A COMPARISON OF HEART RATE AND HEART RATE VARIABILITY INDICES IN DISTINGUISHING SINGLE-TASK DRIVING AND DRIVING UNDER SECONDARY COGNITIVE WORKLOAD

Bruce Mehler, Bryan Reimer, & Ying Wang
Massachusetts Institute of Technology (MIT) AgeLab
& New England University Transportation Center
Cambridge, Massachusetts, USA
E-mail: bmehler@mit.edu

Summary: Heart rate and heart rate variability (HRV) measures collected under actual highway driving from 25 young adults were compared to assess the relative sensitivity of each for distinguishing between a period of single task driving and periods of low and high additional cognitive workload. Basic heart rate, skin conductance and most, but not all, of the HRV indices were significantly different between single task driving and the high secondary demand period. Heart rate and skin conductance were also robust at distinguishing between single task driving and the low added demand period; however, several HRV measures did not show statistically significant differences between these two periods and the remaining HRV measures that did were less robust than basic heart rate as assessed by effect size and observed power. Rather than attempting to argue for the inherent superiority of any one physiological measure, these findings are presented with the intent of encouraging a broader discussion around the conditions under which particular physiological measures may be most useful and/or complementary for detecting different aspects of workload and operator state.

INTRODUCTION

In their recommendations for measuring mental workload in a test and evaluation environment, Wierwille and Eggemeier (1993) observed that heart rate was one of the most frequently used techniques for assessing aspects of operator workload. While much of this work was initially carried out in aviation (Kramer, 1991; Roscoe, 1992; Veltman & Gillard, 1998), a substantive literature is emerging using physiological measures such as heart rate in driving related research as well (Brookhuis & de Waard, 2001; Collet, Clarion, Morel, Chapon & Petit, 2009; Lenneman & Backs, 2009; Mehler, Reimer & Coughlin, 2010; Wilson, 2002). In addition to basic heart rate, there has also been growing interest in various measures of heart rate variability (HRV).

This interest in HRV has likely grown out of two research threads. One, from the human factors literature (Aasman, Mulder & Mulder, 1987; Mulder, 1992), focuses on changes in the 0.1 Hz component / low frequency band as a measure of mental effort. In parallel, there has been a surge of publications in the psychological and medical literature exploring a range of HRV metrics for gaining insight into the interplay between sympathetic and parasympathetic influences on the heart (Allen, Chambers & Towers, 2007; Malpas, 2002; Task Force, 1996). Numerous studies show situations where one or another of various HRV indices provides useful information not obtainable from mean heart rate data alone. While this is in many ways very exciting, we have noted an apparent trend in recent research papers to deemphasize basic heart rate in analyses or

even to omit reporting heart rate data altogether in studies reporting HRV values. This is a questionable choice since there are studies in the driving literature showing instances where heart rate distinguished specific workload conditions and the reported HRV metric did not (Wilson, 2002). Similarly, we are familiar with situations where the use of heart rate as a mental workload measure has been criticized since HRV has "clearly been established to be superior". The present paper reexamines data collected in an on-road experiment where participants were monitored under single task driving conditions and under varying levels of added cognitive load from a working memory task. Physiological measures in the original report consisted of heart rate and skin conductance level (Reimer, Mehler, Coughlin, Godfrey & Tan, 2009). For this paper, we have reprocessed the data to calculate a range of commonly used HRV metrics and examine the relative sensitivity of basic heart rate and these HRV measures at distinguishing periods of single task driving from periods of secondary workload. In addition, skin conductance level (SCL) is considered as a complementary measure of sympathetic nervous system activation.

METHOD

Subjects

Recruitment was conducted through online and newspaper advertisements. Participants were required to read and sign an approved informed consent form, to present a valid driver's license and attest to having had their license for more than three years, to driving more than three times per week and be in self-reported good health. A research assistant verified that participants clearly understood and spoke English. Individuals were excluded if they had been involved in a police reported accident in the past year or were taking a medication that caused drowsiness. A total of 33 individuals participated; seven cases were excluded from consideration due to heavy traffic, weather or data logging issues. The resulting sample was closely balanced by gender and consisted of 26 drivers between the ages of 22 and 27 (M=23.9; SD=1.6).

