Pertanika J. Sci. & Technol. 25 (S): 55 - 68 (2017)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Crowdsourcing Platform for Collecting and Rating Emotion Elicitation Media

Alghowinem, S.* and Albalawi, A.

College of Computer and Information Sciences, Prince Sultan University, SA

ABSTRACT

Giving computers the ability to understand the user's mood and feelings is the aim for affective computing field. This ability would enhance the interaction between the user and his/her computer to create advanced systems for education, commerce, security and mental disorder diagnosis, among other functions. To achieve this goal, computer software needs to be trained on big data using emotion measures. These emotions should be elicited by a standardised, replicable and validated medium. However, collecting and rating such emotion elicitation media is not a trivial task, as it involves several factors. This research aims at designing a crowdsourcing platform to collect and rate emotion elicitation media. The platform is designed such that registered users can add, recommend and rate emotion election clips, whereas researchers can access and statically analyse data about the rated clips. This crowdsourcing platform can be used by emotion researchers to collect highly- rated emotion elicitation media, and by individuals through social media platform to share emotion elicitation media. The highly-rated clips could be used to elicit emotions, which then could be used to create models for automatic emotion recognition. The automation of emotion recognition will benefit different fields such as health (physical and mental), education and technology.

Keywords: Crowdsource platform, emotion elicitation, emotion recognition, human-computer interaction, media collection

ARTICLE INFO

Article history:
Received: 09 March 2017
Accepted: 18 September 2017

E-mail addresses:
sghowinem@psu.edu.sa) (Alghowinem, S.),
alia.albluwi@hotmail.com) (Albalawi, A.)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Human-Computer Interaction (HCI) in general and in affective computing in particular has an interest in emotion recognition. This interest has increased with a view towards implementing advance intelligent systems that interact with users based on their feelings. Such intelligent systems would improve user perception of software and their provided services (Pantic, Pentland, Nijholt, & Huang, 2006). For example, when including effect-state recognition in a remote educational system, presentation style, content complexity and task difficulty could be adjusted based on the emotional state of the student (Craig, Graesser, Sullins, & Gholson, 2004). Advertisements on e-commerce contexts could be personalised by understanding the clients' preference based on their current mood (Zhou, Ji, & Jiao, 2012). Smart surveillance systems could use people emotions, mood and behaviour to detect abnormal activities for security and safety assurance (Tao & Tan, 2005). Emotion and mood recognition have become an active research area recently to support and assist psychologists in diagnosing mental disorders such as depression (Alghowinem, Goecke, Wagner, Epps, Breakspear, & Parker, 2012).

To give the computer the ability to recognise emotions, big data of emotion measures have to be collected for each emotion. Regardless of the sensors used to collect these measures, the investigated emotions have to be elicited in a replicable way between subjects. Psychologists have investigated several methods to induce emotions (Westermann, Spies, Stahl, & Hesse, 1996). These methods are categorised in three categories: (1) acted or posed emotions, where actors are asked to perform a certain emotion, (2) induced emotions, where stimuli are used to evoke emotional response, such as looking at pictures, listening to music, watching videos and reading texts, and (3) naturalistic displays of emotion involving subjects in conversation (Kory & D'Mello, 2014). These emotion-inducing methods vary by several factors, including emotion sensory measures, intensity and duration of response, reliability and ecological validity (Kory & D'Mello, 2014). Nevertheless, watching film clips has proven to be effective in emotion elicitation for its advantages compared to other techniques (Zeng, Pantic, Roisman, & Huang, 2009; Jerritta, Murugappan, Nagarajan, & Wan, 2011).

For a clip to be suitable for eliciting emotions, several criteria have to be taken into account. The differences between clips have to be controlled so that comparisons of emotion elicitation would be due to the emotions, not to differences in the clips. These criteria include clip duration and intensity of elicited emotion etc. One of the most serious challenges in selecting emotional clips is to ensure that the selected clip only induces the target emotion and not unintended or mixed emotions. Several studies have investigated and validated such emotion elicitation clips. A list of validated clips has been made available in Gross and Levenson (1995) and Schaefer, Nils, Sanchez and Philippot (2010). However, the validation method of selecting these clips used a relatively small number of subjects and did not consider the subjects' cultural background. Since emotions are subjective, a large number of subjects are required to compensate for this issue. Moreover, emotions are influenced by cultural background of subjects. For instance, different emotions could be induced by the same set of elicitation clips among different cultures.

