A Human-Centered Model for Detecting Technology Engagement

James Glasnapp and Oliver Brdiczka

Palo Alto Research Center (PARC)
3333 Coyote Hill Road
Palo Alto, CA, 94304, USA
{glasnapp,brdiczka}@parc.com

Abstract. This paper proposes a human-centered engagement model for developing interactive media technology. The human-centered engagement model builds on previous interaction models for publicly located ambient displays. It is designed from ethnographic observation with the aim of informing technological innovation from the perspective of the user. The model will be presented along with technological mechanisms to detect human behavior with the aim of responsive media technology development.

1 Introduction

This paper proposes a human-centered model to aid interactive media technology development that can detect and thereby engage individuals with this technology. We derived the human-centered engagement model from exploratory ethnography that aimed to develop a broad conceptual model for designing responsive display media technology. The concept of engagement has a wide range of contextual meanings crossing different disciplines [1] [2]. For our purposes, we are referring to engagement as a goal for focused interaction with display technology. [12] This technology is intended to develop a way to sense individuals' presence and subsequently attract them to interact with a display (or other possible technology).

The rise of digital signage as an opportunity space is vast. Several working prototypes exploring expansion of what has been accomplished with digital signage are currently being tested or are in planning [3]. One such system selects ads and other content based on the viewing audience's size and demographics. Interactive displays will be commonplace in the near future. Design frameworks that take into account how people interact with technology will help shape the way this technology is designed to respond and interact with humans [4]. Many current conceptual models for how humans interact with displays are machine centered; they are conceptualizations of users in terms of their interactive relationship with a display. Our research explores engagement with display technology from the perspective of the mobile transient users in public settings. We focus on understanding human behavior so that technology may be designed to respond and engage individuals based on fundamental human interactive practices. [23] Ethnographic observations have been analyzed and serve as basis for a human-centered model of engagement. The different stages of this model

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2009, LNCS 5612, pp. 621–630, 2009.

are intended to be a fine-grained description of the engagement process with a technology device and as a foundation for future detection algorithms. The remainder of this paper is structured as follows. First, we will give a short review of literature on human-computer interaction and human engagement. Our ethnographic data and methodology is then described. Finally, the human-centered model of engagement and possible detection mechanisms are detailed. A short conclusion terminates this paper.

2 Related Work

Related work includes human engagement with robots as well as displays. Sidner and Lee, [5] focusing on human-robot interaction have proposed a three-part model of engagement for developing the ability of robots to engage with an individual in collaborative conversation. Their model consists of, (1) initiating interaction, (2) maintaining interaction, and (3) disengaging. Research in the arena of interaction with display technology has been largely focused on collaborative workspaces and environments. Izadi, et al. [6] explored the design of interactive design in shared and sociable spaces. Beehl et al. [7] studied collaboration in multiple display workspace environments. Bringnull and Rogers [8] observed people's activity patterns around large displays and identified three activity spaces: *peripheral awareness* at the outer boundary; *focal awareness* in an area where one can give attention to the display; and, *direct interaction* where people can straightforwardly interact with the display (Figure 1) Building on this conceptual model, Finke et al. [9] designed a display to encourage user participation in an interactive game. The intent is to move those in the periphery into direct interaction, or from being *bystanders*, to *spectators*, and finally, to *actors*. Benford et al. [10] used a similar framework to focus on the design of the audience experience of public interaction referring to those on the periphery as *spectators* and those who are engaged as *performers*.

Similarly, Vogel et al. [11] proposed a display-centered model of engagement, for sharable, interactive public ambient displays (Figure 1). This display centered framework offers four phases starting with an ambient display phase at the farthest distance from the display where users can see static and generalized information. From there, as users move into closer proximity to the display, they are in the implicit interaction phase and can be recognized by the system by their body position. When the user

