

AFFECTIVE EVALUATION OF A MOBILE MULTIMODAL DIALOGUE SYSTEM USING BRAIN SIGNALS

Manolis Perakakis and Alexandros Potamianos

Dept. of Electronics and Computer Engineering, Technical University of Crete, Chania 73100, Greece

{perak, potam}@telecom.tuc.gr

ABSTRACT

We propose the use of affective metrics such as excitement, frustration and engagement for the evaluation of multimodal dialogue systems. The affective metrics are elicited from the Electroencephalography (EEG) signals using the Emotiv EPOC neuroheadset device. The affective metrics are used in conjunction with traditional evaluation metrics (turn duration, input modality) to investigate the effect of speech recognition errors and modality usage patterns in a multimodal (touch and speech) dialogue form-filling application for the iPhone mobile device. Results show that: 1) engagement is higher for touch input, while excitement and frustration is higher for speech input, and 2) speech recognition errors and associated repairs correspond to specific dynamic patterns of excitement and frustration. Use of such physiological channels and their elaborated interpretation is a challenging but also a potentially rewarding direction towards emotional and cognitive assessment of multimodal interaction design.

Index Terms— Affective interface evaluation, Multimodal interaction, Speech interaction, Graphical user interfaces

I. INTRODUCTION

In human communication, affect and emotion play an important role, as they enrich the communication channel between the interacting parties. Recently there has been much research interest in the CHI community aiming at incorporating affective and emotional cues in the human computer interaction loop. These efforts are known collectively as affective computing [13].

Multimodal spoken dialogue systems are traditionally evaluated with objective metrics such as interaction efficiency (turn duration, task completion, time to completion), error rate, modality selection and multimodal synergy [12], [11]. The methodology proposed in this paper is based on the use of Electroencephalography (EEG), a rich source of information which is able to reveal hints of both affective and cognitive state during an interaction task. In this work, we investigate the use of EEG elicited affective metrics such as excitement, frustration and engagement for the evaluation of interactive systems and multimodal dialogue systems in particular. This, not only provides a more qualitative approach to evaluation, it also provides a better understanding of the interaction process from the user perspective. Extracting robust information from such physiological channels is a challenging but also potentially rewarding task, opening new avenues for the emotional and cognitive assessment of multimodal interaction design. To our knowledge, this is the first effort for the affective evaluation of multimodal spoken dialogue systems using brain signals.

The paper is organized as follows: In Section II a brief introduction to affective computing is presented. Section III describes the EEG apparatus used along with the software developed (affective evaluation studio) to record and analyze the evaluation sessions. Section IV outlines the evaluation procedure, the evaluated system and the evaluation metrics used. Section V presents the affective evaluation results that complement the use of traditional evaluation techniques. We further discuss issues of affective evaluation in Section VI. Finally in Section VII the conclusions and future work are presented.

II. AFFECTIVE COMPUTING

Affective computing studies the communication of affect between humans and computational systems. It is an interdisciplinary area of

research, spanning disciplines such as psychology (study of affect and emotion) cognitive science (emotion and memory, emotion and attention) computer engineering and design (signal processing, affective detection and interpretation, affective design). In the realm of HCI and context awareness, it aims at supporting enhanced interaction experience by utilizing the affective dimension of communication.

The main efforts until now have been concentrated in the fields of affective detection and emotion recognition¹. Affective computing relies on detection of emotional cues in channels such as speech (emotional speech), face (facial affect detection) and body gestures. It also utilizes a number of physiological channels such as galvanic skin response (GSR), facial electromyography (EMG), blood pressure, heart rate monitoring (EKG) and pupil dilation. All these channels have been shown to correlate with certain emotional states such as fear, joy, surprise. Such signals can potentially provide valuable information in the course of interface evaluation. Galvanic skin response (GSR) or skin conductance measures the electrical conductance of the skin, which varies with its moisture level. For example, GSR is used as an indication of psychological or physiological arousal since sweat glands are controlled by the sympathetic nervous system. Previous studies indicate that GSR may correlate not only to emotional (e.g. arousal [9]) but also to cognitive (e.g. cognitive load [15]) activity. Using more than one channels (e.g. facial expression and EKG) is common in the research community and is referred to as multimodal affect recognition.

Advancements in cognitive and brain sciences has recently made it possible to add the brain as another source of rich information. Detection of emotions using brain signals (EEG) is an active research effort. Using brain signals we can also study other cognitive functions that are of high significance in the context of interaction design such as attention, memory and cognitive load.