In this paper, we propose a design for a crowdsourcing platform to collect and rate emotion elicitation media. The proposed platform utilises the power of the crowd to share and rate emotion elicitation media. For individuals, the suggested platform could be seen as a social media platform. For emotion researchers, this platform could be used as a crowdsourcing platform to select highly-rated emotion elicitation media for emotion elicitation studies.

BACKGROUND AND RELATED WORK

Emotion Representation in Psychology

Defining each emotion is a critical first step in collecting eliciting stimuli. Yet, definitions of emotions in psychology literature are highly controversial. Emotion representations can be divided into categorical and dimensional representations. Categorical emotion representations include simple (i.e. positive vs. negative emotions) and discrete (i.e. specified set of emotions) representations. A well-known emotion category has been suggested by Ekman, who categorised emotions as six basic emotions: anger, disgust, fear, happiness, sadness and surprise (Ekman, 1992). Since emotions are complex and mixed in nature, dimensional emotion representations have been proposed. The dimensional emotion theory includes two well-known models: two dimensional (2-D) and three dimensional (3-D) models. The two-dimensional emotion model has two axes, valence and arousal, as proposed by Jaimes and Sebe (2007) and Russell (1979). The three-dimensional emotion model was proposed by Kehrein (2002), Schroder, Cowie, Douglas Cowie, Westerdijk and Gielen, (2001) and Wundt, (2009). The three dimensions are valence, arousal and dominance, and they range from weak to strong. In effective computing and automated emotion recognition, categorical emotion representations are mostly used. That is due to their suitability as labelled classes in the classification task. Therefore, in this study emphasis is given to Ekman's set of emotions, not only for their classification suitability, but also for their universality in cross-cultural contexts (Ekman, 1992).

As mentioned earlier, several techniques exist to elicit emotions, with eliciting emotions by watching video clips being the most effective technique (Westermann et al., 1996; Schaefer et al., 2010). Eliciting emotions by watching video clips has several advantages compared to other methods. Typically, films are an artificial reflection of real life, but without ethical boundaries. Films can elicit strong subjective and physiological changes (Gross & Levenson, 1995). Video clips can induce mixed emotions with different levels of arousal. Therefore, selecting film clips should be done carefully and have to be validated to elicit related emotions.

Several studies have been conducted to collect and validate emotions elicited by movie clips. In an early Western study by Philippot (1993), 12 emotion elicitation film clips were collected to elicit six emotions. Subsequently, Gross and Levenson (1995) collected emotion elicitation films with the help of film critics, video store employees and film buffs for nomination. Over 250 films were evaluated and selected carefully to yield a filtered list of 78 clips. They showed the clips to 494 English-speaking subjects. Based on the subjects' evaluation of these clips, a shorter list of 16 films for eliciting eight targeted emotions was filtered. However, these lists can now be considered outdated. In a recent study, 70 French-speaking movie clips collected from different cultural backgrounds (French, Italian, British and American) were validated (Schaefer et al. 2010). Another recent study presented a validated emotion elicitation Spanish-dubbed set of clips (Fernandez, Pascual, Soler, & Fernandez Abascal, 2011). However, a known gap of the above studies is that the selected population for validating the clips were mostly young, which may have affected the result. Another important limitation is the cultural differences between both the selected clips and subjects, which play a critical role in emotion trigger for audiences.

As mentioned earlier, collecting video clips for emotion elicitation should consider several factors carefully. Duration of the emotion elicitation video is one of the most important factors, due to its effect on the emotion latency, rise time and offset. A short 1- to 2-minute clip is desirable for this purpose. Typically, emotion elicitation video clips are part of a film or a play. Therefore, the selection of the emotion elicitation clip should be self-explanatory and should consider inducing emotions without additional background explanation or additional contexts. Moreover, the selected clip should induce the target emotion and not a mix of emotions or non-target emotions. The intensity of inducing the target emotion should be recorded and used as a measure for clip selection. Most importantly, the selection of the clips should consider cultural acceptance based on the cultural background of the intended subjects.

