Touching the Void: Gestures for Auditory Interfaces

Katrin Wolf
T-Labs (Quality & Usability Lab), TU Berlin
E.-Reuther-Platz 7, 10587 Berlin
katrin.wolf@acm.org

Christina Dicke
HIT Lab NZ
University of Canterbury
Private Bag 4800, Christchurch
christina.dicke@hitlabnz.org

Raphael Grasset
HIT Lab NZ
University of Canterbury
Private Bag 4800, Christchurch
raphael.grasset@hitlabnz.org

ABSTRACT
Nowadays, mobile devices provide new possibilities for gesture interaction due to the large range of embedded sensors they have and their physical form factor. In addition, auditory interfaces can now be more easily supported through advanced mobile computing capabilities. Although different types of gesture techniques have been proposed for handheld devices, there is still little knowledge about the acceptability and use of some of these techniques, especially in the context of an auditory interface. In this paper, we propose a novel approach to the problem by studying the design space of gestures proposed by end-users for a mobile auditory interface. We discuss the results of this explorative study, in terms of the scope of the gestures proposed, the tangible aspects, and the users' preferences. This study delivers some initial gestures recommendations for eyes-free auditory interfaces.

Author Keywords
Gestures, tangible interface, embodied interaction, eyes-free, mobile, auditory display, participatory design.

ACM Classification Keywords
H.5.2 [User Interfaces]: *Haptic I/O*

General Terms
Design, Experimentation, Human Factors

INTRODUCTION
Current smartphones often come equipped with a wide range of advanced hardware and software features including those which can offer spatial sound rendering and even provide access to standard 3D sound APIs like OpenAL [5]. Beyond their potential usage for gaming or musical applications, these features can be exploited for the design of new user interfaces on mobile phones like spatial auditory interfaces. This paper is focused on mobile spatial auditory interfaces, specifically input control and interaction strategies that can be used in a mobile audio context.

TEI'11, January 22–26, 2011, Funchal, Portugal.

Lately, the manipulation of information on mobile devices has moved from keypad interaction to touch input. However, one of the drawbacks of this approach is that it requires a large part of the users' visual, cognitive, and motor attention, and could be harmful in some specific mobile situations (e.g. steering a car or walking in a busy urban environment). Speech recognition is alternative approach, but it is difficult to use in mobile situations due to the high signal-to-noise ratio caused by traffic, other conversations, and constantly changing environmental sound fields and levels.

Figure 1. Gesturing to interact with an auditory interface.

Gesture interaction techniques [3] exploiting inbuilt sensors (e.g. accelerometer, gyroscope, and/or digital compass) can overcome the previously described issues by providing an elegant, eyes-free solution for user input.

The large range of potential gestures that can be executed in a mobile context and the different input factors (DOF, user dexterity, mobile form factor, etc) need to be studied more thoroughly. The explorative study presented in this paper aims at shedding light on the type of gestures users will perform freely and intuitively when interacting with a spatial auditory interface.

For this purpose we asked participants in a qualitative user study to perform several tasks. Their actions and comments were recorded, analyzed, and are summarized in the results section of this paper. Discussions of the results, and suggestions and guidelines for gesture design are given in the discussion and conclusion sections.

RELATED WORK
Over the last few years we have seen an emergence of work on gesture techniques for spatial interaction design. One of the earliest mobile auditory interfaces presented was the

Nomadic Radio introduced by Sawhney et al. [8]. Gestures for mobile auditory interfaces have been explored in different projects and the validity of the approach has been shown in user studies like Pirhonen et al. [6]. In terms of techniques, Brewster et al. [2] presented one of the first spatialized audio systems combined with gestures. The Shoogle system [9] engaged the user to query information by shaking the device and delivered information by vibrotactile and sonified feedback. Similarly, Li et al. [4] presented a set of eyes-free gestures using a mobile keypad for an (non-spatial) auditory interface after surveying the most relevant tasks for end-users.

A recent study by Rico et al. [7] has observed the general social acceptability of some gesture techniques and demonstrated their impact in real use of some of the proposed metaphors. Similarly, Bhandari et al. [1] asked participants to match specific gestures to common tasks realized on mobile devices. Differently from our approach, the gestures were predefined by the authors and not designed by the participants.

