WorldCupinion: Experiences with an Android App for Real-Time Opinion Sharing during Soccer World Cup Games

Robert Schleicher[1*], Alireza Sahami Shirazi[2,4], Michael Rohs[1,3], Sven Kratz[1], Albrecht Schmidt[2,4]
[1]Deutsche Telekom Laboratories, TU Berlin
[2]University of Duisburg-Essen
[3]Ludwig-Maximilians-Universität München
[4]VIS, University of Stuttgart
*corresponding author

ABSTRACT
Mobile devices are increasingly used in social networking applications and research. So far, there is little work on real-time emotion or opinion sharing in large loosely coupled user communities. One potential area of application is the assessment of widely broadcasted television (TV) shows. The idea of connecting non-collocated TV viewers via telecommunication technologies is referred to as Social TV. Such systems typically include set-top boxes for supporting the collaboration. In our work we investigated if mobile phones can be used as an additional channel for sharing opinions, emotional responses, and TV-related experiences in real-time. To gain insight into this area, we developed an Android app for giving real-time feedback during soccer games and to create ad hoc fan groups. We present results on rating activity during games and discuss our experiences with deploying this app over four weeks during soccer World Cup. In doing so, we highlight challenges and opportunities we faced and give an outlook on future work in this area.

ACM Classification Keywords
H5.2 [Information Interfaces and Presentation]: User Interfaces.

General Terms
Design, Experimentation, Human Factors

Author Keywords
Mobile applications, mobile social networking, opinion sharing, real-time feedback, sports.

INTRODUCTION
Mobile devices are increasingly used for mobile social networking. One explanation for this development is that mobile devices are almost always with their users, have continuous wireless connectivity, and feature increasingly capable user interfaces. They can thus serve as ubiquitous input devices and sensors for user reactions, emotional responses, and opinions around large public events (Diakopolous & Shamma, 2010).

The goal of the work presented here is to investigate mobile social software as a tool for research on opinion sharing in large user communities. We picked the soccer World Cup 2010 as a use case for this research because it is an event with extremely high public attention in many parts of the world and many people have a high emotional involvement to (at least some of) the matches. The matches are also synchronized in time with many simultaneous viewers and thus many potential users. We focus on exchanging spontaneous emotional feedback between users who are part of a virtual fan block.

The particular test application, World Cupinion, is an Android application that lets soccer fans express their opinions about events and moments in soccer matches while watching them. Through this application users can support their favorite teams and share their opinions with other fans. As we expected that users' focus of attention is mainly on the match itself and short bursts of usage occur when interesting events happen, the design focus was on simplicity and quick usage. When not actively used, the app mostly served as an ambient display that conveyed the aggregated opinions of the active users.

This work addresses the following aspects and research questions:

- How to share experiences and opinions effectively in real-time across a large number of mobile devices?
- How to design for awareness of group opinion in a loosely coupled ad-hoc group? How to visualize information related to shared experiences?
- How to distribute and maintain a free Android app for ambient mobile communication?

In the following sections we first discuss the concepts of Social TV, real-time opinion sharing, and the utilization of mobile phones as a research tool. We then give an overview of the design and system architecture of our test application and discuss the distribution and publication channels for the application. After that we present results derived from log files as well as from a subsequent online-questionnaire and report on the experiences we made with the public prototype. We conclude with recommendations for research in the large and with giving ideas for future work.

RELATED WORK

Social TV

Various researches have been exploring the idea of using additional communication channels in parallel with watching TV. "AmigoTV" (Coppens et al., 2004) was an early social TV system that used voice chat communication in combination with broadcast TV. It also provided emoticons and a buddy list with online status. Motorola Labs developed a series of prototypes called "Social TV" system (STV), which allowed users to engage in spontaneous communication with their buddies through text or voice chat while watching TV (Harboe, Metcalf et al. 2008; Metcalf, Harboe et al., 2008). The system also included an additional display to convey views of the current TV-watching users. Harboe, Massey et al. (2008) give a comprehensive overview of social TV systems. Further, various user studies investigated the communication modalities. Geerts (2006) as well as Baillie et al. (2008) compared communication via voice with other modalities. Both studies reported that most users believed that voice chat was more natural and easier to use than text chat. However, Huang et al. (2009) conducted a similar study using the STV system. They found that participants preferred text chat and they often communicated about topics unrelated to the TV content. Geerts and DeGrooff (2009) reported a set of comprehensive sociability heuristics for social TV systems.

Media annotation and sharing while watching TV has been studied, too. Diakopoulos and Shamma (2010) analyzed the sentiments of tweet annotations for a presidential debate to find out their relationship to discussed topics and performance of the opponents in the event. Miyamori et al. (2005) proposed and examined a method for generating views of TV programs based on viewer's opinions collected from live chats on the Web. Affective responses to unstructured video commenting systems were evaluated by Nakamura et al. (2008).

All the mentioned social TV systems require the installation of set-top boxes for supporting collaboration. Since set-top boxes are only available in certain locations, users are restricted to particular environments. To overcome this limitation and attract participants we intended a mobile phone application that would give users the chance to use it for sharing their opinions in any context in which watching the event is possible, even in bars, the stadium, or at public places – a requirement indispensable for the sports domain.

Mobile Phone Sports Apps

Mobile phones are particularly suited to support sports fans that attend such events which frequently take place outdoors: *MySplitTime* (Esbjörnsson, Brown et al., 2006) allows users to take pictures of bypassing cars at rallies and obtain additional information about the current ranking of the photographed car. Although not particularly designed for sports events, *coMedia* (Jacucci, Oulasvirta, et al., 2007), an app to create and share digital memories was also tested at a big rally in Finland. *TrottingPal* (Nilsson, 2004) helps spectators at the trotting track to gather additional information to improve their betting and to coordinate with other visitors who might be dispersed across the area. Information retrieval appears the focus of apps that target in-stadium sports: *eStadium* (Ault, Dunlop, et al., 2008) which was later extended to *RISE [Rich Immersive Sports Experience]* provides visitors of football games at Purdue's Ross-Ade stadium with various statistics, replays and other multimedia services (Facwett, Beyer et al., 2009). *TuVista* is a similar app tested at the Estadio Azteca in Mexico City (Bentley & Groble, 2009) *YinzCam*, a spinoff from the Carnegie Mellon University[1] offers related services for various popular college sports including basketball, ice hockey, etc.. While the capabilities of these existing apps may vary in detail and may be extended in the meantime, their general intention appears to be to provide a service for mobile phone users that is comparable to the characteristics of professional TV sports broadcasting i.e., detailed background information and multimedia material to guarantee an exclusive viewing experience. We aimed at supporting sports events viewers in another regard, namely sharing their immediate impressions of the game, which may be one of the major reasons to watch such events in a group.

