Journal of Engineering and Applied Sciences 12 (7): 1864-1870, 2017
ISSN: 1816-949X

Affect Recognition Challenge Bridging Across Audio, Video and Physiological Data

[1]D. Lakshmi and [2]R. Ponnusamy
[1]Department of Computer Science and Engineering, Satyabama University, Chennai, India
[2]Rajiv Gandhi College of Engineering, Sriperumpudur, India

Abstract: Automated analysis of human affective behavior has attracted increasing attention from researchers in psychology, computer science, linguistics, neuroscience and related disciplines. However, the existing methods typically handle only deliberately displayed and exaggerated expressions of prototypical emotions, despite the fact that deliberate behavior differs in visual appearance, audio profile and timing from spontaneously occurring behavior. The goal of the challenge is to provide a common Benchmark test set for multimodal information processing and to bring together the audio, video and physiological emotion recognition communities to compare the relative merits of the three approaches to emotion recognition under well-defined and strictly comparable conditions and establish to what extent fusion of the approaches is possible and beneficial. This study presents the challenge, the dataset and the performance of the baseline system.

Key words: Affective computing, speech, facial expression, emotion recognition, physiological signals

INTRODUCTION

Human-Computer Interaction (HCI) is the study of the interaction between people and computers (Almaev and Valstar, 2013). Such interaction is mainly done at the user interface. One of the major concerns of professional practitioners in the field of HCI is the design of interactive computing systems for human use. As a result, it is a basic goal of HCI designers to make computers more usable and more receptive to the user's needs. To provide the best possible interface within given constraints, the HCI designers are supposed to develop systems that minimize the barrier between the human's cognitive model of what users want to accomplish and the computer's understanding of the user's task (Bone *et al.*, 2014).

Mobile devices play an important role in the modern society. They are being used by people of all walks of life for various purposes. They can be found in the fields of education, entertainment, medicine, communication service, military systems and so on. Due to the multi disciplinary nature of HCI, designing user interfaces for mobile devices poses several interaction challenges (Chen *et al*., 2015). Some of these challenges are hardware-related while the others software-related.

Statement of problem: Owing to the fast development in the digital technology the operation of human-computer interface is becoming more and more complicated. Consequently, to catch up with the speedy and fleeting transformation the user of digital interactive products can only keep on learning various operating interfaces, programming languages and development environments. Now a days, in our daily lives, we can hear more and more people complaining about the bad design in interaction interface. Is this problem caused by the bad design of the interactive products or by the shortage of users knowledge about the logics of the human-machine interaction design?

Most researchers in HCI take interest in developing new design methodologies, experiencing with new hardware devices, prototyping new software systems and exploring new paradigms for interaction. Designs in HCI aim to create user interfaces which can be operated with ease and efficiency. Many digital products that require users to interact with them to accomplish their tasks have not necessarily designed with the users in mind. The designer always claims how usable the products are however, an even more basic requirement is that the interface should allow the user to carry out relevant tasks completely. In other words, the design must be both usable and useful for the user and it must be a user-centered design. Current mobile computing devices such as palmtop computers, Personal Digital Assistants (PAD) and mobile phones have a problem in common attempting to provide users with powerful computing

Corresponding Author: D. Lakshmi, Department of Computer Science and Engineering, Satyabama University, Chennai, India

services and resources through small interfaces (Cronbach, 1951). As is usually the case with mobile devices, limited screen size makes it difficult to efficiently present information and help users navigate to and from the information they want. And since mobile devices are often required to possess multiple functionalities, the convergence of electronics, computing and communication is becoming a must in the mobile industry. In addition because mobile devices need to operate with limited battery charge, how to deal with the power consumption has also become one of the most important issues for system designers. Based on the problems in HCI design for mobile devices mentioned above, this paper addresses the following research questions:

- What are the major challenges of designing interactive products for mobile devices?
- What are the possible solutions to such problems in developing a good user-centered design for mobile computing devices?
- What are the principles of user interface design for mobile devices?
- What are the new trends in mobile industry in the nearfuture?

