non-invasive, objective measure of cognitive load, and may prove valuable in applications in which spoken communication is already extensively used, or as an additional information source to boost the current state-of-the-art.

Most of the papers on the topic follow the same methodology, they present novel handmade voice features as extensions of basic speech analysis feature sets, use them for classification of cognitive load in a particular dataset, and finally present the improvements in classifier scores when using the presented features. In speech research, some of the first papers that used such a framework [4], [5] present Gaussian Mixture Modelling of cognitive load of reading comprehension tasks and Stroop task variations using features like Mel Frequency Cepstral Coefficients (MFCC) and basic prosodic features which were at that time popular in speech recognition tasks. Papers by [6] and [7], [8] extend these results by extending the feature sets with, spectral centroid frequency and amplitude (SCF, SCA) in [6]; cepstral peak prominence (CPP) and harmonic-to-noise ration (HNR) in [8]; and vowel formant features in [7]. More lately, a variety of machine learning methods and feature sets were compared at the [9], presenting further improvements on earlier work in the form of i-vector based classifiers [10] and Deep Rectifier Networks [11], results of which can illustrate how classification scores can be strongly influenced by the features, preprocessing methods, and classification models used. The dataset used in the challenge was another Stroop test based dataset which improves upon the basic dataset used in the cognitive load classification papers above [4], [5]. Other Stroop task based datasets used for cognitive load classification from speech/voice data include [12] and [13]. Examples of papers that didn't use Stroop based tasks include [14] and [15] in which word recall and logical deduction tasks were used, respectively. This paper focuses on presentation of a new dataset that comprises consistently elicited cognitive load responses and with the goal to increase the diversity of possible experimental procedures for cognitive load estimation. Our previous paper [16] aimed at providing preliminary insights into voice-response-based classifiers of cognitive load, but given that the data used was from a larger experiment not specifically designed for cognitive load elicitation, some experiment design features were lacking. In this paper we aimed at rectifying those features. Finally, our additional focus was aimed at researching the multimodality of responses to cognitive load, and combining the more recent speech/voice classification attempts with more established physiologically based classification systems.

The structure of this paper is as follows; firstly, an overview of the complete research method is given (including the experimental procedure, processing techniques and classification procedure), followed by the evaluation of preliminary classifiers trained with the dataset; and finally with a discussion on the multimodal aspect of the results.

II. Method

A. Participants

The participants were 40 college students, ages 22-25, 34 male and 6 female. Written informed consent was signed by the participants prior to the experimental sessions.