Apparatus

The experiment was conducted in an instrumented Volvo XC 90 equipped with a custom data acquisition system designed for time synchronized collection of vehicle, driver and environmental information (Coughlin, Reimer & Mehler, 2009). Physiological data were recorded at a sample rate of 250 Hz. using a MEDAC System/3 instrumentation unit and NeuGraph software (NeuroDyne Medical Corporation, Cambridge, MA). EKG recordings used a modified lead II configuration; the negative lead was placed just under the right clavicle (collar bone), ground just under the left clavicle, and the positive lead on the left side over the lower rib. Isopropyl alcohol was used to clean the skin and standard pre-gelled silver/silver chloride disposable electrodes (Vermed A10005, 7% chloride wet gel) were applied. Skin conductance was recorded utilizing a constant current configuration and non-polarizing, low impedance gold plated electrodes that allowed electrodermal monitoring without the use of conductive gel. Sensors were placed on the underside of the outer segment of the middle fingers of the non-dominant hand and secured with medical grade paper tape.

Secondary task & procedure

The secondary task and experimental protocol have been described in greater detail elsewhere (Reimer, 2009; Remier et al. 2009; Reimer & Mehler, 2010). After approximately 40 minutes of driving (10 in an urban area followed by 30 on an interstate highway) participants were presented with instructions, practice sets, and then 2 minutes of a scored working memory, digit recall task. Participants understood that an incentive payment would be added to their basic compensation based on their performance on the scored trials of each task. This protocol was repeated 3 times, with the difficulty of the task and the number of practice sets increasing at each level. The easiest level (0-back) required participants to repeat out loud randomly ordered single digit numbers (0-9) immediately after they were presented via an audio recording. The most difficult level (2-back) required holding digits in memory for 2 intervals back in the presentation sequence before repeating the value out loud.

Data analysis

Three periods are considered here: 2 minutes of single task driving that ended 30 seconds prior to the instructions for the first task and 2 minutes each of scored task trials for the easy (0-back) and hardest (2-back) secondary demand periods. Individual heart beats were detected using *EKG Wave Editor* release 1.8 (NeuroDyne Medical Corporation, Cambridge, MA), a software package that identifies R-wave peaks in the raw EKG signal and supports correction of peak markers when artifact due to movement or other factors is present. Processed records were reviewed by trained research associates and a limited number of skipped and double beats were edited when present to provide normalized values following general guidelines recommended for heart rate variability analysis (Mulder, 1992). To ensure accuracy and consistency, a second review of all records was performed by the first author. Inter-beat interval (IBI) data were preserved at 250 Hz resolution as well as the equivalent beat per minute (bpm) heart rate values. IBI data from one subject were excluded from analysis due to excessive artifact. Review and processing of the skin conductance recordings is detailed in Reimer et al. (2009). One case was dropped from initial analysis due to excessive artifact and two additional cases were later identified as extreme outliers and excluded. The resulting 26 case analysis set consisted of 25 with IBI data and 23 with skin conductance data.

The heart rate and SCL values described in Reimer et al. (2009) were derived from 250 Hz continuous records and resampled at 10 Hz; the entries in the first 2 rows of Tables 1 & 2 are based on this data. A dataset consisting of discrete consecutive IBI values for each condition and subject were processed using the publically available Kubios HRV analysis package version 2.0 (Tarvainen & Niskanen, 2008). The default settings were used to generate the heart rate values presented in row 3 of the tables and the HRV statistics. A within subject analysis using a repeated measures general linear model (SPSS, ver. 16) was used to carry out comparisons of each of the variables under single task driving vs. the low demand dual task condition (0-back) and vs. the high demand dual task condition (2-back). Gender was considered in an initial assessment, was found to be a non-significant factor ($p<.05$) and dropped from further modeling.

RESULTS

Performance on the secondary tasks and the impact of the associated cognitive demand on visual behavior, driving performance, and heart rate and skin conductance across each of the demand periods and a recovery segment have been reported previously (Reimer, 2009; Remier et al. 2009; Reimer & Mehler, 2010). In brief, engagement in the secondary tasks remained high across the three levels as evidenced in very low error rates (99.9%, 98.1%, and 95.7% respectively). The modest increase in errors with each task level and significant main effects of period (baseline single task driving, 0-, 1-, and 2-back, and recovery) on heart rate and skin conductance provided objective evidence that the secondary tasks increased workload.