Mobile devices: Mobile devices can be defined in different ways when they are looked at from different perspectives. They can be definedin terms of the services they offer or based on the level of functionality connected with the devices. According to Sharp they refer to the devices that are handheld and intended to be used while on the move. Now a days, mobile devices are being used by different people for various purposes.

A mobile device refers to a pocket-sized computing device, typically having a small display screen, a small keypad with miniature buttons or a touch screen with stylus of input. Examples of mobile devices include mobile computers like handheld or palmtop PC and Personal Digital Assistant (PDA), handheld game consoles such as Nintendo DS and game boy advance, media recorders like Digital Still Camera (DSC) anddigital audio recorders and communication devices such as mobile phones, cordless phones and pagers. As is often the case, mobile devices have wireless capability to connect to the internet and home computer systems (Dawson *et al*., 2007). However, wireless capability poses a number of security risks. It takes considerable knowledge of the threats posed to mobile devices to deal with the risks.

Use of mobile devices: A mobile device refers to a pocket-sized computing device, typically having a small display screen, a small keypad with miniature buttons or a touch screen with stylus of input. Examples of mobile devices include mobile computers like handheld or palmtop PC and Personal Digital Assistant (PDA) handheld game consoles such as Nintendo DS and game boy advance, media recorders like Digital Still Camera (DSC) and digital audio recorders and communication devices such as mobile phones, cordless phones and pagers (Ekman *et al*., 2002; Eyben *et al*., 2016, 2013). As is often the case, mobile devices have wireless capability to connect to the internet and home computer systems. However, wireless capability poses a number of security risks. It takes considerable knowledge of the threats posed to mobile devices to deal with the risks.

Operating systems for mobile devices: Mobile devices are increasingly being used by different types of people. In medicine, for instance, PDAs are used to record symptoms for patients and to support the cardiologist in the medical decision-making process and have been proven to help both diagnosis and pharmacy selection (Dawson *et al*., 2007). Besides, they are also used to improve the effectiveness of communication between the patient and the hospital during follow-up treatment.

Literature review

Web accessible database: The SEMAINE Solid-SAL dataset is made freely available to the research community. It is available through a web accessible interface with url http://www.semaine-db.eu the available dataset consists of 25 recordings, featuring 21 participants. Four of these participants play the role of the operator in the sessions but they also appear in the user role in some of the interactions. At time of recording, the youngest participant was 22, the oldest 60 and the average age is 32.8 years old (std. 11.9) 38% are male and the participants come from 8 different countries. Unfortunately, all but one participants come from a Caucasian background, making the dataset ethnically biased.

Organisation: Within the database, the data is organised in units that we call a session. A session is part of a recording in which the user speaks with a single character. There are also two extra special sessions per recording to wit the recording start and recording end sessions. These sessions include footage of the user/operator preparing to do the experiment or ending the experiment. Although, these sessions do not show the desired user/character interaction they may still be useful for training algorithms that do not need interaction such as the facial point detectors of detectors which sense the presence of a user.

Each session has 11 sensor data files associated with it. We call these database entries tracks. Nine of these are the five camera recordings and the four microphone recordings. In addition, each session has two lower-quality audio-visual tracks, one showing the frontal colour recording of the user and the other showing the frontal colour recording of the operator. Both low-quality recordings have audio from the operator and the user. The use of these low-quality recordings lies in the fact that they have both audio and video information which makes them useful for the annotation of the conversation by human raters. To allow annotators to focus on only one person talking we stored the user audio in the left audio channel and the operator audio in the right audio channel. Because most media (Grimm and Kroschel, 2005; Halko *et al*., 2011; Hall *et al*., 2009; Keltner and Lerner, 2010).