The human brain is the more complex human organ. The study and understanding of the brain has benefited from recent advances in sophisticated equipment such as functional magnetic resonance imaging (fMRI). Compared to EEG (that is an older method used in brain studies) most of these methods are very expensive, have low data transmission rates and/or are not ambulatory. EEG is an established and mature technology that can be used outside the lab, has high temporal resolution (which makes it ideal for interaction evaluation) and is relatively cheap. The main drawback of EEG compared to the other methods is the relative poor spatial resolution and the high noise from non cognitive sources called artifacts.

Fig. 1 shows the main areas of the human brain called lobes in different colors along with their main functionality. EEG measures the electric potential of the scalp by detecting the summation of the synchronous activity of thousands or millions of neurons. Using surface electrodes at various scalp locations it can reliably detect even small such changes in the cerebral cortex. A large number of electrodes are usually used in clinical settings to allow for a relative adequate spatial resolution. The placement of the electrodes follows usually the 10-20 standard to allow for reproducibility across a subject's measurements or between subjects. Locations are identified by a letter corresponding to the lobe (e.g. F for frontal, P for parietal, etc) and a number that identifies the hemisphere location.

¹Note that affect is generally considered to be the expression of emotions and as such is used in the context of this paper.

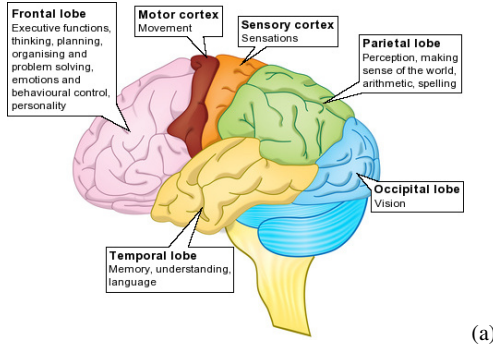


Fig. 1. Human brain areas

The brain activity produces a rhythmic signal which can be divided in several frequency bands: delta (up to 4Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), gamma (>30Hz) and mu rhythm(8-13 Hz). Alpha waves are typical of an alert but relaxed mental state and are evident in the parietal and occipital lobes. Beta waves are indicative of active thinking and concentration, found mainly in frontal and other areas of the brain.

EEG studies more relevant to the domain of HCI and affective computing are those studying emotions and fundamental cognitive processes related to attention and memory. EEG emotion recognition has been an active topic in the last years. There are various representations of emotions such as the wheel of emotions by Plutchik [14] and the mapping of various emotions in the arousal-valence space, one of the most used frameworks in the study of emotions. Arousal is the degree of awakesness and reactivity to stimuli and valence is the positiveness degree of a feeling. According to previous studies, indicative metrics of arousal is the beta/alpha band power ratio in the frontal lobe area. For valence the alpha ratio of frontal electrodes (F3, F4) has been used, as according to [5] there is hemisphere asymmetry in emotions regarding valence e.g. positive emotions are experienced in the left frontal area while negative emotions on the right frontal area.

User state estimation based on cognitive attributes related to attention and memory (in addition to emotions) is also of great importance in the context of HCI research. Memory load, an index of cognitive load [10] is an important index of mental effort while carrying out a task. Memory load classification has thus drawn attention from the HCI research community since it can reveal qualitative parameters of an interface. In [7] authors report a classification accuracy of 99% for two and 88% for four different levels of memory load, by exploiting data from the *N-back* experiment [6] for their classifier. They also argue that previous research findings that high memory loads correlate with increase in theta and low-beta(12-15 Hz) bands power in the frontal lobe or the ratio of beta/(alpha+theta) powers may not hold always true.

III. AFFECTIVE EVALUATION STUDIO

III-A. EEG device

The EEG device used is the Emotiv EPOC², a 14 electrode consumer neuroheadset device (see Fig. 2). The main advantage of the device compared to clinical grade EEG devices prices (tenths of thousands of dollars) is the very low price (\$700 for research edition). In addition, the device is very easy to use and the preparation time is very short (only few minutes to apply saline solution) compared again to clinical EEG systems which require enough time and expertise in order to use. It is also wireless allowing the users to freely move while interacting. Also, the provided SDK provides a suite of detections (affective, expressive and cognitive) which allow people without EEG expertise to integrate them to any application (research edition offers access to raw EEG signals).

²<http://www.emotiv.com>



Fig. 2. Evaluation setting. Depicted counter clockwise is the iPhone device, the GSR apparatus (arduino and breadboard), the Emotiv device, the audio headset and the PlayStation Eye camera.

