Making Tabletop Interaction Accessible for Blind Users

Andreas Kunz

ETH Zurich

Leonhardstrasse 21

8092 Zurich, Switzerland

kunz@iwf.mavt.ethz.ch

Dirk Schnelle-Walka

TU Darmstadt

Hochschulstraβe 10

64289 Darmstadt, Germany

dirk@tk.informatik.tu-darmstadt.de

Ali Alavi

ETH Zurich

Leonhardstrasse 21

8092 Zurich, Switzerland

alavi@iwf.mavt.ethz.ch

Stephan Pölzer

Johannes Kepler-University

Altenbergerstrasse 69

4040 Linz, Austria

stephan.poelzer@jku.at

Max Mühlhäuser

TU Darmstadt

Hochschulstraβe 10

64289 Darmstadt, Germany

max@informatik.tu-darmstadt.de

Klaus Miesenberger

Johannes Kepler-University

Altenbergerstrasse 69

4040 Linz, Austria

klaus.miesenberger@jku.at

Abstract

Tabletop systems and their interaction capabilities are typically a domain for sighted people only. While the content on the tabletop can already be made accessible to blind people, the interaction above the tabletop is still inaccessible. This paper describes our approach towards making the above tabletop interaction accessible to blind people by using LEAP sensors and speech recognition.

Author Keywords

Tabletop interaction; Deictic gestures; Accessibility

ACM Classification Keywords

H.5.2. Group and Organization Interfaces: Computer-supported cooperative work.

Introduction

In the recent years, interactive tabletop systems gain significant importance, since new technologies offer low-latency multi-user interaction. The latter made such systems suitable to be used in smaller team meetings for problem solving and idea generation. For such an idea generation, mind maps are one of the established ways to generate, visualize, structure and classify ideas [18] on a common workspace. Other scenarios for using interactive surfaces include teaching and learning [9]. However, they rely on visual
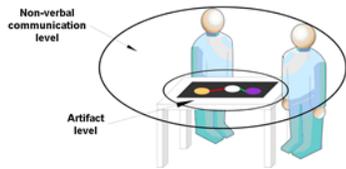
Figure 1 Brainstorming integrates non-verbal communication and artifact level elements [6], [7].

Figure 2 Sketch of the tracking above the PixelSense.

communication and visual perception of the artifacts [1], [19], excluding blind or partially sighted people or hindering their full participation in the meeting.

The use of this visual augmented communication in such settings is described by Ellis et al. [5], stating that explanatory gestures are used above the table on the non-verbal communication level (see Figure 1). Within this group of nonverbal communication elements, pointing gestures – and more specific so-called deictic gestures – are important. Deictic gestures are co-speech pointing gestures [4], which highlight a certain artifacts on the common workspace. Thus, as stated by Schnelle-Walka et al. [17], an inclusive mind map must support (i) access for the blind participants to the artifacts being displayed on the tabletop, and (ii) access to the important non-verbal communication elements.

While many digital content and interactions on the tabletop (the artifact level) can already be accessed by blind people using standard screen readers together with a Braille display and/or speech output [15], non-verbal communication elements above the tabletop are almost impossible to access, if they are not reflected by the artifacts on the tabletop. Suitable means to transfer e.g. deictic gestures in the nonverbal communication level to blind users is still an unsolved issue.

Within this paper, we describe our ongoing efforts to capture these non-verbal communication elements above the tabletop and how to reliably detect deictic gestures. Here, we introduce a first prototype for tracking deictic gestures above the tabletop and to display them together with the content on the tabletop to blind users.
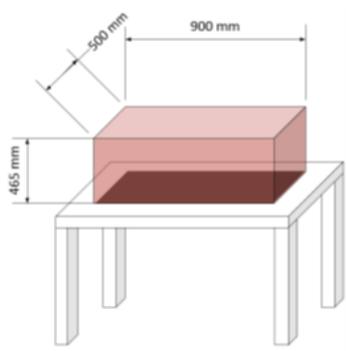
Design Considerations

Our work is based on the evaluation results of a focus group with blind participants as described in [17]. This focus group was in favor of detecting deictic gestures which are known and considered to be the most relevant form of nonverbal communication elements. However, frequent use of such gestures may result in an information overload if these gestures are presented without any filtering. The participants also saw the danger of interpretation errors, distraction and information overflow [9].