MATERIALS AND METHODS

This research integrates design thinking with technology development process for developing a multimodal Human Computer Interface (HCI) for Smart TVs. Figure 1 shows the comprehensive structure of the integration of design thinking concept with technology development an inter-disciplinary approach (Fig. 2). The comprehensive structure of inter-disciplinary integration of design thinking with technology development for mobile phone.

This structure has 5 major phases (Fig. 2). The first phase outlines the vision for developing the Smart TV multimodal HCI design and is based on a review of development trends and visions for Smart TVs. The second phase proposes a plan for inter-disciplinary integration of domain experts in technology and interaction design by holding a service design workshop for brainstorming. The third phase which is the section for integration, defines the user scenario and technology benchmark with user-centered design insights. The fourth phase develops the applications and multimodal interaction prototype by integrating technologies into the user interface. The final phase evaluates user experiences with the prototype on a system usability scale that can collect scientific data (e.g., eye-tracking system) for objective analysis. Evaluation results are then feed back to the inter-disciplinary team to modify and adjust the prototype. This study will present the results and evaluation of the first three phases. For practical implementation, this work follows the implementation process (Fig. 3),the details of which are as follows.

Figure 4 shows inter-disciplinary team discussion and character map of service design workshop. This workshop helped the team gain a clear understanding the features of multimodal HCI design. The many ideas generated were then narrowed down from global thoughts into specific and applicable features that meet user requirements and are applicable to technical features development (Knapp *et al*., 2011; Koelstra *et al*., 2012; Picard, 2014; Ringeval *et al*., 2014, 2015, 2013a, b).

Fig. 1: Types of mobile devices

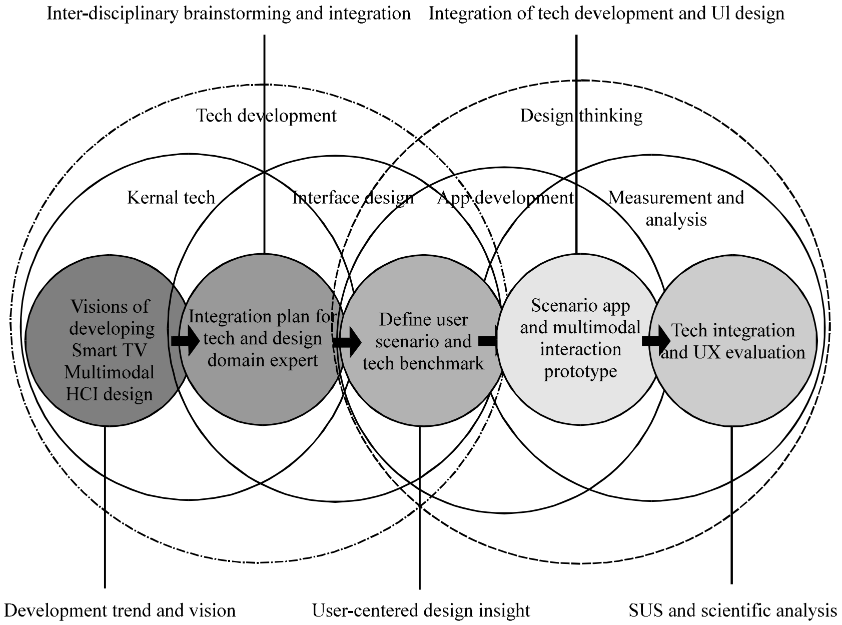

Inter-disciplinary brainstorming and integration
Integration of tech development and Ul design
Tech development
Design thinking
Kernal tech
Interface design
App development
Measurement and analysis
Visions of developing Smart TV Multimodal HCI design
Integration plan for tech and design domain expert
Define user scenario and tech benchmark
Scenario app and multimodal interaction prototype
Tech integration and UX evaluation
Development trend and vision
User-centered design insight
SUS and scientific analysis

Fig. 2: The comprehensive structure of inter-disciplinary integration of design thinking with technology development for mobile phone

