CAPTURING DRIVER RESPONSE TO IN-VEHICLE HUMAN-MACHINE INTERFACE TECHNOLOGIES USING FACIAL THERMOGRAPHY

Michelle L. Reyes[1], John D. Lee[1], Yulan Liang[1],
Joshua D. Hoffman[1], Ritchie W. Huang[2]
[1]The University of Iowa
Iowa City, Iowa USA
[2]Honda R&D Americas, Inc.
Torrance, California USA 90501
Email: mlries@engineering.uiowa.edu; jdlee@engineering.uiowa.edu,
yulliang@engineering.uiowa.edu; jhoffman@engineering.uiowa.edu; rhuang@hra.com

Summary: Measuring driver response to in-vehicle human-machine interface (HMI) systems is critical for the automotive design and evaluation process. Physiological measures provide a useful complement to performance-based and subjective measures because they promise an estimate of the affective response of drivers to an in-vehicle system in a way that requires no overt response by the driver. This research explored how facial temperature might reflect the drivers' response to the demands they confront when interacting with in-vehicle systems. Sixteen drivers completed a series of in-vehicle tasks while driving in a simulator. Facial temperature was measured using an infrared camera. The analyses focus on how the thermal data, aggregated over four facial regions, correlated with both measures of driving performance and subjective ratings of workload and frustration. Facial temperature measures correlated with more driving performance measures of longitudinal control than lateral control, suggesting that thermal measures are sensitive to different cognitive processes than are typically assessed by measures of steering and lane position. Thermal measures aggregated over a 15-second window correlated with subjective ratings. Unlike other measures typically used to evaluate in-vehicle systems that are aggregated over long time windows, thermal measures have temporal specificity and might be able to identify specific interactions that increase workload and frustration. No single facial area or summary measure emerged as the best indicator of driver response; rather, composite measures of facial temperature could be developed that offer a more complete profile of driver response.

INTRODUCTION

In-vehicle information systems are becoming an increasingly visible and important part of the vehicle. Driver response to emerging in-vehicle human-machine interface (HMI) systems may have a substantial influence on drivers' perception of the overall vehicle and on driving performance. As a consequence, measuring driver response to in-vehicle HMI systems is critical for the design and evaluation process. Typical measures include driving performance, performance of interacting with the in-vehicle system, and subjective workload ratings. Although each of these measures can provide useful information to guide design, each has important limits. Importantly, drivers' affective or emotional response is often neglected. Physiological measures provide a useful complement to these measures because they promise an estimate of drivers'

affective response to an in-vehicle system in a way that requires no overt response. This paper explores how facial temperature might reflect the demands and affective response drivers experience when interacting with in-vehicle HMI systems.

Facial skin temperature is an indirect indicator of cognitive state. The physiological response of increased blood flow to the skin in the face is triggered by the autonomic nervous system, which can redistribute blood between the skin and the periphery and the core of the body (Tsiamyrtzis et al., 2007). Past research has shown skin temperature in various facial regions to be sensitive to a variety of psychological constructs, including anxiety, stress, fatigue, deception, frustration, anger, happiness, and fear.