It is well-known that cultures affect the elicitation of emotions from certain clips. Studies investigating different cultures by Mesquita, Frijda and Scherer (1997) and Richerson and Boyd (2008) concluded that emotion responses and triggers differ from one culture to another. Sato, Noguchi and Yoshikawa (2007) discussed cultural influences on emotional responses. The study applied the same video clip stimuli developed by Gross and Levenson (1995) on Japanese participants. It has been reported that non-target emotions have been elicited (i.e. excitement, embarrassment and surprise) while watching the amusement film. One possible reason for this inconsistency is that the amusement clip, which shows a woman pretending to act out a sexual behaviour, is not culturally accepted by the Japanese. A previous cross-cultural study on sexual permissiveness reported that the Japanese are more conservative than Americans. Thus, cultural differences affect the emotional response towards film clips. Collecting and evaluating lists of emotion eliciting film clips should suit the cultural background of the intended subjects.

Automatic Recognition of Emotions

HCI is being increasingly recognised and promoted as an important aspect of software systems and products (Buie, 1997). One of the improvements in HCI is the inclusion of automated emotion recognition. To give the computer the ability to recognise emotions, the full process of pattern recognition steps is required. This includes data collection, data pre-processing, feature extraction and classification. Affective computing measures and techniques have been surveyed by Calvo and D'Mello (2010). Regardless of the emotion model/categories used, emotions can be detected using different modalities such as speech, facial expression, body gesture and physiology signals etc. (Calvo & D'Mello, 2010; Pan-tic & Rothkrantz, 2003; Wu, Parsons, Mower, & Narayanan, 2010). These modalities measure objective physical response to emotions using different sensors/devices (e.g. microphone, camera).

Regardless of the modality used to automatically recognise emotions, the computer has to be trained from a big sample of observations. To obtain accurate automated emotion recognition, supervised machine learning is preferred (Mohri, Rostamizadeh, & Talwalkar, 2012). In supervised learning, the classifier learns from labelled training observations, and then applies the resulted classification model to new unseen observations (testing data). Hence, the performance of supervised learning techniques depends on the quality of the training data and the quality of the labels.

However, emotion labelled training data from experts is expensive and requires time and effort. Moreover, since the emotions are highly subjective, expert labelling of emotions should be validated by a big sample of the population. Tarasov, Delany and Cullen (2010) have shown that it might be impossible or too expensive to acquire actual emotion labels. Thus, they used crowdsourcing platforms to recruit and encourage public annotators to annotate and label emotions from already chosen speech data. Our proposed platform focusses on both collecting and rating elicited emotions from the public using video clips. The following section (Crowdsourcing Platforms for Emotions) elaborates on using crowdsourcing in labelling emotions.

Moreover, most automated emotion recognition systems use emotional stimuli to induce emotions. These stimuli are usually selected from experts and assume the elicitation of the target emotion. As mentioned in the previous section, these stimuli need to be collected, rated and validated for their effectiveness in eliciting the intended emotions.

To sum up, to have accurate automated emotion detection labelled emotion data is preferred for training. While emotion stimuli responses used for automated emotion detection are being measured objectively using sensors, emotion labels are highly subjective. Therefore, a refined and validated stimuli list that induces the target emotions should be collected and rated by experts as well as by big samples of the population. This research aimed to design a crowdsourcing platform for this purpose.

Crowdsourcing Platforms for Emotions. Crowdsourcing is the practice of dividing labour between large numbers of (typically online) workers. It is a promising method for outsourcing (Hupont, Lebreton, Mki, Skodras, & Hirth, 2014). Another definition of crowdsourcing is the act of breaking down work into many small independent units and distributing them among a large number of people, usually over the web (Mohammad & Turney, 2013). Crowdsourcing can be a fast, cheap and effective way to collect data from a wide range of demographics.