978-1-7281-3140-5/19/$31.00

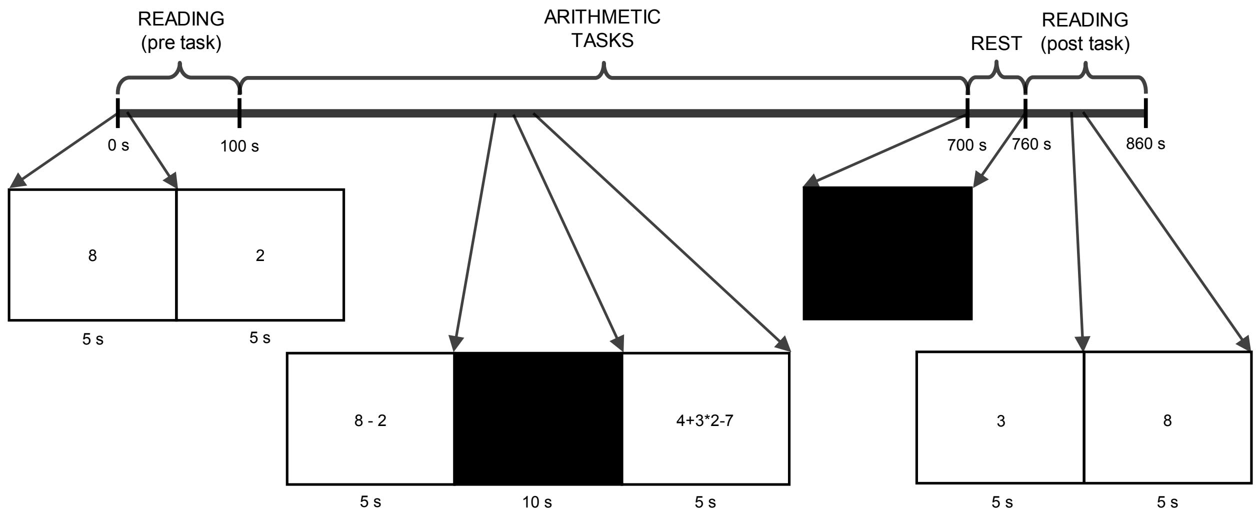

Figure 1. A visualisation of the experimental procedure

B. Recording Setup

The participants were seated in a temperature, illumination and noise controlled room, at a desk with physiological signal recording equipment and a computer monitor used for task delivery.

The BIOPAC MP150 system was used for recording of the peripheral physiology, namely the electrocardiogram (ECG) and electrodermal activity (EDA), at a sampling frequency of 1 kHz. ECG Lead I was measured by electrodes (EL503 from BIOPAC) placed on both wrists and above the right ankle. EDA was measured via two isotonic gel electrodes (EL507 from BIOPAC) placed on the index and ring finger of the participants' non-dominant hand. The signals were visually inspected and the electrodes were considered successfully placed if QRS complexes in the ECG and skin conductance responses (SCRs) in the EDA signals were easily observable.

The participants' speech was recorded with the adjustable microphone integrated in the Sennheiser PC360 headset. The speech was recorded with a sampling frequency of 20 kHz. For each of the participants, the integrated microphone of the headset was manually adjusted 1 cm below the lip line, approximately 1.5 cm away. The microphone was positioned as close to the horizontal centre of the mouth as possible with the sensor directed perpendicular to the lip line.

C. Experimental Procedure

The experimental procedure is presented in Figure 1. It contained three disjoint segments:

- **Reading segment** - a reading segment at the start of the experiment in which the participants were tasked with reading the single digit numbers appearing on the screen for 5 seconds each. The segment included 20 utterances, which comprised two utterances each of every possible single digit number (0-9). The reading tasks were randomly presented across possible digits.
- **Arithmetic segment** - a segment that included 40 basic arithmetic tasks, in which single digit operands were combined with basic operations (addition, subtraction, division and multiplication), and single digit numbers were expected as answers. Two complexities of tasks were present, 2-operand tasks for the low cognitive load condition, and 4-operand tasks for the high cognitive load condition. Within each condition, every possible digit (0-9) was the expected answer twice (10 digits * 2 conditions * 2 times = 40 tasks). The tasks were randomly presented across both digits and conditions. Each task was presented on the screen for 5 seconds after which there was a pause of 10 seconds before the next task. The pause was included to give sufficient time for the participants to give the answer, and to record the slower physiological responses following the response.
- **Rest segment** - a 60 second rest period after the arithmetic segment in which no tasks were presented and the participants were instructed to relax.
- **Reading segment** - a reading segment at the end of the experiment, with the same parameters as the reading segment at the start of the experiment.

D. Utterance Annotation

The beginnings and ends of all utterances (spoken task answers) in the recorded wave files were annotated by hand. Such an annotation strategy was preferred to speech tokenisers because of the need for the highest possible precision for the exact utterance start times, which correspond to response times for each of the tasks. Most speech tokenisers depend both on local and global signal energy thresholds which are both speaker and recording dependent, and may introduce unwanted bias/variance to the response times if they were to be used for such annotation purposes.

E. Paralinguistic Speech Features

An out of the box affective computing feature set was used for extraction of paralinguistic speech features: the Geneva Minimalistic Acoustic Parameter Set (GeMAPS) [17] which aims to provide the minimal amount of essential features for voice research and affective computing. The feature set comprises 88 parameters, including but not limited to, statistical functionals of the usual paralinguistic low-level descriptive signals like the fundamental frequency (F0), formant positions and bandwidths (F1-F4), mel-frequency cepstral coefficients (MFCC), etc. Detailed descriptions of all the parameters comprised in the feature sets can be found in [17]. All features were extracted from the segmented utterances using the openSMILE [18] feature extraction tool.

