“Can we work this out?”: An evaluation of remote collaborative interaction in a mobile shared environment

Dari Trendafilov[1,2], Yolanda Vazquez-Alvarez[2], Saija Lemmelä[1] and Roderick Murray-Smith[2]
[1] Nokia Research Center, Helsinki, Finland
{dari.trendafilov, saija.lemmela}@nokia.com
[2] University of Glasgow, Glasgow, G12 8QQ, UK
{dari, yolanda, rod}@dcs.gla.ac.uk

ABSTRACT

We describe a novel dynamic method for collaborative virtual environments designed for mobile devices and evaluated in a mobile context. Participants interacted in pairs remotely and through touch while walking in three different feedback conditions: 1) visual, 2) audio-tactile, 3) spatial audio-tactile. Results showed the visual baseline system provided higher shared awareness, efficiency and a strong learning effect. However, and although very challenging, the eyes-free systems still offered the ability to build joint awareness in remote collaborative environments, particularly the spatial audio one. These results help us better understand the potential of different feedback mechanisms in the design of future mobile collaborative environments.

Author Keywords

Collaborative virtual environments, mobile shared interaction, tactile, spatial audio, social presence.

ACM Classification Keywords

H.5.2 User Interfaces: Haptic I/O, Prototyping, Interaction styles, Auditory (non-speech) feedback.

General Terms

Design, Experimentation, Human Factors.

INTRODUCTION AND BACKGROUND

Mobile social interaction occurs in personalized virtual communities of friends and many of the well-established web-based social networks now feature a mobile component that allows users to stay connected while on the move. A number of existing commercial social networking applications enable users to communicate via regularly updated location or context information that offers a richer sense of the dynamics within a social group. However, they are still restricted to asynchronous updates emphasizing the indirectness and remote nature of the interaction, whereas a fluid synchronous communication style would create the sense of immediateness and engagement and bridge the physical distance gap.

Distributed collaborative technologies, which have become commonplace with the proliferation of the Internet, could potentially enable a richer and more engaging mobile social interaction. However, the means for remote communication between users in collaborative environments are typically audio and video, which reflect a limited amount of the richness of real world communication and are not always suitable in mobile settings. Furthermore, users of traditional audio-video communication systems often experience difficulties in maintaining awareness of the actions of others and in inferring where a co-worker's attention is directed.

The techniques we present in this study explore the potential of audio and haptics in addressing these issues. They have been suggested in previous research but, to our knowledge, they have not been evaluated formally in a collaborative mobile context. Non-visual interaction methods for mobile devices are rare, although they could bring real benefits to users, especially in a mobile context when visual attention is compromised. Previous research investigating cooperative physical tasks in collaborative virtual environments (CVE) suggests that haptics provide significant performance benefits and a strong interpersonal link, i.e. sense of “togetherness”, between participants [2,7]. More recent work concerning haptic communication shows that haptics significantly improves presence and perceived performance [8]. This topic is not only of theoretical interest, but has great practical importance. Artificially affecting the sense of togetherness could benefit remote communication. Previous work has also shown the benefits of audio in representing location or distance of objects in CVE. Crommentuijn and Winberg [5] suggest using rhythm to represent distance in a virtual haptic 3D environment, combined with discrete directional auditory cues. Winberg and Bowers [11] represented content with pitch and timbre and location with stereo panning. However, their concept used discrete turn-taking steps and did not allow for a fluid dynamic joint interaction. The use of 3D or spatial audio based on Head Related Transfer Function (HRTF) filters [3] provides for a more effective positioning of audio around the user than stereo panning and is capable of mirroring the spatial organization of a visual display.

Joint activities require coordination of both content and process, mediated by common ground which is defined in [4] as a *“state of mutual understanding among conversational participants about the topic at hand”*. Clark’s theory of common ground states that people must have shared awareness in order to carry out any form of joint activity.

MobileHCI 2011, Aug 30–Sept 2, 2011, Stockholm, Sweden.

(Shared) awareness is defined in [6] as *"an understanding of the activities of others, which provides a context for your own activity"*. To achieve shared awareness people need a high sense of presence in the CVE. Presence is defined in [12] as *"the subjective experience of being in one place or environment, even when one is physically situated in another"*. Previous work has identified low latency and high degree of interactivity as some of the factors contributing to a high sense of presence. In a CVE using text-chat, audio or video-conference [9] found perceived performance to be significantly higher with video than text-chat and perceived presence to be lower with text-chat than audio or video.

We present a novel mobile interaction method for remote CVE. This method is based on touch input allowing for a more embodied style of interaction digitally bridging the gap with others in social networks. Our goals were to:

- Determine to what extent users can successfully operate an eyes-free version of the system.
- Determine to what extent users felt connected and present in the shared environment.
- Provide some useful guidelines to designers who are required to implement mobile CVE.