In summary, gesture techniques have been well developed and explored for mobile devices. Yet there are only few studies exploring gestures as an input technique for spatial auditory interfaces, and none of these studies actively involves end-users in a participatory design process.

EXPERIMENT

Our motivation was to investigate how users would interact with basic elements of a spatial auditory interface without restricting them by pre-defined gestures or limited system capabilities. We were especially interested in which concepts and metaphors users would transfer from their everyday usage of computers and mobile phones to the entirely unknown domain of spatial auditory interfaces.

We solicited 10 participants, 4 male and 6 female, with a mean age of 30 (12-49) and a wide variety of professional backgrounds. All participants were familiar with desktop computers and were using them at least once a week for communication, accounting, or gaming. All participants owned a mobile phone but just one owned a smartphone.

Before the experiment began participants were familiarized with synthesized spatial sound by playback via headphones of sound scenes consisting of single and multiple sound sources. Finally, the participants were given a featureless phone dummy made of wood to perform the gesture they would envisage for each of the different task (see Fig 2.).

Users were allowed to perform any gestures with or on the device. The experiment was video recorded and we encouraged participants to verbalize their thoughts during each task (think-aloud method). We also interviewed the participants at the end of the session.

Figure 2. A gesture for moving a sound source (Task 14).

Tasks

The tasks for the user study were based on the concept of a traditional desktop interface (WIMP), using spatial sound to represent applications, menus, folders, and hierarchical file structures. Each of the tasks was explained symbolically and textually to the participant. A total of 20 tasks were structured in three main categories:

Item Selection

Task 1: Select a single sound source.
Task 2: Select a sound source from a list.
Task 3: Skip through sound sources in a list.
Task 4: Deselect a selected sound source.
Task 5: Select several disjointed items from a list.
Task 6: Select several contiguous items from a list.
Task 7: Select all items of a list.

Attention Prioritization

Task 8: Change the distance of a sound source.
Task 9: Maximize/focus attention on one sound source.
Task 10: Undo maximization of one specific sound source.
Task 11: Minimize one specific sound source.
Task 12: Undo minimization of several sound sources.
Task 13: Minimize all sound sources.

Item Manipulation

Task 14: Move a single sound source.
Task 15: Lock a single sound source.
Task 16: Unlock a single sound source.
Task 17: Pause a single sound source.
Task 18: Re-activate a paused sound source.
Task 19: Delete a single sound source.
Task 20: Activate/open a single sound source.

RESULTS

Participants used a total of 254 gestures, 98 of which were 3D movements with the device, 137 gestures were performed on the "touch screen", and 19 were combinations of both 3D and 2D gestures. An overview of the most frequently used gestures is presented in table 1. Essential gestures are illustrated and described in detail along with user comments and some notes on from which domains users transferred gestures to solve the tasks.

Touch-screen and embodied gestures

The pointing gesture (**Point**) was one of the most elemental 3D gestures to select an item (see Fig. 3, left). This gesture often preceded other gestures such as **TiltUp** (see Fig. 3, right) and **TiltDown**, **DoubleTouch**, **Arc**, etc.

