

Evaluation of Affective State Estimations using an On-Line Reporting Device during Human-Robot Interactions

Susana Zoghbi, Dana Kulić, *Member, IEEE*, Elizabeth Croft, *Member, IEEE* and Mike Van der Loos

Abstract— In order to develop a friendly and safe interaction between humans and robots, it is essential for the robot to evaluate user's affective states and respond accordingly. However, affective states are typically assessed using offline questionnaires and user reports. In this paper we investigate the use of an online-device for collecting real-time user reports of affective state during interaction with a robot. These reports are compared to both previous survey reports taken after the interaction, and the affective states estimated by an inference system. The aim is to evaluate and characterize the physiological signal-based inference system and determine which factors significantly influence its performance. This analysis will be used in future work, to fine tune the affective estimations by identifying what kind of variations in physiological signals precede or accompany the variations in reported affective states.

I. INTRODUCTION

IMPROVING productivity and performing dangerous or monotonous tasks have been successfully accomplished by robots in industrial settings. More recently, attention has turned to the use of robots to aid humans outside the industrial environment, in places such as the home or office. For example, as the population in the developed world ages, robots that can interact with humans in a safe and friendly manner, while performing necessary home-care/daily living tasks, would allow more seniors to maintain their independence. Such devices could alleviate some of the non-medical workload from health-care professionals, and reduce growing healthcare costs.

An essential step towards this goal is the acquisition, analysis, and integration of non-explicit user-interaction data within the human-robot control loop, such as user attention and approval. Essentially, the robot should respond to the "body language" of the user. In recent work, the acquired

non-explicit control data from the human user has been integrated into the robot control system to provide a safer and more usable modality for human-robot control. Face direction, body location, and physiological cues such as heart rate, skin conductance, etc. are tracked [1]. Previously, these studies have used off-line user reports to calibrate this information in order to form an estimate of human affective state. In fact, throughout the literature, the most common way of recording the reported mood and intensity of a user for calibration of physiological sensors has been through the aid of pre and post activity questionnaires. This "off-line" information is not ideal when attempting to relate changing physiological responses to reported affective state, and is especially problematic in a dynamic human-robot interaction environment where affective responses may vary substantially throughout the interaction..

Real-time reporting of affective state by the user, is proposed as a more effective way to calibrate user physiologic responses. In this work, we investigate the use of a handheld device that allows people to report their affective state, i.e., valence and arousal, continuously during a human-robot interaction trial. We study how well the online data correlates to both previously employed off-line questionnaire and a physiological signal based affective-state inference system. The main purpose is to evaluate and characterize the physiologically-based estimations and determine which factors significantly influence its performance. The users' reported data can be further utilized to fine tune the affective state estimation system by indentifying the variations in physiological signals that precede or accompany the variations in reported affective states.

II. RELATED WORK

During human-human interaction, non-verbal communication signals are frequently exchanged in order to assess each participant's affective state, focus of attention and intent. Many of these signals are indirect; that is, they occur outside of conscious control. By monitoring and interpreting indirect signals during an interaction, significant cues about the affective state of each participant can be recognized [2]. However it is essential to investigate how well affective inference systems based on physiological signals estimate the actual affective state of the user.

To the authors knowledge, no device has been reported in

Manuscript received March 1, 2009. This work was supported in part by the Canada Natural Science and Engineering Research Council.

S. Zoghbi is with the University of British Columbia, Vancouver, BC V6T1Z4, Canada (e-mail: szoghbi@interchange.ubc.ca)

D. Kulić was with the University of British Columbia, Vancouver, BC V6T1Z4, Canada. She is now with the Nakamura and Yamane Laboratory, Department of Mechano-Informatics, University of Tokyo, Tokyo 113-8656, Japan (e-mail: dana@ynl.t.u.-tokyo.ac.jp).

E. A. Croft is with the University of British Columbia, Vancouver, BC V6T1Z4, Canada (e-mail: ecroft@mech.ubc.ca).

Mike Van der Loos is with the University of British Columbia, Vancouver, BC V6T1Z4, Canada (e-mail: vdl@mech.ubc.ca)

literature that can actively record, on-line, affective response from a human that would be useful for calibration of physiological responses. Devices that record human autonomic responses, such as heart rate, blood pressure, and skin conductance are being used during human robot interactions [3]. In this work, the data collected from these devices was correlated to human emotion through the use of after-the-fact questionnaires.