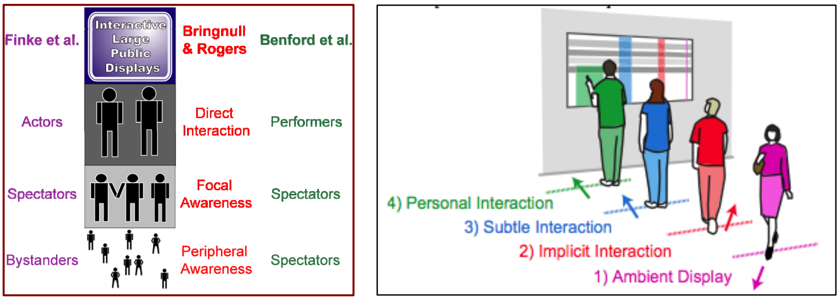


Fig. 1. Conceptual models of how individuals orientate themselves in public settings

pauses, the system enters what is called the subtle introduction phase. During this phase, the system may identify relevant information. Finally, in the personal interaction phase, the user moves closer to the screen and touches items for more information. This framework for interaction allows the user to go from a state of implicit to explicit interaction. Additionally, Vogel et al. conceptualized a list of design principles to guide the development of responsive displays.

These display centric models assist in understanding how individuals orient themselves to displays and how groups of people involved in collaboration or social settings respond and relate to displays. Studies of human interaction and conversation analysis help us to understand the human practices of people in these environments—both within and outside the periphery of the display. Goffman's description of social interaction has implications for understanding how individuals behave within and outside display peripheries. Goffman [12] divides face–to–face interaction into *unfocused* and *focused* interaction where *unfocused interaction* occurs when individuals are merely copresent. In this type of interaction one or both parties might modify behavior because they know they can be observed. Such would be the case when walking in a busy shopping center, or when in close proximity to strangers in the periphery or interacting with a display.

In contrast, *focused interaction* occurs when people "agree to sustain for a time in a single focus of cognitive and visual attention, as in the conversation, a board game, or a joint task [with a] close face-to-face circle of contributors." [12] Goffman refers to social arrangements that involve participants with a cognitive focus of attention as a "focused gathering." [12] In these encounters, an individual's presence is acknowledged through expressive signs and formal rituals. What emerges is a single thing or shared experience that both parties achieve together over time. Encounters can be accompanied by rituals that include ceremonies of entrance and departure, but always provide a central base for communication between parties as well as "corrective compensation" for deviant acts. Benford et al. [13] used the concept of frame in their conceptual framework for promoting game play so that players understand what is within the circle of play. As individuals move from subtle to personal interaction, or from bystander to actor, the encounter with the display could be considered a focused gathering.

Mobile individuals outside the periphery of a display would have a different definition of their situation than those who have a focal awareness of the display, particularly if they are in a close-knit collaborative or social environment. Kendon [14] writes about how individuals agree on a 'frame of the situation' [14]. He also argues that participants establish assumptions about situations through their own interpretive perspectives. Garfinkel [15] suggests these frames are shared because people agree on 'the constitutive expectancies of a situation.' Transitory individuals in mobile settings have a different frame of the situation that is communicated through embodied actions. To be effective, media technology would need to identify these individuals and transition them from one interpretive perspective to another, from a receptive yet mobile experience to an interactive one.

Finally, in regard to the act of engagement itself Goffman [12] desribes engagement as when an individual is "caught up by it, carried away by it, and engrossed in it – to be, as we say, spontaneously involved in it." [12] He specifies engagement as a spontaneous involvement in a joint activity and that an individual becomes an integral part of the situation. This is the very goal of technology developers to create a situation where the user is engaged with the display as an actor or performer in an interactive experience.

