Towards Detection of Side Activities and Emotions of Anonymous TV Viewers through Body Postures

Chiew Seng Sean Tan, Jan Van den Bergh, Johannes Schoning, Kris Luyten
Hasselt Univerity - tUL - iMinds
Expertise Centre for Digital Media, Wetenschapspark 2, 3590 Diepenbeek, Belgium
{sean.tan, jan.vandenbergh, johannes.schoening, kris.luyten}@uhasselt.be

ABSTRACT
TV viewers often perform various types of side-activities, including usage of mobile devices to browse, chat, e-mail or use social networks. These activities can be impacted by or related to TV content. In order to understand the possibility to unobtrusively monitor emotions of TV viewers, we want to get a better understanding of their body postures and side activities and how these relate to emotions. In this paper we explore the role and interconnections between body postures, side-activities and emotions of people watching TV through a questionnaire and a user study. We found that postures do not directly relate to emotions perceived from emotion eliciting video clips. Analysis of emotions and side-activities did reveal some interesting correlations between device usage and the experienced emotions, which can be useful to inform the design of applications on or in relation with the TV. A first user study showed that a subset of the relevant side-activities could be tracked based on posture data, thus enabling activity recognition and maybe even emotion estimation without the need for personal characteristics.

Author Keywords
Affective states, activity recognition, emotion elicitation, television, body posture.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): User Interfaces, Interaction Styles.

MOTIVATION & RELATED WORK
Television is part of our everyday life. The average person worldwide spends at least 3.2 hours a day watching television [2]. In 2003, Schmitt et al. [11] reported that, while watching TV, a person spends an average of 46% of their time on side-activities (based on observations three decades ago). The most common side-activities back then included social interaction, eating and reading. Their study, however, pre-dates the ubiquity of mobile technology, such as laptops, tablets and smartphones. These days, while watching TV, people use various mobile devices to browse the web, play games or socialize with other people. In addition, the devices can also be used to display the same or additional information or enhance the viewing experience with application-specific content (so called second-screen apps) [6]. On the other hand, peripherals connected to the TV, such as gaming consoles and their accessories allow unobtrusive observation of TV viewers. This motivates us to investigate and understand whether their side activities or body postures are related to their emotions, which could be interesting for both broadcasters and television distributors to create more personalized services.

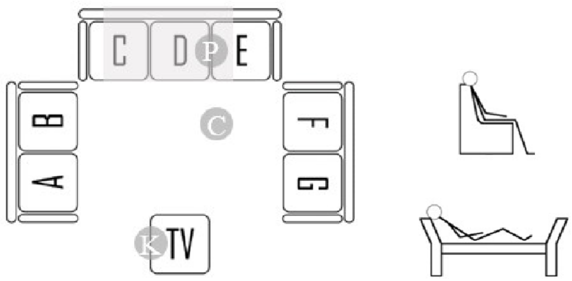
In this paper we explore how body postures and side-activities of TV viewers relate to how people feel while watching TV. We want to understand the relationships between TV viewers' body postures, side-activities, which can be observed through devices such as Kinect, and emotions. Such relations could allow tracking emotions of TV viewers on a large scale without revealing personal characteristics and thus provide in-depth insights of how viewers perceive this content.

Research regarding the emotions experienced while watching television is mostly related to commercials [1, 8]. However, to our knowledge, there is no literature that intertwines body postures, side-activities and corresponding emotions to investigate the way people watch TV. Therefore we conducted two studies to research this. In our first study we interviewed 15 participants regarding their sitting position, body postures, and side-activities while watching TV. The second study investigated the recognition of body postures and side-activities of people watching TV. In addition we explored if body postures could relate to the current emotion of a viewer while watching video clips.

STUDY 1: TV VIEWERS POSTURES AND ACTIVITIES
We conducted a first study to get a better understanding of the context in which people watch television. Fifteen European participants (10 males and 5 females, with a median age of 21 years), were asked about their TV watching habits using a questionnaire. The items in the questionnaire include graphical depiction of preferences on sitting positions and postures, side-activities and their frequency as well as technologies used while watching TV.

All participants reported to have a small viewing angle to their TV (Figure 1a) and respect a distance between 2.5 and

TVX'14, June 25 – June 27, 2014, Newcastle, UK.



