Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 32 (2014) 1016 – 1021

Procedia Computer Science

International Workshop on Enabling ICT for Smart Buildings (ICT-SB 2014)

Affective interaction in smart environments

Maurizio Caon[a,c,]*, Leonardo Angelini[a,b], Omar Abou Khaled[a], Denis Lalanne[b], Yong Yue[c,d], Elena Mugellini[a]

[a]University of Applied Sciences and Arts Western Switzerland, Bd de Pérolles 80, Fribourg CH1700, Switzerland
[b]University of Fribourg, B de Pérolles 90, Fribourg CH1700, Switzerland
[c]University of Bedfordshire, Park Square, Luton LU1 3JU, United Kingdom
[d]Xi'an Jiaotong-Liverpool University, 111 Renai Road, SIP, Suzhou, Jiangsu 215123, China

Abstract

We present a concept where the smart environments of the future will be able to provide ubiquitous affective communication. All the surfaces will become interactive and the furniture will display emotions. In particular, we present a first prototype that allows people to share their emotional states in a natural way. The input will be given through facial expressions and the output will be displayed in a context-aware multimodal way. Two novel output modalities are presented: a robotic painting that applies the concept of affective communication to the informative art and an RGB lamp that represents the emotions remaining in the user's peripheral attention. An observation study has been conducted during an interactive event and we report our preliminary findings in this paper.

Keywords: Emotions; Facial Expression; Computer Mediated Communication; Multimodal Interaction; Smart Environment;

1. Introduction

A smart building is not only about efficiency and automation, but the smart environments of the future should be able to satisfy different human needs and to establish a special relationship with the inhabitants. The human being needs much more than mere improvement of the efficiency or functionality; a big part of the human life is composed of emotions. In the last decades, the computer science neglected this factor but some years ago part of the scientific community understood this needs and founded a new filed of research: the affective computing.

* Corresponding author. Tel.: +41 0264296748.
E-mail address: maurizio.caon@hes-so.ch

doi:10.1016/j.procs.2014.05.527

In this position paper, we introduce the concept of ambient interaction for affective communication. The environment is able to provide multimodal context-aware feedback that makes the user to live an immersive experience. Moreover, the surfaces are smart and they allow acceding information. Also the wearable devices are connected and allow the user to interact ubiquitously.

In this concept, furniture does not assume only a role of décor or functionality; in fact, it allows displaying and creating information in an unobtrusive manner. The distribution of different types of smart furniture allows also distributing the information according to the context. The context awareness grants the possibility to provide the information in the best modality and to display it in the right part of the environment. The different elements of the furniture are connected and can share information.

In order to understand the real potential of this concept, we developed a prototype and conducted an experiment at the Affective Computing and Intelligent Interaction (ACII) 2013[1]. We participated in the demo session where we set up the prototype and recorded with a camera the interaction between the conference attendees and the system. We conducted an observation study in two different times: on the field and after the experiment, watching the recorded video.

The next section describes the prototype that we developed as proof-of-concept and in Section 3 we report our observation study and its outcomes. The last section is dedicated to the conclusion of the paper.

2. Prototype

The prototype has been developed for the computer-mediated communication (CMC) of emotions[2]. In particular, this prototype allows the sharing of emotions on “Social Awareness Streams” (SASs). This choice is due to the observation of how people changed the way of communicating in the digital era. Nowadays, human-to-computer-to-human interaction (HCHI) is becoming more and more important and some psychologists and sociologists claim that the current trend of interpersonal communication is shifting from the physical to the virtual world[3,4]. A consequence of the shift to the CMC is that all the paralanguage of a normal face-to-face communication is missing. The SASs do not allow a variety of different modalities and, usually, they have also a limited number of characters for messages; therefore, the expressivity of emotions in SASs is quite limited. This prototype allows the user to interact with the whole environment to share their emotions in a more natural way. In fact, this system exploits multimodality since the use of natural means of communication facilitates the human interaction with the machine[5]. Several input and output modalities are offered to the user in order to share and receive Twitter messages augmented with emotional contents.

This prototype allows the user to interact with several surfaces distributed in the environment (e.g., tables, walls, TVs, smartphones et cetera). In fact, it adopts the Inter-Face framework, which has been conceived in order to guarantee the compatibility among multiple heterogeneous technologies to turn any object into an interactive surface[6]. In the actual setup used for these preliminary experiments, the user can share a text, an image and his/her personal emotional state using three different interfaces: a Smart TV, a smartphone and a laptop.

The laptop allows the use of the classical Twitter web interface with text and images.