Real-time Emotion Sharing

Taking a closer look at what information the audience/watchers of sport broadcasting actually wished to share with their friends or fan group, it turned out to be mostly the preliminary evaluation mixed with the personal emotional impact of specific events during the game, much less a "cold" rational assessment of the ongoing maneuvers on the field. This was no surprise, as emotions are known to have a strong social component (Ochsner & Schacter, 2000) and probably even developed to provide a fast and immediate way to communicate the momentary state of an organism to the environment (Ekman, 1999).

The sudden onset and strong expressive component of emotions make them an ideal candidate for mobile communication as it allows the user to somehow extend his/her reach beyond the usual radius of face-to-face communication. On the other hand, these properties also impose a number of requirements for any application: feedback should be quick and if possible "analogous," i.e., non-verbal to avoid the necessity of lengthy formulation to describe a simple and transient affective rush. Emoticons appear to be an appropriate way to communicate these states (Derks et al., 2008). In addition, the provided rating scheme should contain domain-specific labels (e.g., "yellow card") as well as domain-independent features (e.g., "like-dislike") (Pang & Lee, 2008). Relying on such a limited set of means of expression is also referred to as *lightweight communication* (Metcalf, Harboe, et al., 2008) or if it is not restricted to a specific location *ambient mobile communication* (Bentley, Kaushik, et al., 2006). The latter authors describe an exemplary prototype, called *Music Presence* that uses a set of domain-independent icons, namely thumbs up/down, and

"!" similar to what we offered in our app. The communicative purpose differentiates these approaches from related work subsumed under *experience sampling methods (ESM)* that also utilize mobile phones as digital diaries (Carter & Mankoff, 2005), but mostly let the user create their entries for later analysis by the researcher, not for conveying them to other users. Our app extends the notion of ambient mobile communication by additionally using the mobile phone (and the corresponding app "market") for acquiring people interested in this form of communication.

Mobile Phone Apps as a Research Tool

The approach to recruit participants via mobile phone applications has been deployed by other research labs as well. In the following we will give a short and eclectic overview pointing out the aspects we consider of interest when using apps for research. A detailed description of the mentioned projects can be found in the references as well in the other contributions of this issue.

Oliver (2010) released a Blackberry application to obtain measures such as average usage frequency and duration from more than 17000 users. Main challenges reported were the increase in power consumption caused by the research app and the reliability of time stamps when the data were logged locally on the device. Michahelles (2010) published various social apps, mostly on the Android portal and gives a brief assessment of the varying success in distribution as well as requirements that may be new for research using apps like the need to constantly add new features to maintain the users' interest in the application.

CenceMe, initially developed for Nokia's N95 (Milluzo *et al*, 2008), later also released for the iPhone (Milluzo *et al*, 2010), allows users to communicate and share their momentary activity and location with others, including a linkage to Twitter and Facebook. As this app infers high-level status information from the mobile phone sensors, the major concern was which analysis should be done locally and which on the backend server, and how this affects battery lifetime. *CenceMe* is a good example of how extensive the information can be that researchers can derive from mobile app usage. On the other hand, *Hungry* (Windows Mobile) / *Feeding* (iPhone) *Yoshi* (McMillan et al., 2010) shows what innovative data collection paradigms are possible with mobile technology: The basic plot is a location-based game in which users have to collect fruits at various locations. The players can gain additional scores by accomplishing tasks or quests which consist of questions asked by the researchers. The researcher can tailor these questions to specific locations or usage patterns. Since the questions appear to be part of the game, this mitigates the issue of inferring user intentions etc. from log data alone in order not to disrupt the game flow. The authors (McMillan *et. al.*, 2010) also discuss the perils of this procedure.

The idea to exploit competition amongst users as a means to increase participation was one of the reasons we chose the World Cup 2010 as a starting point for our research trial. The stimulative nature of this event has obviously also been identified by other researchers: Morrison et al. (2010) offered *World Cup Predictor* for that purpose. Whereas the goal of that app apparently was to forecast the winner of the cup, i.e. the outcome, we focused more on the process of social interaction, namely sharing opinions about ongoing events while watching the game.

APP DESIGN

As mentioned above, the design criteria included simplicity, since the user's focus of attention is on the match itself, and short-term usage, since situations arise quickly an interaction might just involve stating one's opinion about the current event. Moreover one aspect was to visualize the aggregated opinion of a potentially large number of users and to use the screen as an ambient display. The latter feature gives the user the opportunity to observe how the fan-aggregated opinion evolves even though he or she is not actively interacting, but may react to the updates by rating again.

The app is structured in three screens (Figures 1 and 2). The first screen (Figure 1, left) shows the list of upcoming matches with their starting times and dates in the user's local timezone. The timezone played an important role here, since we intended to deploy the app in the Android Market for worldwide distribution. The game selection could have been automatized, except for parallel games during the first phase of the tournament, but we decided to keep the list in order to allow users to plan their viewing times in advance of the games.

After selecting a game the user would enter the "arena" for that game (Figure 1, right). That screen allows the user to give feedback during the game and to see the aggregated opinions of the fans of the own and the other team. Initially the rating buttons are disabled and the user has to select the team he or she wants to support in order to activate the interface. This design decision means that users have to be fan of a particular team in order to provide input. The input buttons cover most of the display area to be easy and quick to press. Their functionality was either to give soccer-specific assessments of events like yellow/red card, whistle/play on, and in the beginning "offside". The remaining buttons served to express the current mood of the user, i.e., thumbs up/down, invoke a vuvuzela sound to express excitement, and later on a "Yippee" button which replaced the offside icon and was accompanied by the sound of applause. Below each button there is a horizontal bar that indicates the average opinion regarding that input category. For example, if the bar below the "thumbs up" button is half filled that means that 50% of the fans have pressed that button during the last 30s.