The effect of stress caused by deception in polygraph-like settings has been investigated by Pavlidis et al. (Pavlidis, Eberhardt, & Levine, 2002; Pavlidis & Levine, 2002; Pollina et al., 2006; Tsiamyrtzis et al., 2007). Skin temperature just under the lower eye lid measured via infrared thermography was sensitive to whether the participants were deceptive or not, and the changes in temperature occurred within about one second (Pollina et al., 2006). More recently, filtering the temperature signal from the periorbital region of the inner eye (see Figure 1) allowed for 87% of the 39 participants to be correctly classified as deceptive or non-deceptive (Tsiamyrtzis et al., 2007).
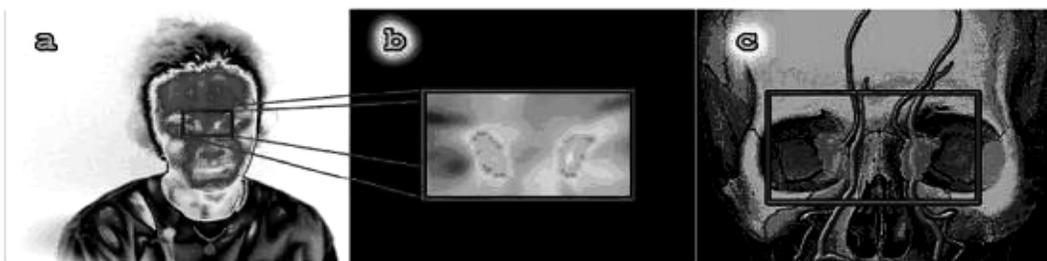
Figure 1. (a) Facial thermal image with the periorbital region highlighted. (b) The selected region of interest enlarged with the 10% hottest pixels marked in pink. (c) The periorbital region of interest superimposed on the ophthalmic arteriovenous complex (reprinted from Tsiamyrtzis et al., 2007 with kind permission from Springer Science and Business Media)

Other studies have used cognitive tasks to induce stress in the form of frustration. One of these was a computerized Stroop test in which the response window decreased over time to make the task increasingly difficult (Puri, Pavlidis, Olsen, Levine, & Starren, 2005). Blood flow and blood volume were derived from forehead temperature for the hottest 10% of pixels the corrugator or 'frowning' muscle on the forehead. Energy expenditure, considered to be the ''ground-truth'' measurement of stress, was found to correlate highly with blood volume ($r = 0.91$ for 11 out of 12 participants). These findings suggest that changes in forehead temperature may provide a reliable index of frustration by measuring the vascular activity associated with the corrugator muscle. However, a study that used a tracking task to induce stress found contradictory results (Genno et al., 1997). Temperature was measured with thermistors in a number of facial locations, and forehead temperature was not sensitive to the induction of stress. The difference in measurement methods might explain the conflicting results because when a thermistor is used, the skin surface area measured is limited to the area under the electrode (Rimm-Kaufman & Kagan, 1996).

Genno et al. (1997) found that while there was no change in forehead temperature under stress, nasal temperature decreased. Using data from previous studies, the authors derived a linear model to predict subjective stress ratings using the difference between the forehead temperature and the nasal temperature. The predicted ratings correlated with the actual ratings with $r = 0.51$.

Thermography has been applied to investigations of emotion. Rimm-Kaufmann and Kagan (1996) asked participants to perform tasks meant to elicit performance anxiety, positive emotions, fear, and anxiety. Temperature of the inner eyes, measured using infrared thermography, was sensitive only to the task meant to elicit anxiety.

Facial temperature has also been used to assess driver workload while driving in simulated city and highway settings with and without a mental arithmetic task (Or & Duffy, 2007). Forehead and nose temperatures were measured immediately before and after each three-minute trial and subjective workload scores were obtained via the Modified Cooper-Harper scale. Forehead temperature remained stable while nose temperature dropped significantly during all conditions. The decrease in nose temperature was significantly greater for the trials that included the arithmetic task and was correlated with the subjective workload scores ($r = 0.32$, $p = 0.009$).

Interactions with in-vehicle HMI systems place demands on drivers that can increase stress, frustration, and cognitive load. These effects can undermine driver performance and enjoyment of the vehicle and are revealed by a complex pattern of physical and physiological behavior. From the review of the facial thermography literature it is clear that no one location on the face is "best" for measuring driver response to in-vehicle technology and different regions are likely sensitive to different affective responses. In addition, most of the literature neglects the temporal dimension in facial temperature changes. This research evaluates changes in facial temperature at several locations and across different time windows for several representative in-vehicle tasks. The relationships between facial temperature and both driving performance and subjective ratings are considered.

