Tangoscope: A Tangible Audio Device for Tabletop Interaction

Jörg Edelmann[1], Yvonne Kammerer[1], Birgit Imhof[1],
Peter Gerjets[1], and Wolfgang Straßer[2]

[1] Knowledge Media Research Center,
Konrad-Adenauer-Straße 40, 72072 Tübingen, Germany
[2] University of Tübingen, Department of Graphical Interactive Systems (GRIS),
Sand 14, 72076 Tübingen, Germany
`{j.edelmann,y.kammerer,b.imhof,p.gerjets}@iwm-kmrc.de`

Abstract. Tabletop installations allow multiple users to playback digital media simultaneously. With public speakers, however, simultaneous auditory content gets superimposed, leading to a confusing and disturbing user experience. In this paper, we present *Tangoscope*, a tangible audio output device for tabletop interaction. As we augmented headphones with a visual marker-based identification mechanism, with the Tangoscope each user is provided with individual auditory content. To allow for an instruction-free and intuitive usage of the audio device we employed the metaphor of a real stethoscope. A first user study indicated the self-explaining use of the Tangoscope.

Keywords: Tabletop interaction, audio interface, individual audio content, natural user interface, multi-user interaction.

1 Introduction

Camera based Tabletop systems like "Microsoft Surface" or the "reactable" [1] allow users to interact with the surface directly by using their fingers and hands. Moreover, these systems enable users to interact with the interface through the manipulation of real objects, taking advantage of the natural physical affordances of these objects [2]. These systems are capable to process multiple simultaneous input and provide therefore a good foundation for multi-user applications [1, 3]. Since user interactions occur directly in screen space on the display, these systems are deemed to be easy-to-learn and intuitive. With large screen estates, several users can simultaneously explore multimedia contents (e.g., videos) displayed on the tabletop system. However, if users simultaneously explore different interaction objects for which audio is provided (e.g., auditory text or audio channels of videos), public audio can be very confusing and disturbing. This is because users are not only exposed to their requested auditory information, but also to those of other users. In order to avoid such interferences, thus, certain auxiliary mechanisms are required to convey private auditory content to individual users of a shared tabletop application [4].

2 Private Auditory Information in Shared Tabletop Applications

One approach to convey private auditory information to each individual user, is to provide separate shielded speakers mounted in front of each user [5, 6]. Although being successful applications for collaborative music generation systems [5] or for the provision of auditory feedback [6], several individual speakers conveying different auditory contents simultaneously might still be disturbing for the exploration of complex multimedia materials as, for example, in a museum setting. Furthermore, users cannot easily listen in to audio content of other users or move freely around the installation.

Morris et al. [4] introduced a system that uses earbuds in a single ear to convey private auditory information (i.e., mp3 songs) to specific users via individual sound channels. By using earbuds users are only exposed to the song they have currently requested. This system, however, employs a DiamondTouch [7] table that allows user identification. Each user is sitting on a custom chair, through which he/she is capacitively coupled to the table, thereby allowing the identification of touch events of a certain user. As a consequence, users are not able to move around freely during tabletop interaction, but have to stay seated on their assigned chair.

With common camera-based tabletop systems, user identification as provided by the DiamondTouch System is not possible, as they merely track the various contact points on the tabletop surface by objects such as fingers (e.g., Microsoft Surface) or visual tags. Therefore, in order to determine the receiver of certain audio content, such vision-based interactive tabletops must be augmented with additional identification mechanisms. This paper describes an approach to convey private auditory information to specific tabletop users for systems that cannot identify the corresponding user of each touch event. We propose a tangible audio device with a visual marker to assign audio requests to a connected headphone and thereby to the designated user. Moreover, in order to make the interaction with the tangible audio device as self-explaining and intuitive as possible, we tried to use a suitable metaphor for audio retrieval, that is, a stethoscope. We chose this metaphor, because the use and function of a stethoscope were assumed to be well-known to everyone, thus enabling users to reuse existing skills for interaction with the new device.

The following section describes our interactive multi-touch tabletop system that aims at providing users with private auditory information through the use of *Tangoscope* – an audio output device built from a real stethoscope (Fig. 1a left and middle). Furthermore, we present a user study which tests the intuitive understanding and use of the Tangoscope by comparing it to a technically equivalent "earbuds-tangible audio output device" which is not conveying the stethoscope metaphor (Fig. 1a right).