Task	2D Gestures	3D Gestures
1	Scanning the screen with a finger + Hold for selection of a "touched" item (1)	Point + Touch (9)
2	Drag&Drop sidewards + Press (3) Drag&Drop sidwards + ScrollDown (3)	Point + TiltUp (2) Point + TiltDown (1) Point + TouchBelt (fig. 3) (1)
3	Drag&Drop sidwards (4) ScrollUp+ ScrollDown (2)	Shake left/right (3) PageFlip (1)
4	ScrollUp (3) Touch (2) Press (1)	Shake left/right (2) Flick (1)
5	n* Drag&Drop sidewards + ScrollDown (5) n* Drag&Drop sidewards + Press (1)	n* {Point + TiltUp}(2) n*{Point+TiltDown}(1) n*{Point + TouchBelly}(1)
6	Hold + n* {Drag&Drop + ScrollDown} (Fig. 4) (6) Touch + n* {Drag&Drop + Press&Hold} (2)	PageFlip (1) TiltUp + PageFlip + TiltDown (1)
7	DoubleTouch (3) Press&Hold (1) Draw a circle (O) (1) Drag&Drop+ScrollDown(1)	Draw a circle (O) (3) PointAll + TouchBelly (1)
8	ScrollUp + ScrollDown(6)	TiltUp + TiltDown (4)
9	DoubleTouch (4) Press&Hold (1)	Device to landscape format (3) Shake (1) Arc towards user (2)
10	DoubleTouch (4) Press&Hold (1)	Shake (4) Arc away from user (2)
11	Multitouch zoom out (1) Draw a dash (\) (1) Flip away (1)	Push away (2) Move downwards (2) Cross out (1)
12	DoubleTouch (2) ScrollDown (1)	Pull close (2) Move upwards (2) Shake (1)
13	Draw a circle (O) + ScrollDown (2) Draw a circle (O) + Draw a dash (\) (2)	Draw a circle (O) + TiltDown (2) PointUp + TiltDown (1)
14	Drag&Drop (3)	PickUp&Drop (8)
15	Press&Hold (3) Draw a circle (O) clockwise (2)	Lock (turn a key in a door) (2) TiltUp (2) DoubleTiltUp (1)
16	Press&Hold (3) Draw a circle (O) counterclockwise (2)	Unlock (turn a key in a door) (2) TiltUp (2) DoubleTiltUp (1)
17	Touch (4) DoubleTouch (2) Draw a dash (\) (2)	TiltDown (1) Move away (1)
18	Touch (4) DoubleTouch (2) Draw a dash (/) (2)	TiltUP (1) MoveCloser (1)
19	Cross out (X) (2) Flip away (1)	Cross out (X) (3) Flick (2) Shoot (1)
20	DoubleTouch (4) Touch (1)	Shake (2) TiltUp (2) TiltDown (1)

Table 1: Most frequently used 2D and 3D gestures by task. (Frequency is indicated by the number in brackets.)

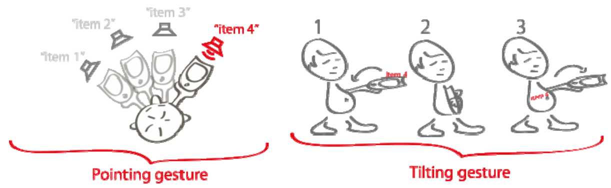

Figure 3. Point (left) and TiltUp + Move (right) gestures.

Some gestures were similar to gestures already implemented on the iPhone or Android (**ScrollUp**, **ScrollDown**, **Drag&Drop**). More original gestures proposed by the participants were:

- **Arc**: The device is moved from point A to B in an arc shaped curve
- **Flick**: A flicking hand movement - like throwing a Frisbee disc
- **PageFlip**: Rotating the device around its vertical axis
- **Shoot**: pretend to shoot with the device

Disjointed item selections (like task 5) were usually realized by repeating the single-selection gestures from task 1 and 2. Most contiguous item selections (like task 6) were split into three different actions: Select 1st item + Browsing the list + Selecting last item using 2D as well as 3D or combinations of gestures. Regarding reversible commands (like undo/redo); participants preferred repeating the original commands (9 and 11) or inversing the gesture (e.g. "**Arc** towards user" and "**Arc** away from user").

Combinations of 2D and 3D gestures

Combined gestures were mostly found in tasks involving 2D object selection (**Hold**) combined with spatial object manipulations such as moving a sound source by 3D **PickUp&Drop**. Fig. 4 shows an example of a combination of a 3D gesture (**Point**) followed by a 2D gesture (**Drag&Drop**).

Figure 4: Selecting items with a combined gesture (Task 6).

Another example of a combined gesture is illustrated in Fig. 5: **TiltUp** (to select all sources) followed by the 2D **ScrollDown** gesture to mute all sound sources.

Gesture associations

We encouraged participants to think aloud during the experiment. Some of these comments offered insights into the participants' associations and intentions:

Windows 7: Shake gesture to quickly minimize every open window except the one shaken: participant shook the

device to focus attention on a sound source.
☟ **Windows**: X-Button to close an application: participants used the X-shape to delete a sound source.
☟ **iPhone**: Slide finger across screen to scroll between different screens: participants used this to skip through sound sources in a list.
☟ **Books**: Flipping pages: participant used this to skip through sound sources in a list.