Figure 1. WorldCupinion screens: Initial screen *(left)* showing the match list and main screen *(right)* used during a game.

A feature that was added one week after publishing the game is to see the aggregated opinions of the fans of the other team as well. The statistics of the own fans are shown in green and the statistics of the other team in blue. The blue bar is located behind the green one and the green bar is not fully opaque to see the blue one behind it.

Another change to the main screen (Figure 1, *right*) was the replacement of an "offside" icon with a "yippee!" icon (Figure 1, *right*, center icon). We made this change after we realized that the "offside" icon was rarely used. At the same time, the app lacked a way to express strongly positive emotions, e.g., when the own team scored a goal. Adding the "yippee!" icon provided a way to express that kind of emotional feedback.

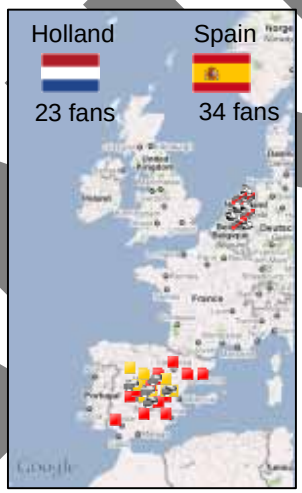

Figure 2. Geographic distribution of fan opinions.

The third screen, "world opinion" (Figure 2), shows the geographical distribution of fan opinions of both teams on the map. The underlying idea is that this visualization shows geographical clusters of users having opposing opinions. The map view is based on the standard Google Maps APIs with icon overlays for the feedback that was given at a particular location. Using Google Maps APIs allows interactive panning and zooming of the map. However, we restricted the maximum zoom level for privacy reasons.

SYSTEM ARCHITECTURE

World Cupinion is implemented as a client-server architecture. The World Cupinion mobile application sends two basic request types to the server, update requests, and input requests. Update requests are used to poll the state of the mobile application's user interface, and input requests are used to send user opinions to the server, as soon as an opinion button has been pressed. The map view sends a further request type, to which the server generates a response containing the user inputs of the last 5 minutes.

The server logs all inputs to a SQLite database and maintains statistics of the user opinions received in the last 30s. These 30s statistics are sent to the mobile clients in response to update requests.

We initially used UDP datagrams for communication, as our protocol does not require an active connection. UDP also imposes a lower load on the server, which is beneficial if there are many simultaneous server requests. However, it soon appeared that certain network firewalls and also mobile network providers may block UDP packets that have non-standard destination ports. To remedy this, our mobile application has a fallback mechanism that automatically switches to HTTP requests if UDP communication is unsuccessful. User input events are always sent via HTTP to ensure that they do not get lost.

Supporting HTTP requests has the further advantage of enabling the implementation of platform-independent web interfaces. Although we did not originally plan to use a web interface, we implemented one for evaluation purposes and to fulfill requests from users that did not use the Android platform. It turned out that this was useful, as our statistics show that a substantial proportion of the input originated from the web interface.

A further important issue of mobile phone application is energy consumption (Oliver, 2010; Miluzzo et al., 2008). Over the 90 minutes of a game (plus the 15 minutes break and an optional 30 minutes extension), the application continuously communicates with the server via the mobile phone network or WiFi. There is a tradeoff between the update rate of the interface and energy consumption. In pilot tests we found that one update every 3 seconds is sufficient. A significant contribution to energy consumption comes from continuously using the device as an ambient display for the opinion state. Even if the user is not interacting with the device the community opinion is updated and shown. This is technically implemented with a "wake lock" that prevents the display from switching off completely. Usually, a NexusOne mobile phone that is fully charged at the start of the game has a battery level of about 50% at the end of a game.

DISTRIBUTION AND PUBLICITY

Distribution via the Android Market

From June 4, 2010 onwards (one week before the start of the World Cup), the World Cupinion app could be downloaded for free from the Android Market. An advantage of using the Android Market as a distribution platform over Apple's iTunes AppStore, for instance, is that published applications appear almost instantly for download, and are not subject to a lengthy reviewing process (see Miluzzo et al., 2010) with the risk of rejection of the application.

The ability to rapidly push new releases of the application to the Android Market allowed us to publish weekly updates containing bug fixes or new features during the actual soccer World Cup.

Public Relations

It of course does not suffice to simply release a new application into the wild. Potential users need to be informed of the application's existence in order for them to download it.

We used a number of channels to make the application known to potential users. In addition to press releases made by the Deutsche Telekom Laboratories and the TU Berlin, we tried to promote the application in internal events (summer party, weekly lab meeting) and external events (lab open house) of the Deutsche Telekom Laboratories. At these events we distributed flyers advertising the app, containing a QR code linking to the app on the Android Market. We also created a website (www.worldcupinion.com) and actively used social media (Twitter and Facebook) and forum entries to reach as large an audience as possible. Finally, we sent emails to a number of mailing lists in our lectures and posted messages about the app on Android developer forums.

Updates

The Android Market allows to easily publish updates. We took advantage of this feature several times. If a long-term study is conducted it allows to carry out several design iterations while keeping the user base of the app. Besides bug fixes, the changes and updates related to (1) replacing the "offside" icon with the "yippee" icon, (2) adding group-generated sounds, (3) showing the opinion of the opposing fans, (4) sending notifications when a game starts (and restarts after the halftime), and (5) the addition of an in-application questionnaire displayed after the conclusion of the World Cup. These changes were not obvious from the start and reflected insights gained from application usage and user comments. The update mechanism provided a convenient way to do these changes. On the other hand, one has to be careful not to confuse users when features change. If some users do not update the application there might be inconsistencies between deployed application versions, e.g., some users might still have the "offside" icon in the place where users of the updated app have the "yippee" icon. We expected that most users would update their app, and for us having the flexibility to try out different versions was more important than version consistency. Unfortunately, the server protocol did not include the version number, so we could not track the percentage of users connecting to the server with outdated application versions. This is something we will clearly consider in the future when deploying similar apps.

RESULTS

Based on the Android portal at the end of the World Cup, we had registered a total of 1645 downloads and 448 "active" installations (=29% of all downloads). The number of active installations denotes the number of users that still had the app installed on their devices at that point.