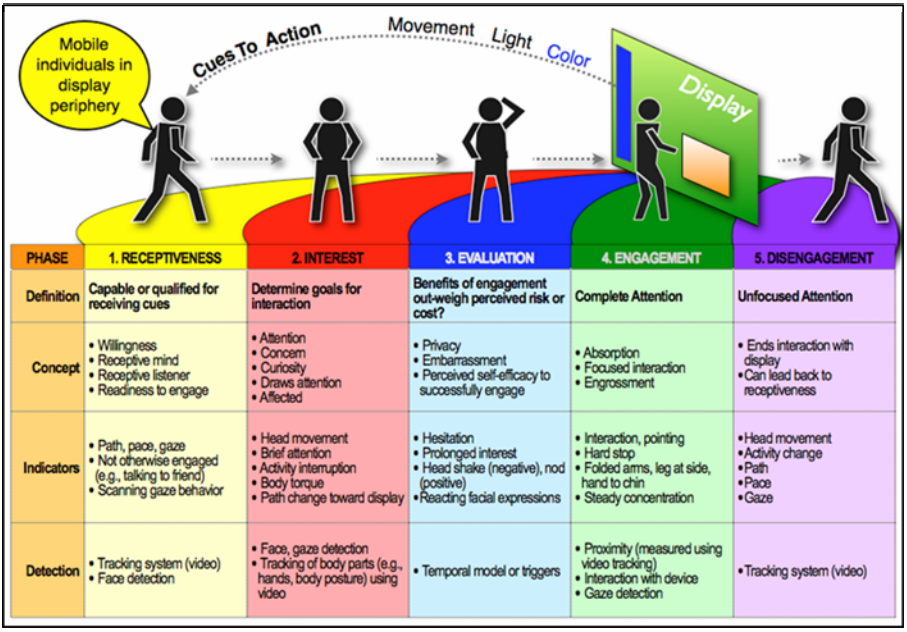
3 Methodology

Our research objective was to describe the natural practices people use as they come into contact with and interact with public displays in transitory and mobile settings. We collected ethnographic video data in public settings with the aim of observing behaviors that demonstrate engagement and disengagement with products or displays. The data includes observations of individuals in public spaces such as an airport, two shopping malls, a movie theater pavilion and a sidewalk in a commercial setting. We specifically selected areas where people had an opportunity to enter the periphery of an interactive or static self-serve kiosk or display. Data collection occurred over the course of three weeks. Approximately 8 hours of video data was collected in addition to 10 hours of observation. Video data was analyzed on a computer desktop. We included individuals in the analysis who entered into the periphery of the display and excluded those who could not be adequately observed or who did have an opportunity to visibly show their attention to the display. We categorized behavior traits into progressive stages toward potential engagement with displays or interactive media.

4 Qualitative Analysis and Results

Based on the collected video data and observations, we inductively derived a conceptual model that we call a human-centered engagement model. This model consists of five stages that guide technological development with the goal of reaching and maintaining full user engagement: receptiveness, interest, evaluation, engagement, and disengagement. The purpose in applying this model to technology design is for technology to assist or promote a user through increasing stages of participation, to achieve and maintain full engagement between the user and machine as long as possible before disengagement. What follows are descriptions of each of the phases including definition, basic concepts, indicators and possible detection. We have adopted computer vision technology as major cue for detecting these phases because it is rather light-weight to be deployed in an existing setting (mounting cameras compared to e.g., installing floor sensors or other more invasive technology) and it does not require users to carry or wear specific technology items to be detected (e.g., sensor badges or dedicated cell phones).

Receptiveness. The first stage of our model describes basic receptiveness, corresponding to the capacity and willingness of a user to receive cues like advertisements or qualified information. An individual might be available to engage with a new activity irrespective of distance as long as the individual is able to observe and approach the technology in question, and that technology is able to sense their presence. Our data set consists of individuals in public areas, particularly areas in which individuals are on leisure time or have idle time (i.e., shopping malls, airports); people can be seen to walk at a more leisurely pace and demonstrate signals that they are available



