Measuring spontaneous vocal and facial emotion expressions in real world environments

Khiet P. Truong, Mark A. Neerincx, and David A. van Leeuwen
TNO Defence, Security and Safety, P.O. Box 23, 3769ZG, Soesterberg, The Netherlands
{khiet.truong, mark.neerincx, david.vanleeuwen}@tno.nl

Affective computing [1] has been introduced in the nineties as a research area that aims at designing systems that can recognize, interpret and synthesize emotions. For the development of these systems, databases containing recordings of natural emotions and descriptions (annotations) of these emotions are needed. In the past few years, there has been much discussion about the use of acted or real emotion data, and on how to annotate and measure emotion. In this paper, we elaborate on how to measure naturally occurring emotions in real world environments and how to establish a description (annotation) of these emotions with the aim to develop automatic emotion recognizers. We focus on vocal and facial expressions which can be measured in a relatively unobtrusive and simple manner. Learning from the findings and difficulties experienced during data collection in three different real world environments, we present a fourth experiment in which we measured and collected, natural vocal and facial emotion data in a multi-player gaming environment in a relatively short amount of time.

We had the opportunity to record vocal (and sometimes facial data) in three different real world situations where emotional expressions and behavior were expected:

1) Emergency situations on a naval ship which needed to be solved by operators.

In [2], an experiment is described in which the goal was to measure cognitive task load by processing several measurements, including vocal and facial expressions. In the ship control center, the operator had to deal correctly with the emergencies that occurred on the naval ship (e.g., fire or platform system failures). Several scenarios were designed to evoke low, medium or high task load with operators. High-quality webcams and head-mounted microphones were used to record video and audio. After each scenario, the participants had to rate task complexity and subjective effort on a five point scale. The idea was to find correlations between task load (or stress) and vocal and facial measurements. Unfortunately, the current vocal and facial analysis tools proved to be insufficiently robust against the realistic environmental conditions; e.g., background noise in audio and video, moving head postures etc., made it difficult to perform a reliable acoustic and facial analysis.

2) A flood (crisis) situation which needed to be dealt with by local authority members

This exercise was organized by the DECIS lab [3]. The 1-day scenario started with a simulated flood disaster in a small community in the Netherlands. During the crisis situation, a crisis policy team consisting of eight persons held five meetings in which the team members had to make time-pressured decisions. Vocal recordings were made using an 8-channel circular microphone array that was positioned at the table. We expected to find vocal expressions related to e.g., stress or frustration. However, this seemed less apparent than expected: the number of emotional expressions found in the meetings was too low to perform vocal and statistical analysis. An alternative way to describe this type of data could be to apply discourse analysis which can give more insight in the interaction between team members.

3) Players immersed in a virtual reality game

Exercise in Immersion (developed by Marnix de Nijs and V2_lab [4]) is an art-game played in an existing physical space. The player(s) wore a head-mounted display and a crash suit and immersed themselves in a virtual world. We expected a lot of vocal expressions, due to the player's immersion in the game and the "fun" factor. High quality voice recordings were made through close-talk microphones. Players were asked to "think aloud". Afterwards, the players filled in two questionnaires that were related to the emotions felt and the amount of presence experienced. For measuring presence, we used the Igroup Presence Questionnaire [5]. We are currently analyzing the data, and we can already observe that the "think aloud" procedure might not be very suitable for voice analysis since the players reported that it felt awkward to "think aloud" during the game which might have affected the way they express their emotions vocally and the naturalness of the speech.

Taking into account the "lessons learned" from the three previous recording sessions (environmental "noise" should be reduced as much as possible, the sparseness and naturalness of emotion expressions should be dealt with), we designed a fourth experiment to acquire annotated, multimodal, natural emotion data in a relatively short amount of time [6]. Participants played a first person shooter video game (Unreal Tournament) against each other in teams of 2x2. In total, we recorded 1120 minutes audiovisual data (28 participants x 20 minutes x 2 sessions). Several steps were taken to evoke as much vocal (and facial) interaction and expressions as possible by 1) asking participants to bring a friend as a team member, 2) manually generating "surprising events" in the game (e.g., the sudden appearance of monsters), and 3) granting bonuses for the winning team and good team collaboration. High quality audio recordings were made through close-talk microphones and recordings of the face were made with high quality webcams (placed at eye level). To obtain emotion annotations in a relatively short amount of time, we decided to use self-reported emotion values as subjective emotion measurements. After playing the video game, the participants had to watch their own video twice, and had to annotate their own emotions in two different ways: 1) choose an appropriate emotion label from a given set of labels whenever applicable (event-based, selection of multiple emotions at the same time was possible), and 2) give an arousal and valence estimation on a scale of 0-100 (continuously, each 10 seconds).

