Assessing Cognitive Workload Using Cardiovascular Measures and Voice

Subjects: Computer Science, Interdisciplinary Applications

Contributor: Eydis H. Magnusdottir , Kamilla R. Johannsdottir , Arnab Majumdar , Jon Gudnason

Monitoring cognitive workload has the potential to improve both the performance and fidelity of human decision making. Cognitive workload can be assessed and monitored using cardiovascular measures and voice. Heart rate and blood pressure signals are combined with voice signal formant features to classify cognitive workload into three levels of difficulty. The feature level fusion combination is carried out using each heartbeat to synchronize the two data sources. This allows the subsequent machine learning mechanism to monitor cognitive workload on a beat-by-beat basis.

cognitive workload cardiovascular signals speech processing emotion recognition

1. Introduction

The cognitive workload of personnel in a workforce, especially those involved in safety critical industries, e.g., airline pilots and air-traffic controllers, is crucial to ensuring both the well being and productivity of the personnel and the broader safety of the public. Consideration of workload and its management is therefore a crucial aspect of any safety management system in a safety critical organisation. Assessing and monitoring cognitive workload is, therefore, of great importance. While there are numerous methods by which to measure workload, e.g., subjective methods, psychophysiological approaches to cognitive workload monitoring, that use signals such as cardiovascular measures and electroencephalography (EEG), have recently shown promise in identifying cognitive workload in laboratory settings [1][2][3][4][5][6]. Increased mental demand is highly correlated with increased cardiovascular reactivity [7][8]. Various classifier methods have also been used to recognize different cognitive workload states (reliably high and low workload) based on combined psychophysiological signal sources [9][10][11]. Despite this, going beyond a binary high/low workload classification has proved to be difficult. The problem might at least partly be methodological. Many current approaches to monitoring cognitive workload fail to consider the individual variation in responses to workload and this shortcoming has been highlighted in the literature [12], though attempts have been made by the authors to address this issue specifically for cardiovascular measures in [13][14]. Furthermore, prior work has not taken into consideration combining the cardiovascular signal with another promising signal source, the individual's speech even though studies suggest that the speech signal may be a good indication of the individual's mental state (e.g., see [15][16]).

1.1. Cardiovascular Measures and Speech

Cardiovascular measures are relatively unobtrusive and well suited for the aviation environment where speech communications is used to solve tasks. Furthermore, with technological enhancements such as wearable devices, these measures are now becoming even less intrusive than before. Short term fluctuations in task demand affects the cardiovascular system [7][8] and these responses can be objectively identified by monitoring the cardiac muscle or the vascular system by observing, for example, the heart rate (HR) and the blood pressure and stroke volume [17][18].

An alternative way of monitoring cognitive workload is through the speech signal. Whilst not applicable in all environments, it may be ideal in situations where speech can be captured and communications can be monitored in real-time without interruptions. Yin et al. [15] performed a trinary classification task using Mel-frequency cepstrum coefficients and prosodic features with a speaker adapted Gaussian mixture model. This feature extraction method was extended [19] to include targeted extraction of vocal tract features through spectral centroid frequency and amplitude features. In both instances, a relatively small data set was used for validation where each participant performed reading and Stroop tasks. These feature extraction and classification schemes indicated a strong relationship between cognitive workload and the speech signal within the experimental framework of the studies.

Cognitive workload experiments using speech have been carried out for real-life tasks in military flight simulation [20]. This approach has the advantage of being closer to real operational situations indicating that the technology is suitable for aviation applications. Mean change in fundamental frequency and speech intensity was used to detect cognitive workload of the participants, but more detailed speech analysis was not performed. Speech analysis has also been used for related tasks of affective speech classification. For example, the Interspeech 2014 Cognitive load challenge (Computational Paralinguistics ChallengE, ComParE) [21] was based on the same principle. A data set of 26 participants provided speech recordings and EEG signals during a reading task and a low-, medium-, and high cognitive load level Stroop tasks. The winning entry used an i-vector classification scheme based on a combined feature set of fused speech streams, prosody and phone-rate [22].