(a) Sitting positions (b) Sitting postures

Figure 1: Most common sitting positions (C-E) and most frequent postures in study 1. Overlays indicate position of couch, participant (P), camera (C) and Kinect (K) in study 2.

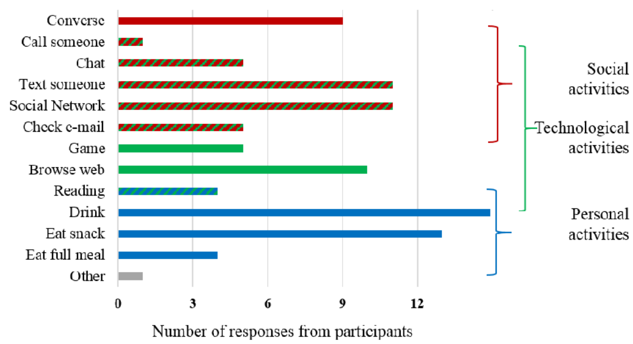


Figure 2: Activities and the devices (using green color-codes) used while watching TV. We can see more social and personal activities involving the use of technology-enabled devices.

4 meters, which is consistent with the results from the Japan Ergonomics society [2] gathered in 6 large countries on 4 continents. Figure 1b shows the most frequently reported sitting postures (out of four predefined postures): leaning back (top) and lying stretched backwards (bottom).

Figure 2 gives an overview of the side-activities and devices used by the participants. Social activities are colored in red tone, whereas technology activities have a green tone. Drinking and eating were the most frequent personal activities (blue tone). Reading was done using a book or a TV-guide (both by 2 persons). Mobile devices (smartphones: 12, laptops: 10, and tablets: 3) were predominantly used to perform asynchronous communication (texting, chatting, social networks, and checking e-mail). 30% of the participants also indicated that they used the Internet for various tasks while watching TV. Ten participants claimed to use synchronous communication and all other activities were performed by 40% of the participants. The results are in line with recent surveys such as the study conducted in Germany [12]. Our findings are also consistent with those of Schmitt et al. [11], but our results show, that more and more personal activities like reading are now transferred to the digital domain (see Figure 2 for details about social activities involving the use with more technology-enabled devices).

STUDY 2: USER OBSERVATION

We conducted a second study to observe body postures and side-activities in a living room lab setting. The goals were:

- Verify whether commonly reported postures and activities can be tracked through an off-the-shelf sensor.
- Explore correlation between, self-reported emotion, automatically tracked emotions and viewers' emotions.
- Explore relations between viewers' emotions and side activities.

Participants & Apparatus

In our living room lab, which we arranged according to our findings from the first study (see Figure 1a). The Kinect was vertically tilted and turned towards the participant in such a way that a person was visible from head to toe when standing exactly in front of the sofa seat at about 2.5 meters distance. Eleven participants were recruited (7 males and 4 females). Their ages ranged between 17 and 35 years old with median age at 26. They consisted of a mix of undergraduate and postgraduate students.

Procedure

After a briefing, participants were seated in a sofa as indicated in Figure 1a. Participants were first asked to perform five activities in random order with four objects (mobile phone, remote control, TV guide, magazine) placed on a table in front of them.

The participants were then connected with a skin conductance sensor on their non-dominant hand and a webcam was used to capture their facial expressions. A three-minute baseline was established by watching a designated emotionally neutral video clip (i.e. stardust simulation). Every participant viewed a series of six emotion elicitation video clips. The video clips were identical to those used by Ray [10], with the exception of the clip that should trigger disgust, for which an animation was used on the same topic. After each video clip, the participants were asked to solve a simple mental quiz. The purpose was to resume a neutral state to the participants by drawing their attention away from the previous video clip. When the participants finished viewing all video clips, they were asked to fill out a short survey in which they rated the intensity of their emotions, as well as valence and arousal on a 9-point Likert scale for each video clip. We then let the participants graphically define relations between emotions and various side-activities through colored lines, which represented three different scales of frequency. Each line connected options in two adjacent columns to complete the sentence: "While watching TV (1), if I feel a certain *emotion* (2), then/and I prefer to use the *device, object* (3), then/for the purpose of the following *activity* (4)".