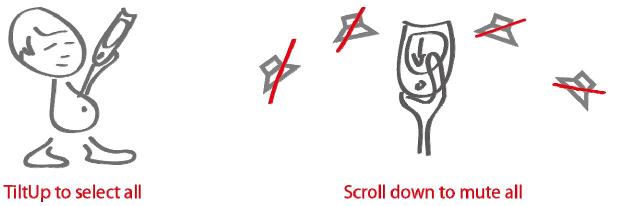

Figure 5: Combined gesture to mute/minimize all sound sources (Task 13).

DISCUSSION

We distinguished three main categories of gestures used by participants: 2D gestures, 3D gestures, and combinations of both. With rising task complexity we noticed a preference of 2D and 3D gesture combinations. Participants often used 2D gestures (e.g. **Hold**) to address an object, and proceeded with a 3D gesture (**Tilt**, **Shake**, **Move**, etc.) to manipulate this object. Half of the users clearly preferred 2D gestures, few people were using solely 3D gestures, and almost all participants combined gestures for certain tasks.

From the think-aloud analysis we learned that feedback for successful (and unsuccessful) gestures is essential. Also participants wanted an overview of available commands executable on an object or container. The gestures were mostly created through associations with interaction techniques known from other devices and tool. Experiences from the real world were mainly used in case the person did not have an association with any technical context (e.g. "lock" a sound source). Although only one user owned a smartphone, several participants slid their fingers over the imaginary screen to skip through sound sources in a list or scroll up and down (iPhone or Android touchscreen gesture). Other noticeable analogies were based on the users' current usage of traditional desktop GUIs (e.g. x-icon to delete a sound source or the mouse shake to focus on a source). The **Touch**, **DoubleTouch**, and **Point** gestures show strong resemblance to using a remote control or computer mouse (DoubleClick = **DoubleTouch**). We also perceived that users favored reusing a set of basic gestures for different tasks (differentiate by a specific context) or using 'inverse' gestures over having a wide range of unique gestures. An alternative strategy to keep the gesture set small and simple is to use context menus, although this may slow down interaction and require more cognitive attention.

We also observed that participants generally preferred performing discreet gestures instead of very expressive gestures. Compared to the results gained in [1], where a preference of 2D gestures for public scenarios is advised, we identified the same tendency. In addition our participants also favored minimalistic 3D gestures.

CONCLUSION AND FUTURE WORK

Our study provides an overview of gestures users would perform to interact with basic elements of a spatial auditory interface. Users chose gestures based on pre-existing knowledge and the ability to translate experiences from other domains to the domain at hand. When the aim is to design an intuitive and user-centered interface for an auditory interface, we recommend using a small, context related gesture set, supporting gesture inversions for do-undo-commands and supporting gestural analogies from other domains. Clear and distinct feedback to actions and information about available commands is crucial to prevent frustration and to guarantee a successful interaction. Future work is required on good feedback strategies and how to create an easy to learn and memorizable gesture set.

REFERENCES

1. Bhandari, S. and Lim, Y. Exploring gestural mode of interaction with mobile phones. In *Proc. CHI 2008*, ACM Press (2008), 2979-2984.
2. Brewster, S., Lumsden, J., Bell, M., Hall, M., and Tasker, S. Multimodal 'eyes-free' interaction techniques for wearable devices. In *Proc. CHI 2003*, ACM Press (2003), 473-480.
3. Karam, M. and Schraefel, M. C. *A Taxonomy of Gestures in Human Computer Interactions*. Technical Report ECSTR-IAM05-009, Electronics and Computer Science, University of Southampton, 2005.
4. Li, K. A., Baudisch, P., and Hinckley, K. Blindsight: eyes-free access to mobile phones. In *Proc. CHI 2008*, ACM Press (2008), 1389-1398.
5. OpenAL API. www.openal.org.
6. Pirhonen, A., Brewster, S., and Holguin, C. Gestural and audio metaphors as a means of control for mobile devices. In *Proc. CHI 2002*, ACM Press (2002), 291-298.
7. Rico, J. and Brewster, S. 2010. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proc. CHI 2010*, ACM Press (2010), 887-896.
8. Sawhney, N. and Schmandt, C. Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. In *Proc. ACM Transactions on Comput.-Hum. Interact. 2000*, ACM Press (2000), 353-383.
9. Williamson, J., Murray-Smith, R., and Hughes, S. Shoogle: excitatory multimodal interaction on mobile devices. In *Proc. CHI 2007*, ACM Press (2007), 121-124.