Due to the expansion of online social networks, the crowd have been utilised for emotion studies. Using crowdsourcing platforms in affective computing could enhance the process of obtaining emotion elicitation media and rating and validating these media and annotating and labelling emotional media. Several studies have been conducted utilising crowdsourcing platforms to label emotional speech (Tarasov et al., 2010), annotate facial responses to online videos (McDu, El Kaliouby, & Picard, 2015), annotate videos (Soleymani & Larson, 2010) and label emotion in response to music (Morton, Speck, Schmidt, & Kim, 2010). In Hupont et al. (2014), reliability of crowd rating emotional images have been investigated, where the rating scales were compared with laboratory-set rating. The results showed that the crowdsourcing platform was efficient and effective. As can be seen, most of these affect studies that utilised crowds focussed on annotation and labelling emotions from different media. None of the studies was aimed at collecting and rating emotion elicitation video clips. Therefore, our proposed platform could help overcome the limitations of previous studies. Nevertheless, utilising crowds to collect and rate emotion elicitation clips is challenging and may introduce uncontrolled variables due to collecting highly subjective perceptions of emotions from a wide range of participants from different cultures, languages, knowledge background etc.

Therefore, this research utilises crowdsourcing to collect and rate emotion elicitation media by designing a crowdsourcing platform. In this platform, participants' general information and demographics will be recorded to allow for accurate filtering and analysis of both clips and participants by emotion researchers. For individuals, this platform can be seen as a social media platform for sharing emotion-related information.

Design of Crowdsourcing for Emotion Elicitation Media. The proposed crowdsourcing platform is aimed at emotion researchers who must select emotion elicitation clips. To achieve this, the emotion elicitation media have to be collected and rated by the crowd. In this study, the participants registered their information and demographics for filtering of the results of the collected clips as well as their rating. The following sections analyse each component of the proposed platform. Figure 1 elaborates on the proposed software architecture from the viewpoint of users.

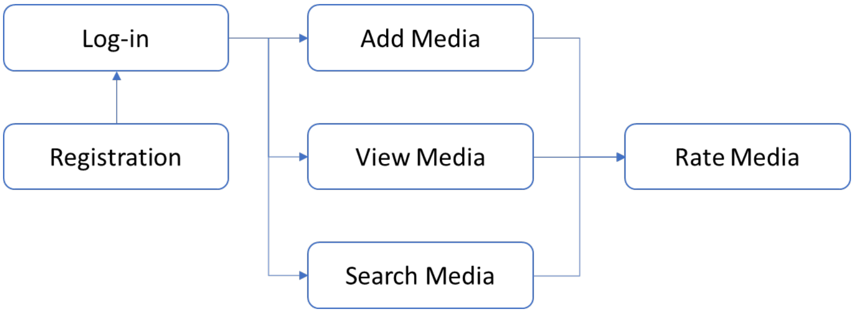

Figure 1. General design of software architecture from the viewpoint of users

Collecting Media. Gross and Levenson (1995) collected emotion elicitation clips for their study from film critics and experts. Even though an expert view is important for such a collection, it may not always be available. Therefore, besides involving experts, film creators and film critics in collecting emotional content for the platform, crowdsourcing was also utilised. The proposed crowdsourcing platform will involve the general public in adding emotional video clips. Clips can be segments from various resources such as films, series, plays or any source that stimulates emotions. However, for privacy and ethical reasons personal or real-life recordings should not be accepted. Participants wishing to add emotional clips could perform this task in two ways. They may refer to the source of the segment (e.g. film or series name, episode number), the start and the end time of the segment and the target emotion. Other participants or experts could then segment and upload the referred emotional elicitation segment. The second way is to directly upload the segment. A functionality for reporting unrelated clips would be made available to focus on rating and validating emotion-elicitation related clips.

Rating Media. The most important functionality of the proposed crowdsourcing platform is to utilise the crowd for rating the collected emotional clips. Since the clips might be mostly selected by experts, rating the clips is a critical component of the proposed platform. The scores for rating the selected media will result in producing a list of clips that elicit certain emotions. After watching each clip, the user rates the emotions elicited and the emotion intensity (arousal).