The Smart TV is touch-enabled and has a graphical interface that allows the manipulation of the digital content. The graphical interface allows entering a message, to browse and select images and to share emotions. Till this day, sharing emotions was limited to the semantic meaning of the text and to the emoticons. This prototype is different, because the user has to perform the facial expression associated to the emotional status in order to share it in the SAS. The emotion capture is based on image processing techniques. In fact, the Smart TV integrates a Microsoft Kinect camera and it has been implemented an algorithm based on the Facial Action Coding System (FACS)[7] in order to recognize five different facial expressions: smiling for happiness, frowning for sadness, scowling for anger, winking for trust and laughing for ecstasy. The prototype is shown in Fig. 1 (a).

The smartphone runs a software application for Android. This application allows the ubiquitous sharing of emotions through text, images and emotions. As the Smart TV interface, the smartphone induces the user to communicate her/his emotional state through the facial expression. In this case, the user’s facial expression recognition is always based on the FACS but it is captured through the embedded front RGB camera, see Fig. 1 (b).

The choice of giving to the user the opportunity to communicate her/his current emotional state using the facial expression aims at reintroducing the use of the paralanguage typical of the non-verbal communication. In fact, the facial expressions and the gestures are very meaningful means of expression in the face-to-face communication and they add more information to the spoken sentences. In particular, affection not only allows experiencing a richer interaction but also helps to disambiguate meaning, facilitating a more effective communication[8].

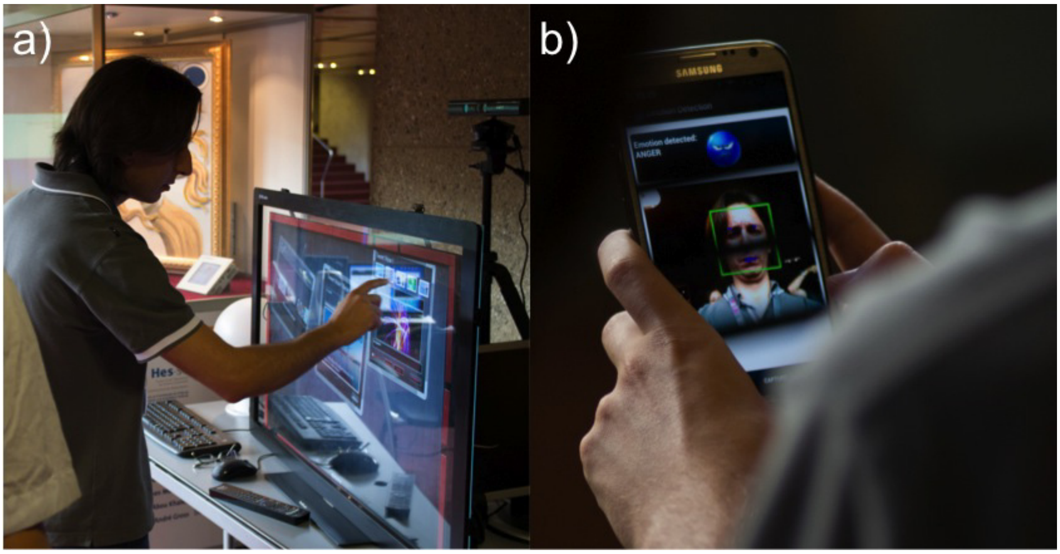

Fig. 1. (a) the Smart TV interface that allows recognizing the facial expression through the Microsoft Kinect camera; (b) a user testing the Android application that allows sharing the emotional state through the facial expression recognition.

The current prototype provides many different modalities of information display. It is composed of several distributed surfaces, which can be made interactive thanks to the Inter-Face framework and that can be used to represent the graphical information as texts, images and emoticons. The text messages can be also communicated either via Text-to-Speech in loudspeakers or via the speakers of the Smart TV; the synthetized voice reads the message shared on the SAS, i.e., Twitter, and the name of the person that shared that information. The representation of the emotions finds two output forms: using Aphrodite and an RGB lamp.