Detecting Pointing Gestures

Within a preliminary study on a predefined mind map, we used a manual transcription of a videotaped session to determine the type and amount of gestures, distinguishing "specific", "area" and "coarse" gestures (see Figure 3).
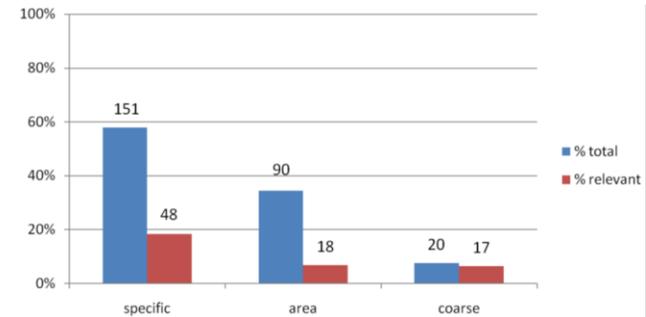

Figure 3 Comparison of relevant and irrelevant pointing gestures.

As it can be seen from Figure 3, the detection of deictic gestures becomes a delicate task, since only specific

deictic gestures can be considered to be relevant for a current task.

In a first approach, we placed the MS Kinect about 820 mm above the PixelSense [11], facing downwards onto the tabletop. Since MS Kinect and MS PixelSense work on the same infrared wavelength of λ = 830 nm, their interference triggered random touches on the PixelSense.

Figure 4 Prototypical setup.

As PixelSense [11] together with MS Kinect were not capable of tracking deictic gestures above the table top, we used LEAP Motion sensors [7] to track such gestures (see Figure 5). However, the use of LEAP Motion sensors has two major drawbacks to work around. First, their limited tracking range of about 500 mm requires to equip all sides of the table (with exception of the blind user's position) with such a sensor (see Figure 4). Second, it requires the use of a dedicated

computer for each LEAP. As an advantage, the observed tracking accuracy is significantly higher than alternative sensors, such as Microsoft Kinect. Given LEAP's inherent depth accuracy of +/- 0.1 mm, a finger pointing angle of 45° and a height above the tabletop of 450 mm result in a pointing accuracy of +/- 0.35% (see Figure 5).
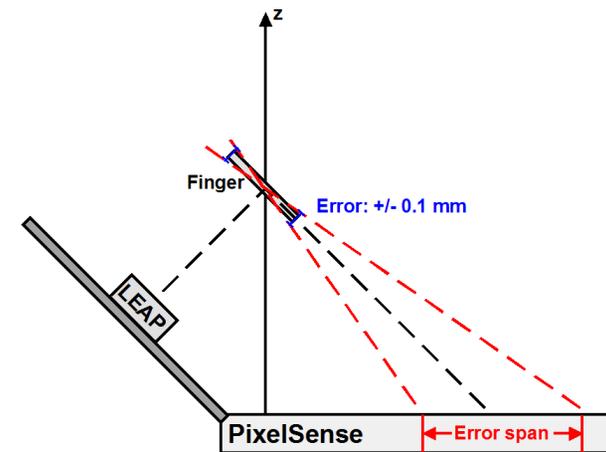

Figure 5 LEAP's measurement setup and position accuracy.

With this setup, it is possible to reliably track pointing gestures to targets on the tabletop, depending on the user's pointing accuracy. To overcome this problem, we show a visual feedback on the screen similar to the approach described by the Majewski et al. for smart homes [9], which encourages the user to align his or her pointing direction with the wanted target.

Figure 6 Kinect's microphone array is used for speech recognition.

Speech Recognition

Although pointing gestures can now be tracked with sufficient accuracy, we need to cope with relevance aspects as mentioned above, meaning to differentiate simple pointing gestures and deictic gestures. In order to avoid too many or false notifications to the blind user, it is important to filter deictic gestures only. Following Eisenstein et al. [4], deictic gestures can be understood as co-speech pointing. In a first approach we implemented word spotting for words like "this", "that", "there", etc.