Our research questions were: 1. Can users achieve shared awareness in an eyes-free environment? 2. Is spatial audio a superior audio technique in such an environment? 3. How efficient and usable is such an interface?

EXPERIMENTAL DESIGN

Prototype systems

Our prototype systems were implemented on Nokia N900 mobile phones linked via WiFi and connected over Bluetooth to an SK7 sensor pack (SAMH Engineering Services). The SK7 offers capacitive sensing from 12 square pads in a 4x3 configuration at 100Hz and a dual vibro-tactile feedback display - one pager-style vibration motor and one pulsed resonant actuator (ALPS Force-Reactor™ S-type). The sensor pack provided a simple one-dimensional scanning touch input for exploring the interaction space.

We implemented three prototypes, i.e. one visual-only and two eyes-free. For the visual-only prototype the mobile phone and the sensor pack were fitted in a custom-made case (Figure 1a). For the eyes-free prototypes the phone was placed on a lanyard around the user's neck with the sensor pack in hand while wearing a pair of Sennheiser M@B40 headphones (Figure 1b). Each prototype corresponded to one of the conditions tested in this study:

1. *Visual-only*: A simple collaborative calendar-like GUI enabled users to browse appointments on one specific day (Figure 1a). Using the one-dimensional capacitive touch input users agreed on a time for an appointment with their partner. Both users' calendars were aligned on the timeline and cursor movement was discretized at 30-minute-slot steps from 8am to 5pm (20 slots in total). This way, users could visually follow the position of both cursors on their own mobile phone display. Free calendar slots were shown in green and busy ones in blue. The calendar content was always opaque and users were told this was due to privacy reasons. We focused on the shared collaborative interaction and not on how the content would affect the interaction.

Figure 1. (a) Visual-only, (b) eyes-free testing setup.

2. *Audio-tactile*: A pitch-tone audio cue indicated both the location and the direction of movement of the partner's cursor. We used the SoundTouch audio processing library (www.surina.net/soundtouch) to create 22 different sounds forming a 'chromatic scale' with 11 semitones in ascending order and 11 semitones in descending order. This chromatic sound scale was mapped to the 20 slots available in the calendar (see Figure 2 top). The resulting sounds were all stereo, 16-bit, sampled at 44 kHz and normalized to 70% of the audio dynamic range, which equals to a normal conversation typically 60-70dB. Using this scale, a high-pitch sound indicated the partner's cursor was to the left and a low-pitch sound indicated the partner's cursor was to the right. The higher the pitch sound, the further to the left the cursor was and *vice versa*. To avoid overload with sound cues, the audio was active only when one of the users stopped browsing and suggested a free slot. At this point, audio was played back at two predefined speeds: 1) quickly: indicated the partner was browsing, and 2) slowly: indicated the partner's cursor was stationary. Additional audio cues were played to both users in case of success or failure. We complemented the auditory display with a dual vibro-tactile display, using slow pulsing pager motor vibration to identify available free slots and a sharp and fast vibration from a pulse resonant actuator to indicate both cursors were located in the same slot, i.e. users had met in the shared environment.

3. *Spatial Audio-tactile*: The design of this prototype was identical to the Audio-tactile except for the mapping of the location and direction of movement of the partner's cursor. Only one low-pitch sound source was used to indicate the

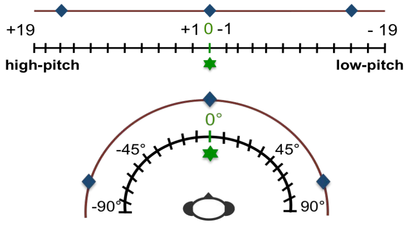

Figure 2. Audio-tactile (top) SpatialAudio-tactile (bottom). Diamonds represent possible partner's cursor locations. The star at '0' represents a proposed free slot.

Condition	Togetherness Median (IQR), *n*	Social Presence Median (IQR), *n*	Affective Benefits Median (IQR), *n*	Perceived Task Performance Median (IQR), *n*
Visual	4 (3 to 5), 26	4 (3 to 5), 26	5 (5 to 5), 26	4 (3.5 to 5), 26
Audio-Tactile	2 (1 to 3.25), 26	3 (2 to 4), 26	4 (3 to 5), 26	1.5 (1.5 to 3), 26
Spatial Audio-Tactile	2 (1 to 3), 26	3 (2 to 4), 26	4 (2.75 to 4.5), 26	2 (1.5 to 3.5), 26

Table 1. Analysis of Questionnaire Data for Togetherness, Social Presence, Affective Benefits and Perceived Performance.

location of the partner's cursor, instead of a chromatic scale, and the cursor movement direction was mapped to a location in a 3D space around the user's head (-90° to +90° azimuth in 20 discrete steps) always 1m away in the frontal horizontal plane (see Figure 2 bottom). The sound source was mono, 16-bit, sampled at 16 kHz and normalized to a conversation audio level. A Maemo version similar to the JAVA JSR-234 Advanced Multimedia Supplements API was used to position the audio sources.