3 System Description

3.1 Tabletop Setup

We implemented a prototype of the audio device on a custom built standing-height tabletop setup with a display size of 1.2x1.35 m. The image is back-projected by two

tiled Full-HD projectors allowing shadow-free user-interaction at a resolution of 1920x2160 pixels. The capture of user-input in terms of touch and tangible interaction is provided by a vision-based approach where the surface is illuminated with infrared light which gets diffused by the projection layer. Objects and fingers placed upon the table are captured from below by an infrared sensitive camera with a resolution of 864x768 pixels at 120 Hz. With the large size and high resolution of the display and the simultaneous interaction possibilities, the system is well suited for multiple simultaneous users. To extract the touch and visual marker information the camera image is processed similar to [8] utilizing the graphics processing unit to provide low feedback latency. The tracking software is able to differentiate between touch input and visual markers as well as between different markers. Like common vision-based setups, the system, however, is not able to distinguish between different users by touch input.

Multiple different audio channels on a single computer require either specialized audio hardware or multiple regular sound cards. Our implementation utilizes four external USB sound cards with each of them being connected via cable to one audio output device. To create working prototypes of the Tangoscope that match the desired metaphor as close as possible we altered real stethoscopes by replacing the regular earplugs with in-ear headphones and attaching a unique visual marker to the membrane of the head (Fig. 1 b).
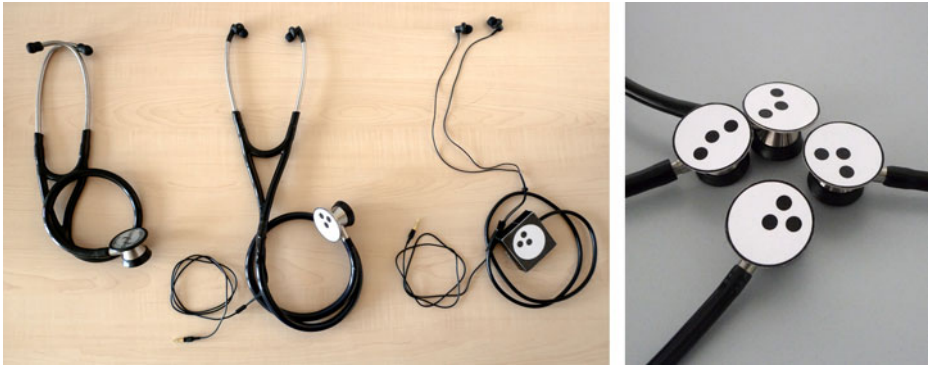

Fig. 1. **(a)** Original stethoscope (left), *Tangoscope* (middle) and earbuds with marker only (right); **(b)** Unique visual markers attached to stethoscope heads

3.2 Interaction Techniques for the Tangoscope

Audio in Scatter-View Applications. Scatter-view specifies a typical presentation mode of a collection of visual material on tabletop systems. The items are presented like cards lying on a table and can be transformed with multi-touch gestures. Though applications usually allow the simultaneous playback of multiple video items, the audio channels interfere with each other. By placing the stethoscope head upon an item, the application can now stream the desired audio channel to the respective headphones, while other videos can continue playing for visual comparison. Users can also switch audio channels by easily swapping the stethoscope head to another item.

The rotation of the head can be additionally used to control other audio parameters such as volume or playback speed.

Spatial Exploration of Virtual Audio Sources. Beside the possibility of starting and stopping single audio channels, the application can also utilize the position of the stethoscope head upon visual material to allow the exploration of spatially distributed audio material. When moving the head between two sources, the audio seamlessly blends from one source to the other. Combined with multi-touch gestures for transformation, the user can spread the virtual space to listen more isolated to a specific audio source or shrink the material to hear a mix of the different sources.

Multi-User Applications. The described applications easily scale with multiple simultaneous users. Each user is then presented his/her private audio channel depending on the position of his/her stethoscope head identified by his/her own unique visual marker (Fig. 2). Users can listen in to other users' content easily by placing the head of their stethoscope next to that of another user.

Fig. 2. Multiple users accessing individual auditory content

3.3   Application Scenarios

Even though ownership of each touch point is not provided by the Tangoscope, audio assignment can be accomplished with this device on systems that cannot identify users. In the following we want to outline some tabletop applications where the Tangoscope can add further interaction possibilities with audio content.

Public Installations: Public speakers are usually not applicable for tabletop systems in public spaces like museums or libraries due to noise disturbance. Ordinary headphones, however, require an additional user identification mechanism. By using

the Tangoscope, users can listen to individual audio content without disturbing other users or visitors. Furthermore, the Tangoscope allows for an instruction-free usage that is especially needed for interfaces and devices in public places.