The affective suite measures several affective metrics such as frustration, engagement and excitement, while the expressive suite detects user's face expressions. The cognitive suite allows the mapping of different cognitive patterns to different actions allowing EPOC to be used as a BCI (brain computer interface) [1] device. The main critique is the low number of electrodes (compared to e.g., 128 electrodes of clinical grade EEG) and the lower sensitivity/higher noise of EEG measurements compared to clinical grade EEG devices. Nevertheless, using advanced noise filtering techniques one can solve most of these issues. As a result, the device has been actively used recently by game developers, individual researchers and HCI labs around the world.

III-B. Affective Evaluation Studio

A dedicated tool was developed to collect in real time, data from the Emotiv device (affective and EEG³) and a video camera. Screenshots of the affective studio are shown in Fig. 3 for the standard and research versions of the Emotiv SDK respectively. The tool was used to capture, record, replay and analyze evaluation sessions of the multimodal interaction system. A short list of its capabilities include: In capture mode, it captures data from Emotiv device and a video camera. This is useful for the examiner to check and resolve any problems such as the correct contact quality of Emotiv or GSR before starting the recording of a new session. In recording mode, affective, EEG and video data are all concurrently saved while also been displayed for the duration of the interaction session. Again the real time information is used by the examiner to ensure for the correctness of the recording sessions. In play mode, data are displayed, annotated and analyzed (e.g., affective annotations and EEG spectrograms and scalpmaps - see Fig. 3(b)) offering valuable insights for the course of an interaction session. The Studio serves as a valuable tool for inspecting in detail how users interact with the system in real time.

One important problem in EEG analysis is that the recorded signal might be affected by various artifacts caused by eye movements, eye blinks and muscle activity. To alleviate this, we have implemented an EEG artifact removal algorithm using the Independent Component Analysis (ICA) method proposed in [8].

IV. AFFECTIVE EVALUATION

IV-A. Evaluated System

The evaluated system is built using the multimodal spoken dialogue platform described in [12] for the travel reservation application domain (flight, hotel and car reservation). The multimodal system is implemented for both the desktop and smartphone environments (running on an iPhone device). Our analysis focuses on the form-filling part of the application.

The user can communicate with the system using touch and speech. Overall, five different interaction modes are evaluated; two unimodal ones, namely, "GUI-Only" (GO) and "Speech-Only" (SO), and three

³Although we focus on EEG based affective metrics, GSR data were also collected during the evaluation.