The results presented in the following are based on data from two sources: on the one hand the logs of user activities during the matches and on the other from the in-application questionnaire provided with the last update.

Usage Statistics

On average 28.6 (+/- 19.1) users were active during the games, with a maximum of 94 and a minimum of 8 users for a single game. Figure 3 shows the number of participating devices for all of the 64 games of the World Cup. Of course this number was highly dependent on the nationality of the teams playing, but a general decay from the first couple of games (when the app was still "new") could be observed. From match 49 the round of sixteen started, which led to a temporary increase in usage. The most prominent game of the World Cup, the final (game 64) had surprisingly low number participants.

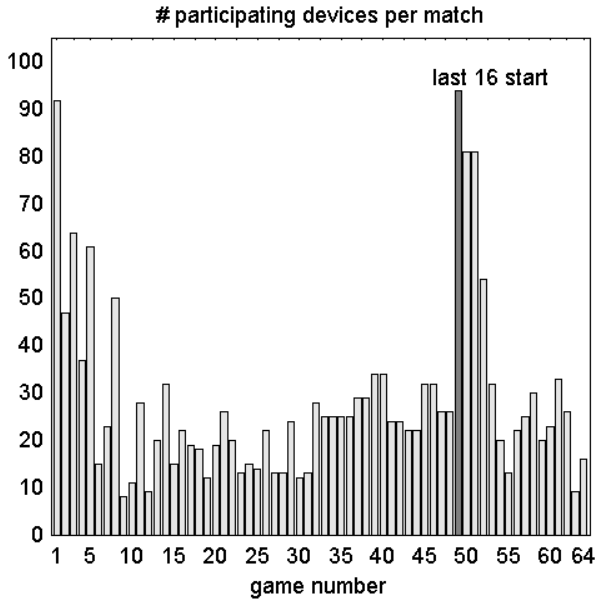

Figure 3. Number of participating devices during the course of the soccer World Cup 2010.

The average participation lasted 681 (+/- 1316.2) seconds during which 17.6 (+/-33.6) actions, i.e., button presses were performed. Based on our database, 71% of inputs were

from the Android client and 29% from the Web-based client. Table 1 shows the basic statistics of average number of inputs during a game, and average session length divided by interface type, i.e., the mobile phone client or web interface.

	Android app		web interface	
	#inputs	session length	#inputs	session length
Mn	15.58	703.79	23.61	615.13
Md	9.00	60.87	9.00	44.86
STD	24.56	1320.22	51.11	1305.39
Min	1.00	.38	2.00	1.44
Max	372.00	6997.30	591.00	6611.56

Table 1: Usage statistics by interface type, i.e. Android mobile phone app vs. web interface.

A multivariate analysis of variance (MANOVA) revealed that there was a significant effect of input interface on these parameters: users of the app tended to give less ratings during longer sessions ($F_{2,820}$=6.584, p=0.001). However, univariate comparison revealed that this difference was only significant for #inputs ($F_{1,821}$=9.063, p=0.003), but not for session length ($F_{1,821}$=0.714, p=0.398).

While the means in number of inputs clearly varied for both interface types, the medians were identical, indicating that in the web interface there were some users with very high number of inputs as also shown by the higher maximum (591 total inputs for the web interface vs. 372 for the app). This might be due to excessive clicking using the mouse. To give a better impression of input activity, Figure 4 shows the click distribution for the first 4 games of the round 16 (games 49-52) in all of which more than 50 users participated. For these four games, no clear difference in input patterns between the mobile phone and the web client could be seen, but both occurred during the whole game time. However, the graphs also show that ratings did not happen at a constant rate, but were closely linked to events in the game. This is exemplified for game 51 (Germany vs England)[2], and the graphs suggest that users understood the app as intended to communicate moments of high relevance to the other participants. For all games there was also some activity during halftime, which we found to be mostly vuvuzela and thumbs up/down assessments after examining the log files. This led to the question which buttons were used how often.