Combining different physiological signals may provide a more prominent, detailed cognitive workload monitoring tool [23]. Most commonly, studies focus on cardiovascular signals combined with electrical brain activity signals either as a pair [24][25] or grouped with signals such as galvanic skin response [26] or oculomotor measures [9][27]. No attempts, to researchers' knowledge, have been reported investigating the supplemental possibility of cardiovascular and speech signals for cognitive workload monitoring.

1.2. Related Work on Cognitive Workload Classification

Various pattern recognition methods have been applied to the task of classifying cognitive workload using psychophysiological measures. It has been pointed out that artificial neural networks are opaque and hard to interpret in terms of how individual variables interact to predict workload [11]. Classification methods have been used such as discriminant analysis and support vector machines [28], as well as logistic regression and classification trees [11]. There is no indication that other classification methods can provide better results for cognitive workload monitoring [11][28]. Recently, however, trinary cognitive workload level classification with cardiovascular signals has been demonstrated with promising results [29].

Few studies have used artificial neural networks to classify cognitive workload states in the field of air-traffic control using combined physiological signals [11][30][31][32][33]. In particular, multiple psychophysiological measures were combined to provide high accuracy in classifying at least a limited number of cognitive load states [31][33]. High binary classification accuracy was achieved for high and low workload states in air-traffic control using neural networks based on EEG and electrocardiography (ECG) signals [33]. However, when the training scenario included four and seven different cognitive load states based both the on complexity and the number of aircraft, the classifier confused adjacent states and was unable to distinguishing between low and medium or medium and high states. A neural network model based on multiple EEG channels, HR, and eye-blink measures produced reliable discrimination between low and high workload and was also able to distinguish between two out of three load-tasks [32] and neural networks also performed well in distinguishing high and low workload particularly at small time intervals [11]. It was pointed out however, that whilst promising, EEG is both complex to use and not easily portable [34]. This work presented a neural network trained on various cardiovascular measures along with performance-based measures and did not manage to reliably classify different cognitive workload states. Other examples of multi-modal fusion for cognitive workload assessments can be found in [35][36].

1.3. Challenges in Assessing Cognitive Workload

The main challenge of assessing cognitive workload is the latent nature of the variable in question. A close proxy of cognitive workload is the task difficulty which is typically used when assessing cognitive workload. Albeit close, the relationship between task difficulty and cognitive workload is complex depending on issues spanning from the nature of the task to the condition of the individual being assessed. Tasks can rely on one or more senses (e.g., sight and hearing) and require one or more motor skills (e.g., touch and voice) and be simple or complex in space and/or time. The condition of the individual brings other variables such as ability and fatigue into the equation.

It has also been known for quite some time now that individuals show different psychophysiological responses to cognitive workload. Measured voice parameters, for example, were found to be different between individuals with respect to workload as far back as 1968 [37]. The matter of individual differences has been noted periodically with Ruiz et al. [38], for example, claiming that more than a single voice parameter needed to be measured as an indication of workload and Grassmann et al. [12] found that integrating individual differences may reduce unexpected variance in workload assessment. Moreover, research has shown that individual working memory capacity may play a critical role in determining how individuals react to changes in cognitive workload [14].

Cognitive workload is also perceived to be a continuous variable although its effect on the individual might be categorical (i.e., fight-or-flight vs. rest-and-digest). Researchers have, however, struggled with this assessment and many have reduced the problem of cognitive workload monitoring to a binary classification of high or low workload [31][33]. How cognitive workload assessment is developed beyond this dichotomy remains an open research question.