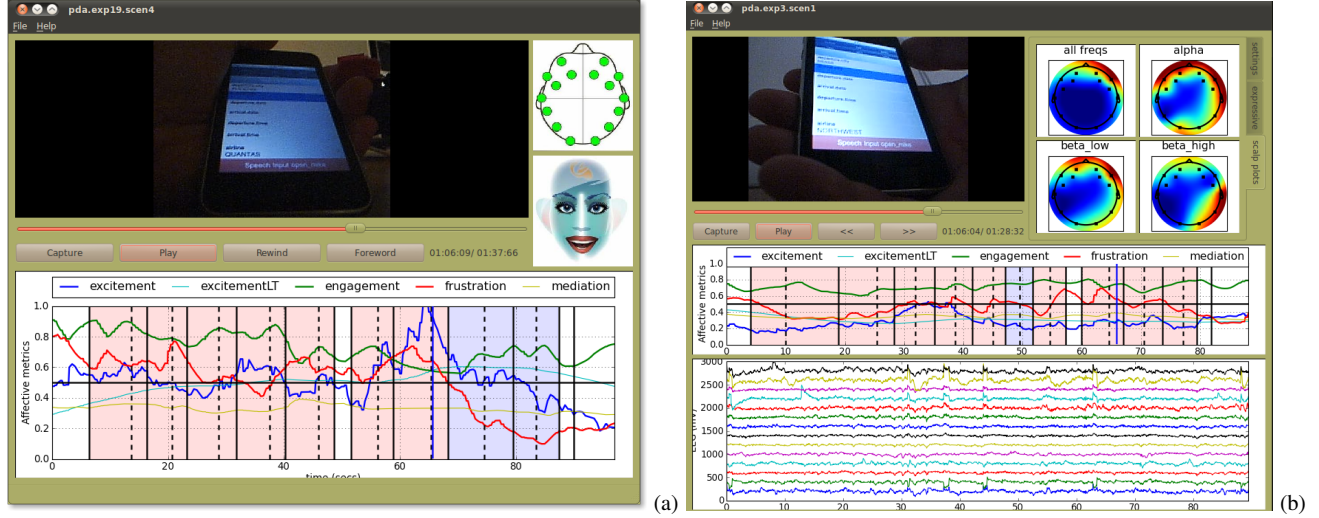


Fig. 3. Screenshots of the affective evaluation studio replaying previously recorded sessions. (a) Standard edition. The two main components depicted are the video and affective plot (see Fig. 4) widgets. The vertical blue line indicates the playing position in the affective data corresponding to video frame displayed. The user can click on any position of the plot to move in that particular moment in the video stream or vice versa using the video slider. The two widgets in the right of the video widget display the 14 electrode contact quality and the user face expression widget. (b) Research edition. Offers additional EEG processing capabilities such as EEG plot (found below affective plot) and single channel analysis plot and spectrogram (shown when selecting specific channel). It also provides real time scalp plots (next to video widget) which show EEG power distribution for selected spectrum bands animated through time.

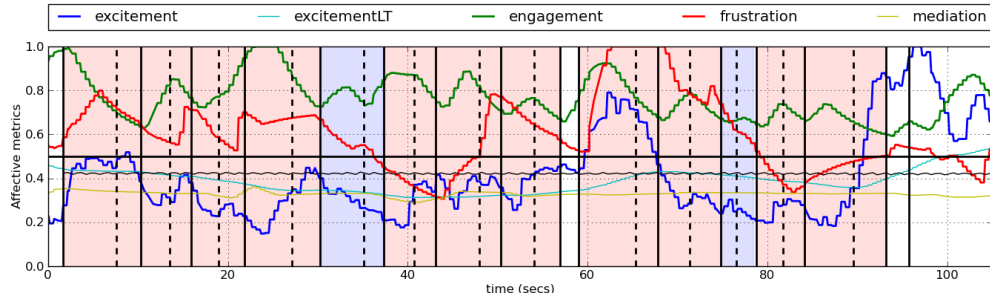


Fig. 4. Example interactive session annotated in the affective plot. Annotation projects the multimodal system’s log file information (turn duration, input type, etc) onto the affected data of a recorded session. The five affective metrics (excitement, long term excitement, engagement, frustration and mediation) provided by EPOC are depicted, along with the GSR values (black horizontal line oscillating around 0.4) in the [0-1] scale. The software automatically annotates the plot showing all interaction turns. A turn is the time period between two thick vertical lines; each dotted vertical line separates a turn into the inactivity and interaction periods. Only fill turns have background color. That color is red for speech turns and blue for GUI turns.

fully multimodal ones, namely, “Click-to-Talk” (CT), “Open-Mike” (OM) and “Modality-Selection” (MS). The main difference between the three multimodal interaction modes is the default input modality. For CT interaction, touch is the default input; the user needs to tap on the “Speech Input” button to override the default input modality. For OM interaction, speech is the default input modality and the system is always listening for a voice-activity event. Again the user can override by tapping anywhere on the screen. MS is a mix of the CT and OM interaction; the system automatically switches between the two multimodal modes depending on interface efficiency considerations (the number of options available for each field in the form). A detailed description of the interaction modes⁴ can be found in [12].

IV-B. Participants and Procedure

For this evaluation study, eight healthy right handed graduate university students participated. They were all briefly introduced to the

nature of the experiment. After wearing the Emotiv headset and the GSR apparatus, they were asked to take a comfortable position and instructed to avoid excess movement. All five interaction modes were used during the evaluation (SO, GO and three multimodal ones CT, OM, MS). Participants tried all different systems at least once in order to get familiar with the systems before starting the evaluation scenarios. For the evaluation scenarios, four different two way trip scenarios were used, that is a total of 20 (5 systems \times 4 scenarios) sessions per user.

IV-C. Evaluation Metrics

The following (traditional) objective evaluation metrics have been estimated: 1) Input modality usage: we measure the usage of each input modality as a function of number of turns, and duration of turns attributed to each modality (touch, speech). 2) Input modality overrides: the number of turns where users preferred to use a modality other than the one proposed by the system. 3) Turn/dialogue duration, inactivity and interaction times: In addition to measuring turn and

⁴For a video demonstration see <http://goo.gl/rIjS3>

dialogue duration in total and for each input modality, we also further refine turn duration into interaction and inactivity times. Inactivity time refers to the idle time interval starting at the beginning of each turn, until the moment the user actually interacts with the system using touch or speech input (interaction and inactivity sum up to turn duration). The above metrics are also estimated as a function of interaction context (type of field that is getting filled) and user.