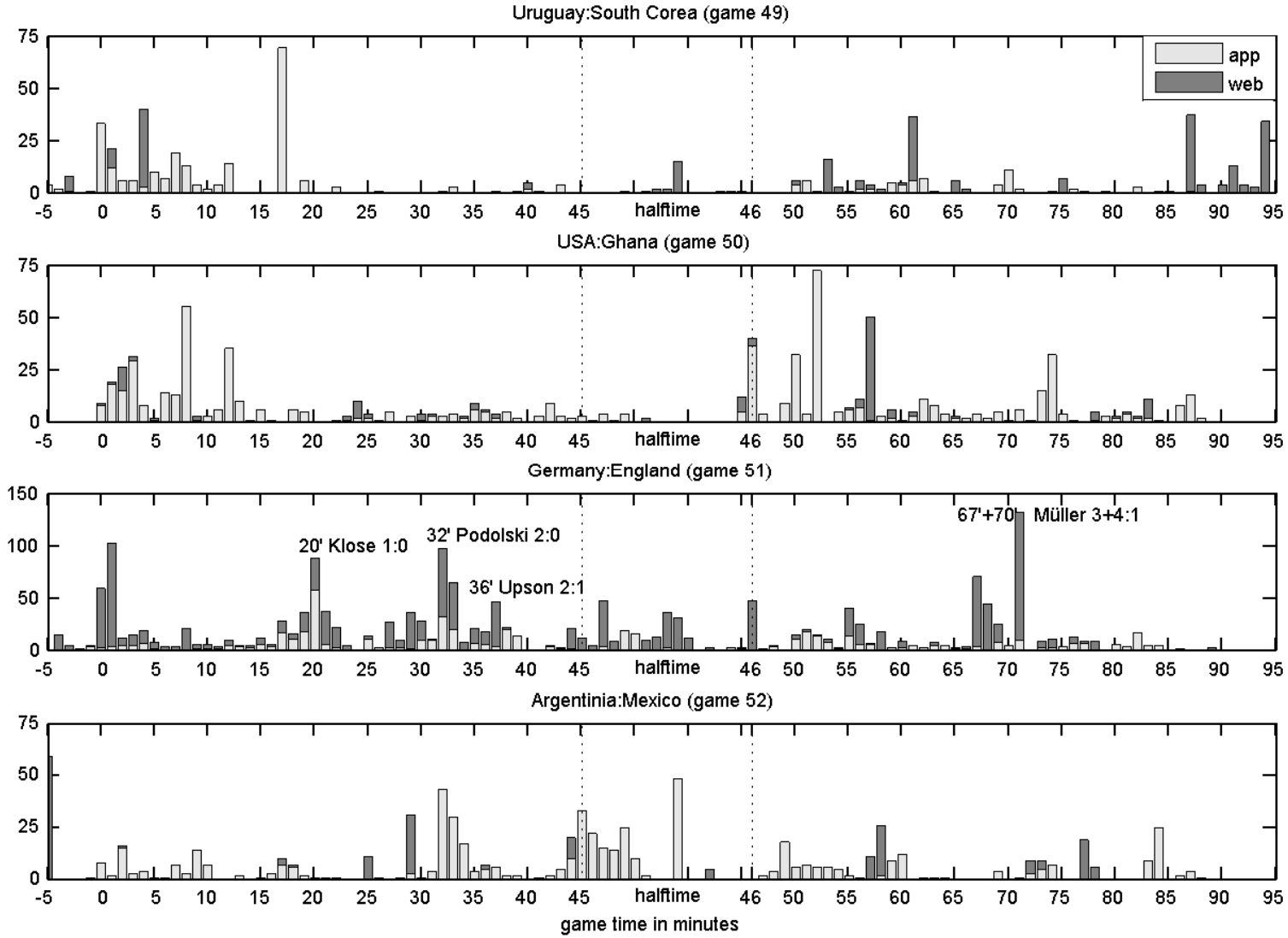

Figure 4. number of inputs during the first four games of the round of 16, distinguished by interface type (Android mobile phone app or web interface), over game time (begin and end of halftime indicated by dotte lines). Bars represent absolute numbers. Please notice the adjusted y-axis limits for game 51 (England : Germany), where also the moment of the goals and corresponding rating bursts are labeled.

Button/Icon Usage

In addition to the amount of activity over time, we were also interested in the usage of the offered icons. Table 2 shows the relative frequency of button clicks per game for games 49 to 52 and across all games. For game 49 to 52 the differences in relative usage frequency where also assessed statistically using a Chi^2 test, which revealed a clear effect of button meaning on usage frequency ($Chi^2(24)=407$, $p<0.001$). Table 2 also indicates to what extent the usage frequency of single buttons varies from game to game and whether this difference is statistically significant (numbers that share the same subscript in a row are not significantly different from each other). While there are slight differences between single games, the vuvuzela was by far the most frequently used button across all games with the highest number in game 51 (Germany : England). The second most frequently used button serves to annotate a typical soccer controversy, namely that the referee should whistle in a specific situation. Yippie!, thumbs up and down were also used regularly with around 9-15% across all games. The remaining buttons that could be used to express boredom with the actual course of the game or whether the referee should give a yellow/red card or let continue to play refer to events that are rather specific to particular moments in the game and, thus, were used less often.

icon	game nr. 49	50	51	52	mean games 49-52	mean all games
	% usage				*% usage*	
vuvuzela	27_a	29_a	36_b	20_c	28	23
whistle!	18_a	19_a	6_b	17_a	15	19
Yippie!	12_a	9_a	21_b	11_a	13	9
thumbs up	9_a	9_a	15_b	$13_{a,b}$	12	15
thumbs down	6_a	$8_{a,b}$	8_a	12_b	9	10
red card	11_a	10_a	2_b	8_a	8	5
boring	5_a	5_a	7_a	6_a	6	6
play on!	7_a	5_a	3_b	6_a	5	6
yellow card	5_a	6_a	3_b	5_a	5	6
Total	100	100	100	100	100	100

Table 2: Relative frequency of button clicks per game in percent (column-wise), for the first 4 games of the round of 16 and across all games. Values are rounded to whole numbers. Percentages with the same subscript letter are not statistically different from each other in the corresponding Chi^2 test when comparing row-wise across games, i.e., 27_a% vuvuzela clicks in game 49 is statistically not different from 29_a% in game 50, but both are significantly lower than 36_b% in game 51. $13_{a,b}$% indicates that 13% is neither different from other numbers in the same row with the subscript *a* nor from ones with the subscript *b*.

Game 51 (Germany : England) is the game that shows most deviation from the other three games with regard to relative button usage. As it was also the game with the highest user activity (see Figure 4), we looked how the relative frequency of button usage varied across fans of both teams. Although the results have to be taken with care as both fan groups differed in terms of participating devices (n=16 for England, n=51 Germany, n=6 devices "switched" teams and voted for both), Figure 5 shows that there is a clear effect of team association on clicking activity ($Chi^2(8)=74$, $p<0.001$) which also is reflected in the pairwise comparisons of button selection. While fans of the winning team Germany used the vuvuzela and the Yippie! button significantly more often, the English fans expressed their disappointment with a clearly higher use of the "thumbs down" button and also voted more frequently for the referee to whistle and react with a yellow card.

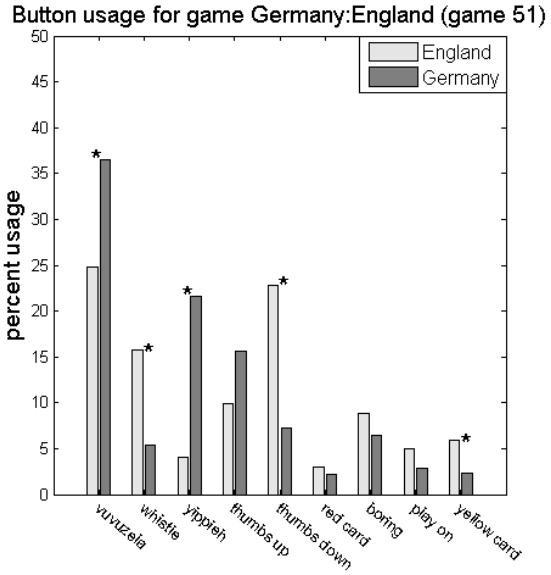

Figure 5: Relative frequency of button clicks during game 51 (England:Germany) split by team. Stars indicate a statistically significant difference between both fan groups ($p < 0.05$).

To avoid misinterpretation about the user intentions and to see whether this idea of opinion sharing was really present, we conducted an additional post-hoc survey which results are described in next section.