PHASE	1. RECEPTIVENESS	2. INTEREST	3. EVALUATION	4. ENGAGEMENT	5. DISENGAGEMENT
Definition	Capable or qualified for receiving cues	Determine goals for interaction	Benefits of engagement out-weigh perceived risk or cost?	Complete Attention	Unfocused Attention
Concept	• Willingness • Receptive mind • Receptive listener • Readiness to engage	• Attention • Concern • Curiosity • Draws attention • Affected	• Privacy • Embarrassment • Perceived self-efficacy to successfully engage	• Absorption • Focused interaction • Engrossment	• Ends interaction with display •Can lead back to receptiveness
Indicators	• Path, pace, gaze • Not otherwise engaged (e.g., talking to friend) • Scanning gaze behavior	• Head movement • Brief attention • Activity interruption • Body torque • Path change toward display	• Hesitation • Prolonged interest • Head shake (negative), nod (positive) •Reacting facial expressions	• Interaction, pointing • Hard stop • Folded arms, leg at side, hand to chin • Steady concentration	•Head movement •Activity change •Path •Pace •Gaze
Detection	• Tracking system (video) • Face detection	• Face, gaze detection • Tracking of body parts (e.g., hands, body posture) using video	• Temporal model or triggers	• Proximity (measured using video tracking) • Interaction with device • Gaze detection	•Tracking system (video)

Fig. 2. Human-Centered Engagement Model

and looking for something of interest. Not all individuals were receptive, some were otherwise engaged with an electronic device, social encounters with friends, or presumably rushing to their destination.

For example, a male and a female subject stroll together past a kiosk, arms at their sides, walking co-jointly at a comfortable, calm stride. Their heads were upright and their attention was unfocused (on any singular object) (Figure 3, below). We also observed this behavior on public sidewalks where individuals appeared more intent on walking as a means of transportation, but still showed signs of interruptability and altered their pace, gaze and in many instances, stopped to observe the informational or eye catching displays in question.[1] These observable embodied actions indicate possible capability for a user to receive cues. They have a receptive mind and are accessible listeners. Indicators of receptiveness are an individual's path, pace, gaze, and the extent to which their attention is unfocused on any single object or event activity. A video tracking system [16] can be used to locate one or several people in real-time using one or several cameras. A face detector [17] applied on different camera images can give indications about approximate gaze and focus of attention.

Technology is capable of sensor-based predictions of receptiveness [18]. The ideal scenario would be similar to the intuitive sales person who knows just how much and when to interact with customers. Our observations indicate that current display

[1] We observed individuals walking down a large boulevard in a major metropolitan city in front of a real estate office and a clothing store. The real estate office displayed non-interactive rotating home information on a large plasma display. The clothing store had two display windows.

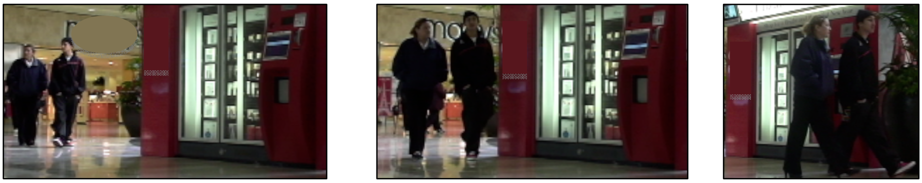


Fig. 3. A couple demonstrates receptive behavior in front of a kiosk

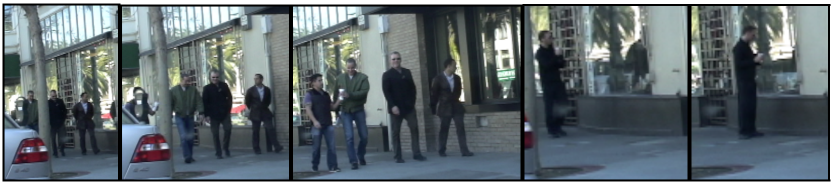


Fig. 4. Receptive groups and individuals strolling on busy boulevard

technology, particularly interactive technology, is lacking the ability to consistently attract individuals who are otherwise receptive to interaction. For example, technology could utilize "cues to action" to move qualified individuals to notice the display and become interested in engaging with it. Assuming that the technology can recognize the qualified user, it uses cues to get users to notice it, leveraging human senses with light, movement, sound, and potentially smell in order to solicit user attention.

In our observations of self-service kiosks, many opportunities to attract a potential user's attention were missed. Most individuals who appeared available for engagement were unaware of the kiosk when passing by. Even if the user had noted the kiosk with peripheral vision, she did not have sufficient information from which to dismiss or make judgment about the target display.