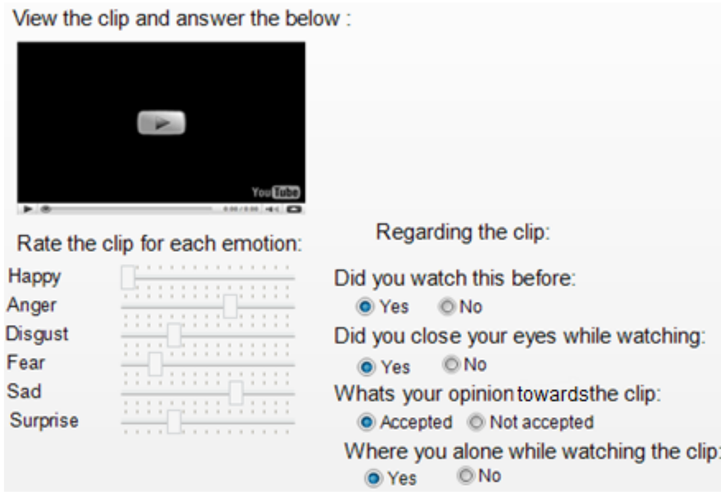

Figure 2. Design interface of the software

The emotions are the six universal Ekman categories: happy, anger, disgust, fear, sad and surprise (Ekman, 1992). The intensity of emotions will be evaluated on a scale of 0-10, with 0 representing emotion that was 'not felt', while 10 was an 'extremely felt' emotion (see Figure 2). Users will be able to rate one or more emotion category; this would help in eliminating clips that induce highly mixed emotions.

The user will answer general questions about the clips that might affect the user's rating of the emotion. For example, if the user is watching the clip with a group of people their rating might be affected (e.g. they might feel embarrassed to show their actual emotion or may experience elevated intensity of amusement as a result of being in a group). Cultural acceptance questions are to be included to give emotion researchers the flexibility to exclude clips with low cultural acceptance.

A threshold of the number of clips to be viewed by users at a time could be set to reduce the effect on emotion rating. Behavioural analysis of user rating could be used to determine if blocking clip viewing from the user is required. If the user rated several clips with high intensity of sadness, the platform blocks the user from viewing clips for a certain time frame.

Participant Registration. Since the main goal of this platform is research use, demographic information is critical as it is needed for the analysis of the collected and rated clips. Therefore, users are required to register, filling in general information about themselves and answering demographic questions for that one time only. The user may then add and rate media without having to re-enter demographic information. The demographic questions take into consideration cultural information and background. Demographic information will be used for filtering and analysing the collected and rated media. Users from age 13 and above may participate and rate all clips. Some specific clips are shown to kids and their rating activities are limited to those clips only. Videos are filtered to suit the age stated in the registration form.

Search Filters. Searching and filtering the clips is enabled for both researchers and users. Users will be able to search and select clips they want to watch based on language, culture and target emotion. For example, a user wishing to watch amusement elicitation clips may filter the clips for amusement emotion. Users with no preference will be shown a random selection of clips. However, with a random view the user has no expectation of what emotion will be viewed. Thus, to avoid the variability of continuous clip order in the analysis, as well as to assure the rating will not be effected by the previous clip, viewing a natural clip has been suggested (Maffei et al., 2014; Bednarski, 2012; Bartolini, 2011). However, in our case, the rating and the questions for each clip could act as a naturalising of emotions.

For emotion researchers, search filters are used for data analysis and media collections. The process where researchers could conduct their evaluation is explained in the following section.

Selecting Emotion Elicitation Media - Research Methodology. The main purpose of the proposed crowdsourcing platform is to facilitate the collection and rating of emotion elicitation media for emotion analysis investigation. The selected media should elicit emotions in a replicable way, and should be evaluated and validated for their capacity of eliciting the target emotion.

As mentioned earlier, emotion elicitation media are to be collected by experts such as film creators, directors and film critics as well as the general public through crowdsourcing platforms. The collection of emotion-inducing media using crowdsourcing platforms will ensure a wide range of demographics; cultural background, for instance would vary greatly. The crowd are expected to rate the emotions and their intensity as elicited by each medium they view. The crowdsourcing platform is used to ensure a wide range of demographic samples not only for collecting emotional clips, but also for rating the effectiveness of the clips in eliciting the target emotion. The proposed platform has the following main functionalities:

- Registration: The user signs up once and then uses his or her username and password for log-in. During registration, the user should fill demographic questions, as described earlier. This will allow for demographic questions to be stored and filled in by the user only once to avoid having to repeat the procedure, which many find tedious. Researchers who use the software will have a different access route, which will allow review of the results and analysis.
- Adding videos: Video clips can be added by both experts and registered users (crowdsourcing).
- Rating videos: Questions will appear after watching each video to rate the elicited emotions. After collecting these ratings, a standard list of clips can be obtained that could then be used to create models for automatic emotion recognition.
- Search video: The user will be able to filter and view clips by different categories.
- Analysing rated videos: The researchers have access to the demographic questions and the rated videos to help them filter the results based on their research needs.