In our system, we employ the Microsoft Speech Server [10] as speech recognizer. The audio is recorded by the Microsoft Kinect that is suited for distant speech recognition thanks to its microphone array, which results in a better audio selectivity in a noisy environment. We place the Kinect on a tripod at one corner of the Pixelsense directed to the opposite corner to avoid interference with the PixelSense and to exploit the capabilities of the microphone array best (Figure 6). By fusing the sensor signals from the speech recognition system and the tracking system, an efficient but simple filtering of deictic gestures is implemented, employing the melting-pot metaphor [12]. Notifications are only given to the blind user if both sensor signals appear with the same time span. Here, we follow the ideas of Nigay and Coutaz to fuse combined use of modalities [12] in the parallel case. We regard information coming from the two different modalities to be in parallel if they appear with a predefined time span. In order to define the duration of this time span, we re-used the transcription of our initial user study. Assuming that a pointing gesture takes longer than speaking to augment an utterance,

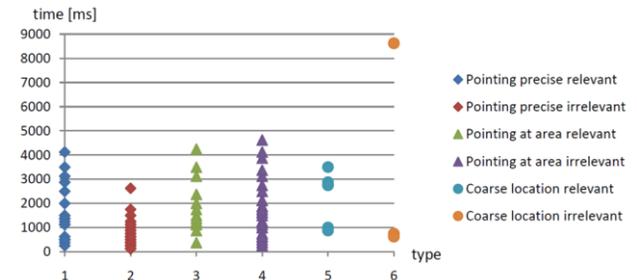we measured the duration of typical gestures (see Figure 7).

Figure 7 Duration of six different gesture types.

These different types of gestures were measured by manually transcribing a videotaped brainstorming session, in which typically a large amount of gestures occur. However, not all of them are considered to be relevant, since they were made by users without being integrated into the verbal communication flow. Moreover, gestures could be further distinguished by their precision, since users highlighted specific targets (elements of the mind map), or just regions on the table (clusters of the mind map). Also very unspecific and coarse gestures occur, referring to the complete interactive surface (the complete mind map).

For defining the duration of a pointing gesture, the movement time towards the target was not considered, but only the time the finger hovered above the target. Based on these measurements, we set the maximum duration of the observation interval to be t = 3.5 s. Only characteristic audio signals (words) within this time span will be further processed in order to indicate deictic gestures.

Presentation to Blind Users

Another crucial role plays the presentation technique for the blind user to make the ongoing interaction above the table-top more accessible. As stated above [9], special care has to be taken not to overload a blind user, so that he/she can still follow an ongoing discussion. Besides the alert message boxes, as they are used e.g. in [1], the blind user has now the possibility to get informed via text-to-speech messages about an occurring pointing gesture. A short beep notifies the blind user about a pointing gesture.

The text of nodes in a mind map tends to be short. An explanatory analysis by Beel and Langer [1] showed that 60% of the nodes have texts with three words at maximum. Currently, we are about to investigate if a rendering of the node contents via text to speech employing a mono-aural headset is suitable to lower distraction while still being able to follow the discussion. The early feedback from blind users for the combination of a beep and a text-to-speech message was positive. The first impression was judged as an unobtrusive possibility to present the pointing gestures to blind users. Further, it allows the blind user to get the important information of the pointing gesture in a quicker way than reading an alert message box.

For now, no empirical evaluation was conducted. However, we tested our prototype in a short brainstorming session of about 15 minutes within a mixed team of blind and sighted users, who were only partly familiar with the goals of the project. Using the serialized representation of the mind map tool[1] which is the first major research result of the project reported in

---

[1] http://come.sourceforge.com

[7], and displaying it on a blind user interface through a screen reader already showed an overall good behavior. The sighted and the blind user could both edit the mind map, while both representations were always consistent [15]. Integrating also the tracked and pre-selected deictic gestures above the tabletop into this presentation scheme seems to be a promising way forward to increase accessibility of non-verbal communication. Especially the synchronization of the individual presentations for the blind and the sighted user was judged to be a good achievement in enabling a visually impaired person to fully take part in a team meeting using a tabletop system. Synchronizing what is going on above the table into this evaluated and approved concept seems to be the right way forward.

Conclusion and Outlook

Within this paper, we presented the procedure how gestures are detected and how they are distinguished between relevant and non-relevant using a speech analyses. This is the core part of overall system description in [7]. Within future work, the whole system will be thoroughly evaluated under various conditions. Moreover, the filtering and smoothing algorithms will be optimized to further reduce the amount of wrong notifications to the blind user.