Interest. The interest phase represents the point at which the user observably demonstrates curiosity or interest in an object. Once a cue from an interactive media technology has drawn the user's attention and awareness, the user will show interest through abrupt physical changes in body movement that demonstrates at least a minimum level of concentration, albeit sometimes brief. We observed patterns of behavior that we later classified as low interest (e.g., head turns, changing the velocity of pace, changing course) and high interest (such as stopping). The technological goal for this phase is to identify potential users at the point of curiosity, increasing the likelihood of transitioning to engagement. For example, in Figure 5 below, we see two individuals exhibiting interest in a kiosk. The elderly woman pushing a cart passes by the kiosk and looks (low interest) while the man is briefly stopped in front of the sample scent tester. Reducing pace enables more precise face and gaze detection [19], which are good indicators for upcoming interest. Tracking body parts like the hands [20] and body posture which can be provided by a video tracking system further enhance the detection of interest.

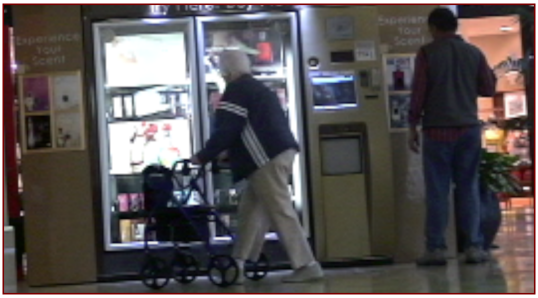


Fig. 5. Low and high levels of Interest

Evaluation. Evaluation is the stage where the user will determine her goal for any potential interaction. This stage is the linkage between noticing a display and engaging with the display. This evaluation can be brief, but we include it because of the importance from the technological receptiveness perspective. The ability to detect and respond to individuals at a decision point will increase the probability of their transition from showing an interest to engagement. Based on the interest detection using face, gaze, and body parts, simple temporal triggers or more complex temporal models like hidden Markov models may indicate the transition to the evaluation phase. Three important aspects of the evaluation stage (and thus of the user's considerations) are privacy, self-efficacy, and social norms. Each is important to the development of the user interface, location, and overall design. Privacy concerns are a potential design consideration, particularly in busy public settings which has been noted and the focus of other work [21]. Another consideration is whether individuals feel they have the capacity to successfully engage with unfamiliar technology; that is one's own considered opinion about their personal capability, or self-efficacy [22]. If users show interest in a display, but immediately determine that further attention would fail, then there is a failed opportunity. Finally we consider the issue of social norms in public spaces. In the case of shopping areas where individuals are either alone or shopping with groups, altering the existing interaction or behavior to engage in something potentially different or beyond that which one is accustomed could deter the user from initiating the engagement.

We observed individuals who demonstrated interest, but physically displayed frustration, which is the result of a negative evaluation. These cases included situations in which the display was too small to approach closely when others were located in front of the display and head shaking. We also noted behavior that indicated embarrassment.

Engagement. Engagement is the point at which the user is focusing her attention. We observed individuals alone and in groups with focused interaction with an object or display and they were unfocused on everything else for a period of time. Physical characteristics that exhibited concentration could be observed: arms folded, arm to the chin, pointing, head slanted to the side. The body position also changed from other phases to a stance: arm perched to the waist, one leg bent to the side. In the figures below, note the children completely engrossed in an interactive floor display (Figure 6). Next to them, onlookers disrupt their movement and show interest in the display while the mother (to the right) tries to coax the children off the display. In the

