

On the Relation of Ordinary Gestures to TV Screens: General Lessons for the Design of Collaborative Interactive Techniques

Oskar Juhlin

Mobile Life @ Stockholm University
DSV Forum 100, 164 40 Kista, Sweden
oskarj@dsv.su.se

Elin Örnevall

Mobile Life @ Stockholm University
DSV Forum 100, 164 40 Kista, Sweden
elin@mobilelifecentre.org

ABSTRACT

We present an interaction analysis based on ethnographic fieldwork of how physical movements, including gestures, are produced by viewers in front of television screens in a sports bar. Understanding ordinary life and specifically television watching in social situations will benefit the discussion of the potential of gesture techniques for controlling interactive televisions in various locations. Challenges for system design include body movement recognition, since movements can have many different purposes and are directed simultaneously at the screen and co-viewers. Moreover, gestures as elements of conversation are sometimes negotiated and overlapping. Since these ordinary movements are hard to automatically track and analyse, suggested systems might lead to demands on viewers to restrain their accustomed movements and adapt them in ways that might be considered awkward. We also reveal new design opportunities that draw upon the ways viewers' gestures are influenced by ongoing broadcast.

Author Keywords

TV viewing; interaction analysis; ethnomethodology; gestures; interactive television; everyday practice; group watching; overlaps; gesture tracking adaptation

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Human Factors

INTRODUCTION

Watching TV is a frequent and ordinary activity in many people's lives. Although conventional TV sets provide little support for interaction beyond handling the remote control, emerging technology is stimulating closer studies of interaction in front of these screens [24]. We analyse how

groups of viewers watching the 2010 Olympics in Vancouver produce gestures in front of TVs'. The focus is on gesturing in naturally occurring viewing situations. The purpose of the study is to broadly influence the discussion on the potential of gesture interaction in situations sometimes attended by groups of people, as well as the specific design of gesture recognition as a means of controlling TV sets. In recent years, there has been a growing interest in this area, fuelled by the widespread success of this mode of controlling commercial games, such as the Nintendo Wii [25] and Xbox Kinect [41]. The success of this technology has inspired research on how to appropriate similar technologies for interaction with television content. Various techniques have been suggested, which either draws on handheld devices that interact with the screen or systems that recognise the viewer's visual and bodily activities in front of the screen [38].

In this research domain, an important motivation for the use of gesture recognition technology has focussed extensively on the 'naturalness' of gesture-based control. Gestures are represented as a 'natural' form of communication [1, 6, 17, 37, 42] because we spontaneously move fingers, hands, bodies and heads to convey information and interact with others. These movements could be used to interact with the TV and the disc player. The success of gesture interaction in gaming has yet to be realised in interaction with television content [1, 5, 15]. The question remains as to how to ensure smooth communication, whether by making the TV able to interpret viewers' 'natural' gestures and relate them to specific commands or by designing new kinds of gestures currently not found in TV viewing practices. Gestures for this domain are often designed to reflect the limitations of the technology [40] and the design process for working out recognisable patterns by using a number of test persons in a lab [9]. Gesture interaction research has focussed on either cataloguing preferred gestures [17, 21], or providing means, such as learning environments [1, 2, 5], that enable the viewer to perform gestures that are recognisable by the screen. This has proved to be a surprisingly challenging task [1]. The design experiments also focus on individual use situations [2,3,5,6,7,15,17,19,21,28,38], although it is argued that television viewing is often a collaborative activity [4,11,12]. Promising to deliver gesture interaction to viewers, often in social settings, further extends the de-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2013, April 27–May 2, 2013, Paris, France.

Copyright © 2013 ACM 978-1-4503-1899-0/13/04...\$15.00.

mands of technologies that promise ‘natural’ gesture interaction.

We argue that accounting for naturally occurring interaction in front of TV screens is needed to support the dominant approach to making use of ordinary gestures that already occur in conjunction with TV viewing. In a sense, the TV has to be able to visually relate to, or look out upon, the living room and make sense of natural physical movements made by the viewers sitting in front of it [9]. This issue was raised in relation to public interaction around screens for games [27], but it is equally important when it comes to TV watching, where it has not yet been on the agenda [9]. Wexelblat [39] also raised the need as a general concern for ‘naturalness issues’ in research on gesture recognition to account for ‘continuous’ gesturing.

Moreover, as we define and introduce new types of gestures, it would be useful to understand how they fit with and will influence current social mores. Studies of ordinary gesturing suggest what might be socially acceptable in this domain [32]. Our study provides data on everyday physical movements in front of TV screens, as well as the role of gestures in social interaction. Following a tradition of interaction and gesture analysis in sociology [24], particularly within ethnomethodology [14, 23, 13], we focus on how gestures are performed in the situation at hand by video recording viewers and then detailed analysis of a single case. Specifically, we discuss a situation where a number of viewers perform overlapping [35] gestures in front of television sets. Overlaps are a common feature in everyday social interaction, which counters the understanding of interaction as a sequential formulation of messages between interacting participants, that is, one party speaking at a time. Such overlaps occur for various purposes, such as interruptions or unproblematic conversation ‘management’ elements [35]. Gesturing in front of screens displays overlaps that make it hard to distinguish and demarcate specific spatial gestural forms and then eventually list them in a library [17,21]. They are continuously created in and through negotiation among the viewers for various purposes in the situation at hand, also accounting for screen content. We argue that the design of gesture-based interface interaction with TV sets would benefit from accounting for these characteristics when balancing the demand for ordinary interaction with expectations of available technology. We discuss the need to account for what we term ‘gesture-tracking adaptation’, which might occur in the living room, in the wake of systems that account for social and situated gesturing. As well, how gestures are produced in the broadcast itself seems to influence how similar physical movements are played out by television viewers.

