Structuring Shared-Collaborative Interaction Using Layered and Localised Auditory Feedback[1]

Victor Bayon

Ambiente Research Division, Fraunhofer-IPSI,
642942 Darmstadt, Germany
bayon@ipsi.fhg.de

Abstract. This paper reports 2 case studies where Auditory Feedback was used to reveal interaction in co-located collaborative settings. As users collaborated with a range of mobile devices sharing small and large displays, multiple foci of attention emerged. As users needed to be aware of each other in order to collaborate, sound emerging from different locations and devices was used to reveal and structure interaction around the physical space and help users to be more aware of the collaborative background activities.

1 Introduction

Auditory Interfaces (AFs) are a fundamental part of designing usable, multimodal and multi-sensory interfaces [4]. With the development and adoption of mobile devices such as Personal Digital Assistants (PDAs), Smart Phones (SPs) and other portable technologies (such as music players), sound is not only used to support visual interfaces for basic background un-attended notifications. There are an ever increasing number of applications that use auditory interfaces that go beyond using sound just as a background feature, such as audio-based experiences [2] or audio based haptic feedback [3].

The PDA and SP user experience can be enhanced with auditory interfaces in order to provide alternative and complimentary information, interaction and notification channels. As PDAs and SPs have a reduced set of input/output capabilities compared to a desktop computing environment and can be used in different contexts (i.e. on the move), it is necessary to think about designing interfaces that are not 100% visual-centric (like the desktop) and exploit multimodalities such as sound input and/or output [7].

However, the potential use of AFs has its limits. On the one hand, speech based interaction requires more mental resources and can be slower than visual based inter-

[1] This work has been carried out across the ESE i3 29310 "Kidstory" and the IST-2000-26089 "View of the Future" project. The author would like to thank to everyone that contributed to the work presented in this paper. This work has also partly been supported by the ERCIM Fellowship Contract nr 2003-20. The author would like to also thank the members of the Ambient Research Division at Fraunhofer-IPSI for their support.

action [10]. On the other hand, when users are surrounded by other people (i.e. open plan offices/ or public transport), sound coming from other people's devices can be annoying and distracting. In physical shared spaces, the use of sound should be restricted (i.e. do not use speech as input and use headphones for speech output) or suppressed by enabling alternative alerting channels such as vibration.

Besides these basic "common sense" and "politeness" guidelines for the use of AFs in public or semi-public co-located spaces, there is little work that reports on the use of auditory interfaces in shared multiuser spaces. Recent work based around shared public-displays (i.e see [5]) focuses mainly on exploring the visual aspects of the interaction (as happened previously with the desktop based interfaces).

For instance, in terms of sound output, Müller-Tomfelde designed custom sound feedback to enhance interaction and awareness of other peoples' interaction whilst using large and small public displays [8]; in terms of input (speech recognition), 2 users using speech and PDAs as an input had to carefully co-ordinate when they were talking to the "machine" and when they were talking to other each other [9].

In co-located collaborative settings where users have to work towards shared goals, the sound "ambience" result of the byproduct of interactions (i.e. moving in the physical spaces, pressing buttons, etc) provides foreground and background feedback, helping individual users to coordinate their own actions towards the shared goals [6]. In this paper, we present briefly 2 setups where AF was used to support and structure shared interaction.

As both setups provided multiple points of attention (i.e. the big screen, the small screens, the users talking, the devices, etc, see Fig 1 and 2), AF, combined with visual feedback was used to provide a sound aura of "what was happening" with the system, the space and other users, revealing the structure of the interaction.

In this case, the concept of sound used as a (private and personal) minimal attention interface is extended towards multi-user settings, where the minimal attention interfaces apply to all users and are shared. Although the areas of application were different, both set-ups shared several characteristics in common:

- The systems were designed to support from 2 to 4-5 users, each one with a different or similar sort of device.
- Both systems supported synchronous/asynchronous interaction, where different users could interact at the same time or in a sequential manner, or discrete/continuous, where interactions would be one-off or required some time to process.
- Both systems allowed overlapped interactions, where one, two or more tasks could be active at any single time.
- Multiple attention points. Users have to balance their own interactions and collaboration with others.

2 The Tangible Set-Up

With the Tangible Set-Up (TS), children could create and retell stories. The TS used KidPad, a Storytelling platform based on a zoomable canvas where contents could be dropped and organised into sequences. By means of linking and zooming

and multimodal interaction, children could create and manage narrative structures. Children could create content with a PDA and "drop it" into KidPad, drawings could be made with pen and paper and could be scanned in, pictures uploaded using a web-cam, sounds recorded or a picture book with all the content previously recorded could be printed out. For more information about Kidpad and TS see [1].