Questionnaire Results

The questionnaire was introduced with the last update as a link to a web page which people could access from within the app on their mobile phone. It consisted of altogether 22 questions ranging from simple demographics (age, sex) to open suggestions for improvements. For all evaluative questions, a five-point Likert scale was offered.

In total 46 users (mean age = 20.1 +/- 9.2 years, 6 female users) replied to the questionnaire, of whom 37% followed the World Cup matches frequently and 50% watched occasionally. 55% of the participants considered themselves as knowledgeable fans and 30% as experts (knowledgeable of players' details). 73% of them stated that they normally watched the matches at home and 65% watched with the family or buddies.

World Cupinion Usage

18% of the participants stated that they used the app for most matches, 40% used it regularly, and 42% occasionally or just once. Also, those who considered themselves as knowledgeable or expert fans used the app more frequently. 60% of those who watched the matches in a group (with family, buddies, or crowd) still used the app regularly or for most matches. Additionally, participants rated the app in general. The average rating was 3.6 out of 5+/- 1.1 (Median=4). Those who used the app more regularly rated the app higher.

Connectedness vs. Fun

Participants were asked to rate if the level of fun and connectedness to other fans changed while using the app. 11% mentioned that the fun aspect did not change at all. 30% believed that the fun aspect increased sometimes and 59% reported to have more fun most of the time or (almost) always. None of the responses indicated that the app reduced the fun of watching. Also, 7 out of 46 participants (15%) did not feel connected to other fans at all. 32% felt (very) strong connection and 53% average/little bit connection. Those who had more fun felt more connected to the other fans (Spearman's $\rho = .636$, $p<0.00$).

DISCUSSION

Taking up the questions from the introduction, the overall intention of this research trial was to explore how to design a mobile phone app that would allow users to form a loosely coupled ad-hoc community and to engage in social interaction, namely expressing their opinion, within this group. There were basically two challenges: on the one hand the technical realization that should be capable of dealing with a large number of devices and provide instant feedback. Our solution to these requirements were described in the first part of this paper, and besides occasional crashes of the map view, the app as well as the server worked stable and reliable. One thing we would probably add to the log data next time is the app version of each user.

The other challenge was to find an area of application and offer an appropriate form of communication that would be encouraging enough for users to interact without additional (financial) compensation. At the same time, they should produce data that would be of use for later analysis. We chose the World Cup 2010 because this event received high public attention, was of high relevance to many people and broadcasted simultaneously. As the rating course for the four games of the round of sixteen showed (Figure 4), the concept of online rating succeeded and already the analysis of rating frequencies alone allowed identifying relevant moments in the game. Thereby, the four expressive icons "Vuzuzela", "Yippie!" and "thumbs up/down" comprise more than half of all clicks across all games (57%, see table 2), which we take as an indicator that the app was used as intended to share experiences/opinions. Among the soccer-specific icons, the "whistle!" button was used most frequently (19% in table 2). Fans of opposing teams clearly vary in term of how they evaluate a specific game (see Figure 5). This result is not surprising but served as a "sanity check" and was the first step in the direction of identifying group-specific rating patterns. For finer analysis, e.g., evaluative meaning of ratings in reaction to activity of the opposing fan group or depending on momentary location, a larger data set would be desired, which leads to shortcomings of this study.

Shortcomings of this study

The number of 1645 users who downloaded the app is a sample size that most researchers in HCI domain would appreciate. However, the fact that only 448 of that downloads were still "active" at the end of the World Cup alreadey points out that a high dropout rate has to be taken into account.

As Figure 3 shows, usage tended to decrease over time even if as in our case the events became more and more thrilling. This appears to be another example for the novelty effect thas has been reported before for Social TV (Huang et al., 2009) and mobile phone apps alike (Consolve et al., 2008). In principle the developer has two possibilities to ensure a more or less constant number of participants: either frequently introduce new features to keep the existing users interested or engage in extensive public relations to attract new users. Both measures require appropriate ressources, the latter probably even the help of a public relations professional as our attempts to attract interest of sport's magazines or newspaper editorial boards were not successful and we were informed that they expect a more or less pre-written description that they can publish immediately. As we did not ask it in the subsequent questionnaire, we are not aware how users specifically learnt about our app and could thus not infer which promotion strategy was most successful. We would clearly revise that in a future field trial, and also suggest to other developers to pay attention to this issue right from the beginning on to be able to allocate promotional ressources more efficently in the later stages of a study.

The decision to introduce a web interface and thereby to some extent adulterate the idea of mobile phone app research was a reaction to several requests to also provide the app to other platforms and grounded in the intention to include as many users possible in this first trial. A web interface enabled users outside the Android world to participate. Although rating frequency in general varied between both interfaces, we could not see a clear bias in rating patterns as again exemplified by Figure 4. The fact that the web inter-

face users tended to have shorter sessions (although not statistically significant) was an indicator to us that for opinion sharing while watching TV a mobile phone app might be the more convenient interface as it can be used in almost any viewing situation, e.g. on the couch, in the bar etc. without having to scroll or zoom in a browser.

However, without additional interview or observational data, this assumption cannot definitely be confirmed. In general we had the impression that the data collected via an app that is released into the wild is much more difficult to interpret than data obtained in a controlled laboratory setting. For certain analyses, an incorporation of other (qualitative) data like interviews or observation in addition to the quantitative information obtained from log files appears to be inevitable. Jacucci, Oulasvirta, et al. (2007) may serve as an example how to combine these different methods in a field test. At the same time, their sample size of n=8 also indicates that some kind of pre-selection, which users of an app might be most representative, is required when adopting their approach for research in the large. Our intention was to see whether the recruitment of anonymous participants and data collection of their communication is feasible per se. We will summarize the observations we made and feedback we got beyond the log files and questionnaire data in the next section called "experiences."