The following section presents the related work. Section three describes our methods and settings. Section four contains an analysis of our empirical data. This is followed by a discussion of our findings in section five, including our discussion of general design choices.

RELATED WORK

The study was influenced by research in three areas: studies of gesturing as part of social interaction, television gesture control and studies of television watching.

Gestures as social interaction

Recognising that TV viewing depends on many modalities in addition to gaze, our study belongs to the body of work focussing on the transmodality of interaction and the identification of gesture and body position [24]. The study of gestures, defined as physical movements that are produced within a context and directed at people nearby [36] has a long tradition in the social sciences [24] and the field of HCI [14,9, 32]. Only physical movements that are intended to create meaning and add to language should be understood as gestures [18]. They can be used in several ways to add meaning, including ‘... enactment [...], the use of body parts as model of things [...], and the use of moving hands as if they are sketching diagrams or shapes in the air [...]’. Speakers can also point to things, persons, or locations as a way of bringing these in as referents [18].’ Conversation analysts [35] see gestures as means in an interaction to stress a particular point in conversation or means to organise and negotiate turns in a conversation. Importantly, embodied conduct, such as the raising of eyebrows, is used to create understanding between collaborating individuals in way similar to how talk [34] accounts for both the situation at hand and the sequentiality or ‘nextness’ of the social interaction [23].

In summary, although gesturing in front of TV screens has not yet been analysed within these traditions, we already have some understanding of how gestures are used in other mundane settings.

TV gesture control

It is argued that how we commonly interact with TV sets is going to change, since the service itself will develop and require a more active user [38]. Choi et al [7] predict that ‘the TV in the future will become a terminal for many interactive applications (...) In response to the new needs of this future TV, and replacing the conventional array of buttons, new remote control designs incorporating alternative input technologies, such as joystick, touchpad, or direct-pointing, are being explored and evaluated.’ Furthermore, the way we interact will change because emerging gesture recognition technologies are more natural and user-friendly [1, 17, 37, 42]. Saffer for example argues that gesture interaction is preferable because ‘[h]uman beings are physical creatures; we like to interact directly with objects. [...] Interactive gestures allow users to interact naturally with digital objects in a physical way, like we do with physical objects [32].’

Currently, there are two different ways of interacting with a TV screen from a distance other than fingering a standard remote control: using pointing devices (such as a touchpad or joystick) or tracking devices that ‘are typically used to

locate the position or motion of a user in space' [38]. It is argued that handling a remote control requires users' visual attention [38] and they thus momentarily lose sight of the screen [17]. Tracking devices are often suggested as a promising technical avenue that would allow interaction while looking at the screen [6], such as gestures performed by moving an arm up and down, left and right, or pushing and pulling, were examined as ways to change the channel or volume and switch the device on and off [17]. There is an emerging body of research that takes up this challenge and aims to identify viewers' preferred or 'natural' gestures. However, this research specifically focuses on studying preferred gestures by individuals standing in front of screens and on interaction with the TV [17, 21]. This sets it apart from how TV watching is often done in social settings (see section below). The only exceptions here seem to be a study on gesture interaction for multiplayer gaming on large public displays [26].

The social nature of TV watching

Studies of TV watching indicate that it is frequently done in groups [4, 11, 12]. A majority of users watch TV in the evenings as part of a social activity among family and friends [20]. It is conducted in people's homes as well as public places like sports bars. The social situation brings with it a need to negotiate what program to choose [8] or otherwise socialise. This talk must be aligned with the ongoing content. Viewers' conversations near TV sets are either synchronous or asynchronous: 'people can either talk or interact while watching television, talk about television afterwards or even before watching television [12].' Kubey [20] argues that a 'vast majority of respondents report that their viewing is accompanied by other activities and it is our conclusion that contrary to the views of others, very few people consistently watch television to the total exclusion of other activities.' Although the TV might be central to the activity, other matters are also part of the watching context. TV watching in general is thus something that needs to be socially negotiated for various reasons. For example, a group of viewers might need to discuss what to watch or what they think about the broadcast, as well as coordinate the viewing with other things happening at the same time. These commonplace practices differ from the set-up of gesture control experiments we previously mentioned, which are usually single-viewer situations.

METHOD AND SETTING

We examine physical movements in front of screens, based on empirical data on TV watching collected through ethnographic observation at sports bars, as well as in private living rooms, in a northern European city. These bars show broadcasts from European football tournaments, as well events like the 2010 Winter Olympic Games in Vancouver, Canada. They typically consist of small rooms that can hold up to seven big screens (see picture 1). The chairs swivel, allowing viewers to turn between screens and friends.

Sports bars were chosen as setting in our study for ease of

access to undisturbed everyday social television viewing, since that type of viewing is done in public and not in the confines of private homes [31]. We conducted six observations of approximately four hours each. A Samsung mobile phone camera with HD video capture capacity was used to record the visual viewing practices, although leaving out their conversation. The viewers were unaware of the recordings and we ensure their anonymity by providing no recognisable features from the video.