Measuring and combining many different psychophysiological measures also presents a set of challenges that researchers have grappled with [39]. Cognitive workload presents differently through the systems being measured (e.g., heart-rate, speech or the brain's electrical activity) so making more than one of these data sources available for the assessment should make the monitoring more robust and accurate. The most straightforward method of combining feature sets from different sources would be simply to concatenate them. There are, however, a few issues that need to be addressed before a concatenated feature set can be successfully used as a pattern recognition classification input. The sampling rates of the two or more signals might not be the same hence some sort of resampling and alignment needs to take place. Alignment in time has to be ensured during data recordings as well as their correspondence after individual feature extraction is concluded.

2. Cognitive Workload Experiments

The cognitive workload experiments are set up so that the relationship between task difficulty and workload is close and with three difficulty levels to reflect the non-binary approach taken. The amount of data obtained from each individual should be sufficient in order to model each participant's response to cognitive workload separately. The data collected are cardiovascular data and voice recordings, analysed separately and combined in a trinary classification of cognitive workload.

2.1. Experiment Design and Configuration of Tasks

Figure 1 depicts a chart of the progress of tasks, instructions, self-assessment questionnaires and resting periods implemented in the experiment. The flow chart depicts all elements included in the experiment such as the OSPAN test (see [40]) and reading task, but the focus in this particular entry is on the Stroop tasks.

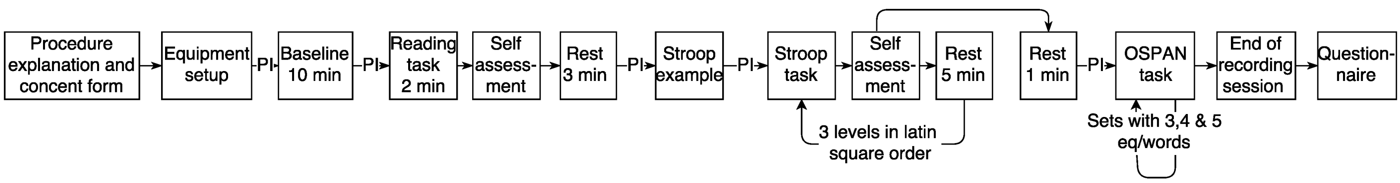

Figure 1. Chart of the flow of tasks and resting periods for the whole duration of one experiment. Progress Instructions PI depicts instructions given to the participants in between tasks to be read out loud.

Cognitive workload levels was introduced through the well-established cognitive stimuli word/color task published by Stroop in 1935 [41]. Throughout the Stroop task a set of either incongruent (e.g., red in blue color) or congruent (e.g., red in red color) color names appears in front of the participant. In this specific setup the Icelandic color names Blue, Green, Brown, Red and Pink were used with the last color name of each set always being Black, with 36 (35 + 1) words appearing in a 6 × 6 matrix on each screen. This design was included to indicate to the researcher, controlling the flow of the screens, that the participant had finished the current screen *j*. The participant's assignment was to say the colors of the words aloud but not to read the words (of the colors). Three cognitive workload levels were induced with the settings of congruence, incongruence and time limits as follows:

- Level 1—Seven congruent sets of screens with all 36 color names appearing on the computer screen at the same time.
- Level 2—Six sets of screens with alternating incongruence levels of 0.3 and 0.7 with all 36 color names appearing on the computer screen at the same time.
- Level 3—Eight sets of screens with one word appearing at a time at randomly timed intervals of 0.75 s and 0.65 s. Here, the same incongruence set-up was applied as in Level 2 and the same number (36) of color names as in Level 1 and 2.

The number of screens in each cognitive workload level were chosen in advance to ensure approximately the same duration of levels. The cognitive workload levels were introduced in six different orders to the participants using the Latin square technique.

The participants were introduced to the task by having them read detailed instructions, appearing on the computer screen, aloud. As depicted in **Figure 1**, each of the three Stroop sessions contained the cognitive workload task, self-assessment questionnaire and resting period, repeated three times for each level with the total number of $J_p=21$ screens, for each participant p and the screen index $j=\{1,2,\cdots,J_p\}$.

The different resting periods and their strategic positions are shown in **Figure 1**. These periods were designed to ensure that the participant had sufficient time to recover between tasks and reduce its influence on succeeding tasks.