F. Peripheral Physiology Features

Raw ECG signals were processed using a highly robust and accurate automatic QRS detection algorithm [19] and a MATLAB-based semi-automatic tool for checking, correction and fine-tuning of detected heartbeats, resulting with accurately detected heartbeat times on a 1-ms precision scale. High precision of detected heartbeat times is crucial for accurate assessment of the variations in the obtained inter-beat interval (IBI) time series, known as heart rate variability (HRV). HRV is generally a direct result of autonomic regulation of the cardiovascular system, in response to various environmental and psychological challenges [20]. In the context of cognitive load assessment, we were particularly interested in HRV that reflects autonomic imbalance generated by projections from higher brain regions to medullary cardiovascular control centres. This brain-heart interaction is reflected on HRV in an array of other states, besides cognitive load, like stress, anxiety, depression, PTSD, fatigue, negative emotions etc., all of which produce autonomic imbalance. For example, mental workload was positively correlated with heart rate [21], while psychological stress produced significant reduction in HRV [22]. Raw EDA signal was downsampled to 10 Hz and preprocessed using a state-of-the-art EDA analysis algorithm cvxEDA [23]. The algorithm decomposes EDA into the tonic and phasic component, while estimating the EDA driver signal, or the sudomotor nerve activity (SMNA) signal which is interpreted as a direct sympathetic activity estimate.

1) ECG features: While an array of time-domain, frequency-domain and nonlinear measures are generally used to assess HRV [20], we extract seven short-term time-domain HRV features based on heartbeats detected in a ± 3 seconds time window before and after the answer for each utterance. These include:

- IBI range calculated as the difference between longest and shortest IBI.
- Mean IBI value.
- Standard deviation of IBI values, also known as SDNN.
- Root mean square of the successive differences (SD) in IBI timeseries (RMSSD).
- RMSSD calculated using normalised successive differences. The normalised successive differences are percentages of the average duration of 2 corresponding adjacent IBI values.
- Ratio of the mean IBI value calculated on the 3 seconds time window before the answer and the mean IBI value calculated on the 3 seconds time window after the answer.
- Difference between the mean normalised SD value calculated on the 3 seconds time window after the answer and the mean normalised SD value calculated on the 3 seconds time window before the answer.

2) EDA features: Using the decomposed EDA components we calculate six short-term time-domain EDA-based features in a ± 3 seconds time window before and after the answer for each utterance. These include:

- Summed (integrated) SMNA estimate over the entire time window.
- Maximum phasic EDA value.
- Number of skin conductance responses (SCRs) measured as the number of peaks in the raw EDA signal.
- Slope of the tonic EDA component.
- Slope of the raw EDA signal in the 3 seconds time window before the answer.
- Slope of the raw EDA signal in the 3 seconds time window after the answer.

G. Data Exclusion Criteria

Two participants were excluded from all analyses based on aspects of their performance on the basic arithmetic tasks:

- One participant had a response rate 3 standard deviations lower than the sample average (45 % response rate, while the second lowest was 95%)
- One participant had an average response time for the low cognitive load condition which was 3 standard deviations higher than the sample average

Six participants were excluded from peripheral physiology analyses based on specific modality criteria and recording conditions, specifically, 6 participants were labeled as EDA non-responders due to no observable phasic changes in EDA, and 1 participant had a faulty ECG recording. These participants were not excluded from speech-based or response time analyses.

H. Response Times and Task Complexity Validation

The response times were calculated as the difference between the timepoint of task presentation and the start of the task answer utterance. As stated in II-D, the starting timepoints of answer utterances were all hand labelled as to get the most precision out of the response times. Response times are usually used to compare the complexity of cognitive load tasks, e.g. in Stroop tasks, incongruent Stroop tasks (higher complexity) will have higher response times. To test for differences in response times between the experimental conditions of a repeated measure experimental design like this one, we've averaged the response times across all subjects and conditions, and subsequently t-tested them between all categories.