Video recordings make it possible to analyse the details of viewers' gestures in everyday practice [10]. This is useful due to the fleeting and fine-grained nature of interactional work between members of a group, which is more easily unpacked if repeated analysis is possible on a single case.

All recordings were repeatedly viewed in team analysis sessions and core events were transcribed and categorised. Our analysis, presented here, focuses on instances where several viewers are visibly gesturing in front of the screen, which were found both in the data from sports bars and living rooms. We have selected a specific part of a video recorded in the spring of 2010.



Picture 1. The sport bar

Four men in their thirties, who appear to be friends, are sitting in one corner of the room. They are watching the Winter Olympic Games in Vancouver. The national favourite downhill skier is in third place as the last skier takes off. The national favourite had a skiing accident the day before this competition, which intensifies the excitement. The vignette captures the body movements within this group, which are used to unpack how gestures in front of TV screens are produced. The vignette was selected because it consists of an extended sequence with several interactional turns among the participants and because it visualises many interesting aspects of gesturing. The use of a single data point is typical of interaction analysis [10]. While it may not cover the full breadth of behaviours, it permits in-depth examination of the impact of practices, technologies and contextual elements that have an effect on social interaction and media consumption. We recognise that it can be hard to generalise from a single case study, but this nevertheless stands as a perspicuous study [8] that

focuses attention on elements surrounding important aspects of collaborative viewing.

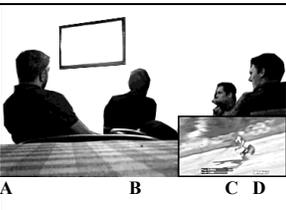
While video recording is increasingly used in data collection in workplace studies in HCI and CSCW, there is as yet no common transcription coding scheme for remote, multi-participant activity similar to that used in conversation analysis. Consequently, we have developed a coding scheme that explicates the empirical material relevant to the analysis. This includes time stamps, screenshots of television broadcasts, transcriptions of commentary (translated to English); descriptions of viewers' physical movements and screenshots of ethnographic video.

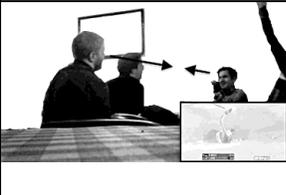
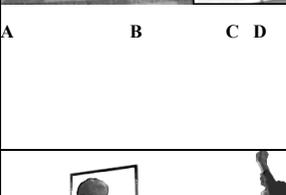
The original images from the video were resized to black and white. We pared the background to more clearly show how the subjects were gesturing, and made sure that the individuals cannot be identified. The individuals are labelled A to D, which can be seen in the screen shots from the ethnographic video. The replication of the broadcast screen shots in the lower right corner is residuals from the video analysis.

ANALYSIS

We studied viewers' physical movements in order to unpack how gestures are related to the screen and the group of viewers. The vignette is presented in three sequential excerpts that bring out various aspects of the role of gestures in interaction, such as how they *overlap* and are continuously shaped and *negotiated* and how the screen content is used *as norms to indicate appropriate gesturing*.

Excerpt 1. Cheering for a bronze

No.	Commentators' voice and broadcast image	Physical movements	Pictures
1 00:49		Viewers are watching the content without making any gestures.	 A B C D
2 00:50	 no!:: she straddles!	D sits up in his chair; starts clapping. A looks at D and lifts his arms at the same time.	

No.	Commentators' voice and broadcast image	Physical movements	Pictures
3 00:50		A raises his arms. D moves his eyes from the screen to look at the others.	
4 00:51	 and she <u>does</u> what we talked about.	A lets his arms fall into his lap when D stretches his arms even higher into the air.	
5 00:51		A sinks down. C's mouth is open. He takes his hand away from his mouth for a second.	
6 00:51		C turns to A with a big smile. A and C exchange looks. D still has his arms in the air. B now turns to the table.	
7 00:52		A turns back to the screen. C puts his head in his hand. B looks towards D and smiles. D looks at B and his arms are straight up in the air.	 A B C D
8 00:53	 she <u>straddles</u> .	D looks towards some other tables and keeps his arms up.	
9 00:59	 Anja takes the bronze medal.	C still has his hand in front of his face. D starts to clap. B turns back to his original screen.	

No.	Commentators' voice and broadcast image	Physical movements	Pictures
10 01:00	 this was the only thing that	C waves his hand back and forth in front of his face. A looks at him, tilting his head and smiling.	 A B C D
11 01:03	 at this time could	D continuously makes celebratory arm movements. C hides his face in his hand	

Overlapping gesturing

The gestures in this excerpt overlapped and were related both to the television content and the ongoing interaction among the viewers. At the start (line 1), the TV viewers watch the skier who is competing with their national favourite for a medal, as they lean back in their chairs. Goodwin [13, p. 1500] uses the term ‘contextual configuration’ to denote a subset of all possible semiotic fields available in a situation ‘relevant to a particular organisation of action’, which is in this case the viewers’ common orientation towards the television. At this moment in the event, a medal can be won only by one of them and the competitor’s race will decide it all. When she straddles a gate and falls (line 2), the race is over. The commentator cries out “No!:: she straddles” (line 2), displaying sympathy on her behalf. There is a great deal of visible physical movement among the viewers. D starts to frantically clap (line 2) and shows less compassion. A begins to cheer by lifting his arms forwards and upwards (line 2 and 3). Their gestures occur at the same time in the cheering. This is something Schegloff [34] calls ‘choral phenomena’, where participants’ expressions are unproblematically overlapping each other.