Acknowledgements

References

[1]   Beel, J., and Langer, S. An exploratory analysis of mind maps. In: *Proceedings of the 11th ACM symposium on Document engineering*, (2011), 81-84.

[2]   Budd, J.W. Mind maps as classroom exercises. In: *Journal of Economic Education*, (2004), 35(1), 35-46

[3]   Chow, J., Aug, K., Lichti, D. and Teskey, W. Performance Analysis of a Low-cost Triangulation-based 3D Camera: Microsoft Kinect System. In: *International Society for Photogrammetry and Remote Sensing Congress (ISPRS)*, Vol. 39, (2012), 175-180.

[4]   Eisenstein, J. and Christoudias, C.M. A Salience-based Approach to Gesture-Speech Alignment. In: *HLT-NAACL*, Vol. 4, (2004), 25-32.

[5]   Ellis, C.A., Gibbs, S.J. and Rein, Groupware: Some Issues and Experiences. In: *Communications of the ACM 34.1* (1991), 39-58.

[6]   Kunz, A., Alavi, A., Sinn, P. Integrating Pointing Gesture Detection for Enhancing Brainstorming Meetings Using Kinect and PixelSense. In: *8th International Conference on Digital Enterprise Technology*, (2014), 1-8.

[7]   Kunz, A.; Miesenberger, K.; Mühlhäuser, M.; Alavi, A.; Pölzer, S.; Pöll, D. and Heumader, P. Accessibility of Brainstorming Sessions for Blind People. In: *14th International Conference on Computers Helping People with Special Needs*, Springer Lecture Notes in Computer Sciences 8547 (2014), 237-244.

[8]   LEAP Motion: https://www.leapmotion.com (accessed 29.7.2014).

[9]   Majewski, M.; Dutz, T. and Wichert, R. An Optical Guiding System for Gesture Based Interactions in Smart Environments. In: *Distributed, Ambient, and Pervasive Interactions,* (2014), 154-163.

[10] Microsoft Speech Server (accessed 26.8.2014); http://www.microsoft.com/de-ch/server-cloud/products/windows-server-2012-r2/default.aspx

[11] MS PixelSense (accessed 26.8.2014); http://www.microsoft.com/en-us/pixelsense/default.aspx

[12] Nigay, L. and Coutaz, J. A generic platform for addressing the multimodal challenge. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, (1995), 98-105.

[13] Nigay, L. and Coutaz, J. A design space for multimodal systems: concurrent processing and data fusion. In: *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, (1993), 172-178.

[14] Pölzer, S. and Miesenberger, K. Presenting Non-verbal Communication to Blind Users in Brainstorming Sessions. In: *14th International Conference on Computers Helping People with Special Needs*, Springer Lecture Notes in Computer Sciences 8547 (2014), 214-219.

[15] Pölzer, S.; Schnelle-Walka, D.; Pöll, D.; Heumader, P. and Miesenberger, K. Making Brainstorming Meetings Accessible for Blind Users. In: *Proceedings of the 12 European AAATE Conference*, (2013), 653–658.

[16] Roth, W.-M. Gestures: Their Role in Teaching and Learning. In: *Review of Educational Research*, Vol. 71, No. 3 (2001), 365-392.

[17] Schnelle-Walka, D., Alavi, A., Ostie, P., Mühlhäuser, M. and Kunz, A. A Mind Map for Brainstorming Sessions with Blind and Sighted Persons. In: *14th International Conference on Computers Helping People with Special Needs*, Springer Lecture Notes in Computer Sciences 8547 (2014), 214-219.

[18] Shui, W. and Le, W. Mindmap-ng: A Novel Framework for Modeling Effective Thinking. In: *3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT 2010)*, Vol. 2, IEEE (2010), 480-483.

[19] Wang, H.-C., Cosley, D. and Fussell, S.R. Idea Expander: Supporting Group Brainstorming with Conversationally Triggered Visual Thinking Stimuli. In: *Proceedings of the 2010 ACN Conference on Computer Supported Cooperative Work (CSCW 2010)*, ACM Press (2010), 103-106.