Aphrodite is a robotic painting that depicts Venus' head from Botticelli's "the Birth of Venus". The robotic painting provides a multimodal feedback composed of mimicked facial expressions and synthetized voice for the representation of emotional states that have been shared by the user's selected group on the social network. It is not the first time that visual art is used to display information. Actually, in 2000 Redström et al. coined the term "informative art" that they described as "computer augmented, or amplified, works of art that not only are aesthetical objects but also information displays, in as much as they dynamically reflect information about their environment" [9]. With Aphrodite, we introduce the concept of emotional design applied to the informative art for affective communication. At first glance, Aphrodite seems to be only a decorative artwork but it becomes animated when the user receives a message containing emotional information and it displays the emotional states of a social group in a calm, unobtrusive manner. Some researchers demonstrated that anthropomorphic artificial agents are very effective for emotion communication[10] and for this reason we chose to create this painting depicting a woman's head. There are 5 emotions available in the set and they are directly connected to 5 different facial expressions: smiling for happiness, frowning for sadness, winking for trust, laughing for ecstasy and scowling for anger (see Fig. 2). These facial expressions have been mapped to Plutchik's wheel of emotions[11].

The lamp provides a visual feedback that can be perceived with only the peripheral attention. It will remain hidden in the periphery of the users' attention, in order not to intrude and distract the inhabitants from their main tasks. We mapped the 5 emotions with 5 colors following the Plutchik's wheel of emotions: light yellow for serenity, blue for sadness, green for trust, dark yellow for ecstasy and red for anger. The choice of mapping the colors on the

Plutchik's wheel is justified by the absence of a universally accepted color model for the representation of emotions[12].

The concept of transforming all the smart environments in a seamless interface for a more natural interaction and to facilitate the communication of emotions encounters a problem when it is the time to display the information. As mentioned, this prototype provides many different interfaces: many surfaces that are able to display graphical information and to allow gestural interactions; many speakers that can reproduce audio and to read messages thanks to the Text-to-Speech service; smart furniture that can augment the interaction experience adding different types of information representation, in this case, the robotic painting that can mimic the facial expressions and the lamp that changes color according to the shared emotional state. The vast variety of types of feedback can become an issue when the representation of the information is not opportunely dispatched. For this reason, we developed a framework called NAIF[13], which is able to interconnect heterogeneous devices that are present in the smart environment. These devices are all interconnected creating a network; they can offer different resources for the information representation to the user and can contribute to the sensing of contextual information. NAIF is able to manage all the contextual information that is sensed by the different devices distributed in the environment and is able to make choice to deliver opportune feedback using the best available modalities. Moreover, each device can send communication intents to the framework, which automatically performs the context-aware multimodal fission of the message content. NAIF allows defining many rules to manage the information in different manners: the fission can be based on spatial information or with reference to privacy concerns or based on the environmental condition. In this prototype, there were three defined rules: personal login for private communication, lighting condition and environmental noise.

The privacy is a big issue in a smart environment where all the surfaces and the furniture are interactive and can display information in many different modalities. In fact, all the inhabitants can access to the information in such environments but an interactive room should facilitate the interaction and, at the same time, it should be able to preserve personal data. In this case, the privacy is granted thanks to the possibility to log in a specific device through the use of a personal object. This object has an integrated RFID tag and when placed on the log in area of a device, the system will convey all personal information of that specific user only to the device that she/he is using, avoiding that other people can access it involuntarily.

The context awareness allows selecting the best modality with reference to the environmental condition. For instance, in this prototype two specific sensors were connected: a light sensor and a microphone. The lighting condition has been chosen for the reasoning of the graphical representation of the information. In fact, many displays are connected in this prototype and the light plays a key role. The rule engine was programmed to select the projected screen in case of low light condition, avoiding the emotional representation through Aphrodite since the inadequate illumination makes hard to recognize the facial expression. In case of high luminosity, the information was dispatched to the Smart TV and Aphrodite avoiding the projected screen; in fact, the projection was not very clear in case of excessive illumination. The microphone was used to estimate the environmental noise and the system was set to allow the Text-to-Speech function only if the room was quiet.

Fig. 2. Aphrodite can mimic several facial expressions that are mapped on the Plutchik's wheel of emotions: a) is the neutral expression with no associated emotion; b) smiling for happiness; c) frowning for sadness; d) scowling for anger; e) winking for trust; f) laughing for ecstasy.

3. Experiment

We participated in the demonstration session of the ACII 2013 that was held at International Conference Centre of Geneva in Switzerland on 4 September 2013. We set up the prototype and connected all the different devices as depicted in Fig.3 (a). The setup was composed of the touch-enabled Smart TV with the Microsoft Kinect for the facial expression recognition, the smartphone running the application we developed, a laptop with the RFID reader for the login through a personal object (in this case, a key-ring), the Aphrodite painting in a showcase, the RGB lamp connected on Wi-Fi, and the sensors of luminosity and noise.