At the University of Hertfordshire, some research and testing was performed to obtain the real-time comfort level data with a one degree of freedom hand-held device [4, 5]. The obtained data was analyzed in combination with video material so the level of comfort/discomfort of a test subject to a moving robot could be determined with respect to robot action, proximity, and motion relative to the test user.

A Squeezemouse prototype [6, 7] was used to log user dissatisfaction with computer-performed tasks as described in the experiment procedures for human-computer affective interactions. The Squeezemouse operated on pressure sensitivity to indicate the level of frustration experienced by a software user.

Both of the devices listed above reported a one-dimensional level of a specific type of affect, comfort, and frustration.

Physiological monitoring systems have previously been used to extract information about the user's reaction, both for human-computer and human-robot interaction [8-12]. Signals proposed for use in human-computer interfaces include skin conductance, heart rate, pupil dilation and brain and muscle neural activity. Rani et al. [9, 11] used heart-rate analysis and multiple physiological signals to estimate human stress levels during video game playing, using a fuzzy inference engine. Kulic and Croft [13] used a fuzzy inference engine to estimate affective state response to articulated robot motions. Picard et al. [8], and Kim et al. [14] used Support Vector Machines to estimate user affective state for human-computer interaction. Liu et al. [15] compared the effectiveness of several machine learning methods for estimating affective state.

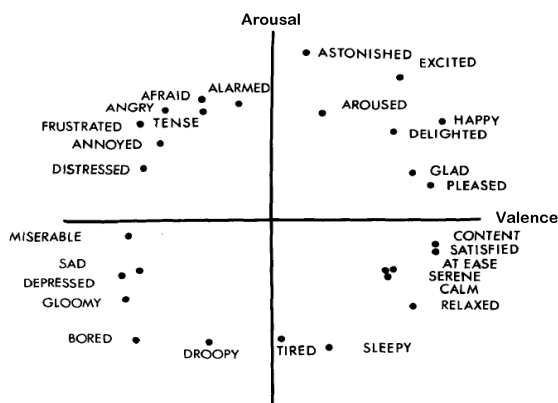


Fig. 1. Circumplex model of affect [23]

All these approaches consider the instantaneous value of various signal features (for example, the rate of change of the skin conductance) for estimating the affective state.

However, physiological response is not characterized by instantaneous changes in value, but rather, by a signal sequence, or waveform. For example, for skin conductance, a stimulus will be followed by a rise in conductance, followed by a slow decay. A good model of the physiological data should capture this time-domain behavior.

In this paper we investigate the use of an online-device for collecting on-line user reports of affective state during interaction with a robot. These reports are compared to previous survey reports taken after the interaction, and the use of these reports to validate a fuzzy inference engine for estimating affective state is considered.

III. METHODOLOGY

Herein, our study proposes and investigates an online method for collection and calibration of physiological signals in an HRI (human-robot interaction) context in order to improve the estimation of the users' affective state response to the interaction. Two instruments were used to collect the data, an in-house developed modified joystick [16] called an Affective State Reporting Device (ASRD) and a questionnaire [17]. This experiment builds on methodology from our previous studies [18] which measured user affective response to HRI, and then utilized the responses in a closed loop system [19].

A. On-line Affective State Reporting Device

For this work, a hand-held affective state reporting device (ASRD) was used to record affective states expressed by each user. A two-dimensional map is used to represent human affective states as valence (pleasantness-unpleasantness) and arousal (activation-deactivation) [23]. This space can be thought of as a Cartesian space with valence on the x-axis and arousal on the y-axis, as shown in Fig. 1. The valence/arousal representation adopted herein appears adequate for the purposes of robotic control. This representation system has also been favored for use with physiological signals and in psychophysiological research [2, 20-22].

A usability study has been performed previously in [16]. The handheld device has two inputs, as shown in Fig. 2 [16]. By using the forward/backward motion, the indication of valence is registered. The forward motion relates to positive affect and the backward motion to negative affect. By squeezing the handle, the level of arousal is recorded where no squeezing represents neutral arousal and forceful squeezing represents high arousal, as only the upper half plane of the circumplex model of affect were mapped. The concepts of valence and arousal were explained to the participants and Fig. 1 was presented to them before using the handheld, as these concepts may not be commonly employed by the average user.

The signals collected from this device were transformed into scores from 1 to 5. For the arousal signal, a score of 1 indicates low arousal level and 5 is high. For the valence signal, 1 indicates negative valence, 5 indicates a positive

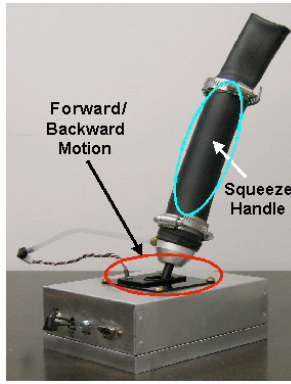


Fig. 2. Hand-Held Affective-State Reporting Device (ASRD)

valence and 3 is neutral.