I. Participant-based Feature Scaling

Given that affective representations are inherently participant dependent, both in the case of speech and physiological responses, participant-based feature scaling was applied [24]. The two techniques applied were:

- Participant based standardisation - the feature mean was subtracted and the resulting values were divided by the feature standard deviation for each feature and participant independently. This technique is most suitable for data with distributions close to normal.
- Participant based robust normalisation - the feature median was subtracted and the resulting values were divided by the feature interquartile range for each feature and participant independently. This technique is most suitable for non-normally distributed data

J. Cognitive Load Classification

Support vector machine (SVM) classifiers were trained for each of the feature sets available, namely, the heart rate variability features (HRV), skin conductance features (EDA) and paralinguistic speech features (GeMAPS). Also, the HRV and EDA feature sets were combined into a joint peripheral physiology feature set (PHYS), and a multimodal feature set combining the PHYS features with GeMAPS features was evaluated (MMOD). The hyper-parameters tuned for the SVM classifier included C (the penalty for the error) and gamma (the kernel coefficient).

The classifiers were trained using nested leave-one-out subject cross-validation (LOSO), which incorporates both hyper-parameter tuning and model evaluation. The procedure was as follows:

1) Divide the data into N folds, where N is the number of participants, and each fold comprises utterances/responses by a single participant
2) Reserve one of the folds for test
3) Reserve one of the N-1 remaining training folds for validation
4) For all the possible combinations of the hyper-parameters, train on the N-2 remaining training folds and evaluate on the validation fold
5) Repeat the last step N-1 times, once for each of the training folds as the validation fold
6) Choose the hyper-parameter combination that minimises the average training error over the N-1 folds. Use that combination to evaluate on the test set.
7) Repeat N times from step 2, once for each of the original N folds in turn as the test fold
8) Report the mean and standard deviation of the evaluation measure over the N test folds

The classifiers were trained and evaluated for two binary classification cases:

1) (Low vs. High) The low complexity arithmetic tasks were classified vs. the high complexity arithmetic tasks
2) (Pre vs. High) The low complexity arithmetic tasks were substituted with the pre-artihmetic task reading responses classified vs. the high complexity arithmetic tasks, given that reading is also low cognitive load task.

Table I
RESPONSE TIME AVERAGES FOR THE 4 EXPERIMENTAL CONDITIONS.

	Reading (Pre)	Reading (Post)	Arithmetic (Low)	Arithmetic (High)
Response time [s]	0.504	0.544	0.931	3.301

III. RESULTS

A. Response Times and Task Complexity

A boxplot/swarmplot of all the utterance response times, grouped by the 4 experimental conditions is shown in Figure 2. The average response times for the conditions are presented in Table I. The t-tests between the samples of averaged participant response times have all shown significant differences at a p<0.01, except between the pre-task reading and post-task reading response times signifying that the arithmetic tasks differ in complexity from the reading tasks, but not between themselves. However, the relation between the task complexity, when viewed through the response times isn't linear, with the low complexity arithmetic task being significantly closer to the reading tasks than the high complexity arithmetic task. This enables us to use the reading utterances as a possible alternative to the low complexity arithmetic tasks, in a low-vs-high cognitive load scenario. This is also what deterred us from building a 3-level cognitive load classifier that included all three experimental conditions as possible low, medium and high cognitive load labels.

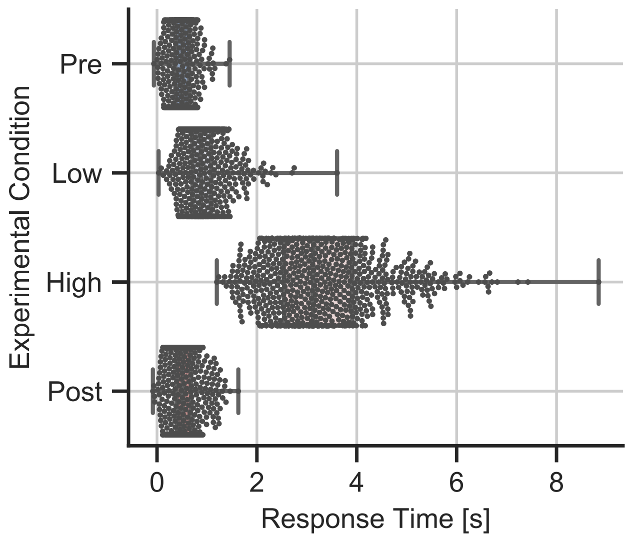

Figure 2. Response times for all four experimental conditions.