Participants in the experiment performed on the operation span task (OSPAN). The OSPAN task is a working memory task that measures the working memory span by having participants solve simple equations and remembering a word at the same time [42]. In this task, participants read out loud an equation (e.g., is (8×3)+2=25) and answer whether the equation is correct or incorrect. The equation is followed by a word (e.g., car) that also is read out loud. There are 12 sets of 3 × 2 words/eq, 3 × 3 words/eq, 3 × 4 words/eq and 3 × 5 words/eq in total. The participants' task is to remember the presented words in the correct order for each set. The total number of words to be remembered is 42; hence, the participants could get a maximum score of 42 and a minimum score of 0. One point was given for a correct word in the correct order and higher scores indicate higher working memory capacity. The results for the OSPAN task were not used in the current entry.

2.2. Two Cohorts

The method was developed on two sets of participants: volunteers who visited the laboratory of Reykjavik University (university cohort) and pilots from a commercial airline, Icelandair, that had just completed a simulation exercise at TRU Flight Training Iceland (pilot cohort). The university cohort had a total number of $P_1=97$ participants with average age of 25.2 ± 5.78 and a gender ratio of 27.83% male to 72.17% females. The pilot cohort had a total number of $P_2=20$ participants with average age of 41.35 ± 9.36 and a gender ratio of 90% male to 10% females.

References

1. Besson, P.; Dousset, E.; Bourdin, C.; Bringoux, L.; Marqueste, T.; Mestre, D.R.; Vercher, J.L. Bayesian Network classifiers inferring workload from physiological features: Compared performance. In Proceedings of the 2012 IEEE Intelligent Vehicles Symposium, Madrid, Spain, 3–7 June 2012; pp. 282–287.
2. Aricò, P.; Borghini, G.; Di Flumeri, G.; Colosimo, A.; Bonelli, S.; Golfetti, A.; Pozzi, S.; Imbert, J.P.; Granger, G.; Benhacene, R. Adaptive automation triggered by EEG-based mental workload index: A passive brain-computer interface application in realistic air traffic control environment. Front. Hum. Neurosci. 2016, 10, 539.
3. Blanco, J.A.; Johnson, M.K.; Jaquess, K.J.; Oh, H.; Lo, L.C.; Gentili, R.J.; Hatfield, B.D. Quantifying cognitive workload in simulated flight using passive, dry EEG measurements. IEEE Trans. Cogn. Dev. Syst. 2016, 10, 373–383.
4. Wu, E.Q.; Zhu, L.M.; Li, G.J.; Li, H.J.; Tang, Z.; Hu, R.; Zhou, G.R. Nonparametric Hierarchical Hidden semi-Markov model for brain fatigue behavior detection of Pilots during flight. IEEE Trans. Intell. Transp. Syst. 2021, 23, 5245–5256.
5. Byrne, E.A.; Parasuraman, R. Psychophysiology and adaptive automation. Biol. Psychol. 1996, 42, 249–268.
6. Wilson, G.F. An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. Int. J. Aviat. Psychol. 2002, 12, 3–18.
7. Mehler, B.; Reimer, B.; Zec, M. Defining workload in the context of driver state detection and HMI evaluation. In Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Portsmouth, NH, USA, 17–19 October 2012; pp. 187–191.
8. Scerbo, M.W. Stress, workload, and boredom in vigilance: A problem and an answer. In Stress, Workload, and Fatigue; Hancock, P.A., Desmond, P.A., Eds.; Lawrence Erlbaum Associates Publishers: Mahwah, NJ, USA, 2001; pp. 267–278.
9. Ryu, K.; Myung, R. Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. Int. J. Ind. Ergon. 2005, 35, 991–1009.
10. Deng, L.; Li, X. Machine Learning Paradigms for Speech Recognition: An Overview. IEEE Trans. Audio Speech Lang. Process. 2013, 21, 1060–1089.
11. Fong, A.; Sibley, C.; Cole, A.; Baldwin, C.; Coyne, J. A comparison of artificial neural networks, logistic regressions, and classification trees for modeling mental workload in real-time. In

Proceedings of the Human Factors and Ergonomics Society Annual Meeting; SAGE Publications: Los Angeles, CA, USA, 2010; Volume 54, pp. 1709–1712.