Tasks

For each trial per condition, a pair of users performed a collaborative target acquisition task consisting of finding a shared free slot on their calendars. In order to find a shared free slot, users had to individually explore their own free slots and suggest a free slot to their partner by maintaining the cursor at the slot location and wait till their partner found this slot. At this point, a success or a failure audio cue would play depending on whether that time slot was free for both users, i.e. shared, or not. For each trial, both users had three 30-minute free slots in their calendars but only one of these free slots was shared. Positions of the free slots were randomized per trial. Each session lasted up to 5 minutes or until five trials were successfully completed. The successful completion of these tasks was logged.

Methodology

Twenty-six participants (16 male, 10 female, aged 18 to 54) were allocated into 13 pairs, all reported normal hearing. We used a within-subjects design and conditions were randomized. Trials were always presented in the same order. Conditions were tested in a walking lab environment with pairs separated into two different rooms. Participants had no other way to communicate except the shared virtual environment. The experiment consisted of a short introduction followed by a training session and the three conditions.

In the training session, which was exclusively devoted to familiarize users with the interaction method, each participant was sitting in a separate room and interacting in pairs using an integrated version of the system consisting of a combination of visual, audio and vibro-tactile feedback. The training session lasted 10 minutes and was divided into two parts: in the first half users could experience the pitch-tone audio and in the second half the spatial audio. In all three conditions the pairs had to perform the same tasks for up to 5 minutes. At the end of each condition they were asked to complete questionnaires for sense of togetherness [7], affective benefits (ABC-Q) [1], perceived social presence and perceived task performance [9]. Once all conditions were completed, a user experience questionnaire was also filled in. The experiment took about 1 hour in total and participants were allowed to rest between conditions.

RESULTS

Presence

A non-parametric Friedman test showed a significant effect on type of feedback per condition on a 6-point Likert scale for the perceived sense of togetherness ($\chi2 = 19.471$, df=2, $p < 0.001$, N=26), the affective benefits ($\chi2 = 27.184$, df=2, $p < 0.001$, N=26), the perceived social presence ($\chi2 = 13.241$, df=2, $p < 0.005$, N=26), and on a 5-point Likert scale for the perceived task performance ($\chi2 = 38.716$, df=2, $p < 0.001$, N=26). Pair-wise Wilcoxon signed ranks tests showed that overall participants perceived a significantly higher Sense of Togetherness ($p < 0.001$), Social Presence ($p < 0.005$), Affective Benefits ($p < 0.001$) and Perceived Task Performance ($p < 0.001$) in the visual condition than in the eyes-free conditions (see Table 1).

Performance

Pair-wise Wilcoxon signed ranks test of Perceived Task Performance showed (Fig. 3a) a significantly higher performance result for the visual condition ($p < 0.001$). In addition, for the eyes-free conditions, participants perceived they had performed significantly better in the Spatial Audio-tactile condition than in the Audio-tactile condition ($p = 0.036$ with Bonferroni correction). This is consistent with the recorded user performance (Fig. 3b), and although that difference is not significant, a more positive trend was observed for the spatial audio condition. Informal feedback from the free form questions also showed the same positive trend in user preference towards Spatial Audio-tactile as opposed to Audio-tactile (11:5).

User experience

Results from the user experience questionnaire show (Fig. 4) the potential in learning this novel interaction method, which was also visible in the performance data of few pairs. However, our short experiment did not provide significant evidence of a marked learning effect and so this data will not be presented here in detail. User experience results also

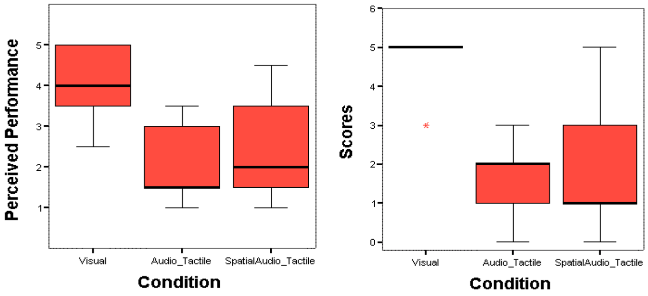

Figure 3. (a) Perceived performance, (b) Targets acquired.

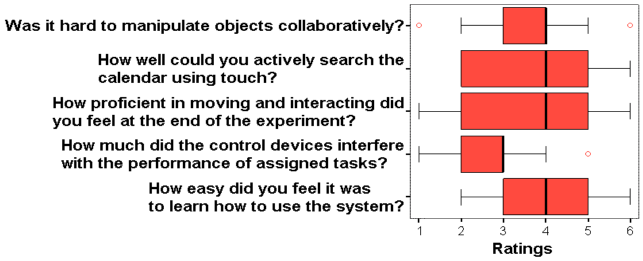

Figure 4. User experience questionnaire (6-point Likert scale).

showed some usability issues, particularly in the control of the input device, which was mainly due to its small size.