Tabletop Games: Multi-player tabletop games like "Dungeons & Dragons" lack the possibility of private information for individual participants, as the whole screen is visible for all players. Here, the Tangoscope can be used to provide players with private information like mission briefings or status messages.

Group Tasks: For collaborative tasks on a tabletop system like [3] where co-located group members work in parallel, the Tangoscope can provide easy access to individual audio information without disturbing other users. Furthermore auditory content can be easily shared with arbitrary other group members who can listen in by placing the stethoscope head on the same content.

Collaborative Music Creation: In music creation applications like the reactable [1], the Tangoscope could provide monitoring functionality for the different participants. The device can be used further to listen to different mixtures by adjusting the volume of the instruments depending on the distance to the stethoscope head on the surface.

4   User Study

Stethoscopes are commonly known as medical devices which are usually used to monitor a person's heartbeat. The Tangoscope incorporates this metaphor to provide an intuitive approach to individual audio content for tabletop interaction. In a first user study we therefore addressed the question whether users are able to transfer without any further instructions the familiar functionality of a real stethoscope from a medical scenario to audio retrieval on a tabletop system. Since the benefit of individual audio channels in a multi-user scenario was already studied by [4] we focused on the value of the stethoscope metaphor. To create a technically comparable control condition we attached a box with a visual marker to in-ear headphones similar to the ones used for the Tangoscope (Fig. 1a). This device provides indeed the same interaction possibilities from a technical point of view, but as only difference to the Tangoscope omits the stethoscope metaphor.

4.1   Methods

Across two groups we contrasted the use of the Tangoscope against the earbuds-tangible device, resulting in a between-subjects experimental design. Twenty doctoral students and staff from a German research institute (average age: 28.9 years; 11 female) volunteered for this study and were randomly assigned to either the Tangoscope or the earbuds-tangible device condition (10 per condition). None of the participants had prior experience with the audio devices used in this study. Twelve participants had limited prior experiences with interactive tabletops.

Twenty music videos of the current MTV charts, were presented as still images on the tabletop surface (Fig. 2). Videos could be started by placing the visual marker upon one of the images. During video playback the corresponding audio was provided

through the headphones of the Tangoscope or the standard earbuds, respectively. If a user removed the tangible object from the surface, video and audio were stopped. Furthermore, users could move, rotate, and scale the video-items with multi-touch gestures. Time-stamped records of all user interactions with the interactive tabletop were logged by the application. Furthermore, a video camera was used to record participants during task processing.

Participants' task was to explore the music videos on the tabletop and to select their personal top 5 videos. In order to analyze their spontaneous exploration of the system, they were neither explained how the videos could be played, nor were they told about the existence and usage of the audio device. Tabletop interaction time was limited to five minutes per participant. The experiment was conducted in a laboratory setting, with participants being tested in individual sessions. After having received task instructions, participants were brought to the laboratory. There they encountered the interactive tabletop with 20 images of the music videos distributed across the tabletop surface. Either the Tangoscope or the earbuds-tangible device was lying on the right-hand side of the tabletop. The experimenter, who stayed in the back of the room during task processing, did not give any comments or advice, with one exception: if participants did not consider the audio device at all in the first two and a half minutes, the experimenter told them that they could use the device. After five minutes participants' interaction with the tabletop was interrupted.

To examine participants' interaction behavior we measured the events related to the audio device, that is, the number of times video/audio was played, time to first playback, and the total duration of video/audio playback. In addition, video data allowed analyzing the time until the audio device was picked up by the participants and the time taken from pickup to the first playback. Finally, participants were asked to rate the ease-of-use with the audio device on a 5-point scale (5 = highly agree).

4.2   Results

To compare participants' interaction behaviors and subjective ratings between the Tangoscope condition and the earbuds-tangible device condition 2-tailed t-tests were conducted. Table 1 shows the mean values of the measures for both conditions. T-tests revealed several significant differences with regard to audio usage in the two conditions. During the five minutes of exploration, Tangoscope users played video/audio significantly more often than users in the earbuds-tangible device condition (earbuds users), $t(18) = -2.23$, $p = .04$. Moreover, time to first playback was significantly shorter for Tangoscope users than for earbuds users, $t(18) = 2.08$, $p = .05$. In contrast, users did not differ with regard to the total duration of video/audio playback, $t(18) = -1.48$, ns. Furthermore, video data showed no differences between conditions regarding the time needed until the audio device was picked up by the users, $t(18) = -1.33$, ns. Four Tangoscope users and six earbuds users received the hint that they were allowed to use the audio device, as they did not consider the audio device at all in the first two and a half minutes of exploration, $\chi^2(1) = 0.80$, ns. Once having picked up the audio device, however, it took Tangoscope users significantly shorter to play video/audio than earbuds users, $t(10.34) = 2.76$, $p = .02$. Finally, users in both condition considerably agreed that the audio device was easy to use ($t(18) = -1.59$, ns).