Experiences

The first users using the app during an actual game were friends, colleagues and students who were made aware of it by our public relations activity. Here the focus was almost more some kind of beta testing and providing feedback to the developers than using it for rating. However, the Vuvuzela functionality was quickly used to emphasize important moments or echo the sound coming from the TV, which then led to a more comprehensive usage of the app to comment the ongoing game. Although we were well aware that there were several other "Vuvuzela apps" available in the market, we still thought that this feature might serve as some kind of door opener to attract the user's attention initially. Our observations within our peer group confirmed that assumption. Probably other research projects might also benefit from this approach. However, there is one caveat when adapting popular "recreational apps" or functionalities for research applications: we found that users expect a running software and will give low ratings if an app is not polished or crashes during usage. There were occasional crashes of the map view in our app, which only affected the map view and not the other parts of the application. Some users appeared to be very critical about this, as is documented by one user comment mentioning this particular issue. Moreover, a few handset types had problems for running the Android app. Even though platform fragmentation is a small problem for the Android platform than for other systems, this issue appeared. Of course, providing an industry-strength app as a research prototype is not feasible for most research labs as it requires more development resources than are typically available.

We were surprised to what extent users apparently download and install an application without actually using it for a longer time. It appears that the abundance of available mobile phone applications let them become a disposable article like promotional gifts, an observation that has also be made by McMillan et al. (2010). When research apps are released on the market, they have to compete with these existing products and apparently cannot expect a "research in progress" bonus that people sometimes implicitly grant when testing prototypes etc. in laboratory settings.

In a similar vein, a few weeks after our app was released on the Android market and while the tournament was still going on, we noticed that there was another, unrelated app for soccer fans made available on the iTunes store that featured some of the functionalities we also offered, including the "thumbs up/down" icons and the map view ("world opinion," see Figure 2). Again, these challenges might be pretty new to people who usually conduct laboratory research. Finally, since the app is released to the public, large number of users might use the app which leads to high traffic and performance issues. Users expect that the app works at any moment. Therefore 24/7 maintenance and monitoring are crucial and should already be considered during evaluation.

CONCLUSION

Exploiting spontaneous lightweight communication with a mobile phone during TV shows for annotating these events relies on a critical mass of users. We tried to gain a large number of users for sharing opinions in large user communities in real-time, by picking a popular topic in which people have emotional involvement and simultaneously follow a shared event. The soccer World Cup 2010 provided a good setting for the first trial, because it is quite popular in many parts of the world and extends over four weeks, which allowed us to do several design iterations. We tried to gain as much user attention as possible by publishing the app on the Android Market and announcing it to make it known to potential users. The market update facility allowed us to try out design modifications easily and improve the prototype based on usage statistics and user feedback. The rating course for the first four games of the last 16 showed that the concept of sharing personal opinions about sports events online is feasible. However, conducting an uncontrolled study in the wild has its own shortcomings. The quality of the data obtained as well as the user experience stands or falls by the number of users involved. We described our experience with the measures we took to attract users in the previous paragraph and will now point out what future work needs to be done.

Outlook

So far, the focus of existing sports apps appears mainly to provide additional statistics and multimedia material to the user as described in "Related Work" Section. If there is any intention to promote a continuous interaction during match

time, the lightweight communication we described might be the proper way. The option to comment or vote and present these sentiments to the own as well as the opposing fan group would enrich the so far predominately individual experience of using a sports app with a social aspect. Integration in social media platforms, such as Facebook and Twitter nowadays appears to be standard for applications that aim at social interaction (Michahelles, 2010; Miluzzo et al., 2008), thus, we also intend that for future apps. As Facebook users in particular appear to be willing to share information (Miluzzo et al., 2008) and as such are of special interest for researchers, it would probably be advantageous at some point if that platform could offer a special API to researchers (like *Google scholar* in addition to the standard *Google*) to facilitate integration and to indicate applications free of commercial interest. A connection to Facebook might also answer two other questions: the integration of a chat functionality that was mentioned by some of our users for boring parts of a game. We were not sure whether such a chat extension would distract users too much from the ongoing game (see Huang et al., 2009) and interfere with the idea of rating synchronously to the timeline of the TV show. Facebook already provides a chat, and for setting up a contact beyond the game this platform might be more appropriate. The second question is what benefit the app could provide to the user, a crucial aspect for the success of any research app as Michahelles (2010) points out. Morrison et al. (2010) offered prizes of £500 for the winners of their *World Cup Predictor*. So far, we solely trusted in the presumed fun of rating while watching. However, the option to gain some kind of "expert" status with an attached "rating profile" as a consequence of frequent ratings which could then be published somewhere (e.g., Facebook) would probably provide additional motivation. Feelings of competence are among the motives that are currently being discussed as the causes for positive experiences when using interactive products (Hassenzahl et al., 2010). A motivational stimulus like this could be even more important if the idea of a simple shared online rating via the mobile phone is extended to other domains: for example, online rating of sneak previews of movies, where again the reputation of being an expert whose opinion is subsequently publicly available might encourage users to participate. For a stronger focus on research purposes, Diakopoulos and Shamma (2010) analyzed the tweets of a presidential debate. Of course short text messages provide more information than button clicks, but the concept of "thumbs up/down" could in principle also be applied here and if combined with the location information obtained from the mobile phone's GPS sensor enable analysis which statements where of special relevance to viewers in a certain area. Accompanying demographic questions to complete the picture could be presented within the application, following a recommendation of McMillan et al. (2010) "to stay in the app" to increase the response rate.

REFERENCES

Ault, A., V., K. J., Dunlop, S. R., & Coyle, E. J. (2008). eStadium: The Mobile Wireless Football Experience. In *Proceedings ICIW 2008* (pp. 644-649).

Baillie, L., Frohlich, P., & Schatz, R. (2007). Exploring Social TV. In *Proceedings ITI'07* (pp. 215-220).

Bentley, F., & Groble, M. (2009). TuVista: Meeting the Multimedia Needs of Mobile Sports Fans. In *Proceedings MM'09* (pp. 471-480) ACM Press.

Bentley, F., Kaushik, P., Narasimhan, N., & Dhiraj, A. (2006). Ambient Mobile Communications. In *Proceedings CHI 2006* (pp. 1-3).

Carter, C., & Mankoff, J. (2005). When Participants Do the Capturing: The Role of Media in Diary Studies, *Proceedings CHI'05* (pp. 1-10).

Consolvo, S., Klasnja, P., McDonald, D. W., Avrahami, D., Froehlich, J., LeGrand, L., Libby, R., Mosher, K., & Landay, J. A. (2009). Flowers or a Robot Army? Encouraging Awareness & Activity with Personal, Mobile Displays. In *Proceedings UbiComp'08* (pp. 54-63).