B. Classifier Evaluation

The evaluation scores of all the learned classifiers are presented in Table II. Given that the classifiers performed binary classification and that the dataset was balanced across conditions, we have chosen simple classifier accuracy as the score metric. The reported scores comprise mean and standard deviation values of all the classifiers learned/evaluated in the nested LOSO cross-validation scheme across all of the conditions described in II-J. Given that in each of the folds of the nested LOSO cross-validation scheme the classifiers were tuned and tested on disparate participants, the resulting scores enable us to evaluate both the general suitability of a feature set for cognitive load classification purposes (from the mean classifier scores) and it's behaviour in the context of inter-participant variation regarding cognitive load responses. The participant count imbalance between feature sets (described in II-G) is also addressed by such score reporting. Additionally, we can also observe the effects of the two normalisation strategies used which were included because of the different feature distributions within the feature sets.

From Table II we can observe that when it comes to single modality feature sets, the EDA feature set is superior both to HRV features and GeMAPS speech features, for both the Pre vs. High and Low vs. High evaluations, with classification scores in the low 70-ies. There is variation between HRV and GeMAPS features, with the GeMAPS speech features having the lowest score in the Low vs. High evaluations, while outperforming the HRV in the Pre vs. High evaluations.

Classifier score gains are generally obtained when combining the individual modality feature sets together, with the highest score on the full multimodal MMOD feature set, over the Pre vs. High evaluation, and the highest score in the Low vs. High evaluation for the full physiology feature set PHYS. Scores are mostly higher for the Pre vs. High evaluations, across all of the feature sets except the HRV feature set, which is congruent with reading being the less complex of the two tasks (check section III-A), and in turn more easily differentiable. The score gains for the MMOD Pre vs. High evaluation, which are not present for the Low vs. High evaluations (PHYS feature set exhibits superior results) may be due to effects of the added GeMAPS features, which seem to be more informational in the Pre vs. High evaluation.

The standard scaling consistently outperformed the robust scaling even though the features included in the datasets are both normally and non-normally distributed, especially in the case of speech features. When looking at the standard deviations of the nested LOSO classifier scores, and given that the standard deviations of the cross-validated classifiers are measures of how well the classifier has generalised over ”never-before-seen” participants, careful consideration should be taken when analysing these results. The lowest standard deviation can be seen in the case of the GeMAPS speech classifier for the Low vs. High evaluation, which is also the worst in general performance. Coincidentally, the second lowest variation can be seen in the multimodal MMOD classifier for the Pre vs. High evaluation, which is not only the highest performing classifier, but also seems to be performing most consistently, and is score-wise in the mid 70-ies. To visualise the differences in the score distributions of the evaluated classifiers, Figure 3 includes distributions of the fold scores for the multimodal MMOD feature set, standard scaled, nested LOSO fold scores for both the Pre vs. High and Low vs. High evaluations (scores in the second to last line of Table II). It can be seen in the figure that the Low vs. High classifier has a substantially higher number of LOSO participant test folds for which it was performing nearly at a random level.

Direct comparison to the state-of-the-art in similar research is hindered by the varying experiments and many evaluation/reporting strategies, e.g., the results in [10] reported a myriad of recall scores varying between the low 60-ies and high 70-ies, on a wide variety of cognitive load tasks, speech feature sets and normalisation strategies, which is why direct comparisons were not included in this paper.

Table II
EVALUATION SCORES FOR THE TRAINED SVM CLASSIFIERS

		Pre vs. High		Low vs. High	
Feature set	Normalisation	mean	std	mean	std
HRV	Standard	63.22	13.82	64.44	13.28
	Robust	62.60	13.37	63.90	13.48
EDA	Standard	72.50	14.16	70.88	11.58
	Robust	71.94	14.16	71.81	11.66
GeMAPS	Standard	67.92	10.36	58.17	8.87
	Robust	65.59	10.51	54.54	8.08
PHYS	Standard	75.25	13.29	**73.62**	**12.37**
	Robust	75.01	12.54	73.13	12.38
MMOD	Standard	**76.66**	**8.74**	71.75	10.65
	Robust	74.47	9.38	70.88	12.12

