INL/CON-08-14336
PREPRINT

Exploring the Potential of Short-Time Fourier Transforms for Analyzing Skin Conductance and Pupillometry in Real-Time Applications

52^{nd} Annual Meeting of the Human Factors and Ergonomics Society

Roger Lew
Brian P. Dyre
Steffen Werner
Brian Wotring
Tuan Tran

September 2008

EXPLORING THE POTENTIAL OF SHORT-TIME FOURIER TRANSFORMS FOR ANALYZING SKIN CONDUCTANCE AND PUPILLOMETRY IN REAL-TIME APPLICATIONS

Roger Lew, Brian P. Dyre, Steffen Werner, & Brian Wotring
University of Idaho
Moscow, Idaho

Tuan Tran
Idaho National Laboratory
Idaho Falls, Idaho
Currently at Sprint/Nextel in Kansas City, Kansas

The development of real-time predictors of mental workload is critical for the practical application of augmented cognition to human-machine systems. This paper explores a novel method based on a short-time Fourier transform (STFT) for analyzing galvanic skin conductance (SC) and pupillometry time-series data to extract estimates of mental workload with temporal bandwidth high-enough to be useful for augmented cognition applications. We tested the method in the context of a process control task based on the DURESS simulation developed by Vincente and Pawlak (1994; ported to Java by Cosentino,& Ross, 1999). SC, pupil dilation, blink rate, and visual scanning patterns were measured for four participants actively engaged in controlling the simulation. Fault events were introduced that required participants to diagnose errors and make control adjustments to keep the simulator operating within a target range. We were interested in whether the STFT of these measures would produce visible effects of the increase in mental workload and stress associated with these events. Graphical exploratory data analysis of the STFT showed visible increases in the power spectrum across a range of frequencies directly following fault events. We believe this approach shows potential as a relatively unobtrusive, low-cost, high bandwidth measure of mental workload that could be particularly useful for the application of augmented cognition to human-machine systems.

INTRODUCTION

The promise of augmented cognition hinges on the ability of researchers to identify valid indicators of relevant cognitive functioning. In addition, the timeliness of the availability of indicators is a large practical concern when dealing with real-time augmented cognitive systems (O'Neill, 2006). In this context, a reevaluation of classic physiological measures, such as galvanic skin conductance (SC, also known as galvanic skin response, or GSR) and pupillometry from multiple perspectives might lead to improved indicator measures as compared to traditional analysis techniques (Cacioppo & Tassinary, 1990). In our current study, we are beginning to investigate the role of mental workload, stress, and fatigue on human error probability in complex dynamical control tasks. We hope to improve on existing human performance and risk assessment tools by using non-obtrusive physiological sensors (eye tracking and SC) to assess human information processing errors and limitations for advanced energy systems in the context of the Standardized Plant Analysis Risk-Human Reliability (SPAR-H) method. The obvious goal is to obtain an improved quantitative estimate of risk.

Galvanic skin conductance and pupil data have traditionally been analyzed in the context of stress and workload. Usually, changes in galvanic skin response have been interpreted in numerous ways as indicating increases in anxiety or stress, while increases in pupil diameter have been used as a measure of task difficulty in simple (sentence processing, mental calculations) and complex (user interface evaluation) tasks (Just & Carpenter, 1993; Nakayama & Shimizu, 2002; Nakayama & Katsukura, 2007). In most cases, SC and pupil data has been interpreted by comparing relative SC changes and pupil dilations in the time domain. This can work well quite well and has been successful in differentiating stressful from less stressful tasks and easy from hard processing conditions. It represents the data in an easily-understandable format, but in some cases it may prevent the the viewer from wholly understand the nuances of the data. Recently, research in pupillometry has begun to examine frequency content. For example, Nakayama and Shimizu (2002 & 2004) found the power spectrum density of the pupil size data between 0.1- 0.5 Hz and 1.6 -3.5 Hz to increased with the difficultly mental arithmetic.

In addition, Marshall (2000, 2002) has demonstrated the usefulness of wavelet analysis for extracting frequency content relating to cognitive activity at specific moments in time (Index of Cognitive Activity, ICA). In this paper, we will present an approach similar to Marshall's ICA in that it also transforms the time signal into the frequency domain yet preserves the time domain.