B. Perceived Safety Questionnaire

Perceived safety describes the user's perception of the level of danger when interacting with a robot, and the user's level of comfort during the interaction. Achieving a positive perception of safety is a key requirement if robots are to be accepted as partners and co-workers in human environments.

After each interaction, participants were asked to rate their emotional state in three semantic differential scales: Anxious / Relaxed, Agitated / Calm, Quiescent / Surprised, as shown in Table I. Previous studies [17] have shown that this semantic differential questionnaire can provide a repeatable and reliable measure for assessing user's perceived safety in response to robot motion.

Users' responses to this questionnaire are compared to both the data collected from the handheld device and the fuzzy inference engine.

TABLE I
PERCEIVED SAFETY - SEMANTIC DIFFERENTIAL SCALES

Anxious	1	2	3	4	5	Relaxed
Agitated	1	2	3	4	5	Calm
Quiescent	1	2	3	4	5	Surprised

C. Affective State Inference

The user's affective state is estimated using a Fuzzy Inference Engine based on three measured physiological signals: heart rate (HR), skin conductance response (SCR) and corrugator muscle contraction (EMG), which have been shown to be the most reliable indicators of affective state in psychophysiological research [2, 21, 22]. Details of the fuzzy rule derivation based on psychophysiological research and the physiological sensing system can be found in [24]. The outputs of the fuzzy engine are the estimated valence and arousal.

For the data analysis, the average arousal and average valence over the duration of each trajectory are calculated for both the affective monitoring device and the fuzzy inference engine. Two-Way Repeated Measures ANOVAs are computed to evaluate the effect of speed and type of task on these measures. A correlation analysis is performed

between the signals as well as between the responses from the questionnaire. Finally, we calculate how much the affective states estimated by the fuzzy inference engine deviate from the subjective online responses collected by the handheld device, which allows to characterize the affect inference system.

D. Study Implementation

A group of 30 people participated in this study; however, due to difficulties with experimental equipment only 26 subject data sets were captured ($n=26$). The group of subjects consisted of 7 females and 19 males. The age of the subjects ranged from 20 to 39, with an average age of 27.4. Most of the subjects were graduate students from a wide range of departments (from Music to Physics) in the University of British Columbia.

Each person was asked to read a description of the experiment and sign a consent form. During the experiment, each participant is seated facing the robot and is asked to use a handheld device to report their affective state while the robot is motion. At the same time, physiological signals data is collected using appropriate sensors attached to the person. Fig. 3 shows the setup of the experiment. The robot is initially held motionless for 60 seconds to collect baseline physiological data for each subject. The robot then executes 7 different trajectories. The trajectories are presented to each subject in randomized order. After each trajectory is executed, the subject is asked to rate their affective state according to a questionnaire of Perceived Safety [17]. After the questionnaire is completed, a 1 minute's rest period is enforced before presenting the next trajectory, to ensure that the physiological data returns to baseline.

The physiological data collected is processed by a fuzzy inference engine to estimate the levels of valence and arousal the user experiences. These estimations are compared to both the subject's reported affective states made through the handheld device and the answers provided on the Perceived Safety Questionnaire.

The experiment was performed using the CRS A460 6 degree of freedom (DoF) manipulator. Two different tasks were used for the experiment: a pick-and-place motion (PP), and a reach and retract motion (RR). These tasks were chosen to represent typical motions an articulated robot



Fig. 3. Experimental setup

manipulator could be asked to perform during human-robot interaction, for example during hand-over tasks. For the pick-and-place motion, the pick location was specified to the right and away from the subject, and the place location was directly in front and close to the subject. For the reach and retract motion, the reach location was the same as the place location. For both tasks, the robot started and ended in the “home” upright position. The main difference between the two tasks is the approach direction of the robot. For the PP

TABLE II
DESCRIPTION OF POTENTIAL FIELD TRAJECTORIES USED

Medium Speed Pick & Place task
Medium Speed Reach & Retract task
High Speed Pick & Place task
High Speed Reach & Retract task

task, the robot approaches the subject from the side, while during the RR motion, the robot approaches the subject from head on. Two different speeds were used for each task: medium and high as seen on Table II.

Each subject was tested once over a contiguous time period of approximately 15 minutes. Single trials of multiple subjects were selected over multiple trials of a single subject in order to capture a general response to the robot motions.