In the following, when the contextual configuration is changed to include the viewers themselves, the overlapping cheering becomes problematic. In the middle of his cheering, A turns his attention from the screen towards D and then sinks down while simultaneously dropping his arms to his lap (line 4 and 5). This restraint is visibly related to D, since A turns towards D at the same time as he pulls back his arms. A’s gesture become an ‘intervention’ into D’s cheering [35]. A drops his shoulders in a bodily posture that might indicate disappointment or resignation (line 5). His gaze at D, as he leans back, indicates that he is concerned about D’s actions. This interpretation is further supported by the ongoing interaction in the vignette.

In lines 7 and 8, A turns his attention back to the screen, while D has continued to celebrate, now with his arms straight up in the air. D then turns to the people sitting at other tables (line 8). We interpret this as an acknowledgment that cheering has a collective dimension, and when the friend nearest to him makes his gesturing problematic, he turns to other members of the audience in order to share his enjoyment with others. In this case, however, this might also be understood as a way of seeking approval for his gestures, since A is visibly reacting to his gesturing by first restraining his own expression and then turning his gaze from D to the screen.

Gestures are continuously shaped and reshaped among the viewers as a part of a negotiation about appropriate ways of cheering formed as the interaction unfolds in time. What we are looking at is a set of *overlapping* gestures that is both of a non-problematic orchestral type as well as a problematic type. The formulation of the request for moderation of D’s gestures should not be seen as an attempt by A to ‘interrupt’ [34] him, since they both started cheering in concert. However, A clearly attempts to influence the type of cheering that D is engaged in. This is an attempt to tone down or moderate the scale of the gesture, including its physical shape.

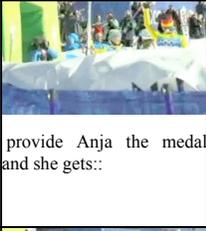
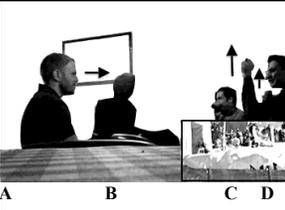
When summing up the actions in this excerpt, it is apparent that the gesturing is a physical interaction both with the TV content and other viewers. First, the viewers react physically to the TV content. Second, they also interact among themselves. The way individuals behave should be seen in relation to the people around them, because each is assessing the behaviour of the others.

Negotiating moderation of celebration

The unfolding of the situation provides additional support for understanding gesture movements as continuously *negotiated* and shaped. During the entire video clip (line 1 to line 22), C holds his hand close to his face, either rubbing it on his nose or leaning his head on it. In line 10, C holds the fingertips of his right hand closely together and repeatedly makes small movements back and forth. It is followed by him hiding his face completely (line 9) in an act often associated with shame. The display of embarrassment becomes a way for C to make a statement about D’s gesturing and thus a comment on appropriate cheering, similar to A’s actions. A looks at C again (line 10) but is now, compared to line 6, leaning back and smiling, confirming C’s visual statement. A continues to engage with D’s gesturing. In line 15, A looks at D’s continued cheering without smiling, but then bursts out laughing (line 16). It is as if he is giving up his attempt to moderate his expression.

What again becomes visually apparent is that gestures are not just performed in relation to what is happening on the screen, but also are interactionally *negotiated* within the group. A’s gestures are focused on negotiating the appropriate expressions of celebration in this situation.

Excerpt 2. Media as resource for deciding appropriate gestures

No.	Commentators' voice and broadcast image	Physical movements	Pictures
15 01:06	 provide Anja the medal and she gets::	D alternately raises his left and right arms. A looks at him without smiling.	 A B C D
16 01:08	 (laughter)	D still does the movement but A starts laughing.	
17 01:09	 waving up to us from below.	A turns to the screen above D. D points to the other screen.	
18 01:10	 Anja cheers. She got a medal after all. Her sixth Olympic medal	A turns around to look at the screen D is pointing at.	

Screen-indicated norms for appropriate gesturing

In the following, we will discuss how broadcast content is used as a resource in negotiating norms that A, C and the commentator reflect and acknowledge in their gesturing.

The gesturing by A and C, performed to influence D, reflects values concerning when and how to cheer. The national favourite's winning of a medal was directly linked to the failure and fall of the competitor. Thus, they were cheering as much for the success of their favourite as for the downfall of the other skier. The latter can sometimes be considered inappropriate.

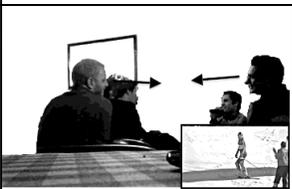
Interestingly, the skiers' broadcasted gesturing behaviour reflects this ambivalence as well. A close-up of the winner's face is shown on the screen immediately after the shot of the falling skier (line 10). She restrains her expression and gives only a slight smile. Then, after three seconds (line 11), she and the other medallists raise their arms to celebrate their achievements. In this initial restraint, we see a parallel to the negotiation among the group of viewers, where the extensive cheering of D was met with restraint by A and embarrassment by C. This ambivalence can also be heard in the commentator's reflections, stating that the na-

tional favourite could only win if the last skier made a mistake (line 9-10).