By dividing the time domain into short epochs (~34s) and applying Fourier analysis to each the physiological signals can be visualized using spectrographs. For years EEG data have been represented in this form showing time, frequency and power in a single graph represented either as color or ridges. This paper explores the potential for adapting this method, based on the short-time Fourier transform (STFT), to pupil diameter and skin conductance. To date, neither pupil

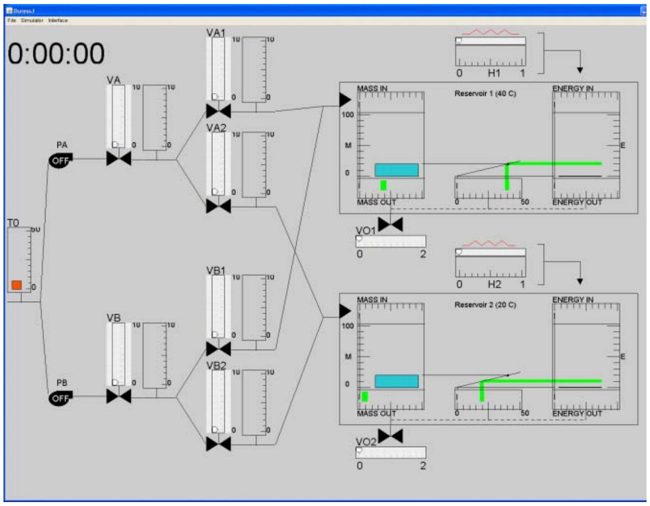

Figure 1. User Interface for the Duress water station simulator. The participant is tasked with matching flow and temperature demands for two outputs as well as reservoir levels.

diameter nor skin conductance have been analyzed in this manner. We examined these measures in the context of the DURESS process control simulation (Vincente & Pawlak, 1994).

METHOD

Participants

Four university students participated in this experiment. All had normal or corrected to normal Snellen visual acuity (20/30 or better). Participant 1 was an experimenter and had extended knowledge of the experiment while the others had only limited knowledge of the experimental measures and manipulations.

Stimuli and Apparatus

Participants operated DURESS (Cosentino,& Ross, 1999; Vicente & Pawlak, 1994), a process-control simulation requiring the control of heat and flow of a dual reservoir system. Their task was to match dynamic water flow and temperature demands by manipulating several pumps, valves, and heaters (see Figure 1). To manipulate workload, each participant was given an identical set of plant failures and events at predefined instances (see Table 1).

The simulation was displayed on a 60 inch rear-projection monitor at a spatial resolution of 1280 x 1024 and a temporal resolution of 60 Hz with a viewing angle of 45° X 33.75°. The display was presented in a darkened room with the participant sitting 1.5m from the display. The participants used a standard optical mouse with their right hand to control the elements of the DURESS simulation.

Table 1. DURESS Fault Events

Time (s)	Event Description
300	Flow rate in valve VA2 changes
420	Demand change for upper reservoir
480	Flow rate in valve VB1 changes
720	Demand changes for upper reservoir
726	Demand changes for lower reservoir
732	Flow rate in VB2 changes
738	Inflow of water to upper reservoir
741	Outflow of water to lower reservoir
750	Output valve outflow increases Inflow temperature changes
756	Demand changes for lower reservoir Flow rate in valve VA2 changes

A model ASL5000 head-mounted eye/head tracker was used to measure gaze direction, pupil diameter, and blink rate at 60 Hz. Skin conductance was measured at a temporal resolution of 256 Hz with a Thought Technologies ProComp5 Infiniti encoder using two finger-mounted sensors placed around the index and ring fingers of the left hand. Participants were instructed to leave their left hand in a stationary position during the course of the trial.

Procedure

Each participant received a short tutorial explaining goals, controls, and caveats of the DURESS program. Participants were then given a 10-minute DURESS practice session with coaching from the experimenter without the eye/head tracking or SC monitor attached.

After training, the eye/head tracker and GSR devices were mounted and calibrated to the individual. The participants then operated and experienced the DURESS fault trial which ended after 15 minutes (900 s).

RESULTS

Currently our treatment of the data has been exploratory rather than inferential. The intent of the study is to examine the relationship between the timing of DURESS faults and activity representing workload and stress induced by these faults in the physiological measures. Future work will involve incorporating inferential tests between power bands before and after events and/or a discriminant network for classifying workload from these physiological measures.

All data analysis and visualization was performed using Python (a high-order, interpreted, open source, programming language) and the SciPy (an open source library of scientific tools for Python). SC, pupil diameter, and duress measures where linearly interpolated to a sampling rate of 60 Hz (the slowest of original sampling rates). Additionally, blinks were removed from the pupil diameter data by holding the last non-blink value during the duration of the blink, and data below three standard deviations of the mean where deemed as unreliable due to measurement error and removed (less than 0.1% of data, typically indicate a partially blink). More