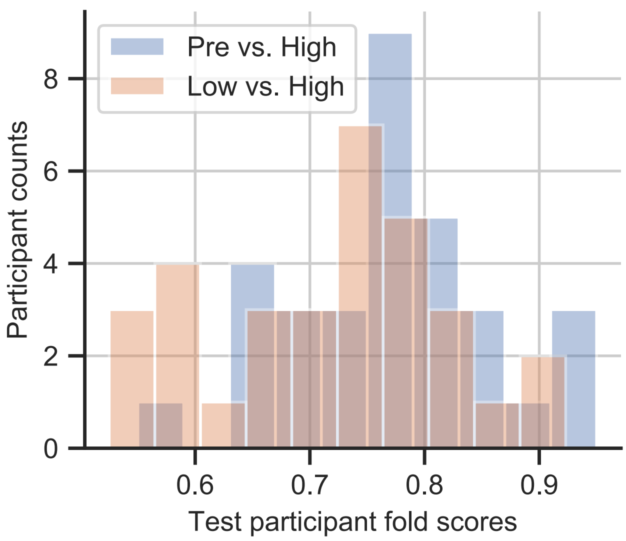

Figure 3. Distributions of the LOSO participant test fold scores for the Pre vs. High and Low vs. High classifier evaluations.

IV. Discussion

The data in this database presents a viable datasource for research of multimodal cognitive load responses. The randomised structure, stratified number of responses across conditions and the appropriate pauses after the task presentation are all improvements over earlier findings in [16]. However, future work may focus on extending the amount of cognitive load levels with additional task complexities or further analysing the existing levels (low complexity arithmetic tasks vs. reading tasks). The dataset is available from the corresponding author on request.

The classifiers built on all the available modalities, and their respective combinations reveal that evaluation scores benefit from classifier input multimodality. In each of the modality cases (especially HRV and speech), future work should base itself around engineering more suitable features for this specific use-case, which should improve the scores presented in the previous section. In the case of speech features, a modality-specific classifier should also be used (like i-vectors) alongside with the handmade features which should also boost their performance. Lower standard deviation of classifier nested LOSO scores, which was obtained by some models that included voice data should also be a design goal for future classifiers.

Acknowledgment

This research is sponsored by NATO's Emerging Security Challenges Division in the framework of the Science for Peace and Security Programme. It is partly supported by: Croatian Science Foundation under the project number IP-2014-09-2625; DATACROSS project under number KK.01.1.1.01.009. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Croatian Science Foundation.

References

[1] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. Van Gerven, "Cognitive load measurement as a means to advance cognitive load theory," *Educational psychologist*, vol. 38, no. 1, pp. 63–71, 2003.

[2] F. G. Paas and J. J. Van Merriënboer, "Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach." *Journal of educational psychology*, vol. 86, p. 122, 1994.

[3] P. Van Gerven, F. Pass, J. Van Merrienboer, and H. Schmidt, "Memory load and task-evoked papillary responses in aging," *Manuscript submitted for publication*, 2002.

[4] B. Yin, F. Chen, N. Ruiz, and E. Ambikairajah, "Speech-based cognitive load monitoring system," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 2041–2044, https://doi.org/10.1109/ICASSP.2008.4518041.

[5] B. Yin, N. Ruiz, F. Chen, and M. A. Khawaja, "Automatic cognitive load detection from speech features," in *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*. ACM, 2007, pp. 249–255, https://doi.org/10.1145/1324892.1324946.

[6] P. N. Le, E. Ambikairajah, J. Epps, V. Sethu, and E. H. Choi, "Investigation of spectral centroid features for cognitive load classification," *Speech Communication*, vol. 53, no. 4, pp. 540–551, 2011, https://doi.org/10.1016/j.specom.2011.01.005.

[7] T. F. Yap, J. Epps, E. Ambikairajah, and E. H. Choi, "Formant frequencies under cognitive load: Effects and classification," *EURASIP journal on advances in signal processing*, vol. 2011, no. 1, p. 219253, 2011, https://doi.org/10.1155/2011/219253.

[8] ——, "Voice source features for cognitive load classification," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5700–5703, https://doi.org/10.1109/ICASSP.2011.5947654.

[9] B. Schuller, S. Steidl, A. Batliner, J. Epps, F. Eyben, F. Ringeval, E. Marchi, and Y. Zhang, "The interspeech 2014 computational paralinguistics challenge: Cognitive & physical load," in *Proceedings of the 15th Annual Conference of the International Speech Communication Association*, 2014, pp. 427–431.