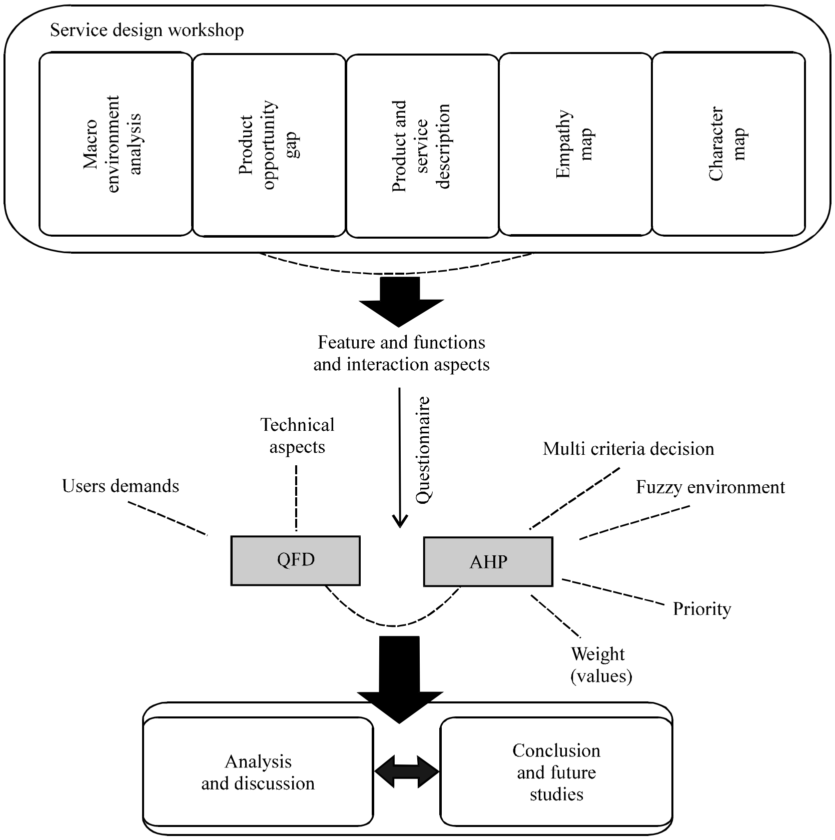
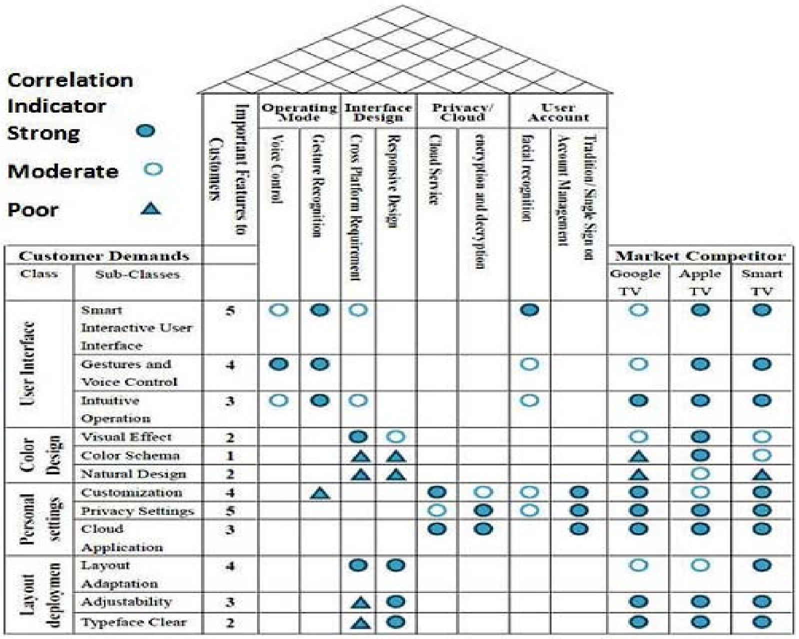