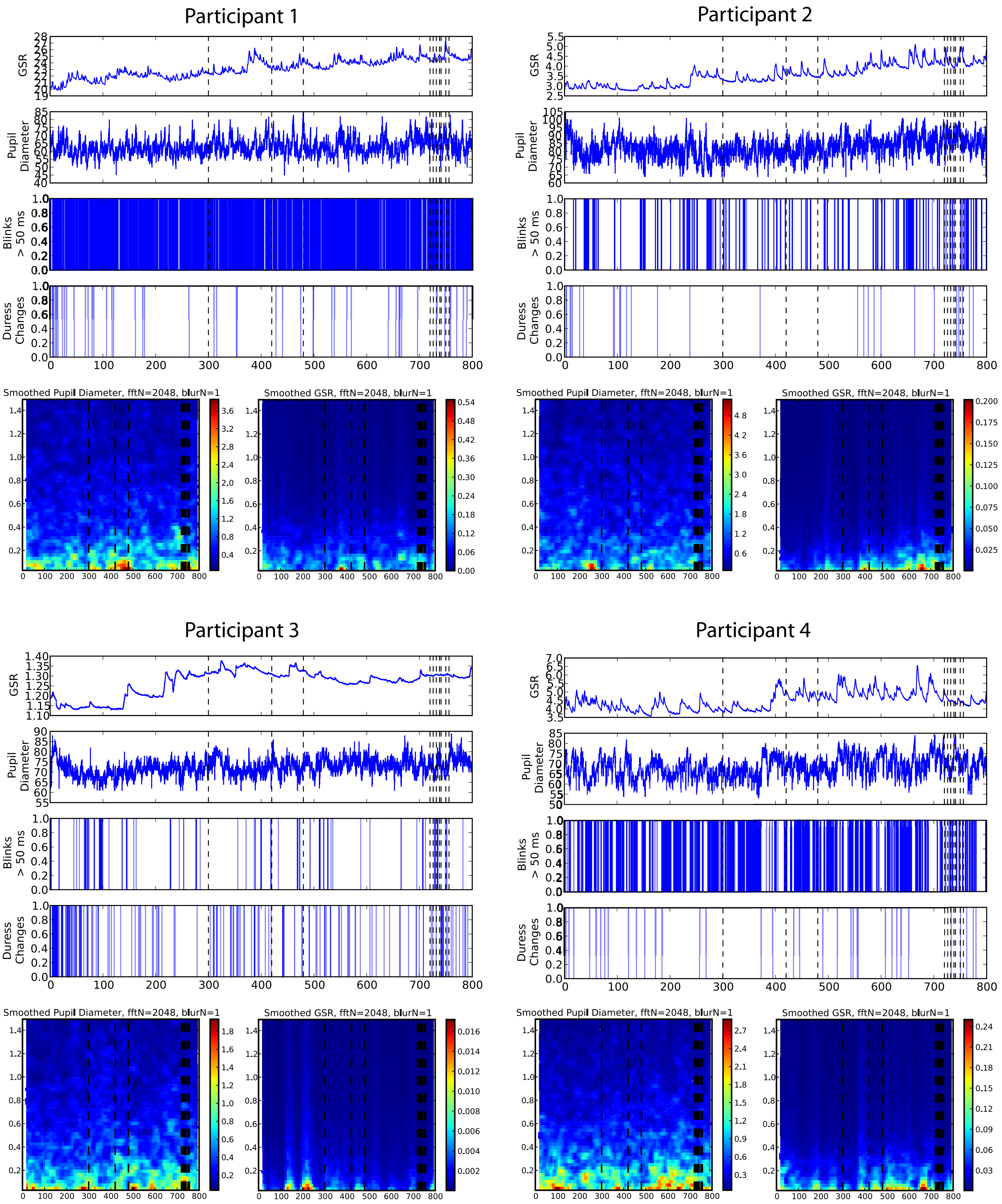
Participant 1
Participant 2
Participant 3
Participant 4
GSR
Pupil Diameter
Blinks > 50 ms
Duress Changes
0
100
200
300
400
500
600
700
800
Smoothed Pupil Diameter, fftN=2048, blurN=1
Smoothed GSR, fftN=2048, blurN=1
0.2
0.4
0.6
0.8
1.0
1.2
1.4

Figure 2.

complex treatments of blinks exist but they typically do not significantly alter the power content below .5 Hz according to Nakayama and Shimizu (2004).

Time-series Representations

The top four panels for each participant in Figure 2 show the SC, pupil diameter, blink rate, and DURESS events plotted as time-series. In each panel the vertical dotted lines indicate DURESS events corresponding to Table 1. SC data in the top panel consist of unsmoothed raw values. Pupil diameter data in the second panel were filtered as described above, but are otherwise unsmoothed. The third panel represents blink occurrences greater than 50 ms in duration. The fourth panel represents occurrences of participant interaction with the DURESS simulation that require a click of the mouse. These interactions include setting heat levels or valve flows to new levels, and turning a pump on or off.