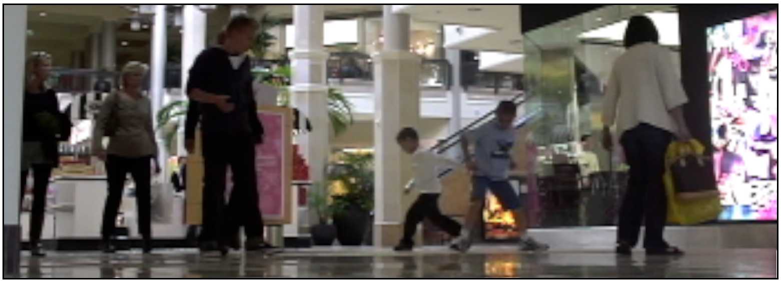


Fig. 6. Two children engaged with floor display

other photo, two men engaged, one pointing to the display while the other puts his hand to his chin (Figure 7). Detecting the engagement phase could be realized by using proximity information coming from the video tracking system, i.e. the distance of a detected entity (person) is below a distance threshold to an object of interest. Additionally, gaze detection using a dedicated camera at the object of interest and direct interaction with the object itself are strong features for detecting this phase.

Disengagement. Disengagement is the point at which the user becomes disengaged. Presumably a user can go from this stage back to the receptiveness stage. Increasing pace, changing path and activity are indicators for this phase. A video tracking system could provide the means for detecting pace and path of moving people.

A few other insights we had were that groups tended to find space around the display as appropriate to view material. Children often created their own space on displays at eyelevel, often mimicking adults and interacting with made-up content. We also noted what we called engagement by default. These secondary users are accompanying individuals who are engaged. We include this insight because technology development might recognize this and provide unique content aimed at this type of user. Finally, we also noted that physical store settings within shopping malls that had brighter lights, open spaces, and easily identifiable and approachable staff attracted and engaged shoppers. We bring this up because the staff can be equated as human sensors and the attractiveness of the stores and interaction they promote are the very goal of the technology we aim to create.

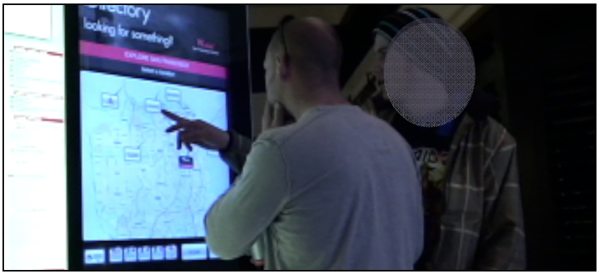


Fig. 7. Two individuals engaged with a display

5 Conclusion and Future Work

We have proposed a human-centered model to describe and model human engagement with the objective of enabling and improving interactive media technology. In contrast to current conceptual machine-centered models aimed at collaborative spaces, our research focuses on the perspective of the mobile transient user in public settings. The five stages of the User-Centered Engagement Model are receptiveness, interest, evaluation, engagement, and disengagement. The first stage is basic receptiveness, corresponding to the capacity and willingness of a user to receive cues. The second stage represents the moment when a cue has drawn attention and curiosity (interest), the user will get affected and briefly determines his goals for interaction. In the third stage, evaluation, the user considers whether to engage in an interaction. The fourth step represents a positive evaluation when the user invests his full attention and gets absorbed by the interaction. The last step constitutes the disengagement: when the interaction has concluded, be it positively or negatively, the user disengages and becomes receptive again.

Our observations indicate that many current systems fail to attract mobile transitory users otherwise available for engagement. We believe that our user-centric model is a keystone to make these systems more appealing to people for interaction. The model is very general and can encompass a wide variety of technology uses and goals. In particular, interaction design and digital content presentation will vary from one system to another and the system deployment will affect and possibly constrain the detection methods that are used for the different phases. Thus, we did not focus on the detection aspects or on the interaction design of an ideal system.

Future work will concern implementation and validation of the different phases of the User-Centered Engagement Model in various settings involving a higher number of users. We are taking first steps towards the implementation of an integrated system combining technological detection of the different phases with interaction design and evaluation. The implementation of a technological prototype for detecting the different phases of the model is foreseen. More focused user studies that reveal each of the proposed phases is also needed.