Table 1 presents mean and standard deviation values for heart rate, SCL and a range of commonly used HRV metrics for the single task driving period, the low secondary demand 0-back period, and the relatively high demand 2-back period. Difference scores between the single task driving baseline period and these secondary task periods appear in the right most columns. Table 2 presents the results of statistical comparisons of each of these secondary task periods against single task driving on each physiological measure. Due to space constraints, readers are referred to a standard reference (Task Force, 1996) or Tarvainen and Niskanen (2008) for descriptions of individual HRV indices.

Table 1. Mean, standard deviation (SD), and change scores for physiological variables

Variable	Baseline[1]		0-back		2-back		Change from Baseline	
	Mean	SD	Mean	SD	Mean	SD	0-back	2-back
Heart Rate I[2]	75.4	9.2	78.4	10.7	84.0	12.8	3.1 bpm	8.7 bpm
SCL[2]	13.7	4.8	15.0	5.5	15.3	5.4	1.1 µmhos	2.4 µmhos
Heart Rate II[2]	76.2	9.2	79.3	10.8	84.9	12.9	3.1 bpm	8.8 bpm
SDNN	49.5	17.1	49.5	22.0	45.1	13.7	-0.1 ms	-4.5 ms
SDSD	33.1	14.7	29.8	14.5	23.3	9.8	-3.3 ms	-9.8 ms
RMSSD	33.0	14.7	29.7	14.5	23.2	9.8	-3.3 ms	-9.8 ms
NN50	23.2	23.1	19.0	19.5	12.7	14.9	-4.2	-10.5
pNN50	16.2	18.3	13.0	15.7	8.3	10.8	-3.2 %	-7.9 %
LF Power[3]	1355	753	1082	642	925	579	-273 ms^2	-430 ms^2
HF Power[3]	564	684	457	629	258	252	-108 ms^2	-306 ms^2
Total Power[3]	2723	1953	2768	2724	2115	1158	45 ms^2	-609 ms^2
LF/HF Ratio	4.29	2.88	3.85	2.31	4.85	3.44	-0.45	0.56

[1]Baseline= single task driving prior to the low demand secondary task (0-back) and high demand (2-back) task
[2]Heart Rate I and skin conductance values calculated from 250 Hz continuous records sampled at 10 Hz; Heart Rate II (and subsequent HRV variables) based on discrete IBI values
[3]Frequency domain / power density spectrum values based on bands (LF= 0.04-0.15, HF=0.15-0.4, Total Power= 0-0.4 Hz) using auto regressive method & default settings in Kubios HRV package, version 2

Heart Rate I values are based on continuous heart rate values sampled at 10 Hz while Heart Rate II values are calculated from discrete IBI values. In the latter case, each individual heart beat is equally weighted in calculating the mean and standard deviation. In the former, longer inter-beat

intervals are given somewhat more weight in the calculations because of the fixed rate sampling of the data. As can be observed in Table 1, this results in slightly lower values for Heart Rate I means. A review of the analyses in Table 2 shows that the two approaches to quantifying heart rate prodce essentially identical results in these assessments.

Table 2. Statistical differentiation of demand periods using various physiological metrics

Variable	Baseline vs. 0-back (low demand)				Baseline vs. 2-back (high demand)			
	F value[1]	sig.	partial eta sq. [2]	observed power[3]	F value[1]	sig.	partial eta sq. [2]	observed power[3]
Heart Rate I[4]	13.1	p=.001	0.354	0.935	44.4	p<.001	0.649	1.000
SCL	11.0	p=.003	0.334	0.887	13.8	p<.001	0.386	0.944
Heart Rate II[4]	13.2	p=.001	0.355	0.936	42.6	p<.001	0.640	1.000
SDNN	0.0	p=.983	0.000	0.050	2.6	p=.117	0.099	0.344
SDSD	6.5	p=.017	0.214	0.688	24.9	p<.001	0.509	0.998
RMSSD	6.5	p=.018	0.213	0.687	24.9	p<.001	0.509	0.998
NN50	3.1	p=.090	0.115	0.396	14.0	p=.001	0.368	0.948
pNN50	3.7	p=.066	0.134	0.457	12.2	p=.002	0.337	0.918
LF Power	6.3	p=.020	0.207	0.671	9.9	p=.006	0.271	0.817
HF Power	3.0	p=.096	0.111	0.384	7.3	p=.012	0.233	0.737
Total Power	0.0	p=.882	0.001	0.052	3.2	p=.088	0.117	0.401
LF/HF Ratio	1.1	p=.314	0.042	0.167	1.1	p=.305	0.044	0.172