When D, on the other hand, points to the screen displaying the cheering medallists (line 17), he shows that his gestures are similar to the way the medallists are now celebrating and thus gets support for his gesturing. The indexical gesture of pointing at the screen might be seen as legitimising or accounting for [30] his previous physical movements. At the same time, the broadcasted gestures do not resolve the ambivalence, since they show both expressive cheering and restrained cheering.

Summing up, by pointing at the screen, the viewers make sure they are referring to the same content. More interestingly, the TV content is used as an example of the type of gestures that are appropriate and legitimate. It provides instructions on how to produce gestures. If they previously used the content as something that triggered their physical interaction, the pointing gesture is here used as a means to account for other movements, that is, as a model of appropriate physical movements in this situation. Thus, the viewers do not only bring a set of normative rules and associated gestures to the social interaction around the screen. The screen itself is displaying a social arena that influences and interacts with the viewers.

Excerpt 3. Reaching agreement

No.	Commentators' voice and broadcast image	Physical movements	Pictures
19 01:26	 but she can enjoy the gold from yesterday.	D looks at A and says something.	
20 01:27		A looks at D and makes a somewhat restrained victory gesture	

Cheering and settling the negotiation

The vignette ends with a sequence where A again celebrates, which indicates how these physical movements do multiple jobs. D ends his continuous celebration (line 16) and points at the screen (line 17). For the first time in this vignette, he makes available a turn in his gesturing interaction. Similar to an event described by Streeck [36, p116], A makes a return gesture to confirm that they have established the viewing as a co-experience (line 20). A turns away from the screen and looks at D (line 19). A then leans forward (line 20), clenches his fist and pumps his bent arm downwards, which could be interpreted as a victory gesture. The screen is showing pictures of the fallen skier who is now standing. A, who previously restrained his celebration,

is now cheering, but in a way that is less public and is orientated towards D. Cheering at this particular moment could be seen as less of bad sportsmanship, since he has already shown D that he needed to curb his expression, and as we argued previously, the broadcast itself indicates that celebration is now appropriate. When A again turns his attention to D (line 19 and line 20) and celebrates, he establishes the cheering as a joint activity and thus the conviviality of the television viewing. This final shared expression can be seen as the closing turn in the entire sequence that starts at line 1.

A's gesture is appropriate since it is temporally organised to occur a few moments after the fall. It has a more restrained shape and, finally, it is indexically directed at D and thus also marks the end of the disagreement. Thus, cheering can be done in various ways, such as putting hands and arms into the air or sitting down and bending the arm. An embodied physical movement can mean different things, such as celebration or an invitation to make up. What is appropriate at one time is not appropriate at another.

DISCUSSION

The study was initiated to unpack how physical movements occur in interaction in front of TV screens, which could influence discussions on design directions in the field of gesture-based interface control mechanisms.

Naturally occurring element of TV watching

It has been argued in previous studies that TV watching is sometimes a social activity [4, 11, 12] in which several people look at the screen together. Our study supports the argument that gestures are a naturally occurring element of such everyday interaction. Our group of viewers did not only converse, but also gestured towards each other and the screen, although looking at a conventional TV set, which does not technically recognise what they are doing. Gesturing was meaningful even in a situation when the audience did not expect feedback from the screens. Moreover, these viewers did not even have access to the remote controls for the screens around them, since they were in a sports bar. Thus, the area in front of the TV is already a context where gestures are made to the TV set, even before the introduction of elaborated interactive services.

An attempt to provide for 'natural' gestural interaction would benefit from drawing upon the type of gestures that commonly occur in front of TV sets. We identified a broad range even in the short vignette discussed here. These gestures include lifting the arms to cheer; waving with hands; pointing with arms and index fingers; covering faces with hands; pumping arms up and down. These sorts of gestures might be candidates for domain catalogues [17, 21].

Temporally unfolding interaction

What complicates the attempt to list and catalogue 'natural gestures' is how the type of movements we describe above are shaped by the ongoing interaction with the broadcast

content and with the co-located group of viewers. The visible gesturing is not a set of easily demarcated signs, either directed by an individual to what appears on the screen or a concise turn in an interaction with a co-viewer. We showed how gestures overlap for various reasons so that they are continuously shaped and formed in the situation and interaction to convey a multitude of meanings e.g. a viewer starts a move to raise the arms but then restrains himself when he sees the body movements of other viewers. The way in which the gestures constitute interaction with other viewers is apparent through a number of turns involving three different persons. The restrained gesture combined cheering with efforts to influence co-viewers to moderate their expressions. Gestures can also occur as accounts of other gestures.

As Wexelblat [39] pointed out, and as we have seen in this study, gestures are continuous rather than discrete. Here we have also seen how they are experiential, sequential [23] and negotiated. This complexity makes it challenging to distinguish meaningful patterns even when one is able to review a short video sequence several times. A gesture recognition system has only one shot.