References

[1] Skocpol, T., Fiorina, M.P.: Civic Engagement in American Democracy. Brookings Institution Press (1999)

[2] Greenwood, C.R., Horton, B.T., Utley, C.A.: Academic Engagement: Current Perspectives in Research and Practice. School Psychology Review 31(3), 328–349 (2002)

[3] Williams, M.: Japanese billboards are watching back. IDG News Service (12/12/2008), `http://www.goodgearguide.com.au/index.php?q=article/270798/japanese_billboards_watching_back&fp=&fpid` (1/15/2008)

[4] Canton, J.: The extreme future - the top trends that will reshape the world in the next 20 years. Institute for global futures, Inc./Penguin books (2006)

[5] Sidner, C., Lee, C.: Engagement rules for human-robot collaborative interactions. Systems (2003)

[6] Izadi, et al.: The iterative design and study of a large display for shared and sociable spaces. In: Proceedings of the 2005 conference on Designing for User ... (2005)

[7] Biehl, J., Baker, W., Bailey, B.: Framework for supporting collaboration in multiple display environments and its field evaluation for .... portal.acm.org (2008)

[8] Brignull, H., Rogers, Y.: Enticing People to Interact with Large Public Displays in Public Spaces. Human-Computer Interaction (2003)

[9] Finke, M., Tang, A., Leung, R.: Lessons learned: game design for large public displays. In: Proceedings of the 3rd international conference on Digital ...(2008)

[10] Benford, S., Crabtree, A., Reeves, S., Flinham, M., Drozd, A., Sheridan, J., Dix, A.: The Frame of the Game: Blurring the Boundary between Fiction and Reality in Mobile Experiences. In: Proceedings of the SIGCHI conference on Human factors in ...(2006)

[11] Vogel, D., Balakrishnan, R.: Interactive public ambient displays: transitioning from implicit to explicit, public to personal .... In: Proceedings of the 17th annual ACM symposium on User ...(2004)

[12] Goffman, E.: Encounters; Two studies in the sociology of interaction. Bobbs-Merrill, Indianapolis (1961)

[13] Benford, S., Crabtree, A., Reeves, S., Flinham, M., Drozd, A., Sheridan, J., Dix, A.: The Frame of the Game: Blurring the Boundary between Fiction and Reality in Mobile Experiences. In: Proceedings of the SIGCHI conference on Human factors in ...(2006)

[14] Kendon: Conducting interaction: patterns of behavior in focused encounters (Studies in Interactional Sociolinguistics). Cambridge Univ. Press, Cambridge (1990)

[15] Garfinkel, H.: Trust and stable actions. In: Harvey, O.J. (ed.) Motivation and Social interactions. Ronald Press, New York (1963)

[16] Zhou, S., Chellappa, R., Moghaddam, B.: Visual tracking and recognition using appearance-adaptive models in particle filters. IEEE Transactions on Image Processing 11, 1434–1456 (2004)

[17] Viola, P., Jones, M.J.: Robust Real-Time Face Detection. International Journal of Computer Vision 57(2), 137–154 (2004)

[18] Hudson, S., Fogarty, J., Atkeson, C., Avrahami, D.: Predicting human interruptability with sensors: a Wizard of Oz feasibility study. In: Proceedings of the SIGCHI conference on Human factors in ...(2003)

[19] Magee, J.J., Scott, M.R., Waber, B.N., Betke, M.: EyeKeys: A Real-Time Vision Interface Based on Gaze Detection from a Low-Grade Video Camera. In: Proceedings of IEEE Computer Vision and Pattern Recognition Workshops (2004)

[20] Kolsch, M., Turk, M.: Robust hand detection. In: Proceedings of Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 614–619 (2004)

[21] O'Neill, E., Woodgate, D., Kostakos, V.: Easing the wait in the emergency room: building a theory of public information systems. In: Proceedings of the 5th Conference on Designing interactive Systems: Processes, Practices, Methods, and Techniques, DIS 2004, Cambridge, MA, USA, August 1-4, pp. 17–25. ACM, New York (2004), `http://doi.acm.org/10.1145/1013115.1013120`

[22] Bandura, A.: Social foundations of thought and action: A social cognitive theory, pp. 390–449. Prentice-Hall, Engelwood Cliffs (1986)

[23] Sacks, A., Schegloff, E., Jefferson, G.: A simplest systematics for the organization of turn-taking for conversation. Language (1974)