Table 1. Mean values of the measures for both conditions

	Tangoscope	**Ear-buds**
# audio playbacks	37.70	24.90
time to first playback (in s)	115.24	176.01
total duration of audio playback (in s)	99.89	73.15
time to first pickup (in s)	104.70	71.20
time to first playback after pickup (in s)	24.60	104.60
ease of use (1-5)	4.50	3.60

5   Conclusion and Future Work

Results of the study indicate the natural and intuitive use of the Tangoscope as an audio output device for tabletop interaction. As compared to earbuds users, Tangoscope users needed less time to the first playback and played video/audio significantly more often. Thus, even without any instructions on how to use the audio device, the Tangoscope unfolded its advantages in comparison to the technically fully equivalent earbuds-tangible device. In both conditions quite a few users, however, completely ignored the audio device until the experimenter told them that they were allowed to use the device. Yet, at least after this hint was given all Tangoscope users understood how to use the device almost immediately, whereas three earbuds users could not find out how to use the device until exploration time was over, not even after the experimenter additionally gave them a second advice to use the "box". Three earbuds users tried to use the "box" as a remote control pressing on the dots of the visual marker (Fig. 1) showing that users usually search for meaningful functions. Particularly, the short time Tangoscope users needed from pickup to the first playback mirrors that users are well able to map the meaning and usage of a stethoscope as a medical device to the scenario of tabletop interaction with digital media.

Since this first user study solely addressed the question whether the metaphor is suitable for an audio device for tabletop interaction, more effort needs to be put in comparing the Tangoscope to other technical solutions like directed speakers in combination with tangible objects or restricted workspaces. For future work we are also planning to explore the Tangoscope device in multi-user scenarios. We assume that in such scenarios users can easily switch between conversation and listening by hanging the device around their neck. Another interesting research question is whether the device could help visually impaired people to access information on tabletop installations by providing auditory feedback of the underlying digital content.

Our current implementation of the Tangoscope needs to be connected via cable to the tabletop system for audio transmission. To make the device more flexible for large tabletop systems and multiple simultaneous users, wireless operations of the headphones of the Tangoscope would be helpful. This might be accomplished by using Bluetooth-headphones which are coupled to the tabletop computer. However, we are not aware of an available Bluetooth stack which allows for multiple

simultaneous audio output devices. Another approach to provide wireless operation could be realized by plugging the headphones of each Tangoscope into a mobile device which receives audio playback commands via Wi-Fi from the tabletop system.

The Tangoscope audio device provides a feasible solution to the problem of superimposed audio output for multi-user tabletop applications. By utilizing the metaphor of a stethoscope it allows for instruction-free usage and we believe it is therefore well suited for tabletop installations particularly in public space.

References

1. Jordà, S., Geiger, G., Alonso, M., Kaltenbrunner, M.: The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In: Proc. TEI 2007, pp. 139–146. ACM Press, New York (2007)
2. Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proc. SIGCHI 1997, pp. 234–241. ACM Press, New York (1997)
3. Morris, M.R., Fisher, D., and Wigdor, D. Search on Surfaces: Exploring the Potential of Interactive Tabletops for Collaborative Search Tasks. Information Processing and Management (Spring 2010) (in press)
4. Morris, M.R., Morris, D., Winograd, T.: Individual audio channels with single display groupware: effects on communication and task strategy. In: Proc. CSCW 2004, pp. 242–251. ACM Press, New York (2004)
5. Blaine, T., Perkis, T.: The Jam-O-Drum Interactive Music System: A Study in Interaction Design. In: Proc. DIS 2000, pp. 165–173. ACM Press, New York (2000)
6. Hancock, M.S., Shen, C., Forlines, C., Ryall, K.: Exploring non-speech auditory feedback at an interactive multi-user tabletop. In: Proc. GI 2005, pp. 41-50, CHCCS (2005)
7. Dietz, P., Leigh, D.: DiamondTouch: A Multi-User Touch Technology. In: Proc. UIST 2001, pp. 219–226. ACM Press, New York (2001)
8. Edelmann, J., Fleck, S., Schilling, A.: The DabR – A multitouch system for intuitive 3D scene navigation. In: 3DTV CON - The True Vision, pp. 1-4 (2009)