12. Grassmann, M.; Vlemincx, E.; von Leupoldt, A.; Van den Bergh, O. Individual differences in cardiorespiratory measures of mental workload: An investigation of negative affectivity and cognitive avoidant coping in pilot candidates. Appl. Ergon. 2017, 59, 274–282.
13. Magnusdottir, E.H.; Johannsdottir, K.R.; Bean, C.; Olafsson, B.; Gudnason, J. Cognitive workload classification using cardiovascular measures and dynamic features. In Proceedings of the 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Debrecen, Hungary, 11–14 September 2017; pp. 000351–000356.
14. Johannsdottir, K.R.; Magnusdottir, E.H.; Sigurjonsdóttir, S.; Gudnason, J. Cardiovascular monitoring of cognitive workload: Exploring the role of individuals' working memory capacity. Biol. Psychol. 2018, 132, 154–163.
15. Yin, B.; Chen, F.; Ruiz, N.; Ambikairajah, E. Speech-based cognitive load monitoring system. In Proceedings of the Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 2041–2044.
16. Meier, M.; Borsky, M.; Magnusdottir, E.H.; Johannsdottir, K.R.; Gudnason, J. Vocal tract and voice source features for monitoring cognitive workload. In Proceedings of the 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Wroclaw, Poland, 16–18 October 2016; pp. 000097–000102.
17. Mehler, B.; Reimer, B.; Wang, Y. A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Iowa City, IA, USA, 27–30 June 2011; pp. 590–597.
18. Malik, M. Heart rate variability. Eur. Heart J. 1996, 17, 354–381.
19. Le, P.N.; Ambikairajah, E.; Epps, J.; Sethu, V.; Choi, E.H.C. Investigation of spectral centroid features for cognitive load classification. Speech Commun. 2011, 53, 540–551.
20. Huttunen, K.; Keränen, H.; Väyrynen, E.; Pääkkönen, R.; Leino, T. Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. Appl. Ergon. 2011, 42, 348–357.
21. Schuller, B.; Steidl, S.; Batliner, A.; Epps, J.; Eyben, F.; Ringeval, F.; Marchi, E.; Zhang, Y. The INTERSPEECH 2014 computational paralinguistics challenge: Cognitive & physical load. In Proceedings of the INTERSPEECH, 15th Annual Conference of the International Speech Communication Association, Singapore, 14–18 September 2014; pp. 427–431.
22. Van Segbroeck, M.; Travadi, R.; Vaz, C.; Kim, J.; Black, M.P.; Potamianos, A.; Narayanan, S.S. Classification of cognitive load from speech using an i-vector framework. In Proceedings of the

INTERSPEECH, 15th Annual Conference of the International Speech Communication Association, Singapore, 14–18 September 2014; pp. 751–755.