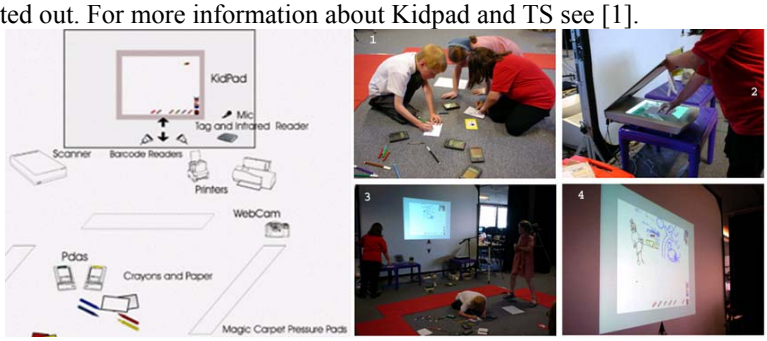

Fig. 1. The Tangible Set-Up

The TS had almost 20 devices in total (input and output, Fig.1 left side) and allowed any multi-modal interaction to be executed in any order, and children (for example) could drop the content of a PDA into KidPad, while the scanner was scanning and at the same time while another child was taking a picture with the webcam. With the tools available, children cold arrange the content into a story as it was captured. Each operation could last from 3-4 seconds of taking a picture with the web-cam, ~10 seconds to get the picture from the PDA (via the infrared transmission), 40-50 seconds of the scanning, to 1 min or more for printing out the stories (depending on the number of scenes).

Fig 1 (right side) illustrate a typical use of the TS. Top left first (Frame Num. 1), clockwise: Children discuss, plan and create content using PDAs and pen and paper; Then (F. Num. 2) a child walks to the scanner and starts scanning the content; while she waits for the scanner to finish, the other children carry on performing other tasks (Frame Num. 3); Once the scanning process is finished, the drawing pops up on the KidPad canvas and children could arrange the content into a story (Num. 4).

During the slack time that the system was busy processing, capturing or printing, children carried on interacting with the system and with each other in order to co-ordinate their actions (i.e. Fig 1, frame Num. 3). As the components were physically distributed, the main focal point of attention was the KidPad canvas with sub-focal points of attention around the other devices and the other children. The physical distribution of the interfaces (Fig 2) and the physical interactions of children "moving around" brought partial awareness of what was happening as a by-product of the interaction itself.

In order to be provided with more feedback in terms of when the devices were busy, the set-up implemented two layers of feedback, one layer local to each device, and another layer global to KidPad. The layers were organised in terms of physical location and distinctive sounds characteristics. It was very important that children could understand what the TS was doing and what the other children were doing in

order to interact efficiently with the TS by simply looking, or by simply "listening around" while they were engaged with content creation or other tasks.

Different feedback layers meant breaking up the physical space into interaction/attention areas and bringing awareness of when they were active to the users in other interactive areas.

Each device implemented a different and distinctive sound to indicate its status. For instance the scanner had a pair of speakers that produced a sound while it was busy scanning. As the scanner was located around one of the sides of the screen, the sound could be localized to that area. The other components such as the WebCamera or the PDA drawing upload implemented similar localised configurations with enough physical separation so that children could easily locate the sound and construct a mental model of the TS. KidPad drove an extra set of speakers localised above the display . While the devices were busy, icon based visual feedback (shown on KidPad) was displayed as well.

A group of 4 children familiar with the individual components but not with the configuration of the TS used it during 3 sessions with the task of creating and retelling stories. The level of collaboration and multi-tasking observed from the first to the third session increased. During the first session children relied more on talking/observing to each other in order to collaborate. By the third session, it was observed informally that children were coordinating their actions less on talking and more on observing and listening to the AF.

3 De-coupled Interaction Setup

In the De-coupled Interaction Setup (DIS), the task was to visualise and interact with an interactive 3D environment of a car and discuss among the participants about possible styling changes (such as the outside colour or the interior using a shared display). The 3D environment was highly interactive and supported a number of animations and modifications to the car (i.e. change textures/colours), so that users could interact and visualise those changes in real time and review and discuss the styling options. Large interactive 3D visualizations are often used for multidisciplinary team meetings and discussion (Fig 2, Frame Num. 1&2).

In the TS, the devices used to enter content into KidPad were localised around or near the main display. With the DIS, all the interaction occurred detached from the main display and users could use PDAs (Fig 2, F. Num. 3), desktop based interfaces (Fig 2, F. Num. 4) or 2D/3D input devices to interact with the environment.