Before conducting the experiments, we verified Emotiv’s affective metrics against manually annotated ratings created using the FEEL-TRACE toolkit [4]. For each session the (raw EEG and) affective metric signals provided by the Emotiv device (engagement, excitement, frustration) are recorded. An annotated example of the affective signals is shown in Fig. 4. We proceed to compute the statistics of these signals per user, input type and interaction mode. Our goal is to investigate how the affective metrics vary as a function of modality input patterns (touch vs. speech), interaction mode (SO, GO, CT, OM, MS), user-initiated modality switches, speech recognition errors and associated error correction sub-dialogues.

V. RESULTS

V-A. Qualitative Evaluation

To illustrate the results, some example evaluation sessions are reported first as shown in Fig. 5. The figure shows three evaluation sessions of participant user 4 including GO, the multimodal system OM and SO (evaluated in the order plotted). As discussed in Fig. 4, a turn is the time period shown between two thick vertical lines; each dotted vertical line separates a turn into the inactivity and interaction periods. Only fill turns have background color; that color is red for speech turns and blue for GUI turns. Note that although EEG data have a constant sampling rate of 128Hz, the affective metrics estimated by the Emotiv device have an average sampling rate of 12Hz; the detections are event-driven and their sample rates depend on the number of expressive and cognitive events. This means that affective metrics have low temporal resolution compared to EEG. The three affective metrics more relevant to this study (and showing more meaningful variation during the interactive session) are frustration, excitement and engagement.

In Fig. 5(a), a typical GO session is shown. The affective metrics do not vary much during the interaction with the exception of turns 7 and 8. In turn 7, frustration raises (and then falls) after the user realizes he entered the wrong value in the previous turn. In turn 8, frustration raises again because the user was confused about which value to select; once the correct value is selected, frustration and excitement start decreasing again. In Fig. 5(b), a typical OM session is shown. Speech recognition errors occur at turns 4 and 5 (back-to-back) resulting in elevated frustration levels around these event. Note however, how using touch input to fix the error in turn 6 results in the rapid decrease of frustration; this is a pattern found frequently in the whole evaluation set.

In Fig. 5(c), we show a SO session with many misrecognitions. Due to many errors, there is significant variability in both frustration and excitement patterns. In Fig. 5(d) we show one more example sessions for users 7 (MS session). User 7 has the lowest overall speech word error rate (WER) and was very confident in using speech. Notice how smooth is the plot for all the affective metrics and that levels of both excitement and frustration are low.

V-B. Quantitative Evaluation

Next we examine how affective metrics relate to input type and the different interaction systems (SO, GO, CT, OM, MS). Table I shows the mean and standard deviation for the three affective metrics according to input type (touch or speech); the last row shows the overall (touch and speech) results. Notice that *for both excitement and frustration speech input has higher levels compared to touch input* by 5% and 6%, respectively. This is most probably due to speech recognition errors causing higher levels of frustration and excitement (as shown previously in the affective plots and quantitatively below). For engagement on the other hand, touch input has slightly higher levels (this could be

Table I. Affective metric statistics per input type.

Input type	engagement		excitement		frustration	
	mean	std	mean	std	mean	std
Touch	0.79	0.11	0.45	0.19	0.51	0.17
Speech	0.76	0.11	0.50	0.19	0.57	0.19
Overall	0.76	0.11	0.48	0.19	0.56	0.19

Table II. Affective metric statistics per interaction mode/input type.

System	engagement		excitement		frustration	
	mean	std	mean	std	mean	std
GO	0.78	0.11	0.44	0.19	0.50	0.15
CT (touch)	0.79	0.11	0.43	0.17	0.50	0.18
CT (speech)	0.78	0.10	0.47	0.17	0.57	0.19
CT (overall)	0.78	0.10	0.46	0.17	0.56	0.19
OM (touch)	0.80	0.11	0.44	0.19	0.52	0.21
OM (speech)	0.76	0.11	0.47	0.17	0.58	0.19
OM (overall)	0.77	0.11	0.46	0.18	0.57	0.19
MS (touch)	0.80	0.12	0.46	0.20	0.54	0.21
MS (speech)	0.76	0.10	0.47	0.17	0.58	0.19
MS (overall)	0.77	0.11	0.47	0.18	0.57	0.19
SO	0.73	0.12	0.54	0.21	0.59	0.20

due to speech being the more “natural” interaction modality). Overall results, show that engagement levels are higher and have less variance compared with frustration and excitement.