IV. RESULTS

A. Subjective response by ASRD

Fig. 4 presents the average responses of arousal and valence given by the participants using the Affective State Reporting Device for each trajectory. Table III shows the corresponding Two-Way Repeated Measures ANOVA tables, where the effects of speed and type of task were analyzed.

Speed was the only significant factor on the subject’s reported average arousal $F(1,25) = 10.50, p = 0.003$. For high speed trajectories, users reported stronger values of arousal than for medium speed; as expected. As well, there is a moderate correlation between average arousal and speed, as it is shown in Table VI.

For the user’s reported average valence, speed was also a significant factor, $F(1,25) = 28.98, p = 0.000$. Users reported lower valence for high speed than for medium. In fact, there is a strong correlation between speed and subjective reported average valence, as seen on Table VI. Additionally, in this case, the type of task was also a significant factor, $F(1,25) = 25.45, p = 0.000$. Participants expressed lower valence for the Reach and Retreat (RR) task than for the Pick a person influences their degree of comfort. For the RR motion, the robot approaches the subject from over head on while during the PP task, the robot approaches the subject from the side. The interaction effect between speed and task type was not significant.

B. Subjective response by Perceived Safety Questionnaire

Fig. 5 shows the mean and standard deviations of responses for the three scales measured: Anxious/Relaxed, Agitated/Calm and Quiescent/Surprise. Table IV presents the two-way repeated measures ANOVA for each scale. Speed and type of task are both significant factors for each of the three scales measured. Participants reported higher anxiety, agitation and surprise for high speed motions as well as for the Reach and Retract task. There is a strong correlation between speed and each of the three scales measured on the questionnaire, Table VI. These results follow a similar behavior as the measures taken from the ASRD, which indicates a correspondence between them. In effect, there are strong correlations between the average valence reported on the ASRD and both anxiety and agitation. Additionally, there are moderate correlations between average arousal as reported by ASRD and each of the three scales measured on the questionnaire.

These results validate the suitability of the ASRD and suggest that it is an effective device for collecting user responses. This finding was corroborated by interviews with the users after the experiment.

C. Arousal response estimated by Fuzzy Inference Engine

Fig. 6 presents the average responses of arousal estimated by the Fuzzy Inference Engine. Table V shows the corresponding Two-Way Repeated Measures ANOVA tables, where the effects of speed and type of task are analyzed. Both factors and their interaction are all significant. Fig. 7 shows this interaction effect.

Breaking down the interaction into simple main effects for each task, it is found on Table VII that, for the Pick and Place task, speed is a significant factor. For high speed motions, the fuzzy inference engine estimated higher average arousal levels. For the Reach and Retreat task, speed is not a significant factor, that is, the average arousal level does not vary significantly for different levels of speed.

Only weak correlations –if any– were found between the average arousal estimated from the fuzzy inference engine and the user’s reported affective states using the handheld device and questionnaire, Table VI.

Given that the estimations don’t seem to follow a similar pattern as the one reported by the user, it is natural to address the question of how much deviation is there from the subjective response. Is it possible to determine whether fuzzy inference engine is over or underestimating user’s responses? Which factors and conditions influence the accuracy of affect estimations? Answering these questions would allow us to characterize the inference system and determine under what conditions they are better used. This could not be done before since there was not a mechanism or device to allow the user to report their affective state continuously at the same time as the robot motions occur and physiological data is collected.

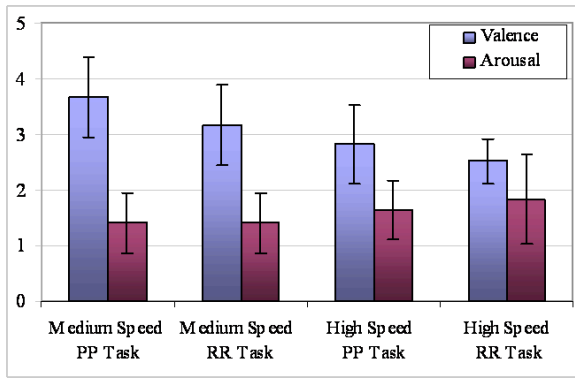


Fig. 4 Average Valence and Average Arousal from ASRD

TABLE III
TWO-WAY REPEATED MEASURES ANOVA FOR AVERAGE AROUSAL AND AVERAGE VALENCE FROM ASRD

Source of Variation	Average Arousal ASRD				Average Valence ASRD			
	df	MS	F	p	df	MS	F	p
Speed	1	2.81	10.50	0.003	1	14.08	28.98	0.000
Error (Speed)	25	0.27			25	0.49		
Task	1	0.24	1.12	0.299	1	4.33	25.45	0.000
Error (Task)	25	0.22			25	0.17		
Speed x Task	1	0.25	1.73	0.200	1	0.24	0.814	0.376
Error	25	0.15			25	0.30		