Moving from sports bars to living rooms

The detailed analysis is made on a case from a sport bar. It was chosen, since the public nature of sports bars gives us access to undisturbed details of interaction around a television set. Although the study was undertaken in such a setting and most viewing occurs in private homes, we argue that the study reveals findings of some validity for both locations. First, we also found situations where groups of viewers displayed co-occurring gestures in these occasions. Second, our identification of e.g. choral gesturing is supported by other studies, mainly in conversation analysis [23, 34], where similar sorts of gestures are found. Thus, some of interactional features that appeared in front of the screen are similar to how social interaction has been discussed in other domains as a means for example of being accountable for actions [30]. This indicates that the physical movements occur as ordinary forms of interaction and that the viewers are not accounting for the sport bars in and of themselves in their interaction. This literature does not however discuss the specific types of gesturing, found in our fieldwork that is done towards a screen.

At the same time, this study and other earlier studies point to the importance of focussing on co-viewing practices. The step from designing for interactive computing, which traditionally focuses on individual humans and individual devices, to interactive TV should be accompanied by increased attention to designing interfaces that account for shared viewing practices.

Broadcasted gesturing as contextual resource

Understanding the meaning of gestures requires access to the context in which the physical movements are done [23, 34]. In this case, the forms and timing of the gesturing were

contextually influenced by the broadcast itself. It is interesting to see how much of the gesturing draws on the content, the gestures made by people in the broadcast. How and when the skiers make their gestures is reflected in how the viewers' organise their movements. Thus, for the analysts, the meaning of the physical movements could be unpacked only by seeing how they were aligned with the television broadcast. Again, availability of the broadcast content made it easier to discern the meaning of the movements. The detailed analysis of a single case displays the complexity of this activity.

Social implications of 'gesture tracking adaptation'

Gesture recognition technology depends on the possibility of recognising meaningful physical movements in all locomotion occurring around a screen, in the living room for instance [9]. This will be difficult, however, given the intricate ways in which we use such language [18]. Thus, other ways to design such systems are open for exploration, that is, for finding ways for people to adapt to what the system can do. This is already an important research task [1, 2, 5], where support is provided to train users to learn to make recognisable gestures. What is at stake in design is thus a balance between the appropriate level of ordinary actions and the extent to which the interaction has to be specially designed, trained and framed for this particular context. How the design of future TV interaction techniques resolves these issues will affect people's everyday lives in front of their TV screens.

Here, we would like to bring up possible social implications of these systems. In everyday life, users' ordinary physical movements might lead to unintended system recognition and undesired feedback. When people move their arms up and down or left and right [17] for reasons other than interacting with the system, the movements may still be recognised as such. This is a problem that will become more critical as the systems apply more elaborated gesture libraries – if, that is, more advanced technology is used to elaborate on the gesture interaction. This will probably lead to behavioural changes in front of television sets, which we term 'gesture tracking adaptation', a situation in which viewers try to minimise their movements.

Viewers might need not only to practice specific gestures, but also to restrain their physical movement or even social interactions in front of the TV set. Such consequences would counteract the promise behind gesture control. Paradoxically, these types of systems could have the effect of constraining gesturing in front of screens, that is, all other sorts of physical movements that are not directed at the system. Such implications might play out in individual use, but are more likely in social viewing, which involves both gesturing towards the screen and gesturing towards co-viewers, as discussed in this paper. On the other hand, simplifications such as adding an "on and off" button, might ease this problem to the expense of decreased user experience of gesture control.

A comparison with contemporary remote control can be revealing. Interacting with a device like a remote control and interacting with gesture recognition technology and tracking devices differs in how the devices support collaborative use. In the first type of technology, people can negotiate control of the television around the technology, for example through discussions among the viewers or by physically withholding the device from each other. In the latter case, ordinary physical movements are also visible to the system and will have to be accounted for in design. Social negotiations conducted around the remote control – although not recognised by it – will be made available as input in the tracking system and thus must be accounted for in design and handled by such systems. Gesture recognition technology tracks complexity and users might therefore feel the technical demand to moderate their usual behaviour in front of the screen.

The social implications of gesture tracking adaptation are an important lesson of this study. We need to discuss the consequences upon ordinary social life and to use them as a guide for further studies in this area.

Interaction and broadcast gestures

Although this analysis argues for moderation of the argument of gestures as a 'natural' way to interact with the television, we would also like to bring up an aspect of our study as potentially inspiring for design. How viewers orient their gesturing to the gesturing displayed on the screen might provide new avenues for the design of future interactive television. The opportunity to design gesture controls that account for gestures displayed in the broadcast might be a line of investigation that users might see as 'natural.' First, the capacity of the system to recognise what viewers are doing might be increased if there were a way for it to recognise the gesturing in the content. This could be accomplished through other techniques, such as manual tagging by content producers. Second, such a path would also open new possibilities for training and teaching viewers to make gestures of a type recognisable to systems. Finally, establishing interaction between broadcasted gestures and viewer gesturing could be used to generate new forms of broadcastable content in a way that is being extensively explored in the gaming area.

Gesturing alone or in groups

The focus of this article is to understand social aspects of TV watching. Although TV watching is frequently done in groups, it is also an activity that is often pursued in solitude. Our study has relevance for group interaction, but does it also have something to say about the design of gesture controls for individuals who are watching TV alone? This has to do with the type of gesturing people who are alone currently do and how they would like to interact with gesture recognition systems. If such gestures are 'natural', in the sense that they would draw upon physical movements as discussed in this paper, it would be important to also consider their continuous nature. Moreover, what we have seen

in this study is how gestures are moulded and continuously shaped to account for norms and appropriate conduct. The question is whether individual viewers would be more distinct and refrain from hesitations, as we have seen here, or whether the solitude would provide a zone free of such considerations.