Table II shows the mean and standard deviation for the three affective metrics for each of the five interaction modes. For the three multimodal modes, results are presented per input (touch/speech) and overall (independent of input type, that is both touch and speech input). Engagement for SO system is lower compared to all other systems; note that for all three multimodal modes touch input has slightly higher engagement compared to speech input as shown in the previous table. Excitement is much higher for SO compared to GO system (0.54 & 0.44 respectively). Multimodal modes as a mixture of SO and GO systems have average excitement values lying between these two values and closer to that of GO. Similarly, for frustration, SO values are much higher compared to GO system (0.59 & 0.50 respectively) while multimodal modes have average values of around 0.57.

Table III shows the mean and standard deviation for the three affective metrics for all eight users. Notice the differences between users. For example user 7 has by far the lowest excitement and frustration levels (ASR WER 6%) while user 8 with the higher WER, has the highest levels for excitement and second highest for frustration.

In Fig. 6, we show the variation of the affective metrics in the neighborhood of speech recognition error and associated repairs. Specifically in Fig. 6(a) we show average engagement, excitement and frustration, two turns before and three turns after a speech recognition error. Results shown are averaged over 266 misrecognitions. As expected, frustration levels rise when a speech recognition error occurs but falls relatively quickly over the next two turns (once errors are fixed). Excitement follows a similar pattern (rise and fall) but with approximately a delay of one interaction turn, probably associated with the effort required to repair the error. Engagement stays relatively constant through speech recognition error events (the slightly lower engagement at the turn where the error occurs could be explained as the cause for making this error). Speech error repairs in Fig. 6(b) (either using the speech or touch input modalities) show complementary patterns. Error fixes are associated with higher frustration and excitement in the two previous turns due to the occurrence of the error and associated effort to correct it. Excitement remains high also for the turn where the error is repaired. Both frustration and excitement fall quickly following a successful error repair.