Once the media have been collected and rated for emotional elicitation effect, emotion researchers may analyse the emotion elicitation rating. This analysis would evaluate the emotion-inducing media to filter a desired set of clips that are effective in inducing the target emotion or combination of emotions.

DATA ANALYSIS AND EVALUATION

As mentioned earlier, the proposed platform utilises a crowdsourcing environment to collect and rate emotion stimuli. Several data analysis functionalities and filtering queries will be provided by the platform to aid emotion researchers to evaluate the effectiveness of the clips in inducing the target emotion(s).

The evaluation is done by filtering the big pool of collected media from the experts and the crowd to a refined list based on the elicitation rating of target and non-target emotions, that is, the media have to be analysed based on target vs. non-target emotions. For example, media that elicit mixed emotions could be eliminated for studies that focus on single-emotion elicitation. This could be done based on the statistical analysis results, where the clips with the minimum mix of non-target emotions will be categorised as single emotion. Statistical analysis could indicate the clips with the maximum mix of certain emotions, such as clips that can elicit the feelings of fear and disgust at the same time.

For gender difference evaluation, the T-test, for example, could be used; this is provided by the platform. Alternatively, comparing emotion groups could utilise ANOVA and/or chi-squared statistical tests to analyse the correlation between the subjects' emotion elicitation rating and the clip's emotion category. Other different variables could be analysed and used as filtering queries of emotion analysis based on the demographic sample, such as age, cultural background, educational level, interests etc. An example of emotion analysis by Sato et al. (2007) saw them analysing clips by discrete emotions by viewing the film, 'When Harry Met Sally' to measure the amusement emotion, as seen in Figure 3, for target and non-target emotions.

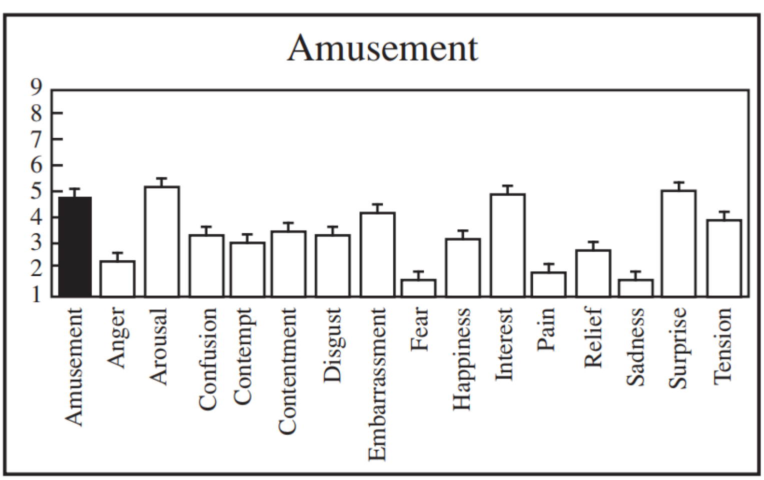

Figure 3. Example of emotion analysis from amusement elicitation clip (Sato et al., 2007)

Validation

Based on the data analysis and evaluation, emotion researchers would be able to refine a set of emotion elicitation clips that is customised for their research. Depending on the researchers' investigation, validating the refined clips could use different approaches. The validation step is a critical step for measuring the effect of the emotion-inducing media on a small sample, before these media are used for collecting big data of emotional measures.

To measure emotional response to the refined stimuli, several devices may be used. Besides using a camera to measure facial responses to emotion stimuli, different wearable devices may be used to measure physiological responses. Such devices could measure heart rate, body temperature and skin conductance level. The signals from each of these sensors should be pre-processed for extracting features that facilitate the recognition of emotions. The classification performance would measure the accuracy of recognising the labelled rated elicited emotions from the clips with the physiological responses.

A variety of sensors have been used in emotion recognition literature. For catching facial expression a device called Microsoft Kinect and its SDK (Software Development Kit) have been used for extracting facial responses (Zhang, 2012). Previous studies have used, for instance, the Tobii Eye Tracker device to measure eye responses to stimuli to recognise positive and negative emotions (Alghowinem, Alshehri, Goecke, & Wagner, 2014). For skin conductance level and temperature, some studies have used a Q-sensor (Al-Mutairi, Alghowinem, & Al-Wabil, 2015). While the mentioned devices are only examples, several other devices have been used, as cited in emotion recognition literature, for detection of a variety of emotion categories and emotion models as mentioned earlier (Calvo & D'Mello, 2010).