[1] $F_{(1,24)}$ degress of freedom except for SCL(1,22); repeated measures GLM model. Corresponding sigificance levels in the next column are displayed in bold when $p<.05$

[2] Partial Eta squared is a measure of effect size that ranges from 0 to 1 and has some similarity to a correlation coefficent. In SPSS it is the proportion of the effect plus the error variace that is attributable to the effect

[3] Observed power ranges from 0 to 1. Observed power values of .80 or higher (bolded) are generally considered desirable to have confidence that a real difference will be detected if it exists

[4] Heart Rate I in top row calculated from 250 Hz continuous records sampled at 10 Hz; value in lower row based on discrete IBI values

As can be observed in Table 2, basic heart rate and the other time domain based HRV measures, with the exception of SDNN, all provide fairly robust indicies for detecting the relatively substantial increase in cognitive workload between single task driving and the added demand of the 2-back task. It should be noted that no adjustments have been made in p-values to compensate for conducting multiple tests of significance; the intent here is primarially to compare the relative sensativity of the individual measures. While there are very large effect sizes (partial eta squared) for each of these measures (again excepting SDNN), basic heart rate shows the largest effect size value at 0.6. In the frequency domain, LF and HF power both show substantive decreases under the 2-back condition and provide very respectable measures of the change in workload. Nonetheless, the significance values, effect sizes and observed power statistics combine to indicate that they are not as sensitive as basic heart rate under the conditions monitored in the drive. Perhaps most striking is that the frequently used LF/HF ratio was not at all useful for differentiating the two conditions; this will be considered more in the discussion.

For differentiating the workload associated with single task driving and that of driving with the modest added demand of the simple auditory aquistion / verbal repetion 0-back task, basic heart rate appears to have clear advantages over the HRV metrics in terms of sensitivity. SDSD,

RMSSD and LF power all provide statistically significant differentiation of the periods, but at lower levels of significance, much smaller effect sizes and with lower observed power. Moreover, SDNN, NN50, pNN50, HF power, and the LF/HF ratio metrics all fail to differentiate conditions.

DISCUSSION

This paper examines the relative utility of heart rate compared to standard HRV metrics for detecting a change in cognitive workload in the rather challanging environment of driving an automobile. In the case of the working memory task studied here, mean heart rate and skin conductance clearly distinguished both the high and low demand secondary task periods from single task driving. Most, but not all, of the HRV metrics were also capable of differentiating the high demand period from single task driving. However, over half of the HRV metrics did not differentiate the change in workload when considering the low demand period. Those HRV measures that did show a statistically significant difference, had lower effect sizes and observed power than basic heart rate and SCL. These are relevent considerations since the capacity to detect subtle changes in workload is essential if physiological measures are to prove useful in assessing in-vehicle HMI and other demands (Lenneman & Backs, 2010; Mehler et al. 2009).

The most robust of the HRV measures in this assessment were SDSD and RMSSD from the time domain and LF power from the frequency domain. The LF power measure corresponds to the 0.1 Hz component advocated as a measure of mental effort (Aasman, et al, 1987; Mulder, 1992). It is useful to note that the LF/HF ratio measure was not at all sensitive to changes in workload in this study. The LF/HF ratio has proven useful in differentating individuals with and without various medical conditions and in other research contexts; however, there are various reasons why it might not be as sensitive as a primary workload measure under actual driving conditions.