We also investigated the correlation between objective metrics (du-

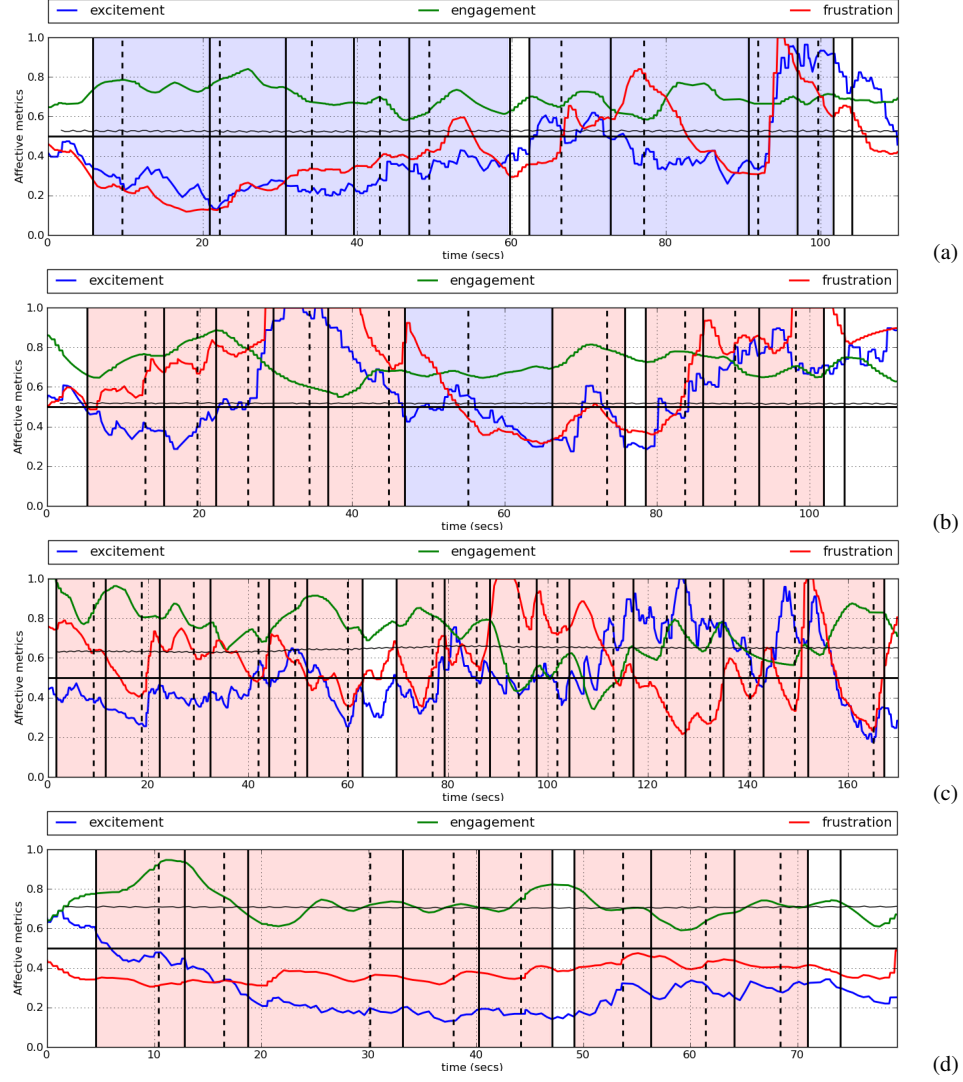


Fig. 5. Sample evaluation sessions for user 4: (a) GO, (b) OM, (c) SO, (d) user 7 MS session

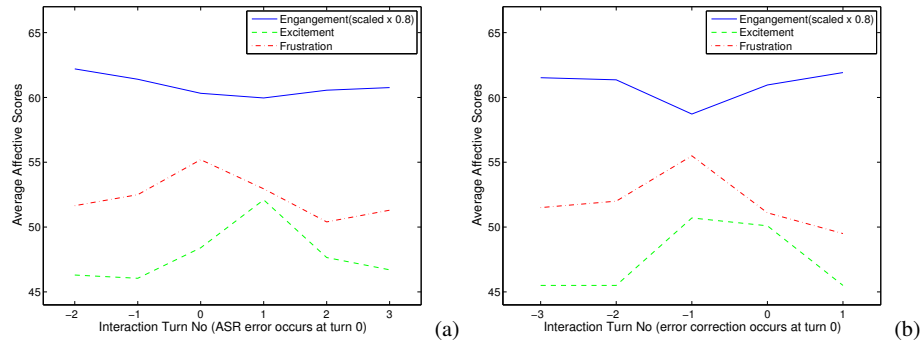


Fig. 6. Variation of affective metrics in the neighborhood of speech recognition errors and repairs: (a) affective metric patterns averaged over 266 speech recognition errors, (b) affective metric patterns averaged over 153 error repairs. Affective metrics are averaged over each interaction turn (turn 0 corresponds to the error or repair event).

ration, errors, modality usage) and the affective metrics. We observed weak correlation (approx. 0.2) between frustration and inactivity time,

which could be due to hot-spots in the interaction where both the cognitive load (associated with inactivity time) and frustration are

Table III. Affective metric statistics per user.

User	engagement		excitement		frustration	
	mean	std	mean	std	mean	std
usr1	0.80	0.09	0.51	0.20	0.61	0.18
usr2	0.74	0.09	0.47	0.15	0.54	0.19
usr3	0.82	0.14	0.51	0.24	0.55	0.20
usr4	0.73	0.10	0.45	0.17	0.54	0.19
usr5	0.79	0.09	0.46	0.19	0.54	0.19
usr6	0.79	0.07	0.46	0.14	0.56	0.17
usr7	0.75	0.10	0.38	0.16	0.44	0.13
usr8	0.71	0.12	0.54	0.21	0.58	0.18

expected to be higher at the same time. In addition, we investigated the affective patterns for modality switches, i.e., when the user overrides the default input modality speech or touch. There is no observable affective pattern during modality switches other than the higher value of the within-turn variance for all three affective metrics. Overall, speech recognition errors and repairs seem to have high affective content.

VI. DISCUSSION

Incorporation of biosignals such as EEG in the user experience design provides a rich amount of data not previously available. Yet the correct association of these data to underlying emotional or cognitive states is a challenging endeavor for the research community. The recent release of the Emotiv device allowed developers and researchers outside of the neuroscience community to exploit EEG technology. In the context of HCI research, several demonstrations and research efforts have emerged mostly towards using the device as a BCI modality [2] by exploiting either expressive or cognitive events (such as P300 ERP). Although verification and validation of such efforts is relatively straightforward, validation of Emotiv's affective metrics (or for any other similar metric for that matter) is more difficult because quantification of emotional states is an open research question. Validity of affective metrics and their use in complex settings make evaluation a challenging task. However we have found that: i) a good agreement between Emotiv's affective metrics and manually annotated ratings exists. ii) both excitement and frustration may increase in the case of speech errors or user confusion (recall affective plot examples) and that this particular pattern pertains throughout the dataset. Note that the affective metrics reported here are averaged over multiple turns and/or users (reducing the estimation variance). iii) also, as shown clearly in Section V, on average there are consistent differences in frustration and excitement levels for different input modality (touch vs speech), interaction modes (especially for SO, GO) and around misrecognition/repair events.