[10] M. V. Segbroeck, R. Travadi, C. Vaz, J. Kim, M. P. Black, A. Potamianos, and S. S. Narayanan, "Classification of cognitive load from speech using an i-vector framework," in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014, https://doi.org/10.1.1.728.2977.

[11] G. Gosztolya, T. Grósz, R. Busa-Fekete, and L. Tóth, "Detecting the intensity of cognitive and physical load using adaboost and deep rectifier neural networks," in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[12] E. H. Magnusdottir, M. Borsky, M. Meier, K. Johannsdottir, and J. Gudnason, "Monitoring cognitive workload using vocal tract and voice source features," *Periodica Polytechnica Electrical Engineering and Computer Science*, vol. 61, no. 4, pp. 297–304, 2017, https://doi.org/10.3311/PPee.10414.

[13] M. K. MacPherson, D. Abur, and C. E. Stepp, "Acoustic measures of voice and physiologic measures of autonomic arousal during speech as a function of cognitive load," *Journal of Voice*, vol. 31, no. 4, pp. 504–e1, 2017, https://doi.org/10.1016/j.jvoice.2016.10.021.

[14] S. Boyer, P.-V. Paubel, R. Ruiz, R. El Yagoubi, and A. Daurat, "Human voice as a measure of mental load level," *Journal of Speech, Language, and Hearing Research*, pp. 1–13, 2018, https://doi.org/10.1044/2018-JSLHR-S-18-0066.

[15] K. R. Scherer, D. Grandjean, T. Johnstone, G. Klasmeyer, and T. Bänziger, "Acoustic correlates of task load and stress," in *Seventh International Conference on Spoken Language Processing*, 2002.

[16] I. Mijić, M. Šarlija, and D. Petrinović, "Classification of cognitive load using voice features: A preliminary investigation," in *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 2017, pp. 000 345–000 350, https://doi.org/10.1109/CogInfoCom.2017.8268268.

[17] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan *et al.*, "The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190–202, 2015, https://doi.org/10.1109/TAFFC.2015.2457417.

[18] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 2013, pp. 835–838, https://doi.org/110.1.1.446.1445.

[19] M. Šarlija, F. Jurišić, and S. Popović, "A convolutional neural network based approach to qrs detection," in *Image and Signal Processing and Analysis (ISPA), 2017 10th International Symposium on*. IEEE, 2017, pp. 121–125, https://doi.org/10.1109/ISPA.2017.8073581.

[20] F. Shaffer and J. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 5, p. 258, 2017, https://doi.org/10.1109/CogInfoCom.2017.8268268.

[21] S. Magnusson and P. Berggren, "Dynamic assessment of pilot mental status," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 46, no. 24. SAGE Publications Sage CA: Los Angeles, CA, 2002, pp. 1997–2001, https://doi.org/10.3389/fpubh.2017.00258.

[22] J. Delaney and D. Brodie, "Effects of short-term psychological stress on the time and frequency domains of heart-rate variability," *Perceptual and motor skills*, vol. 91, no. 2, pp. 515–524, 2000, https://doi.org/10.2466/pms.2000.91.2.515.

[23] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxeda: A convex optimization approach to electrodermal activity processing," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 797–804, 2015, https://doi.org/10.1109/TBME.2015.2474131.

[24] C. Busso, S. Mariooryad, A. Metallinou, and S. Narayanan, "Iterative feature normalization scheme for automatic emotion detection from speech," *IEEE transactions on Affective computing*, vol. 4, no. 4, pp. 386–397, 2013, https://doi.org/10.1109/T-AFFC.2013.26.

Author Index

Organizing Institution

University of Zagreb
Faculty of Electrical Engineering and Computing
Croatiay

Auspices

Croatian Academy of Sciences and Arts

Technical Co-Sponsors

European Association for Signal Processing (EURASIP)
IEEE Signal Processing Society
IEEE Croatia Sectiony
IEEE Croatia Section Signal Processing Chapter

IEEE Catalog Number CFP19504-ART (online)
IEEE Catalog Number CFP19504-USB (USB)

ISBN 978-1-7281-3140-5 (online)
ISBN 978-1-7281-3139-9 (USB)

ISSN 1849-2266 (online)

USB
version