The DIS supported a different range of operations (continuously navigating the environment or discretely changing the colour) that made each device more suitable for a specific task. For example, while a user moved the current viewpoint with the 2D/3D input device (a wireless gyroscopic mouse, Fig 2, Frame 1), another user could trigger via a PDA the "Open Doors" animation (Fig 2, F. Num 3) while another one could take "screenshots" of the car with a PDA, and yet another user could manage the desktop interface of the created media. These media materials could be used afterwards via an intranet as a visual archive of the meeting and to encourage further discussion.

Fig. 2. DIS.

The diversity of the interfaces and interaction techniques made it possible that each user could take a specific interaction role in order to perform the different tasks. The "work load" of interaction could be distributed and delegated among the participants.

Feedback was available as well via visual icons on the main display. For instance, when a user selected "Take Screenshot" from a PDA, an icon was shown on the display, and sound feedback was generated from the PDA. In the case of the "video recording" functionality, the PDAs produced a continuous sound while the recording was activated so that other users could be aware of the functionality being engaged and who triggered it.

Although the design approach of the sound feedback was similar to the TS (two layers, one global to the main display and another localised to the mobile devices), in this case the PDAs generated their own sound feedback. Each device had a unique and distinctive sound in order to facilitate device differentiation.

For example, the desktop user in Fig. 1 F. Num. 4, was not directly facing the main display (Fig1. F. Num 5). However, she could co-ordinate her actions thanks to the sound and visual feedback produced by the whole set-up.

4 Conclusions

This paper has briefly presented the approach to integrate and design user-scalable, localised and layered AF to multi-user and multi-device shared displays. We exploited minimal attention AF based interfaces and explored its use in multi-user settings.

When interacting, users took different interaction roles and created dynamically different attention zones within the space. Users needed to be aware of each other and each zone and needed to know the status of the system to effectively collaborate

AF was used to reveal interaction within the different zones and to notify the other users of what was happening. AF helped to create an audio "aura" that connected the different zones. The AF also helped to reveal the structure and flow of interactions for all users. With the use of the AF, users required less foreground attention towards the environment (visually) and could rely on background AF as well to be aware of the others.

During the iterative design, development and evaluation of the set-ups, we informally observed that the approach of introducing shared and layered AF channels facilitate collaboration.

References

1. Bayon, V., Wilson, J. and Stanton, D., Mixed Reality Storytelling Environments. Journal of Virtual Reality. Vol. 7 No.1 (2003)
2. Benford, S., Savannah: Designing a Location Based Game Simulating Lion Behaviour. Advances In Computer Entertainment Technology. Singapore. (2004)
3. Crossan, A., Williamson, J. and Murray-Smith, R., Haptic Granular Synthesis: Targeting, Visualisation and Texturing. Information Visualisation (IV'04) July 14 - 16, London, England (2004) 527-532
4. Gross, T. Universal Access to Groupware with Multimodal Interfaces. In Proc .of the second Int. Conf. on Universal Access in HCI: Inclusive Design in the Information Society (June 22-27, Crete, Greece). Lawrence Erlbaum, Hillsdale, NJ (2003)1108-1112
5. Izadi, S, Brignull, H., Rodden, T., Rogers, Y. and Underwood, M., Dynamo: A public interactive surface supporting the cooperative sharing and exchange of media. Proc. UIST03, Vancouver, ACM 2003 159-168.
6. Heath, C. and Luff, P., Technology in Action. Cambridge University Press. (2000)
7. Lumsden, J. and Brewster, S.A. A Paradigm Shift: Alternative Interaction Techniques for Use with Mobile & Wearable Devices. In Proc. of 13th IBM Centers for Advanced Studies Conference. Stewart, D.A Ed. (2003) 97 – 110.
8. Müller-Tomfelde, C., Streitz, N. A. and Steinmetz, R., Sounds@Work - Auditory Displays for Interaction in Cooperative and Hybrid Environments. In: C. Stephanidis, J. Jacko (Eds), HCI: Theory and Practice (Part II). Lawrence Erlbaum Publishers. Mahwah, N.J., (2003) 751-755
9. Stedmon, A., Griffiths, G. and Bayon, V., Single or Multi-User VEs, Manual or Speech Input? An Assessment of De-coupled Interaction in Virtual Environments. (2004) In Proc. of VR Design and Evaluation Workshop, 22-23, Nottingham UK (2004)
10. Shneiderman, B. (2000). The limits of speech recognition. Communications of the ACM. Volume 43 , Issue 9 (2000) 63-65