VII. CONCLUSIONS

We have shown that the use of affective signals estimated via an EEG device can provide significant insight in the multimodal dialogue design process. The affective metrics and physiological signals can be used in conjunction with traditional evaluation metrics offering a more user-centric evaluation view of the interactive experience. Specifically we have investigated the affective patterns of speech recognition errors and associated repairs. We have shown that speech recognition errors lead to increased frustration levels, followed by higher excitement levels. In addition, affective interaction patterns vary significantly by input modality. Speech input (when compared with touch input) is associated with higher excitement and frustration (probably due to speech recognition errors) and lower engagement (probably due to speech being the more natural interaction modality).

Incorporation of user affective state and other cognitive features such as attention or cognitive load can prove valuable tools in both the evaluation and design of interactive systems. Here is a preliminary list of interesting future research directions: 1) Robust estimation of affective metrics from the raw EEG signals. 2) Investigate the relation between error related negativity [3] potentials (ERNs) and speech

recognition errors/repairs. 3) Another important attribute that could be estimated using physiological signals is cognitive load that relates to the perceived mental effort. 4) It would be also helpful to integrate information from an eye-tracker to better estimate visual attention and overall engagement, and also investigate how attention is divided between the audio and visual channels in the multimodal system. Crossmodal attention and multisensory integration are in the forefront of neuroscience research and are topics of study that could benefit the multimodal research community too. Overall, this is a first step towards exploring the relevance of physiological signals and associated affective metrics in spoken dialogue evaluation and design.

VIII. REFERENCES

- [1] B. Blankertz, M. Tangermann, F. Popescu, M. Krauledat, S. Fazli, M. Donaczy, G. Curio, and K.-R. Muller. The Berlin Brain-Computer Interface. *Computational Intelligence Research Frontiers*, 5050:79–101, 2008.
- [2] A. Campbell, T. Choudhury, S. Hu, H. Lu, M. K. Mukerjee, M. Rabbi, and R. D. Raizada. Neurophone: brain-mobile phone interface using a wireless EEG headset. In *Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds*, MobiHeld '10, pages 3–8, 2010.
- [3] R. Chavarriaga and J. D. R. Millan. Learning from EEG error-related potentials in non invasive brain-computer interfaces. *IEEE Transactions on Neural and Rehabilitation Systems Engineering*, 18(4):381–388, 2010.
- [4] R. Cowie, E. Douglas-Cowie, S. Savvidou*, E. McMahon, M. Sawey, and M. Schröder. 'FEELTRACE': An instrument for recording perceived emotion in real time. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*. Citeseer, 2000.
- [5] R. J. Davidson. What does the prefrontal cortex “do” in affect: perspectives on frontal EEG asymmetry research. *Biological Psychology*, 67(1-2):219–233, 2004.
- [6] A. Gevins and M. E. Smith. Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, 4(1):113–131, 2003.
- [7] D. Grimes, D. S. Tan, S. E. Hudson, P. Shenoy, and R. P. Rao. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, Florence, Italy, pages 835–844. ACM, 2008.
- [8] T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(2):163–178, 2000.
- [9] P. Lang. The emotion probe: Studies of motivation and attention. *American psychologist*, 50:372–372, 1995.
- [10] F. Paas, J. Tuovinen, H. Tabbers, and P. Van Gerven. Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38(1):63–71, 2003.
- [11] M. Perakakis and A. Potamianos. Multimodal system evaluation using modality efficiency and synergy metrics. In *Proceedings of the 10th international conference on Multimodal interfaces*, ICMi '08, pages 9–16, New York, NY, USA, 2008. ACM.
- [12] M. Perakakis and A. Potamianos. A study in efficiency and modality usage in multimodal form filling systems. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(6):1194–1206, 2008.
- [13] R. W. Picard. *Affective Computing*. MIT Press, 1997.
- [14] R. Plutchik. The nature of emotions. *American Scientist*, 89(4):344–350, 2001.
- [15] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen. Galvanic skin response (GSR) as an index of cognitive load. In *CHI '07 extended abstracts on Human factors in computing systems*, CHI '07, pages 2651–2656, 2007.