Fig. 3: Implementation process

Fig. 4: Inter-disciplinary team discussion and character map of the service



Correlation Indicator
Strong ●
Moderate ○
Poor ▲

Customer Demands		Important Features to Customers	Operating Mode		Interface Design		Privacy/ Cloud		User Account		Market Competitor		
Class	Sub-Classes		Voice Control	Gesture Recognition	Cross Platform Requirement	Responsive Design	Cloud Service	encryption and decryption	facial recognition	Tradition/Single Sign on Account Management	Google TV	Apple TV	Smart TV
User Interface	Smart Interactive User Interface	5	○	●	○				●		○	●	●
	Gestures and Voice Control	4	●	●					○		○	●	●
	Intuitive Operation	3	○	●	○				○		●	●	●
Color Design	Visual Effect	2			●	○					○	●	○
	Color Schema	1			▲	▲					▲	●	○
	Natural Design	2			▲	▲					▲	○	▲
Personal settings	Customization	4		▲			●	○	○	●	●	○	●
	Privacy Settings	5					○	●	○	●	●	●	●
	Cloud Application	3					●	●		●	●	●	●
Layout deploymen	Layout Adaptation	4			●	●					○	○	●
	Adjustability	3			▲	●					●	●	●
	Typeface Clear	2			▲	●					●	●	●

Fig. 5: Quality function deployment matrix results

RESULTS AND DISCUSSION

Quality function deployment matrix results: Figure 5 shows the QFD matrix results. Based on QFD matrix analysis. The smart interactive user interface and privacy settings are two of the most important features of Smart TVs followed by gesture and voice control, customization of personal settings and layout adaptation. These visualized results show that the multimodal interaction design is very important to Smart TVs.

In comparison with technical features, gesture recognition and facial recognition are highly prized by respondents. Privacy via encryption and decryption and traditional/single sign-in on account management are also required by customers. Respondents agreed that Apple TVs and Smart TVs have user-friendly interfaces. The privacy feature has already been developed by Apple TV, general Smart TV and Google TV. The QFD matrix results comprehensivelyshow a significant role to help the development. These results are also evaluated and calculated via the AHP. Each criterion is compared to another such that the importance weight is derived.

CONCLUSION

The AHP, based on the hierarchy principle, assumes consecutive decomposition of multiple aims with degree increasing toward lower levels. Hierarchy development conforms with the principles of system approaches toward task analysis and can facilitate the process of creation and formalization of Participatory Technology Development (PTD) priorities. One main advantage of the AHP is the determination of subjective criteria and scores based on pairwise comparisons. Another advantage involves the structural organization of problem components. The AHP provides consistent assessment tools, analyzes alternative sensitivities, uses relatively simple mathematic equations and allows participation of different specialists or groups.

A strong point of the AHP is the independence of its application from the activity sphere. The AHP results show the service design approach is an efficient way for communication among interdisciplinary team members. The proposed two-stage decision-making processes qualitatively analyze and quantitatively assesses the priority and relevance of features derived from service design process. The technique team can then develop a prototype that demonstrates multimodal interaction with confidence, thereby fulfilling user demands.

RECOMMENDATIONS

Three possibilities directions exist for future study:

- Include raw prices (retail prices) in the QFD matrix method and AHP. This would be comparative as this criterion may affect user demands. For example, if the Kinect Sensor price is excessive and users think it is not as effective as say the motion leap sensor embedded in a remote control could be considered as a criterion to be evaluated
- In-depth understandings of current and existing demands are essential. Failure probability still exists as the AHP does not work well when evaluating quantitative values; it is much better at creating qualitative values
- Implementing these methods is acceptable. For future work could identify Smart TV features. If field report results could be evaluates and joined with questionnaire results this project would generate relevant and effective content

REFERENCES

Almaev, T.R. and M.F. Valstar, 2013. Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition. Proceedings of the Conference on Affective Computing and Intelligent Interaction (ACII), Humaine Association, September 2-5, 2013, IEEE, Geneva, Switzerland, pp: 356-361.

