

Mobile Interactive Image Sonification for the Blind

Torsten Wörtwein¹, Boris Schauerte¹, Karin Müller²,
and Rainer Stiefelhagen^{1,2}(✉)

¹ Computer Vision for Human-Computer Interaction Lab (CVHCI),
Karlsruhe Institute of Technology, Karlsruhe, Germany
`rainer.stiefelhagen@kit.edu`

² Study Centre for the Visually Impaired (SZS), Karlsruhe Institute of Technology,
Karlsruhe, Germany

Abstract. Web-based, mobile sonification offers a highly flexible way to give blind users access to graphical information and to solve various everyday as well as job-related tasks. The combination of a touch screen, image processing, and sonification allows the user to hear the content of every image region that he or she indicates with his/her finger position on a tablet or mobile phone. In this paper, we build on and expand our previous work in this area and evaluate how six blind participants can perceive mathematical graphs, bar charts, and floor maps with our sonifications and tactile graphics.

1 Introduction

Nowadays, two common ways of presenting spatial information to blind users are tactile graphics and sonification. Tactile graphics offer intuitive access to spatial information for blind people, especially if they have experience with braille. However, in recent years, we can observe that more and more people use electronic devices and software such as screen readers instead of printed braille to access text, since especially younger people tend to prefer electronic text on computers. For example, in Britain, it is nowadays assumed that only 15–20 thousand people use braille out of approximately two million blind and low vision individuals [18]. One major disadvantage of printed braille and tactile graphics is that they require special hardware such as tactile embossers. These devices are often complex, only provide help for specific tasks, and are produced in small numbers, which leads to high costs, low flexibility, and a low availability, since no blind person wants to carry a costly array of various devices around at all times. In contrast, smartphones and tablets have become omnipresent, mass produced, multi-function devices. In other words, they are the digital Swiss Army Knives of our era. Astonishingly, despite the fact that blind people are unable to see the content shown on the touchscreen, smartphones have become an essential tool in the lives of many blind people. For example, all participants in our evaluation use their smartphone every day and several of them own tablets and smart watches



Fig. 1. Experiment setting: two blind participants equipped with a headphone trying to identify a box plot (left) and a mathematical function plot (right) by moving their finger on the tablet surface

as well. In fact, blind people have become so apt in operating these devices that some apps are exclusively developed for – and even by [6] – blind users.

In our work, we use modern web-technologies to implement an interactive image sonification platform that can be accessed with a web-browser on devices that many blind users use on a daily basis (most importantly, smartphones, tablets and desktop computers). Hence, our system can be used almost everywhere, see Fig. 1 for the minimal setup. Since we can implement various sonification algorithms and switch them depending on the user’s context, we can use these sonifications to allow blind users access to a wide variety of information that is typically presented in the form of images (e.g., maps, plans, graphs, or charts). Furthermore, electronic devices allow flexible interaction with the user to adapt to his needs. For example, a tactile graphic does not change once it is printed, whereas an electronic representation allows to change its scale, the level of details, the way of sonification, or even output modalities (e.g., sound, speech, or vibration).

2 Related Work

In recent years, various assistive systems for blind people have been developed (e.g., [2, 23, 25]) that rely on a variety of different technologies such as, e.g., computer vision [20], crowdsourcing [5], and touchscreens [13, 24]. For example, it is possible to use computer vision to detect the walkable area in front of a blind person and then transform the information about obstacles into haptic or auditory feedback to guide a blind user around detected obstacles (see [15, 20]). Turning information into auditory signals is called sonification (see [11]; e.g., [1–3, 12, 17, 19, 23, 25]) and is a very common mean to present visual information such as, e.g., maps to blind people [23]. Sonification systems may use different techniques of sound synthesis, which we can differentiate into two main categories: Entirely synthetic sounds and sounds composed with musical instruments (e.g., piano and guitar sounds) [1, 2]. As for sonification for visually impaired people, there also exist two main categories when it comes to image sonification: Low-level sonification of arbitrary images (e.g., [1, 2, 8]), *i.e.* sonification of basic