METHODS

Sixteen drivers between the ages of 35 and 55 each drove in a fixed-base DriveSafety (v. 1.6.2) driving simulator with full cab and a 50-degree field of view for a total of approximately two hours. Each driver completed a series of 60-second tasks with each of two interfaces, a jog dial and a touch screen, while driving on a multi-lane suburban roadway. The tasks were radio tuning, CD switching, address entry for a navigation system by hand and by voice command, and selecting songs on an iPod™. Data were also collected for baseline periods of driving only but are not included in this analysis which focuses on discriminating between interfaces rather than between baseline and task periods. Driving performance data, eye movements, and physiological measures (facial temperature, ECG, GSR, and respiration) were collected. After completing each task, the drivers gave verbal ratings for each of the six NASA-TLX dimensions (Hart & Staveland, 1988), ease of the driving task, and ease of the in-vehicle task. The analyses presented here focus on how the thermal data aggregated in four facial regions (forehead, nose, inner eye, and entire face) and for three time windows (first 15, first 30, and all 60 seconds of each task) correlated with both measures of driving performance and selected subjective ratings.

A ThermoVision® A40M Infrared Camera – Researcher by FLIR Systems (thermal sensitivity of 0.08°C at 30°C) recorded temperature at a frame rate of 30 Hz. The drivers wore small metal markers on their foreheads about 1.5 inches above the pupil of each eye. The markers were visible in the thermal image and anchored the regions of interest (see Table 1 and Figure 2). The mean temperature in each region was calculated for each frame during the task trial, and then the mean from the first frame was subtracted from all subsequent frame means. Because head movements were found to affect the temperature readings, a linear model was used to correct artifacts by predicting temperature changes related to head movements. The mean and standard deviation of the residual temperatures (which reflect the change in temperature that could not be attributed to head movement) were calculated over the duration of the task for each region.
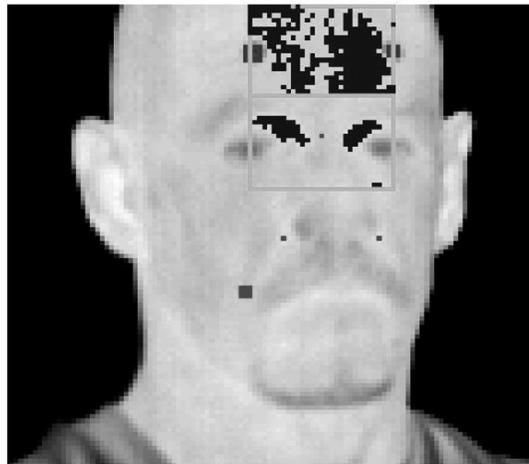

Figure 2. Facial regions of interest overlaid on thermal image

Table 1. Definition of regions of interest

Region	Upper boundary	Lower boundary	Horizontal boundary	Warmest pixels retained
Overall	-	-	-	60%
Forehead	15 pixels above markers	10 pixels below markers	Marker	50%
Nose	Midpoint between inner eye regions	Determined by inspection	Markers or centers of inner eye region	100%
Eye	Bottom of forehead area	10 pixels below vertical midpoint between markers and bottom of nose	Markers	20%

RESULTS

Three measures of lateral control input driving performance were considered: mean and standard deviation (SD) of steering error (as calculated by Nakayama, Futami, Nakamura, & Boer, 1999) and root mean square (RMS) of steering wheel angle. Measures of lateral control output were SD of lane position and minimum time-to-line crossing (van Winsum, Brookhuis, & de Waard, 2000) for each lane boundary. Longitudinal input measures were mean and SD of accelerator pedal position (calculated in the same way as steering error) and RMS accelerator pedal position, and longitudinal output measures of driving performance analyzed were median speed, RMS

speed, and SD of acceleration. All driving performance measures were aggregated over the 60-second task period.

The mean temperature in the nose region is correlated with all three lateral control input driving performance measures (see Table 2). The lateral control output measures are not significantly correlated with any of the thermal variables. All of the thermal measures are significantly correlated with at least one of the longitudinal control measures (see Table 3). Specifically, the mean nose temperature is sensitive for both the mean and SD of accelerator pedal position error, and the mean eye temperature is sensitive to mean accelerator pedal error. A consistent pattern of significance is seen for the longitudinal control output measures as well as RMS accelerator position; these measures are correlated with the SD temperature in all four facial regions. Finally, median speed is significantly correlated with all of the thermal measures.