One feature of these plots is that there is often no visually obvious correlation between the changes measures and the DURESS events. In fact, one problem with associating time-derivatives of these measures with events is that one needs to have a good idea of the time scale over which to calculate the derivatives. An alternative approach that does not require this a priori knowledge of the time-scale of changes a spectral analysis.

Short-time Fourier Spectra

Examining the data in the frequency domain provides a mechanism by which changes in physiological responses to events in a complex task can be estimated on all time scales at once, in effect a “brute force” approach to the problem of determining the appropriate time scale. To do this we first transform short samples of the time-series measures into the frequency domain. The time-series “windows” defining these samples may overlap to provide a more continuous measure of the change in the spectral characteristics of the measures.

To examine the power spectrum as a continuous function of time the SC and pupil diameter data were examined using short time Fourier transform (STFT) with time windows of approximately 34 s (2048 samples) with each window overlapping its neighbor by 50%. For each time slice the power spectrum was estimated using a fast Fourier transform (FFT) algorithm. Finally, spectrograms (color plots of power amplitudes by time and frequency) were created. The spectrographs make low frequency variations in the time domain easily distinguishable from noise. This treatment is especially useful for long trial durations (15 min.) where distinguishing nuances in time domain plots is not possible due to limited resolution.

The spectrographs (see Figure 2, bottom panels for each participant) represent time (in seconds) on the x-axis and frequency (in Hz) on the y-axis. Color is used to represent the power (in decibels) for each particular time and frequency sample. Areas of high amplitude changes in the measures appear red, while areas of low amplitude appear dark blue. Intermediate amplitudes are represented along the rainbow color spectrum. Consistent with the time-series plots, the vertical black dotted lines represent DURESS fault events. The spectrographs are smoothed by convolving the 2D matrix of amplitudes with a Gaussian kernel ($\mu=0$, $\sigma=1$).

DISCUSSION

Inspection of the spectrographs reveals that visually-obvious increases in power across a fairly broad range of frequency occur commonly at the beginning of the trial while the participant is actively adjusting the simulator to achieve steady state and just after the DURESS fault events. In contrast, during time-sequences with relatively little activity, the spectrograms appear dark blue, indicating very little change in the physiological measures. Across several participants “hot spots” appeared shortly after DURESS fault events which we interpret as promising indications that these measures could prove to be a reliable and sensitive real-time indicators of mental workload and stress when analyzed using STFT.

When interpreting the spectrographs it is important to note that fault events were often not noticed by participants. Faults did not trigger visual or auditory alarms; only the unresponsiveness of the flow and temperature states to system changes indicated system failures. Events potentially went unnoticed for several seconds or even minutes. We know when the events occur but not when the participants become cognitively aware of them. Current work is examining these measures with a control task where changes in workload are more salient. Such a task makes performing inferential analyses on these measures a viable option.

REFERENCES

Cacioppo, J.T. & Tassinary, L.G. (1990). Inferring psychological significance from physiological signals. *American Psychologist, 45*(1), 16-28.

Cosentino, B.& Ross*,* A (1999*).* Design and Implementation of a Java Version of the DURESS II Simulator. *Toronto: University of Toronto, Cognitive Engineering Laboratory*.

Jones, E., Oliphant, T., Peterson, P., & others (2001). SciPy: Open source scientific tools for Python. *http://www.scipy.org/*

Just, M.A. & Carpenter, P.A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. Canadian Journal of Experimental Psychology, 47, 310-339.

Nakayama, M., & Katsukura, M., (2007). Feasibility of assessing usability with pupillary responses. *Proc. of AUIC 2007, 15-22*

Nakayama, M., & Shimizu, Y. (2002). An estimation model of pupil size for ‘blink artifact’ and it’s applications. *In Proceedings of the 10th European Symposium on Artificial*, d-side, Evere, Belgium, M. Verleysen, Ed., 251-256.

Nakayama, M., & Shimizu, Y. (2004). Frequency analysis of task evoked pupillary response and eye-movement. *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, 71-76.

Marshall, S. (2000). Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity, *US Patent 6,090,051*.

Marshall, S. (2002). The index of cognitive activity: Measuring cognitive workload. *IEEE 7th Human Factors Meeting,* (7.5-7.9).

O'Neill, K. (2006). Modular software for an augmented cognition system. In D.D. Schmorrow (Ed.), *Foundations of Augmented Cognition* (pp. 309-330). Mahwah, NJ; London: Lawrence Erlbaum Associates Publishers.

Vicente, K. J., & Pawlak, W. S., 1994. Cognitive work analysis for the DURESS II system, (CEL 94-03), *Toronto: University of Toronto, Cognitive Engineering Laboratory*.

Research funded by the Department of Energy through the Idaho National Laboratory.