We conducted an observation study both on the field and with the recorded video. Watching the video we calculated the number of attendees: 85 people stopped to watch the demo and 29 of them actually interacted with the prototype. The public generally demonstrated a large interest in the prototype. The interaction through facial expression was appreciated but considered too artificial. In fact, this kind of interaction implies that the user expresses an emotional state consciously; some participants suggested to integrate an automatic facial expression recognition that detects the unconscious expressions and to propose to the user to share the detected emotional status on the SAS. However, the largest issue that we were trying to test in this interactive event was to understand if performing facial expression could be a problem in terms of user acceptability. Since nobody refused to perform facial expressions for the communication of the emotional state, we can consider this preliminary result as encouraging, although we are aware that thorough tests are needed.

The output modalities turned out to be the most interesting part of the concept for the public. In particular, Aphrodite captured the public's attention and many attendees took pictures of it. The surprising fact is that Aphrodite aroused positive feelings in the participants; in fact, we could observe that most of the participants (almost everyone) started smiling when they saw the painting moving to mimic an expression. Some showed also an initial reaction of surprise, usually followed by smiling. The valence of the emotions aroused by the interaction with the painting was positive independently from the valence of the mimicked facial expression. Another interesting fact was that people watching the demo usually indicated toward the painting engaging a dialogue with the neighbors and some also took a picture of it with a smartphone or a camera, as depicted in Fig. 3 (b). The effect of social catalyst induced by Aphrodite was unpredicted. Probably, this reaction of arousal and social interaction is due to the novelty of the device and in a scenario of daily living, where Aphrodite is part of the routine, it would not be the focus of the general attention. Nevertheless, an observation study of the use of Aphrodite in a daily living scenario would be very interesting and it could also provide evidence of whether this kind of emotional feedback could improve the CMC. The other output modalities such as the Text-to-Speech and the lamp were quite neglected because of Aphrodite. Moreover, the Text-to-Speech was not always understandable because of the noise in the common room; indeed, we had also to deactivate the noise rule to show that audio feedback. The lamp has been designed to provide feedback remaining in the periphery of the user's attention and the crowded stand at the conference was not the ideal condition for such test. An additional interpretation of the participants' behavior in relation to the different input and output modalities suggests that people prefer to interact with life-like objects[14], especially with those that have an anthropomorphic aspect. While participants found awkward performing facial expressions in front of the graphical interfaces of the Smart TV, they naturally reacted with gestural and facial expressions in front of Aphrodite. The surprise generated by the unusual robotic painting offers additional stimuli for reflection: people interact longer with objects that behave in an unpredictable manner[14], especially in the context of emotional interactive systems. Indeed, not only the form of life-like objects is important, but also their behavior matters, which should be similar to living beings.

This project started with the purpose of creating a *smart* living room, i.e., a living room augmented with technology in order to offer to the user a better interaction experience. The results of this experiment shifted our attention to an important aspect of our smart *living* room: its life-like behavior. The room is not only a place where to live, but should have its own life. While anthropomorphic or zoomorphic forms can facilitate the user willingness to interact with the environment, life-like and unpredictable behavior is also necessary in order to preserve this willingness over time. With these considerations in mind, new directions arise for our system. First, the almost ignored RGB lamp could obtain more attention by taking, similarly to Aphrodite, an anthropomorphic form. The round lamp could resemble to a human head, which in turn can mimic facial expressions. Since people are more likely to express emotions in front of an anthropomorphic object, both systems can be augmented with facial expression recognition capabilities. Touch gestures performed on Aphrodite or on the lamp can become an additional natural input modality for the system. Also interactive surfaces can be ameliorated with the introduction of life-like agents that can better exploit the text-to-speech output modalities.

4. Conclusion

In this paper, we have presented a concept about smart buildings that goes beyond the automation for efficiency. This concept explores how smart environments can help the human being of the digital era in the HCHI. The surfaces and the furniture of an environment will turn in a seamless and pervasive interface. In particular, we have explored the possibility to enhance the CMC providing the possibility of sharing emotions on the social networks in a more natural way. The prototype we developed allows sharing the emotions just recognizing the user's facial expression from both a smartphone and a Smart TV. Moreover, this prototype provides different modalities of output: many screens distributed in the environment, a Text-to-Speech service, a lamp that changes color according to the shared emotion and a painting that mimics the human facial expressions. The output modalities are managed by the NAIF framework, which executes the context-aware multimodal fission in order to opportunely represent the information. The observation study conducted during the interactive demonstration at the ACII 2013 has allowed us to extract some preliminary findings that will help us to enhance the system we are developing.