Table 2. Temperature in nose region correlated with measures of lateral control input ($p < 0.05$ in bold)

		Nose Mean
Mean steering error	r	0.13874
	p	**0.0478**
SD steering error	r	0.14041
	p	**0.0452**
RMS steering angle	r	0.14085
	p	**0.0445**

Table 3. Thermal summary measures aggregated over 60 seconds correlated with longitudinal measures of driver performance (p < 0.05 in bold)

		Overall		Forehead		Nose		Eye	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mean accelerator position error	r	0.1363	0.0345	0.1315	0.0341	0.2159	0.0590	0.1481	0.0371
	p	0.0520	0.6246	0.0608	0.6285	**0.0019**	0.4016	**0.0345**	0.5980
SD accelerator position error	r	0.1098	0.0606	0.1290	0.0618	0.1524	0.0987	0.0987	0.0609
	p	0.1180	0.3893	0.0659	0.3797	**0.0295**	0.1600	0.1601	0.3872
RMS accelerator position	r	-0.0364	0.2375	-0.0427	0.2410	0.0825	0.2361	-0.0135	0.2322
	p	0.6054	**0.0006**	0.5443	**0.0005**	0.2405	**0.0007**	0.8479	**0.0008**
Median speed	r	-0.1990	0.2815	-0.1994	0.3028	-0.1520	0.2809	-0.2212	0.2928
	p	**0.0043**	**<.0001**	**0.0042**	**<.0001**	**0.0299**	**<.0001**	**0.0015**	**<.0001**
RMS speed	r	-0.0681	0.2527	-0.1375	0.2604	-0.0452	0.2366	-0.1058	0.2447
	p	0.3334	**0.0003**	**0.0498**	**0.0002**	0.5209	**0.0007**	0.1319	**0.0004**
SD acceleration	r	0.0043	0.1773	0.0024	0.1795	0.1162	0.1791	0.0314	0.1729
	p	0.9511	**0.0112**	0.9724	**0.0102**	0.0981	**0.0104**	0.6559	**0.0134**

Various thermal measures aggregated over the 15-second windows are significantly correlated with the subjective ratings (see Table 4; in the interest of space the results for only four ratings reported). In general, the SD is the more sensitive measure for the eye, forehead, and overall regions while the mean temperature is more sensitive for the nose region. The SD for the overall and forehead regions are correlated in one direction for negative scales (a higher rating indicates a worse or more negative situation, e.g., frustration) and in the opposite direction for positive scales (a higher rating indicates a better situation, e.g., ease of driving). The SD for the eye region and the mean of the nose, on the other hand, are very sensitive for only the negative scales. Though space does not allow for a direct comparison of the three time windows, the

thermal data are most significantly correlated with the subjective ratings when they are aggregated over the shortest time window (i.e., 15 s).

Table 4. Thermal summary measures aggregated over the first 15 seconds correlated with selected subjective ratings (p < 0.05 in bold)

		Overall		Forehead		Nose		Eye	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mental demand	r	0.0619	0.1989	0.0688	0.2043	0.3236	0.0173	0.2204	0.2366
	p	0.3664	**0.0034**	0.3152	**0.0023**	**<0.0001**	0.8014	**0.0011**	**0.0005**
Frustration	r	0.0707	0.1496	0.0473	0.1268	0.1825	0.0305	0.1095	0.1657
	p	0.3019	**0.0283**	0.4902	0.0636	**0.0073**	0.6561	0.1094	**0.0150**
Ease of driving	r	0.0572	-0.1805	0.0764	-0.2671	0.0159	0.0448	0.0468	0.0209
	p	0.4043	**0.0080**	0.2645	**<.0001**	0.8171	0.5135	0.4948	0.7601
Ease of in-vehicle task	r	0.0795	-0.1678	0.1145	-0.2582	-0.0211	0.0505	-0.0154	-0.0171
	p	0.2456	**0.0137**	0.0941	**0.0001**	0.7589	0.4613	0.8227	0.8028