Once the refined list of emotion elicitation media has been validated with classification and physiological measures, the final step is to collect big data of emotional measures. This data are to be used to create a trained supervised model for each emotion using specific physiological measures. These models are then used to give the computer the ability to recognise the user's feelings. Because emotions are highly subjective, having a big sample of different demographics of subjects would allow for creating emotion models specific for different variables. Specificity of training emotion models would increase the accuracy of emotion recognition.

CONCLUSION

Acknowledging the importance of recognising and detecting users' emotions to enhance user interaction with a variety of systems, system developers have been investigating methods of emotion recognition. Such intelligent systems could have a positive influence in several fields including Human-Computer Interaction, psychology, neurobiology, sociology and marketing. To give computers the ability to recognise emotions, they should be trained using big data of emotion measures. To obtain emotion measures from big samples of subjects, a validated emotion elicitation trigger should be collected. Several emotion elicitation methods have been investigated, and video clips have proven to be the most effective. Several studies have collected emotion elicitation clips, yet these studies used only a limited demographic sample. Moreover, the selected clips from these studies might not have a universality effect on different cultures. Therefore, this research aimed to design a crowdsourcing platform to collect and rate emotion elicitation media from a wide demographic sample with consideration of cultural background. The main aim of this crowdsourcing platform is to help emotion researchers to collect and evaluate highly-rated emotion elicitation media. For individuals, this platform could be seen as a social-media platform for sharing and rating emotional media. Using filtering and analysis functionality of the platform, emotion researchers can select, evaluate and validate the rated media. Evaluation of the media could be performed to select, for example, the media that highly elicit the target emotion and to eliminate the media that elicit mixed emotions. Another example is to select the media that elicit a certain combination of emotions. For validating the evaluated media, measuring physiological signals and reaction (e.g. facial response) could be performed using wearable devices. Once the emotion elicitation media have been validated for their effectiveness, data collection of emotion measures can be performed.

ACKNOWLEDGEMENT

This work was supported by the Human-Computer Interaction Research Group, Prince Sultan University, Riyadh, Saudi Arabia [RG-CCIS-2017-06-01].

REFERENCES

Alghowinem, S., AlShehri, M., Goecke, R., & Wagner, M. (2014). Exploring eye activity as an indication of emotional states using an eye-tracking sensor. *Intelligent systems for science and information, 542*, 261-276.

Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., & Parker, G. (2012, May). *From Joyous to Clinically Depressed: Mood Detection Using Spontaneous Speech*. In FLAIRS Conference.

Al-Mutairi, N., Alghowinem, S., & Al-Wabil, A. (2015, August). Comparison of User Responses to English and Arabic Emotion Elicitation Video Clips. In *International Conference on Cross-Cultural Design* (pp. 141-152). Springer, Cham.

Bartolini, E. E. (2011). *Eliciting emotion with film: Development of a stimulus set*.

Bednarski, J. D. (2012). *Eliciting Seven Discrete Positive Emotions Using Film Stimuli* (Doctoral dissertation, Vanderbilt University).

Buie, E. A., & Vallone, A. (1997). Integrating HCI engineering with software engineering: A call to a larger vision. In Design of Computing Systems: Social and Ergonomic Considerations (*Proceedings of the Seventh International Conference on Human-Computer Interaction*) (Vol. 2, pp. 525-530).

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing, 1*(1), 18-37.

Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media, 29*(3), 241-250.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3-4), 169-200.

Fernández, M. C., Pascual, M. J., Soler, R. J., & Fernández-Abascal, E. G. (2011). Spanish validation of an emotion-eliciting set of films. *Psicothema, 23*(4), 778-785.

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion, 9*(1), 87-108.

Hupont, I., Lebreton, P., Mäki, T., Skodras, E., & Hirth, M. (2014, July). Is affective crowdsourcing reliable? In *Communications and Electronics (ICCE), 2014 IEEE Fifth International Conference on* (pp. 516-521). IEEE.