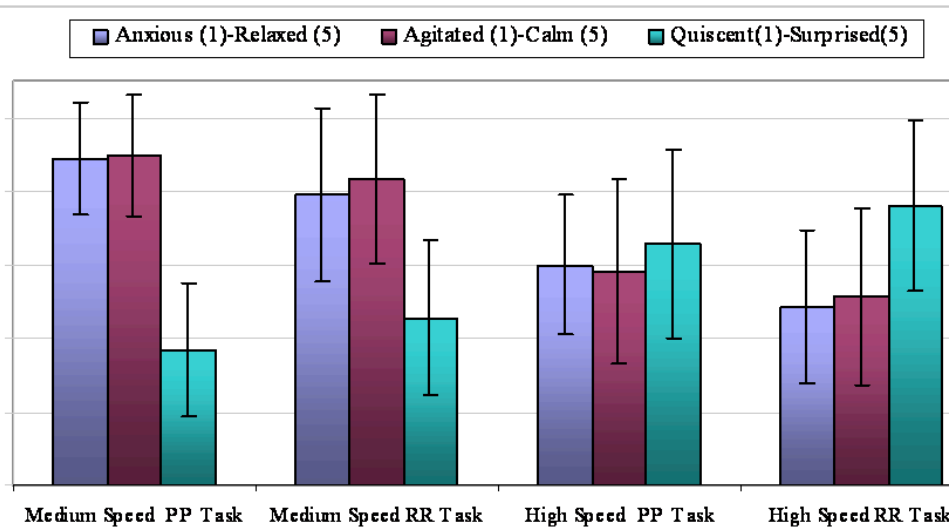


Fig. 5 Descriptive Statistics of Perceived Safety Questionnaire

TABLE IV
TWO-WAY REPEATED MEASURES ANOVA FOR PERCEIVED SAFETY QUESTIONNAIRE

Source of Variation	df	<i>Anxious-Relaxed</i>			<i>Agitated-Calm</i>			<i>Quiescent-Surprised</i>		
		MS	F	p	MS	F	p	MS	F	p
Speed	1	54.76	78.51	0.000	62.41	65.58	0.000	54.76	35.29	0.000
Error (Speed)	24	0.70			0.95			1.55		
Task	1	6.76	9.15	0.006	2.89	5.19	0.006	5.76	5.27	0.031
Error (Task)	24	0.74			0.56			1.09		
Speed x Task	1	0.04	0.08	0.775	0.01	0.02	0.775	0.04	0.07	0.788
Error (Speed x Task)	24	0.48			0.43			0.54		

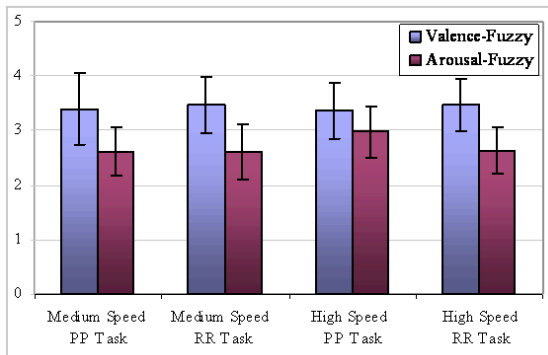


Fig. 6 Average Valence and Average Arousal estimated by the Fuzzy Inference Engine

TABLE V
TWO-WAY REPEATED MEASURES ANOVA FOR AFFECTIVE STATES FROM FUZZY INFERENCE ENGINE

Source of Variation	Average Arousal ASRD				Average Valence ASRD			
	df	MS	F	p	df	MS	F	p
Speed	1	0.86	7.29	0.012	1	0.009	0.082	0.777
Error (Speed)	25	0.12			25	0.109		
Task	1	0.82	5.00	0.035	1	0.238	3.35	0.079
Error (Task)	25	0.16			25	0.071		
Speed x Task	1	0.76	6.02	0.021	1	0.006	0.041	0.841
Error (Speed x Task)	25	0.13			25	0.140		