CONCLUSION

This research explored how facial temperature might reflect the demands that drivers confront when interacting with in-vehicle HMI systems. Facial temperature measures correlated with both measures of driving performance and subjective ratings of workload. The relatively strong relationship between the thermal measures and driving performance measures of speed control rather than lane position suggests that thermal measures are sensitive to different cognitive processes than are typically assessed by measures of steering and lane position. Thermal measures correlated with subjective ratings of workload and these correlations were most significant when aggregated over a 15-second window rather than 30 or 60 seconds. This suggests that, unlike other measures typically used to evaluate in-vehicle systems that are aggregated over long time windows, thermal measures have temporal specificity and might identify specific interactions that increase workload and frustration. Additionally, these results showed that facial temperature aggregated in one area of the face with one particular measure cannot alone capture a complete picture of driver response. The interactions with the in-vehicle systems likely invoked complex affective responses that consisted of various degrees of mental effort, frustration, accomplishment, and other emotions. Future research might induce various states in a more controlled fashion and correlate those states to specific cognitive processes. However, segregating an individual dimension of driver state (e.g., mental effort) from other related dimensions (e.g., frustration and anger) is a challenging and exceptionally complicated task. Rather, composite measures of facial temperature could potentially be developed that offer a more complete profile of driver response. Although aggregating the raw facial temperature data into the thermal measures required substantial effort and further refinement of the data reduction process is required, this research demonstrates that facial thermal measures can provide insight into the demands drivers experience and their affective responses when interacting with in-vehicle HMI systems.

ACKNOWLEDGEMENTS

REFERENCES

Genno, H., Ishikawa, K., Kanbara, O., Kikumoto, M., Fujiwara, Y., Suzuki, R., et al. (1997). Using facial skin temperature to objectively evaluate sensations. *International Journal of Industrial Ergonomics, 19*(2), 161-171.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of experimental and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: North Holland.

Nakayama, O., Futami, T., Nakamura, T., & Boer, E. R. (1999). SAE Technical Paper Series: Development of a steering entropy method for evaluating driver workload. *Human Factors in Audio Interior Systems, Driving, and Vehicle Seating, SP-1426*.

Or, C. K. L., & Duffy, V. G. (2007). Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement. *Occupational Ergonomics, 7*, 83-94.

Pavlidis, I., Eberhardt, N. L., & Levine, J. A. (2002). Seeing through the face of deception. *Nature, 415*(6867), 35.

Pavlidis, I., & Levine, J. (2002). Thermal image analysis for polygraph testing. *IEEE Engineering in Medicine and Biology Magazine, 21*(6), 56-64.

Pollina, D. A., Dollins, A. B., Senter, S. M., Brown, T. E., Pavlidis, I., Levine, J. A., et al. (2006). Facial skin surface temperature changes during a "Concealed Information" test. *Annals of Biomedical Engineering, 34*(7), 1182-1189.

Puri, C., Pavlidis, I., Olsen, L., Levine, J., & Starren, J. (2005, April 2-7). *StressCam: Non-contact measurement of users' emotional states through thermal imaging.* Paper presented at the ACM CHI 2005 Conference on Human Factors in Computing Systems, Portland, OR.

Rimm-Kaufman, S. E., & Kagan, J. (1996). The psychological significance of changes in skin temperature. *Motivation and Emotion, 20*(1), 63-78.

Tsiamyrtzis, P., Dowdall, J., Shastri, D., Pavlidis, I. T., Frank, M. G., & Ekman, P. (2007). Imaging facial physiology for the detection of deceit. *International Journal of Computer Vision, 71*(2), 197-214.

van Winsum, W., Brookhuis, K. A., & de Waard, D. (2000). A comparison of different ways to approximate time-to-line crossing (TLC) during car driving. *Accident Analysis and Prevention, 32*(1), 47-56.