Jaimes, A., & Sebe, N. (2007). Multimodal human–computer interaction: A survey. *Computer Vision and Image Understanding, 108*(1), 116-134.

Jerritta, S., Murugappan, M., Nagarajan, R., & Wan, K. (2011, March). Physiological signals based human emotion recognition: a review. In *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on* (pp. 410-415). IEEE.

Kehrein, R. (2002). *The prosody of authentic emotions*. In Speech Prosody 2002, International Conference.

Kory, J., & D'Mello, S. K. (2014). *Affect elicitation for affective computing*. The Oxford Handbook of Affective Computing.

Maffei, C., Roder, E., Cortesan, C., Passera, F., Rossi, M., Segrini, M., ... & Fossati, A. (2014). Kinematic elicitation of basic emotions: A validation study in an Italian sample. *Psychology, 5*(09), 1065.

McDuff, D., El Kaliouby, R., & Picard, R. W. (2015, September). Crowdsourcing facial responses to online videos. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on* (pp. 512-518). IEEE.

Mesquita, B., Frijda, N. H., & Scherer, K. R. (1997). Culture and emotion. *Handbook of Cross-cultural Psychology, 2*, 255-297.

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence, 29*(3), 436-465.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of machine learning*. MIT press.

Morton, B. G., Speck, J. A., Schmidt, E. M., & Kim, Y. E. (2010, July). Improving music emotion labeling using human computation. In Proceedings of the acm sigkdd workshop on human computation (pp. 45-48). ACM.

Pantic, M., Pentland, A., Nijholt, A., & Huang, T. (2006). Human computing and machine understanding of human behaviour: A survey. *Proceedings of the 8th International Conference on Multimodal Interfaces* (pp. 239–248).

Panic, M., & Rothkrantz, L. (2003, September). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE, 91*(9), 1370–1390.

Philippot, P. (1993). Inducing and assessing differentiated emotion feeling states in the laboratory. *Cognition and Emotion, 7*(2), 171–193.

Richerson, P. J., & Boyd, R. (2008). *Not by genes alone: How culture transformed human evolution.* University of Chicago Press.

Russell, J. A. (1979). Affective space is bipolar. Journal of Personality and Social Psychology, 37(3), 345–356. Retrieved from http://dx.doi.org/10.1037/0022-3514.37.3.345

Sato, W., Noguchi, M., & Yoshikawa, S. (2007). Emotion elicitation effect of films in a Japanese sample. *Social Behaviour and Personality: An international Journal, 35*(7), 863–874.

Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion eliciting films: A new tool for emotion researchers. *Cognition and Emotion, 24*(7), 1153–1172.

Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., & Gielen, S. (2001). Acoustic correlates of emotion dimensions in view of speech synthesis. In *Seventh European Conference on Speech Communication and Technology.*

Soleymani, M., & Larson, M. (2010). *Crowdsourcing for affective annotation of video: Development of a viewer reported boredom corpus*. Retrieved from https://www.ischool.utexas.edu/~cse2010/slides/Soleymani_CSE_SIGIR2010.pdf

Tao, J., & Tan, T. (2005, October). Affective computing: A review. In *International Conference on Affective computing and intelligent interaction* (pp. 981-995). Springer, Berlin, Heidelberg.

Tarasov, A., Delany, S. J., & Cullen, C. (2010). *Using crowdsourcing for labelling emotional speech assets*.

Westermann, R., Stahl, G. U. N. T. E. R., & Hesse, F. (1996). Relative effectiveness and validity of mood induction procedures: analysis. *European Journal of Social Psychology, 26*, 557-580.

Wu, D., Parsons, T. D., Mower, E., & Narayanan, S. (2010, July). Speech emotion estimation in 3D space. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on* (pp. 737-742). IEEE.

Wundt, W. (2009). *Outlines of psychology (1897).* In Foundations of psycho-logical thought: A history of psychology (pp. 36–44). Sage Publications, Inc.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(1), 39–58.

Zhang, Z. (2012). Microsoft kinect sensor and its effect. *MultiMedia, IEEE, 19*(2), 4–10.

Zhou, F., Ji, Y., & Jiao, R. J. (2013). Affective and cognitive design for mass personalization: status and prospect. *Journal of Intelligent Manufacturing, 24*(5), 1047-1069.