TABLE VI
CORRELATION ANALYSIS

		Anxious-Relaxed	Agitated-Calm	Quiescent-Surprised	Speed	Average Arousal ASRD	Average Arousal Fuzzy	Average Valence ASRD	Average Valence Fuzzy
Anxious-Relaxed	Corr.	1							
	p-value								
Agitated-Calm	Corr.	0.843	1						
	p-value	0.000							
Quiescent-Surprised	Corr.	-0.735	-0.741	1					
	p-value	0.000	0.000						
Speed	Corr.	-0.591	-0.581	0.555	1				
	p-value	0.000	0.000	0.000					
Average Arousal ASRD	Corr.	-0.398	-0.261	0.323	0.253	1			
	p-value	0.000	0.009	0.001	0.011				
Average Arousal Fuzzy	Corr.	-0.163	-0.172	0.226	0.213	0.037	1		
	p-value	0.105	0.086	0.024	0.033	0.718			
Average Valence ASRD	Corr.	0.528	0.505	-0.449	-0.468	-0.346	-0.170	1	
	p-value	0.000	0.000	0.000	0.000	0.000	0.091		
Average Valence Fuzzy	Corr.	-0.122	-0.082	0.073	-0.017	0.107	0.195	-0.168	1
	p-value	0.225	0.418	0.468	0.866	0.289	0.052	0.096	

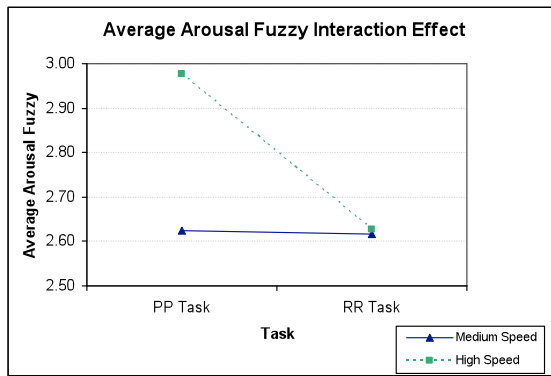


Fig. 7 Average Arousal estimated by Fuzzy – Interaction Effect

TABLE VII
BREAKING DOWN INTO SIMPLE MAIN EFFECTS

Source of Variation	df	MS	F	p
Speed at PP Task	1	1.62	15.95	0.001
Error	25	0.101		
Speed at RR Task	1	0.002	0.012	0.913
Error	25	0.143		

TABLE VIII
DESCRIPTIVE STATISTICS IMSE

	n	Mean	Std Dev
Speed Medium – PP Task	26	2.99	1.66
Speed Medium – RR Task	26	3.52	1.93
Speed High – PP Task	26	4.21	1.96
Speed High – RR Task	26	2.90	1.34

In order to characterize the estimated signal, we compare the data collected from the user and the estimations generated by the fuzzy inference engine.

Calculating how much fuzzy estimation deviates from the user's reported affect state will give a measure of how accurate the estimations are. We define the Normalized Integrated Mean Square Error as,

$$IMSE = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2, \quad (1)$$

where X is the estimated affective state calculated by the fuzzy inference engine, Y is the affective state reported by the user and N is the number of measurements for a given trajectory. The average $IMSE$ and standard deviation per trajectory are presented in Table VIII.

One can note that, for the scale used in both fuzzy estimations and user's data, the lower and upper boundary values for $IMSE$ are given by

$$0 \leq IMSE \leq 16. \quad (2)$$

Table IX presents the Two-Way Repeated Measures ANOVA, where the effects of both speed and task on the value of $IMSE$ are assessed. The interaction between speed and type of task is found significant. Fig. 8 depicts this interaction effect.

Breaking down the interaction effect by treating each task separately, it can be seen that for the Pick and Place task there is a significant effect of speed on the $IMSE$, Table X. There is a larger $IMSE$ for high speed than for medium. However, for the Reach and Retreat task, speed is not a significant factor, that is regardless of speed level, the $IMSE$ values tend to be about the same. Thus the fuzzy inference engine tends to perform better for the Reach and Retreat task at both speed levels and for the Pick and Place task at

TABLE IX
TWO-WAY REPEATED MEASURES ANOVA FOR IMSE

Source of Variation	df	MS	F	<i>p</i>
Speed	1	2.31	1.08	0.308
Error (Speed)	25	2.13		
Task	1	3.98	2.17	0.153
Error (Task)	25	1.84		
Speed x Task	1	22.18	8.61	0.007
Error (Speed x Task)	25	2.58		