Frustration seems to be the most frequent emotion that was reported by the participants themselves (see Figure 1). Figure 2 presents the self-reported arousal and valence values: we can observe that high arousal is often accompanied with high valence. These self-reported emotion measurements need further investigation: we have to deal with both the personality of the participant and the subjectivity of the annotator (which in this case is equivalent to the participant).

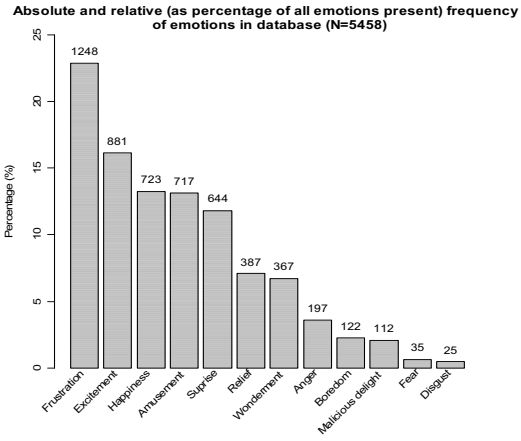

Figure 6. *Categorical emotion labeling*

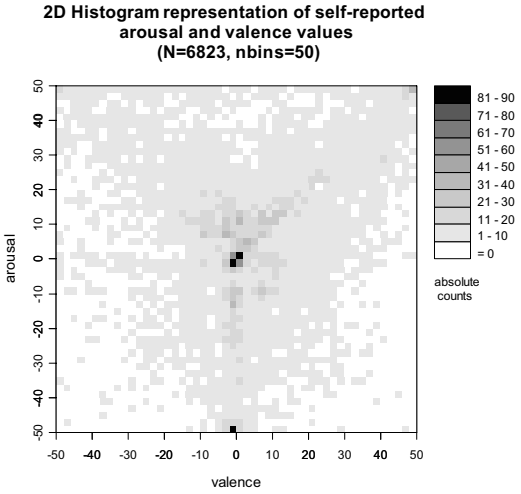

Figure 5. *Continuous arousal and valence labeling (darker areas indicate higher absolute frequency)*

To summarize, we can identify two main challenges for building automatic emotion recognizers: 1) the development of databases containing natural emotions that are annotated and measured in a reliable way, and 2) the development of robust automatic emotion recognizers for real-time emotion sensing in the real world. Our gaming experiment seems to provide a solution for the first problem (see Figure 3 for some example screenshots). However, we still need to validate our subjective emotion annotations which can be done by applying the "theory on context effects" (introducing a stressor in the game evokes a stress reaction) or by using multiple annotators (high agreement between the annotators can indicate high reliability). The second challenge also implies that emotion recognizers should be able to deal with "gradations" of emotions (which are very common in realistic emotions) instead of only the extremes. In future research, we will take up these challenges and measure emotions in complex, hectic environments (e.g., process control) with the aim to develop real-world emotion sensing systems.

Acknowledgements

This study is supported by MultimediaN, a Dutch BSIK-project (http://www.multimedian.nl).

References

1. Picard, R.W. (1997). *Affective computing*. MIT Press, Cambridge, MA.
2. Grootjen, M., Neerincx, M.A., Weert, J.C.M. van, and Truong, K.P. (2007). Measuring cognitive task load on a naval ship: Implications of a real world environment. *Proceedings of 12th HCI International (HCII) 2007, Beijing, China, July 22-27 (LNAI 4565),* 147-156.
3. http://www.decis.nl *[website]*
4. http://www.v2.nl *[website]*
5. http://www.igroup.org/projects/ipq/ *[website]*
6. Merkx, P.A.B., Truong, K.P. and Neerincx, M.A. (2007). Inducing and measuring emotion through a multiplayer first-person shooter computer game. *Proceedings of the Computer Games Workshop 2007, Amsterdam, The Netherlands*, 231-242.
7. http://www.noldus.com/site/doc200705001 *[website]*
8. Uyl, M. den; Kuilenberg, H. van (2005). The FaceReader: Online Facial Expression Recognition. *Proceedings of Measuring Behavior 2005,* 598-590.

Figure 7. *Screenshots of facial expressions recorded in gaming experiment, emotion classification provided by FaceReader [7,8]*