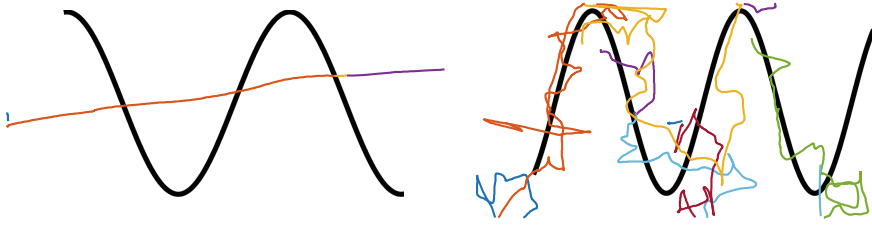


Fig. 2. Finger trajectories with function value sonification (left) and distance sonification (right)

image properties like color [8] and edges [25], and task-specific sonification (*e.g.*, [23]). Furthermore, we have to differentiate between sonification methods that sonify the image as a whole and present it to the user in form of a pre-calculated audio clip (*e.g.*, [16]) and systems that allow user interaction (see, *e.g.*, [10]), *e.g.* by sonifying the image area under the mouse cursor (*e.g.*, [1, 25]). Web and mobile technologies have also been explored to assist blind people. Most prominently, Bigham *et al.* incorporate sighted persons (“the crowd”) to help blind people find specific objects [4, 5] on a picture taken with their iPhone.

3 Interactive Image Sonifications

In the following, we briefly describe the evaluation of several sonification methods for three tasks: mathematical graph identification, proportion estimation in bar charts, and path finding on floor maps.

3.1 Mathematical Graphs: “Identify the Underlying Function of the Presented Graph.”

The *distance sonification* sonifies the distance to the plotted function. The frequency becomes lower the closer you are to the function. Hence, it encourages a free exploration of the whole *image*. We used an adapted depth-first-search to calculate the shortest distance to the graph.

Similar to the sonification by Grond *et al.* [9], the *value sonification* sonifies the value of the function at the x-location of the user’s finger. Accordingly, the user will hear a high tone, if the function’s value is high and vice versa. As a consequence, it is only necessary to move your finger from left to right, see Fig. 2.

The following functions are used: a linear, hyperbolic, parabola, sine, square-root, and an exponential function. The function value is represented by the position of the first non-background pixel per column.

3.2 Bar Charts: “Interpret the Relative Dimensions Between the Bars in a Bar Chart”

The *semantic map sonification* closely resembles a tactile graphic. Each texture is mapped to a different frequency (for the background and for each bar).

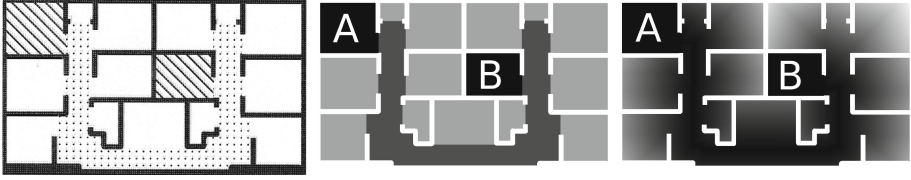


Fig. 3. Floor map as tactile graphic (left) and heat maps of frequencies of the semantic map sonification (middle) as well as the guided sonification (right)

The *value sonification* again sonifies the function value or in other words the height of the bars. Accordingly, the output frequency at an image location corresponds to the height of the bar at this location.

3.3 Floor Maps: “Find the Way Through a Corridor from One Room to Another on a Floor Map”

The *semantic map sonification* again resembles the tactile graphic, see Fig. 3. The source (A) and destination (B) room, corridors, other rooms, and walls are mapped to distinct frequencies. Thus, this sonification denotes the semantic information of a floor map.

To assist the task to find a path from room A to room B, the *guided sonification* guides the user to the destination. Therefore, a path from the source to destination is computed. The further you are away from the path the higher the frequency. Walls are denoted with the highest frequency, while the source and destination room share the lowest frequency. We use an A* algorithm to find the shortest path between the two rooms. Since the shortest path is along the walls and not centered in the corridor, we penalize pixel close to a wall based on a depth-first-search. After we calculated the path, we compute the distance to the path in the same manner as in *distance sonification*.

3.4 Frequency Mapping

All distances and function values are not directly mapped to output frequencies, because a twice as high frequency is not perceived as being twice as high by the human auditory perception. Therefore, we use the Mel transformation [22] to have a meaningful auditory representation. However, too low frequencies are hardly heard, we add a constant offset just before the Mel transformation to have perceivable frequencies. Thereby, we lose absolute relations but equal distances are still perceived equally. Frequencies range from 51.49 Hz (80 Mel) to 254.93 Hz (350 Mel).