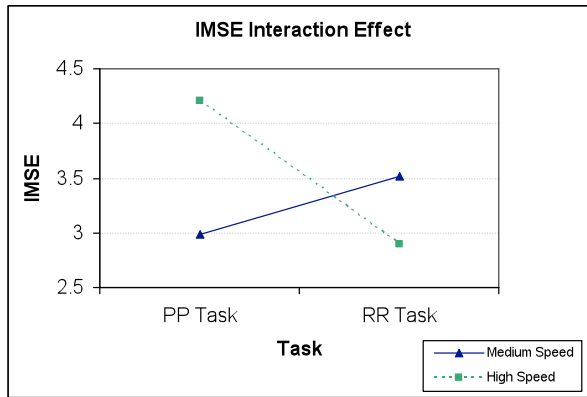


Fig. 8. IMSE Interaction Effect

medium speed. For the Pick and Place task at high speed, the fuzzy inference engine tends to overestimate more the levels of arousal.

From Fig 4 one can note that, the lowest arousal levels reported by the user occur at medium speed for both tasks, whereas the highest arousal occurs for the Reach and Retreat task at high speed. For the Pick and Place task at high speed, the user's level of arousal lies in between the maximum and minimum values. Hence, the fuzzy inference engine tends to perform better when the user's reported arousal is either high or low. However, it produces larger *IMSE* when users reported intermediate arousal levels.

TABLE X
TWO-WAY REPEATED MEASURES ANOVA FOR AROUSAL IMSE -
BREAKING DOWN THE INTERACTION EFFECT

Source of Variation	df	MS	F	<i>p</i>
Speed at PP Task	1	19.40	7.35	0.012
Error	25	2.64		
Speed at RR Task	1	5.09	2.47	0.129
Error	25	2.07		

D. Valence Response estimated by Fuzzy Inference Engine

The affective state inference engine also attempted to estimate valence, based on the corrugator muscle EMG signal and heart activity. Corrugator muscle activity has been reported to have strong correlation with negative valence in existing studies, and had shown promise during earlier studies with the inference engine using images as stimuli. However, for most of the trajectories tested, the estimated valence was very neutral. Corrugator muscle

activity was primarily at baseline for all subjects tested during robot motions. Heart rate and skin conductance response did not allow for discrimination of valence either [9]. Therefore, no further analysis was performed.

V. DISCUSSIONS AND CONCLUSIONS

In this paper we have presented the implementation of a new handheld device that permits the users to report their affective state, i.e. valence and arousal, continuously during a human-robot interaction trial. The Affective State Reporting Device was shown to be an effective tool for collecting subjective responses and its data was correlated to a previous questionnaire of perceived safety. Earlier findings were verified. Participants tend to feel more discomfort for high speed motions where the robot approaches the subject from over head.

The ASRD provides a new source of information that allows us to evaluate and characterize the performance of an affective state inference system., by calculating how much deviation it exists between the estimations and the actual affective state the user experiences at every time step during the motion. By analysing the behaviour of deviations, it is possible to determine under which conditions the inference system can be best used and which factors influence its performance. In particular, we calculated the deviation between the estimated signals from a fuzzy inference engine and the users' reported states, by using the Integrated Mean Square Error over the duration of the motion. The analysis of the *IMSE* showed that the fuzzy inference engine is likely to perform better when the user's reported arousal is either high or low. However, it produced larger deviations when users reported intermediate arousal levels.

Being able to determine when the participants experience a change in their comfort level and its intensity will help identify what kind of variations in physiological signals -and their extracted features- precede or accompany the variations in reported affective states. This would lead to fine tune the fuzzy inference engine (or analogously, other affective inference systems) so that deviations -i.e. *IMSE*- from the users' reported signal are minimized.

As a final remark, it is important to comment on the potential drawbacks when using the ASRD. Participants may fail to recognize their affective state continuously, get distracted or become startled by the robot motions, thus failing to accurately report their affect. Through observations during the experiments, it became clear that it is also important to take into account subject individual differences. For example, one male subject reported high levels of anxiety, agitation and surprise in the questionnaire, yet he reported positive levels of valence using the ASRD. When asked why, he said that he enjoyed being surprised and agitated by the robot. Another example, a female subject reports in a previous interview that she really likes robots. When she performs the experiments, she's noticeably startled by the robot motions, yet she reports high levels of relaxation, calmness and quiescence. Accordingly, she also

reported high levels of positive valence using the ASRD. It seemed she had consciously decided she would like the robot before doing the experiment. These examples underline the importance of individual normalization for both the affective inference system and the online tool for collecting user data. A more thorough training protocol for using the ASRD would also lead to improvements.