23. Chen, F.; Zhou, J.; Wang, Y.; Yu, K.; Arshad, S.Z.; Khawaji, A.; Conway, D. Robust Multimodal Cognitive Load Measurement; Human–Computer Interaction Series; Springer International Publishing: Berlin/Heidelberg, Germany, 2016.
24. Stikic, M.; Berka, C.; Levendowski, D.J.; Rubio, R.F.; Tan, V.; Korszen, S.; Barba, D.; Wurzer, D. Modeling temporal sequences of cognitive state changes based on a combination of EEG-engagement, EEG-workload, and heart rate metrics. Front. Neurosci. 2014, 8, 342.
25. Borghini, G.; Astolfi, L.; Vecchiato, G.; Mattia, D.; Babiloni, F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci. Biobehav. Rev. 2014, 44, 58–75.
26. Zhang, H.; Zhu, Y.; Maniyeri, J.; Guan, C. Detection of variations in cognitive workload using multi-modality physiological sensors and a large margin unbiased regression machine. In Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 2985–2988.
27. De Rivecourt, M.; Kuperus, M.N.; Post, W.J.; Mulder, L.J.M. Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. Ergonomics 2008, 51, 1295–1319.
28. Russell, C.A.; Wilson, G.F. Feature saliency analysis for operator state estimation. In Proceedings of the 11th International Conference on Human-Computer Interaction. Foundations of Augmented Cognition, Las Vegas, NV, USA, 22–27 July 2005; Volume 11.
29. Magnusdottir, E.H.; Borsky, M.; Meier, M.; Johannsdottir, K.; Gudnason, J. Monitoring Cognitive Workload Using Vocal Tract and Voice Source Features. Period. Polytech. Electr. Eng. Comput. Sci. 2017, 61, 297–304.
30. Hope, R.M.; Wang, Z.; Wang, Z.; Ji, Q.; Gray, W.D. Workload classification across subjects using EEG. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting; Sage Publications: Thousand Oaks, CA, USA, 2011; Volume 55, pp. 202–206.
31. Wilson, G.F. Real-Time Adaptive Aiding Using Psychophysiological Operator State Assessment; Ashgate Publishing Company: Farnham, UK, 2001.
32. Wilson, G.F.; Estepp, J.; Davis, I. A comparison of performance and psychophysiological classification of complex task performance. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting; SAGE Publications: Los Angeles, CA, USA, 2009; Volume 53, pp. 141–145.
33. Wilson, G.F.; Russell, C.A. Operator Functional State Classification Using Multiple Psychophysiological Features in an Air Traffic Control Task. Hum. Factors J. Hum. Factors Ergon.

Soc. 2003, 45, 381–389.

34. Kaber, D.B.; Perry, C.M.; Segall, N.; Sheik-Nainar, M.A. Workload state classification with automation during simulated air traffic control. Int. J. Aviat. Psychol. 2007, 17, 371–390.
35. Mijić, I.; Šarlija, M.; Petrinović, D. MMOD-COG: A database for multimodal cognitive load classification. In Proceedings of the 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), Dubrovnik, Croatia, 23–25 September 2019; pp. 15–20.
36. Debie, E.; Rojas, R.F.; Fidock, J.; Barlow, M.; Kasmarik, K.; Anavatti, S.; Garratt, M.; Abbass, H.A. Multimodal fusion for objective assessment of cognitive workload: A review. IEEE Trans. Cybern. 2019, 51, 1542–1555.
37. Hecker, M.H.L.; Stevens, K.N.; von Bismarck, G.; Williams, C.E. Manifestations of Task-Induced Stress in the Acoustic Speech Signal. J. Acoust. Soc. Am. 1968, 44, 993–1001.
38. Ruiz, R.; Legros, C.; Guell, A. Voice analysis to predict the psychological or physical state of a speaker. Aviat. Space Environ. Med. 1990, 61, 266–271.
39. Kraft, A.E.; Russo, J.; Krein, M.; Russell, B.; Casebeer, W.; Ziegler, M. A systematic approach to developing near real-time performance predictions based on physiological measures. In Proceedings of the 2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA), Savannah, GA, USA, 27–31 March 2017; pp. 1–7.
40. Conway, A.R.; Kane, M.J.; Bunting, M.F.; Hambrick, D.Z.; Wilhelm, O.; Engle, R.W. Working memory span tasks: A methodological review and user’s guide. Psychon. Bull. Rev. 2005, 12, 769–786.
41. Stroop, J.R. Studies of interference in serial verbal reactions. J. Exp. Psychol. 1935, 18, 643.
42. La Pointe, L.B.; Engle, R.W. Simple and complex word spans as measures of working memory capacity. J. Exp. Psychol. Learn. Mem. Cogn. 1990, 16, 1118.