We also need to ask whether viewers are alone, if they are the only person in a living room. The influence on broadcast content on gestures, as discussed here, indicates that solo viewers are still sometimes part of a social interaction, which would make them orient to social negotiations when forming their gestures.

Generalizations from detailed analysis

Although the findings of our study provide implications for design, the use of detailed analysis of a single case also make it difficult to generalise beyond the cases discussed here i.e. from gesturing towards a TV in one place to another, and from one TV show to another. The way the participants do gesturing in this case might not apply to e.g. news programs. Still the findings could also be seen as non-particular since we found co-occurring gesturing both in private and public settings and previous studies identified their occurrence in non-TV settings. Live sport is also one of the most popular genres. Still, subsequent research is needed to establish generalizability of the peculiarities of gesturing towards a screen across a number of relevant categories. Generalizations and the way in which the findings in single case studies could be of relevance for other settings is also a common concern among single case studies in HCI, and has been addressed in literature. We argue therefore along conversational analysis that: “[t]he claim of regularity, however, is not the news, or value, of the analysis.”[29]. Rather, the aim is to show how, using this micro-oriented approach, we can identify a number of problems that can occur during social watching. In sum, the challenge of discussing a topic with potential relevance, based on a limited empirical scope, needs to be recognized. The careful study of one type of watching in a particular setting is revealing, and might even be troubling, but it needs to be complemented with other studies of ordinary gesturing, focusing on a plethora of situations and a multitude of methods.

CONCLUSION

Gesture control of TV content is motivated by recent success in the domain of digital games, as well as the understanding of gestures as an ordinary form of interaction. This has for example inspired research to catalogue appropriate gestures and investigate technology. This study focuses specifically on challenges in social viewing. This everyday experience raises the stakes of applying ‘natural’ or ordinary gesture elements in interaction. We have shown how this introduces the problem of making sense out of a very ambiguous set of physical movements. Social gesturing is negotiated, continuous and overlapping. If gesture recogni-

tion technology fails to work out how to make sense of this interaction, it might instead lead to demands on viewers to discipline their behaviour in living rooms and sports bars, a situation we refer to as ‘gesture recognition adaptation.’ We also show how social interaction includes the interaction between viewers and the media content, which could be further explored in design.

ACKNOWLEDGMENTS

We would like to thank Alexandra Weilenmann and colleagues at the Centre for useful help. The research was made possible by a grant from the Swedish Governmental Agency for Innovation Systems to the Mobile Life VinnExcellence Center, in partnership with Ericsson, Microsoft, Nokia, IKEA and the City of Stockholm. It was also made possible by EU Fascinate integrated program.

REFERENCES

1. Ashbrook, D. and Starner, T. MAGIC: A Motion Gesture Design Tool. In *Proc. CHI 2010*, ACM Press (2010), 2159-2168.
2. Bau, O. and Mackay, W. E. OctoPocus: A Dynamic Guide for Learning Gesture-Based Command Sets. In *Proc. UIST '08*. ACM Press (2008), 37-46.
3. Bavelas, J. B., Chovil, N., Lawrie, D. A. et al. Interactive Gestures. *Discourse Processes 15-4* (1992) 469-489.
4. Bernhaupt, R., Obrist, M., Weiss, A., et al. Trends in the living room and beyond: results from ethnographic studies using creative and playful probing. *Computers in Entertainment (CIE), Social Television and User Interaction, Vol. 6, No. 1*, ACM Press (2008), 5:1-5:23.
5. Bragdon, A., Uguray, A., Wigdor, D., et al. Gesture Play: Motivating Online Gesture Learning with Fun, Positive Reinforcement and Physical Metaphors. *Proc. ITS 2010*, ACM Press (2010), 39-48.
6. Chen, M-Y., Mummert, L., Pillai, P., Hauptmann, A. and Sukthankar, R. Controlling Your TV with Gestures. *Proc. 11th ACM Multimedia Information Retrieval 2010*, ACM Press (2010), 405-408.
7. Choi, S., Han, J., Lee, G., et al. RemoteTouch: Touchscreen-like interaction in the TV viewing environment. *Proc. CHI 2011*, ACM Press (2011), 393-402.
8. De la Flor, G. Luff, P., Jirotko, M. et al. The Case of the Disappearing Ox: Seeing Through Digital Images to Analysis of Ancient Texts. *Proc. ACM CHI*, (2010) 473-482.
9. Fothergill S., Mentis H., Kohli P. et al. Instructing People for Training Gestural Interactive Systems. *Proc. CHI 2012*, ACM Press (2012), 1737-1746.
10. Garfinkel, H. Studies of the Routine Grounds of Everyday Activities. *Social Problems, 11-3*, (1964) 225-250.