Method & Apparatus

Skeleton Tracking & Activity Recognition

We implemented an activity recognizer using a depth camera to detect the users' postures in an unobtrusive manner. By creating a virtual skeleton using the OpenNI

Figure 3: T-test results showed that average emotions as (on a 9-point Likert scale) as measured and reported by the participants matched those of Ray [10]. The difference for *Leg surgery* could be because an animation was shown rather than real footage.

framework, we then used the method proposed by Kleinsmith et al. [5] to define skeleton features for detecting different poses (such as arms crossed, lean side wards, etc.). For the recognition of side activities, we used a moving window algorithm with 100ms time interval. Every trained activity is defined as a sequence of changes in body postures. The defined postures were accurately detected from skeleton tracking. We plotted the F-measure value of all participants for every threshold value for posture detection between 65% and 85%. The maxima of the curve for F-measure is reached at a threshold value of 66%. Using this threshold for posture detection, the activity recognition yield an accuracy percentage of 72.9%.

Affective State Recognition

The affective state recognizer uses a ProComp Infiniti to measure the value of skin conductance, which can be approximated as a proportion to the arousal of the participants [9]. To avoid attaching EMG sensors on the participants' face, we opted to use an unobtrusive face reading software called FaceReader to automatically analyze facial expressions for the valence value. The values of arousal and valence were then normalized so that they could be generalized over all the participants. We used an approach based on Mandryk [7] to determine the affective state by transforming the values of valence and arousal in a 2-dimensional emotion space. We used Weka to train a Naive Bayesian classifier on the 5-seconds window averages of valence and arousal of each participant. We applied a filter on the average window valence in the interval [-0.05, 0.05], corresponding to the neutral region in the arousal-valence space. An accuracy percentage of 66.3% was obtained from the trained classifier using the filtered data.

Results

Emotions perceived when watching video clips

The average rating for all emotions per video clip was calculated over all participants. Statistical testing using a T-test is then used to compare the results to those reported by Ray [10]. Figure 3 depicts the comparison of measured affective states, self-reported emotion ratings and the emotions reported in literature as reference values for the video clips. Ray defined the neutral emotion as the absence of other high-rated emotions. We observe that the measured affective states and self-reports of our questionnaire survey follow the results from Ray [10] except for *leg surgery*.

Body Postures and their relations to Emotions

We are interested to seek a relation between affective states and the detected postures during video clips. By reviewing the video recordings of the body postures, we found that eight of the participants sat in a laid back position (confirming with our questionnaire results) and practically did not move during the series of video clips besides re-sitting on the couch (with the exception of one participant leaning forward during the video clip on *Leg surgery*, which was intended to elicit disgust). This prohibits us of making assumptions about correlations between postures and the viewers' emotion.

Relations between Emotions and Activities

By cross-referencing the emotions, devices and their corresponding side-activities from the participants, we are able to determine the relational patterns of how the participants indicated the six basic emotions that would trigger a device mediated activity as presented in Figure 4.

The number of responses in this graph are the number of paths from column 1 (TV show) to column 4 (activity) accumulated from all study participants. No clear trends could be observed regarding the frequency of the activities in relation to specific emotions.

DISCUSSION

Consistent with literature, the participants took a relaxed sitting posture within the observable range of Kinect. The results indicate that it should be possible to detect high-level person-independent postures in a relatively reliable

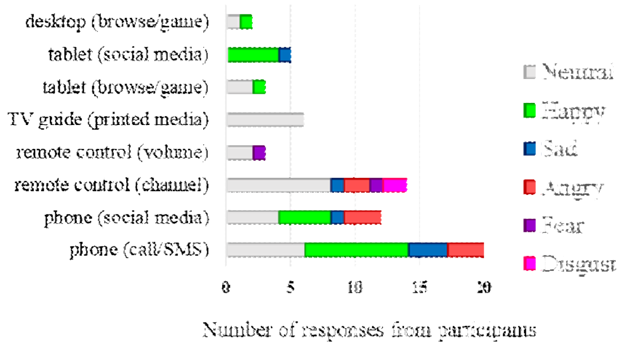


Figure 4: Activities performed in response to an emotion.

manner from skeleton information. We also showed that these postures can be used to detect activities with reasonable accuracy. It seems plausible that these detection rates could be even further improved when combined with properties of body movement, similar to the work presented by Kleinsmith et al. [5].