Coppens, T., Trappeniers, L., & Godon, M. (2004). AmigoTV: towards a social TV experience. In *Proceedings EuroITV'04* (pp.1-3).

Derks, D., Bos, A. E., & von Grumbkow, J. (2008). Emoticons in computer-mediated communication: social motives and social context. *Cyberpsychology & Behavior, 11*(1), 99-101.

Diakopoulos, N. A. & Shamma, D. A. (2010). Characterizing Debate Performance via Aggregated Twitter Sentiment. In *Proceedings of CHI 2010* (pp. 1195-1198).

Ekman, P. (1999). Basic Emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of Cognition and Emotion.* (pp. 45-60). Sussex, UK: John Wiley & Sons.

Esbjörnsson, M., Brown, B., Juhlin, O., Normark, D., Östergren, M., & Laurier, E. (2006). Watching the cars go round and round: designing for active spectating. In *Proceedings CHI 2006* (pp. 1221-1224).

Fawcett, J., Beyer, B., Hum, D., Ault, A., & Krogmeier, J. (2009). Rich Immersive Sports Experience: A Hybrid Multimedia System for Content Consumption. In *Proceedings of CCNC 2009* (pp. 1-5).

Geerts, D. (2006). Comparing voice chat and text chat in a communication tool for interactive television. In *Proceedings NordiCHI'06* (pp. 461-464).

Geerts, D. & De Grooff, D. (2009). Supporting the social uses of television: sociability heuristics for social TV. In *Proceedings CHI'09* (pp. 595-604).

Harboe, G., Massey, N. Metcalf, C., Wheatley, D., & Romano, G. (2008). The uses of social television. *Computers in Entertainment* 6, 1 (2008), 1-15.

Harboe, G., Metcalf, C.J. Bentley, F., Tullio, J. Massey, N., & Romano, G. (2008). Ambient social TV: drawing people into a shared experience. In *Proceedings CHI'08* (pp. 1-10).

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products − Facets of user experience. *Interacting with Computers, 22,* 5, 353-362.

Huang, E.M., Harboe, G., Tullio, J. , Novak, A., Massey, N., Metcalf, C. J., & Romano, G. (2009). Of social television comes home: a field study of communication choices and practices in TV-based text and voice chat. In *Proceedings CHI'09* (pp. 585-594).

Jacucci, G., Oulasvirta, A., Ilmonen, T., Evans, J., & Salovaara, A. (2007). CoMedia: Mobile Group Media for Active Spectatorship. In *Proceedings CHI 2007* (pp. 1273-1282).

Luyten, K., Thys, K., Huypens, S., & Coninx, K. (2006). Telebuddies: social stitching with interactive television. In *CHI'06 extended abstracts* (pp. 1049-1054).

McMillan, D., Morrison, A., Brown, O., Hall, M., & Chalmers, M. (2010). Further into the Wild: Running Worldwide Trials of Mobile Systems. In *Proceedings of Pervasive 2010* (pp. 210-227).

Metcalf, C., Harboe, G., Tullio, J., Massey, N., Romano, G., Huang, E. M., & Bentley, F. (2008). Examining Presence and Lightweight Messaging in a Social Television Experience. *ACM Transactions on Multimedia Computing, Communications and Applications, 4*(4), 1-16.

Michahelles, F. (2010). Getting closer to reality by evaluating released apps? In *Research in the Large,Workshop at Ubicomp 2010*.

Miluzzo, E., Lane, N. D., Fodor, K., Peterson, R., Lu, H., Musolesi, M., Eisenman, S. B., Zheng, X., & Campbell, A. T. (2008). Sensing Meets Mobile Social Networks: The Design, Implementation and Evaluation of the CenceMe Application. In *Proceedings of SenSys 2008* (pp. 337-350).

Miluzzo, E., Lane, N. D., Lu, H., & Campbell, A. T. (2010). Research in the App Store Era: Experiences from the CenceMe App Deployment on the iPhone In *Research in the Large,Workshop at Ubicomp 2010*.

Miyamori, H., Nakamura, S., & Tanaka, K.. (2005). Generation of views of TV content using TV viewers' perspectives expressed in live chats on the web. In *Proceedings International Multimedia Conference'05* (pp. 853-861).

Morrison, A., Reeves, S., McMillan, D., & Chalmers, M. (2010). Experiences of Mass Participation in Ubicomp Research. In *Research in the Large, Workshop at Ubicomp 2010*.

Nakamura, S., Shimizu, M., & Tanaka, K. (2008). Can social annotation support users in evaluating the trustworthiness of video clips? In *Proceedings. 2nd Workshop on Information credibility on the web* (pp. 59-62).

Nilsson, A. (2004). Using IT to make place in space: evaluating mobile technology support for sport spectators. In *Proceedings ECIS 2004* (paper 127).

Ochsner, K. N., & Schacter, D. L. (2000). A Social Cognitive Neuroscience Approach to Emotion and Memory. In J. C. Borod (Ed.), *The Neuropsychology of Emotion* (pp. 163-193). Oxford: Oxford University Press.

Oehlberg, L., Ducheneaut, N., Thornton, J., Moore, R.J., & Nickell, E. (2006). Social TV: Designing for distributed, sociable television viewing. In *Proceedings EuroITV'06* (pp. 25-26).

Oliver, E. (2010). The Challenges in Large-Scale Smartphone User Studies. *In HotPlanet 2010*. San Francisco, USA: ACM.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval, 2* (1-2), 1-135.

Preece, J. & Shneiderman, B. (2009). The Reader-to-Leader Framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction, 1,* 1 , 13-32.

Schleicher, R. & Trösterer, S. (2009). The 'Joy-of-Use'-Button: Recording Pleasant Moments While Using a PC. In *Proc. of INTERACT 2009. Lecture Notes In Computer Science*, vol. 5727. Springer-Verlag, pp. 630-633, 2009.

Weisz, J.D., Kiesler, S., Zhang, H., Ren, Y., Kraut, R.E., & Konstan, J.A. (2007) Watching together: integrating text chat with video. In *Proceedings CHI'07* (pp. 877-886).

ENDNOTES

[1]http://www.yinzcam.com/about.html

[2] Events were retrieved from:http://g.sports.yahoo.com/soccer/world-cup/ [accessed Oct. 2010]