DISCUSSION

Results show that performance in the eyes-free systems dropped significantly compared to the visual one, which was expected, however Social Presence and Affective Benefits dropped less dramatically (see Table 1). This is interesting given the completely synthetic non-verbal audio and tactile displays. Related previous work is mostly based on live video and audio streams, which create a completely different experience in CVE. A similar study [10], using visual and tactile feedback, highlighted the challenges of eyes-free systems and suggested that predefining collaborative strategies could improve performance. However, the real potential could only be explored in longitudinal studies including more training. Prior to the experiment, pairs did not have the chance to agree verbally on a strategy, which made the task even harder. Our focus was on exploring the spontaneously emerging roles in the course of interaction, which are however not in the scope of this paper.

Addressing our research questions, we can state the following. Users were able to build shared awareness in the proposed mobile CVE. They did so to a higher degree in the visual than when eyes-free, which supports results of related previous work. Within the extreme eyes-free conditions spatial audio worked better overall. Efficiency in the visual system was however much higher, together with a strong learning effect. Eyes-free interaction increased cognitive load, decreased performance and in the spatial audio case only showed signs of a slower learning effect.

Technology and design related usability issues affected our prototype systems. Since most participants usually covered the capacitive surface with their thumb, they reported difficulties in controlling the discrete transitions at this specific resolution level. Spatial audio presentation was more beneficial at the two extremes – left and right – and less so in the middle. The range of pitch-tones in the audio version was considered too narrow and the mapping of high-low pitch to left-right direction not obvious and hard to learn. Further work is required to establish the efficiency potential of spatial audio in this type of eyes-free interface, however our results suggest certain advantages for mobile CVEs.

CONCLUSIONS

In this paper we presented an evaluation of a novel collaborative interaction method in a mobile shared environment, in visual and eyes-free conditions. We have shown that although an exact translation of a visual interface is not possible using other modalities such as audio or haptics, it is still possible to design equivalent interfaces enabling the interaction even when eyes-free. This is becoming increasingly important when building more robust mobile interfaces, which enable users to maintain interactions by shifting their attention between modalities when vision is disturbed. Thus, the results presented here are important for the community to better understand the potential of different feedback mechanisms in the creation of richer interpersonal communication systems.

REFERENCES

1. Baren, J., IJsselsteijn, W., Markopoulos, P., Romero, N., and Ruyter, B. Measuring Affective Benefits and Costs of Awareness Systems Supporting Intimate Social Networks. In: Nijholt, Nishida (Eds.), 2004. Social Intelligence Design, pp. 13-19.
2. Basdogan, C., Ho, C., Srinivasan, M., and Slater, M. An experimental study on the role of touch in shared virtual environments. ACM TOCHI 7/4 (2000), pp. 443-460.
3. Blauert, J. Spatial Hearing: Psychophysics of human sound localization. The MIT Press, (1999).
4. Clark, H., and Brennan, S. Grounding in communication. In: Resnick, Levine, Teasley (Eds.), 1991. Perspectives on socially shared cognition, pp. 127-149.
5. Crommentuijn, K., and Winberg, F. Designing auditory displays to facilitate object localization in virtual haptic 3D environments. In Proc. Assets '06. ACM Press (2006), pp. 255-256.
6. Dourish, P., and Bellotti, V. Awareness and coordination in shared workspaces. In Proc. CSCW '92, ACM Press, pp. 107-114.
7. Jordan, J., Mortensen, J., Oliveira, M., Slater, M., Tay, B., Kim J., and Srinivasan, M. Collaboration in a Mediated Haptic Environment. In Proc. PRESENCE 2002.
8. Sallnäs, E. Haptic Feedback Increases Perceived Social Presence. LNCS, 2010, Vol. 6192/2010, pp. 178-185.
9. Sallnäs, E. Effects of Communication Mode on Social Presence, Virtual Presence, and Performance in Collaborative Virtual Environments. Presence, 14/4 (2005), pp. 434-449.
10. Trendafilov, D., Lemmelä, S., and Murray-Smith, R. Negotiation models for mobile tactile interaction. In Proc. Mobile SSP '10. To appear in Springer LNCS.
11. Winberg, F., and Bowers, J. Assembling the senses: towards the design of cooperative interfaces for visually impaired users. In Proc. CSCW '04. ACM Press (2004).
12. Witmer, B., and Singer, M. Measuring presence in virtual environments: A presence questionnaire. Presence: Teleoperators and Virtual Environments, 7/3 (1998), pp. 225-240.