In the user study, we observed that participants rarely changed their posture during the series of video clips. This could be attributed to the fact that the feelings were not highly intense nor in group. When both factors occur, people are known to show (reactions to) their emotions in movement [3]. In general, the measured and self-reported emotions were consistent with those reported by Ray [10]. One exception is the *Leg surgery*. This might be attributed to the fact that we used an animation instead of a real-life amputation video segment causing a less intense emotion.

The relations between emotions and activities (Figure 4) can provide a promising direction for future research. Some notable relational patterns are: (1) the TV guide (printed media) was only associated with neutral emotion, while the remote control was only associated with both neutral and negative emotions, and (2) when people felt angry, they indicated to use either the phone (to engage in social contact) or the remote control (to switch channels). Channel switching occurs when experiencing a neutral emotion, although also negative, but not positive, emotions can be a trigger. Fear and disgust were only related to remote control usage, which could indicate that people do not wish to share these emotions (immediately). Anger, happiness and sadness were primarily related to social activities. Similar results, although not as explicit have been found in a study by Kim et al. [4] on emotions shared on Twitter independent of TV watching.

These findings can be applied for building interfaces that automatically analyze TV viewers' body postures or side activities on platforms that allow adaptive television services, such as XBox Live. The detected postures or side-activities could inform whether people in front of TV are still watching or are performing another activity and could even be used to infer likely emotions (Figure 4). Furthermore, these calculations could be made without revealing personal characteristics such as height or size.

We believe our study is a first step to provide a better understanding of what can be automatically inferred from body postures, side-activities and emotions of TV viewers although further research is necessary with larger groups and more emotions (such as boredom).

ACKNOWLEDGMENTS

We would like to thank Dries Vanmeert for his contributions. Part of this work was supported by the ITEA2 "Empathic Products" project, ITEA2 1105, co-funded by IWT 120171.

REFERENCES

1. Goldberg, M. E., and Gorn, G. J. Happy and sad tv programs: How they affect reactions to commercials. *Journal of Consumer Research* (1987), 387–403.
2. Japan Ergonomics Society, Special Committee on FPD-TV Ergonomics. Ergonomic Design Guidelines for Flat Panel Display Televisions, 2012.
3. Juhlin, O., and Onnevall, E. On the relation of ordinary¨ gestures to tv screens: General lessons for the design of collaborative interactive techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, ACM (New York, NY, USA, 2013), 919–930.
4. Kim, S., Bak, J., and Oh, A. H. Do You Feel What I Feel? Social Aspects of Emotions in Twitter Conversations. In *ICWSM* (2012), 495–498.
5. Kleinsmith, A., Fushimi, T., and Bianchi-Berthouze, N. An incremental and interactive affective posture recognition system. In *Workshop on Adapting the Interaction Style to Affective Factors* (2005).
6. Lochrie, M., and Coulton, P. Sharing the Viewing Experience through Second Screens. In *In Proc. of EuroiTV '12*, ACM (2012), 199–202.
7. Mandryk, R. L., and Atkins, M. S. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *Int. J. Hum.-Comput. Stud. 65*, 4 (Apr. 2007), 329–347.
8. Myers, J., and Sar, S. The influence of consumer mood state as a contextual factor on imagery-inducing advertisements and brand attitude. *Journal of Marketing Communications* (2013), 1–16.
9. Nakasone, A., Prendinger, H., and Ishizuka, M. Emotion recognition from electromyography and skin conductance. In *Proc. of the 5th International Workshop on Biosignal Interpretation* (2005), 219–222.
10. Ray, R. D. Emotion elicitation using films. *Handbook of emotion elicitation and assessment* (2007), 9–28.
11. Schmitt, K., Duyck Woolf, K., and Anderson, D. Viewing the Viewers: Viewing behaviors by children and adults during television programs and commercials. *Journal of Communication 53*, 2 (2003), 265–281.
12. TNS Emnid and IP Deutschland. DigitalBarometer -
13. Parallelnutzung: Interaktivitat beim Fernsehen,¨ December 2011.