11. Gauntlett, D. and Hill, A. *TV Living: Television, Culture and Everyday Life*. Routledge, London (1999).
12. Geerts, D. and De Grooff, D. Supporting the Social Uses of Television: Sociability Heuristics for Social TV. *Proc. CHI 2009*, ACM Press (2009), 595-604.
13. Goodwin, C. Action and embodiment within situated human interaction, *Journal of Pragmatics* 32 (2000) 1489-1522.
14. Heath, C., Hindmarch, J. and Luff P. *Video in Qualitative Research. Analysing Social Interaction in Everyday Life*. London: SAGE publication (2010).
15. Hopmann, M., Salamin, P., Chauvin, N., et al. Natural Activation for Gesture Recognition Systems. *Proc. CHI 2011*, ACM Press (2011), 173-183.
16. Höök, K. Affective Loop Experiences: Designing for Interactional Embodiment. *Phil. Trans. R. Soc. B* 12 December 2009, 364:3585-3595.
17. Kela, J., Korpipää, P., Mäntujärvi, J., et al. Accelerometer-based gesture control for a design environment. *Personal and Ubiquitous Computing 2006*, 10-5, (2005), 285-299.
18. Kendon, A. Gesture. *Annual Review of Anthropology* 26, (1997), 109-128.
19. Kray, C., Nesbitt, D. and Dawson, J. User-Defined Gestures for Connecting Mobile Phones, Public Displays, and Tabletops. *Proc. Mobile HCI '10*, ACM Press (2010), 239-248.
20. Kubey, R. W. and Csikszentmihalyi, M. *Television and the Quality of Life: How Viewing Shapes Everyday Experience*. Lawrence Erlbaum Associates, Inc., Publishers, New Jersey, 1990.
21. Loke, L., Larssen, A. T. and Robertson, T. Labanotation for Design of Movement-Based Interaction. *Proc. Second Australasian Conference on Interactive Entertainment* (2005), ACM Press (2005), 113-120.
22. Michael, M. *Constructing Identities: The Social, the Nonhuman and Change*, Sage, 1996.
23. Mondada, L. Understanding as an embodied, situated and sequential achievement in interaction, *Journal of Pragmatics*, 43 (2011) 542-552.
24. Murphy K. *Shaping Things to Come: Temporality and Transmodality in Swedish Design*, Draft, Univ of California Irvine.
25. Nintendo Wii (2011). <http://www.nintendo.com/wii> (accessed 2013-01-07)
26. O'Hara, K. and Glancy, M. Watching in Public: Understanding Audience Interaction with Big Screen TV in Urban Spaces. In *Social Interactive Television: Immersive Shared Experiences and Perspectives*. Cesar, P. Geerts, D. and Chorionopoulos, K. (Eds.). London: IGI Global, 2009.
27. O'Hara, K., Glancy, M. and Robertshaw, S. Understanding Collective Play in an Urban Screen Game. *Proc. Computer Supported Cooperative Work*. ACM (2008), 67-76.
28. Pirker, M., Bernhaupt, R. and Mirlacher, T. Investigating Usability and User Experience as Possible Entry Barriers for Touch Interaction in the Living Room. *Proc. EuroITV 2010*, ACM Press (2010), 145-154.
29. Pomerantz, A. 1990. Conversation analytic claims, *Communication Monographs*, 1479-5787, 57:3 3, 231-235.
30. Randell, R. Accountable technology appropriation and use, *Proc. NordiCHI 2004*, ACM Press (2004), 162-170.
31. Reeves, S., Sherwood, S. and Brown, B. Designing for Crowds. *Proc. NordiCHI 2010*, ACM Press (2010), 393-402.
32. Rico J. and Brewster S. Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability, *Proc. CHI 2010*, ACM press (2010), 887-896.
33. Saffer, D. *Designing Gestural Interfaces*. Sebastopol: O'Reilly Media, Inc., 2009.
34. Schegloff, E. Accounts of Conduct in Interaction, in *Handbook of Sociological Theory*, Turner J.H. (ed.) Kluwer Academic publishers New York, 2002.
35. Schegloff, E. 'On some gestures' relation to talk', in Atkinson J.M. (ed.) *Structures of Social Action: Studies in Conversation Analysis*, Cambridge University Press, 1985, 266-296.
36. Streeck, J. *Gesturecraft. The Manufacture of Meaning*. John Benjamin's Publishing Company, Amsterdam/Philadelphia. Vol. 2, 2009.
37. Shimada, K., Suka, H., Ban, H. and Ozaki, T. Toward Human-oriented Home. *Hitachi Review* 58 (4), (2009).
38. Sweetser, J., Grunnet-Jepsen, A. and Panchanathan, G. Absolute Pointing and Tracking based Remote Control for Interactive User Experience. *Proc. UXTV'08*, ACM Press (2008), 155-164.
39. Wexelblat, A. Research Challenges in Gesture: Open Issues and Unsolved Problems, In *Gesture and Sign Language in Human-Computer Interaction, Lecture Notes in Computer Science*, 1998, Vol. 1371/1998, 1-11.
40. Väänänen, K. and Böhm, K. Gesture Driven Interaction as a Human Factor in Virtual Environments: An Approach to Neural Networks. *Virtual Reality Systems*. R.A. Earnshaw, M.A. Gigante and H. Jones. London, Academic Press Limited (1993).
41. Xbox Kinect (2011) <http://www.xbox.com/kinect> (accessed 2013-01-07)
42. Zou, S., Xiao, H., Wan, H. et al. Vision-based Hand Interaction and its Application in Pervasive Games. *Proc. VRCAI 2009*, Japan, ACM Press (2009), 157-162.