It is not our intent to argue that any particular physiological measure of workload is inherently superior to another. While heart rate provided the most robust measure here, there are limited but known situations in which heart rate can be expected to remain low or even drop under conditions of hightened arousal (Lenneman & Backs, 2009; Mehler, Reimer, Pohlmeyer & Coughlin, 2008). We believe that data presented provide additional evidence for the position that various measures have varying utility depending on the conditions underwhich testing is taking place and the nature of the workload that is being assessed; multiple measures are certainly to be valued in research contexts such as combining heart rate and skin conductance. That being said, heart rate is a particularly attractive measure for real-time driver state detection because is it is impacted less by common cardiac arythmias and events such as occassional skipped or double beats that are generally delt with in research by post-collection processing and editing.

ACKNOWLEDGMENTS

REFERENCES

Aasman, J., Mulder, G., & Mulder, L.J.M. (1987). Operator effort and the measurement of heart-rate variability. *Human Factors, 29(2),* 161-170.

Allen, J.J.B., Chambers, A.S., & Towers, D.N. (2007). The many metrics of cardiac chronotropy: a pragmatic primer and a brief comparison of metrics. *Biological Psychology, 74,* 243-262.

Brookhuis, K. A. & De Waard, D. (2001). Assessment of drivers' workload: Performance and subjective and physiological indexes. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, Workload, and Fatigue* (pp. 321-333). Mahwah, NJ: Lawrence Erlbaum Associates.

Collet, C., Clarion, A., Morel, M., Chapon, A., & Petit, C. (2009). Physiological and behavioural changes associated to the management of secondary tasks while driving. *Applied Ergonomics, 40*(6), 1041-1046.

Kramer, A. E. (1991). Physiological metrics of mental workload: a review of recent progress. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 279-328). London: Taylor & Francis.

Lenneman, J.K. & Backs, R.W. (2009). Cardiac autonomic control during simulated driving with a concurrent verbal working memory task. *Human Factors, 53*(3), 404-418.

Lenneman, J.K. & Backs, R.W. (2010). Enhancing assessment of in-vehicle technology attention demands with cardiac measures. *Proceedings of the Second International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutoUI 2010),* November 11-12, Pittsburgh, Pennsylvania, USA.

Malpas, S.C. (2002). Neural influences on cardiovascular variability: possibilities and pitfalls. *American Journal of Physiology. Heart and Circulatory Physiology, 282*, H6-H20.

Mehler, B., Reimer, B., Coughlin, J.F., & Dusek, J.A. (2009). The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record, 2138, 6-12.*

Mehler, B., Reimer, B., & Coughlin, J.F. (2010). Physiological reactivity to graded levels of cognitive workload across three age groups: An on-road evaluation. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*, San Francisco, CA, 2062-2066.

Mehler, B., Reimer, B., Pohlmeyer, A.E. & Coughlin, J.F. (2008). The association between heart rate reactivity and driving performance under dual task demand in late middle age drivers. *Advances in Transportation Studies an International Journal*. *Special Issue,* 53-70.

Mulder, L. J. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology, 34*(2-3), 205-236.

Reimer, B. & Mehler, B. (2010). The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. Manuscript under review.

Reimer, B., Mehler, B., Coughlin, J. F., Godfrey, K. M., & Tan, C. (2009). An on-road assessment of the impact of cognitive workload on physiological arousal in young adult drivers. *Proceedings of the AutomotiveUI*, Essen, Germany.

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J.F. (2010). The impact of systematic variation of cognitive demand on drivers' visual attention across multiple age groups. *Proceedings of*

the 54th Annual Meeting of the Human Factors and Ergonomics Society, San Francisco, Sept. 27-Oct. 1, 2010, 2052-2056.

Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration. *Biological Psychology, 34*(2-3), 259-287.

Tarvainen, M.P. & Niskanen, J. (2008). Kubios HRV User's Guide. Verison 2.0. University of Kuopio, Kuopio, Finland. (http://kubios.uku.fi/media/Kubios_HRV_Users_Guide.pdf)

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *European Heart Journal, 17*, 354-381.

Veltman, J. A. & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics, 41*(5), 656-669.

Wierwille, W.W. & Eggemeier, F.T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors, 35(2),* 263-281.

Wilson, G. F. (2002). Psychophysiological test methods and procedures. In S. G. Charlton & T. G. O'Brien (Eds.), *Handbook of Human Factors Testing and Evaluation* (pp. 127-156). Mahwah, NJ: Lawrence Erlbaum.