ACKNOWLEDGMENT

The authors wish to acknowledge the invaluable collaboration of Yuh-Jia Lim, whose diligent work has allowed us to complete this paper.

REFERENCES

- [1] D. Kulic and E. A. Croft, "Real-time safety for human - robot interaction," presented at Robotics and Autonomous Systems, 2005.
- [2] R. Picard, *Affective computing*, MIT Press, Cambridge, Massachusetts, 1997.
- [3] D. Kulic, "Safety for Human-Robot Interaction," *Ph.D. Thesis*, 2005.
- [4] K. L. Koay, M. L. Walters, and K. Dautenhahn, "Methodological Issues Using a Comfort Level Device in Human-Robot Interactions," *Proceedings IEEE of RO-MAN 2005 (Nashville TN, August 2005)*, pp. 359-364, 2005.
- [5] K. L. Koay, K. Dautenhahn, S. N. Woods, and M. L. Walters, "Empirical Results from Using a Comfort Level Device in Human-Robot Interaction Studies," *Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, Salt Lake City, Utah, USA*, pp. 194-201, 2006.
- [6] D. Kirsch, "The Sentic Mouse: Developing a tool for Measuring Emotional Valence," *Bachelor of Science in Brain and Cognitive Sciences, MIT*, 1997.
- [7] C. Reynolds and R. W. Picard, "Designing for Affective Interactions," *Proceedings from the 9th International Conference on Human-Computer Interaction, New Orleans*, 2001.
- [8] R. Picard, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175 - 1191, 2001.
- [9] P. Rani, N. Sarkar, C. A. Smith, and L. D. Kirby, "Anxiety detecting robotic system - towards implicit human-robot collaboration," *Robotica*, vol. 22, pp. 85-95, 2004.
- [10] T. Partala, M. Jokiniemi, and V. Surakka, "Pupillary Responses to Emotionally Provocative Stimuli," presented at Eye Tracking Research and Applications Symposium, pp. 123 - 129, 2000.
- [11] P. Rani, J. Sims, R. Brackin, and N. Sarkar, "Online stress detection using psychophysiological signals for implicit human-robot cooperation," *Robotica*, vol. 20, pp. 673-685, 2002.
- [12] J. Scheirer, R. Fernandez, J. Klein, and R. Picard, "Frustrating the User on Purpose: a step toward building an affective computer," *Interacting with Computers*, vol. 14, pp. 93 - 118, 2002.
- [13] D. Kulic and E. Croft, "Anxiety Detection during Human-Robot Interaction," presented at IEEE International Conference on Intelligent Robots and Systems, pp. 616 - 621, 2005.
- [14] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical & Biological Engineering & Computing*, vol. 42, pp. 419 - 427, 2004.
- [15] C. Liu, P. Rani, and N. Sarkar, "An Empirical Study of Machine Learning Techniques for Affect Recognition in Human-Robot Interaction," presented at IEEE Conference on Intelligent Robots and Systems, pp. 2051 - 2056, 2005.
- [16] A. Lanz and E. Croft, "On Line- Affective State Monitoring Device Design " presented at the 2007 ASME International Mechanical Engineering Congress and Exposition, Seattle, Washington, USA, 2007.
- [17] C. Bartneck, D. Kulic, E. Croft, S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots". *International Journal of Social Robotics*, 2009.
- [18] Kulic, D. and E. A. Croft. "Affective State Estimation for Human-Robot Interaction." *IEEE Transactions on Robotics* 23(5): 991-1000., 2007.
- [19] Kulic, D. and E. A. Croft (2007). "Pre-Collision Safety Strategies for Human-Robot Interaction." *Autonomous Robots* 22(2): 149-164.
- [20] M. M. Bradley, "Emotion and Motivation," in *Handbook of Psychophysiology*, J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, Eds., 2 ed. Cambridge: Cambridge University Press, 2000, pp. 602 - 642.
- [21] M. M. Bradley and P. J. Lang, "Measuring Emotion: Behavior, Feeling and Physiology," in *Cognitive Neuroscience of Emotion*, R. D. Lane and L. Nadel, Eds. New York: Oxford University Press, 2000.
- [22] P. J. Lang, "The Emotion Probe: Studies of Motivation and Attention," *American Psychologist*, vol. 50, no. 5, pp. 372 - 385, 1995.
- [23] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161, 1980.
- [24] Kulic, D. and E. A. Croft. "Experimental Validation of Affective State Estimation and Safe Motion Planning during Human Robot Interaction." *Robotica* 25(1): 13-27, 2007.