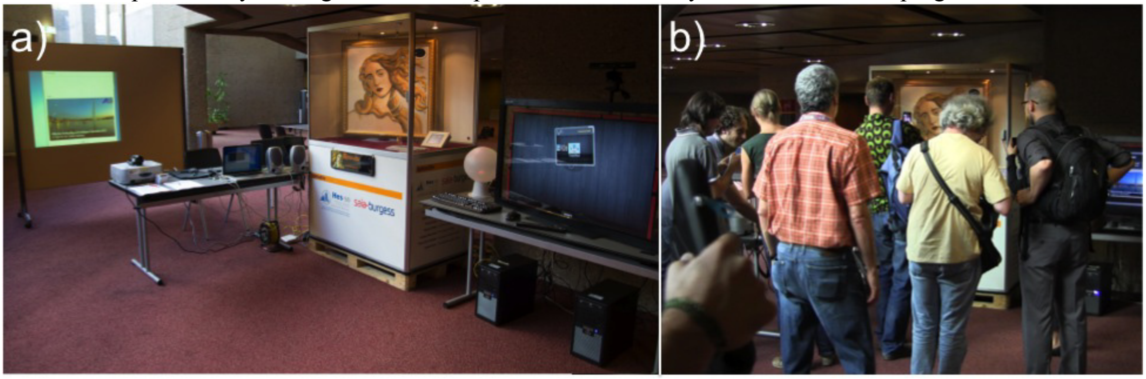

Fig. 3. (a) the setup of the prototype for the interactive demonstration at the ACII 2013; (b) people interacting during the demonstration, image taken from the recorded video.

References

1. Caon M, Abou Khaled O, Mugellini E, Lalanne D, Angelini L. Ubiquitous Interaction for Computer Mediated Communication of Emotions. *In Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, IEEE Computer Society, 2013.
2. Caon M, Angelini L, Yue Y, Abou Khaled O, Mugellini E. Context-Aware Multimodal Sharing of Emotions. *In Human-Computer Interaction. Towards Intelligent and Implicit Interaction, pp. 19-28*. Springer Berlin Heidelberg, 2013.
3. Turkle, S. *The Tethered Self: Technology Reinvents Intimacy and Solitude*. Continuing Higher Education Review, 75 (2011): 29.
4. Clubb OL. Human-to-Computer-to-Human Interactions (HCHI) of the communications revolution. *interactions, vol. 14, p. 35*, Mar. 2007.
5. Dumas B, Lalanne D, Oviatt S. Multimodal interfaces: a survey of principles, models and frameworks. *Human Machine Interaction, 2009*.
6. Mugellini E, Abou Khaled O, Pierroz S, Carrino S, Chabbi Drissi H. Generic Framework for Transforming Everyday Objects into Interactive Surfaces. *In Proceedings of the 13th Int. Conf. on Human-Computer Interaction. Part III: Ubiquitous and Intelligent Interaction*, 2009.
7. Ekman P, Friesen WV. Facial action coding system: A technique for the measurement of facial movement. Consulting Psychologists Press. Ellsworth, PC, & Smith, CA (1988). *From appraisal to emotion: Differences among unpleasant feelings. Motivation and Emotion 12* (1978).
8. Brave S, Nass C. Emotion in human–computer interaction. *The Human–Computer Interaction Handbook, pp. 81–93*, 2003.
9. Redström J, Skog T, Hallnäs L. Informative art: using amplified artworks as information displays. *In Proceedings of DARE 2000 on Designing augmented reality environments, pp. 103-114*. ACM, 2000.
10. Moser E, Derntl B, Robinson S, Fink B, Gur RC, Grammer K. Amygdala activationat 3T in response to human and avatar facial expressions of emotions. *Journal of neuroscience methods, 161, 126–33*, 2007.
11. Plutchik R. The Nature of Emotions. *American Scientist, 89(4), 344- 350*, 2001.
12. Alkawaz MH, Basori AH. The effect of emotional colour on creating realistic expression of avatar. *In Proceedings of the 11th ACM SIGGRAPH Int. Conf. on Virtual-Reality Continuum and its Applications in Industry - VRCAI '12, 2012, vol. 1, no. 212, p. 143*.
13. Perroud D, Angelini L, Abou Khaled O, Mugellini E. Context-Based Generation of Multimodal Feedbacks for Natural Interaction in Smart Environments. *In Proceedings of the 2nd Int. Conf. on Ambient Computing, Applications, Services and Technologies, 2012, no. c, pp. 19–25*.
14. Schmitz M. Concepts for life-like interactive objects. *In Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction - TEI '11*, (2011), 157.