Bone, D., C.C. Lee and S. Narayanan, 2014. Robust unsupervised arousal rating: A rule-based framework withknowledge-inspired vocal features. IEEE. Trans. Affective Comput., 5: 201-213.

Chen, M., Y. Zhang, Y. Li, M.M. Hassan and A. Alamri, 2015. AIWAC: affective interaction through wearable computing and cloud technology. IEEE. Wireless Commun., 22: 20-27.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika, 16: 297-334.

Dawson, M., A. Schell and D. Filion, 2007. The Electrodermal System. In: Handbook of Psychophysiology. Vol. 2, Cambridge University Press, Cambridge, pp: 200-223.

Eyben, F., F. Weninger, F. Gross and B. Schuller, 2013. Recent developments in openSMILE, the munich open-source multimedia feature extractor. Proceedings of the 21st ACM International Conference on Multimedia, October 21-25, 2013, ACM, New York, USA., pp: 835-838.

Eyben, F., K.R. Scherer, B.W. Schuller, J. Sundberg and E. Andre *et al*., 2016. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE. Trans. Affective Comput., 7: 190-202.

Grimm, M. and K. Kroschel, 2005. Evaluation of natural emotions using self assessment manikins. Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding, November 27, 2005, IEEE, San Juan, ISBN: 0-7803-9478-X, pp: 381-385.

Halko, N., P.G. Martinsson, Y. Shkolnisky and M. Tygert, 2011. An algorithm for the principal component analysis of large data sets. SIAM J. Sci. Comput., 33: 2580-2594.

Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I.H. Witten, 2009. The WEKA data mining software: An update. SIGKDD Explorat. Newslett, 11: 10-18.

Keltner, D. and J.S. Lerner, 2010. Emotion. In: Handbook of Social Psychology. Fiske, S., D. Gilbert and G. Lindzey (Eds.). John Wiley & Sons Inc., Hoboken, New Jersey, pp: 317-331.

Knapp, R.B., J. Kim and E. Andre, 2011. Physiological signals and their use in augmenting emotion recognition for human-machine interaction. In: Emotion-Oriented Systems. Cowie, R., C. Pelachaud and P. Petta (Eds.). Springer Berlin Heidelberg, Heidelberg, Germany, ISBN: 978-3-642-15183-5, pp: 133-159.

Koelstra, S., C. Muhl, M. Soleymani, J.S. Lee and A. Yazdani *et al*., 2012. Deap: A database for emotion analysis; using physiological signals. IEEE Trans. Affective Comput., 3: 18-31.

Picard, R., 2014. Affective media and wearables: surprising findings. Proceedings of the 22nd ACM International Conference on Multimedia, November 3-7, 2014, ACM, New York, USA., ISBN: 978-1-4503-3063-3, pp: 3-4.

Ringeval, F., A. Sonderegger, J. Sauer and D. Lalanne, 2013a. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), April 22-26, 2013, IEEE, Shanghai, China, ISBN: 978-1-4673-5545-2, pp: 1-8.

Ringeval, F., A. Sonderegger, B. Noris, A. Billard and J. Sauer *et al*., 2013b. On the influence of emotional feedback on emotion awareness and gaze behavior. Proceedings of the Conference on Affective Computing and Intelligent Interaction (ACII), Humaine Association, September 2-5, 2013, IEEE, Geneva, Switzerland, pp: 448-453.

Ringeval, F., F. Eyben, E. Kroupi, A. Yuce and J.P. Thiran *et al*., 2015. Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data. Pattern Recognit. Lett., 66: 22-30.

Ringeval, F., S. Amiriparian, F. Eyben, K. Scherer and B. Schuller, 2014. Emotion recognition in the wild: Incorporating voice and lip activity in multimodal decision-level fusion. Proceedings of the 16th International Conference on Multimodal Interaction, November 12-16, 2014, ACM, New York, USA., ISBN: 978-1-4503-2885-2, pp: 473-480.