4 Experimental Evaluation

We performed experiments with six blind participants. Four of them were female. The average age was 38.2 years ($SD = 8.1$). All participants have normal hearing

Table 1. Average and standard deviation of time required by the participants to complete each task for all evaluated methods (in seconds).

Method/task	Math. Graphs	Bar Charts	Floor Maps
<i>Tactile graphics</i>	9.5 (4.9)	28.8 (12.6)	49.4 (8.6)
<i>Semantic map sonification</i>		98.9 (28.3)	131.1 (80.5)
<i>Value sonification</i>	25.0 (6.7)	49.9 (9.1)	
<i>Distance/guided sonification</i>	106.0 (43.0)		114.5 (70.2)

and use touchscreens on a daily basis. All except one participant have at least passed the German *Abitur* (A-Levels) examinations. Only two participants use tactile graphics on a regular basis (several times a month). The participants did not have any prior training with our sonification system nor were they involved in the development.

To exclude variations due to different hardware and software configurations (*e.g.*, in stimulus size), all participants used the same platform: A first generation Samsung Nexus 10 tablet (10-inch touchscreen) with Android 5.1, AKG K 701 circumaural headphones and Google Chrome 47 as browser, see Fig. 1.

4.1 Evaluation Methodology

All participants answered first a demographics questionnaire. To ensure a reproduction of this experiment, the same instructions were read to all participants. In addition, they received the instructions in Braille without contractions (*i.e.*, *Deutsche Blindenvollchrift*). Our experiment consists of three tasks: mathematical function plots, bar charts, and floor maps. To solve each of them, three methods were evaluated: first using tactile graphics followed by two sonifications in random order. After each method, participants were asked the *After-scenario Questionnaire* (ASQ) on a five-point likert scale (third question is not applicable) to assess the usability of the methods [14]. After completing each task, participants were asked to rank the usability of the three used methods. The overall usability of the system was evaluated after all tasks on the positive *Software Usability Scale* (SUS) [7]. Each task and method had to be solved by several examples (six function plots, three bar charts, and three floor maps). To avoid a learning effect between the methods and to have an equal difficulty for each method, we varied the examples (function plots were scaled, bar charts were permuted, and floor maps were flipped as well as rotated). The order of the examples per method and task was randomized as well as whether and how the example was varied. For all examples, we measured the time to complete the current example and recorded the finger trajectory. In case of function plots, participants were asked to name or precisely describe the function. For bar charts, participants had to count the number of bars, order them by size, and name the size relations relative to the smallest bar (*e.g.*, “the left bar is twice as high as the smallest bar”). Floor maps were evaluated whether they found the

destination and whether they were able to describe a valid path between the source and destination room. Since many participants were not familiar with mathematical functions, they were told the name of the function after each tactile example.

Table 2. Average and standard deviation of *ASQ* for all tasks and methods.

Method/task	Math. Graphs	Bar Charts	Floor Maps
<i>Tactile graphics</i>	5.0 (0.0)	4.8 (0.6)	4.3 (0.8)
<i>Semantic map sonification</i>		3.8 (0.8)	3.1 (1.4)
<i>Value sonification</i>	4.0 (0.8)	3.7 (0.6)	
<i>Distance/guided sonification</i>	2.2 (0.9)		2.7 (1.0)

4.2 Results

Tables 1 and 2 summarize the required time and the *ASQ*. Let us now briefly present our results. If we compare our new sonifications (*i.e.*, graph value, bar value, and guided floor map) to the sonifications that we used in our prior publications (*i.e.*, graph distance, semantic bar map, and semantic floor map), we can see that the users require less time to solve the tasks with our new sonifications. The average task completion time reduced to 25.0 from 106.0, to 49.9 from 98.9, and to 114.5 from 131.1 s for mathematical graph, bar chart, and floor map understanding, respectively. For example, it can easily be seen in Fig. 2 that the value sonification requires only few, simple finger swipes to understand the graph. But, this does not always lead to a higher user satisfaction as measured by the *ASQ*. The graph value sonification is significantly better for the graph sonification (3.9 on the *ASQ* scale of 1 to 5) compared to the graph distance sonification (2.2). But, the *ASQ* score of our new sonifications is lower for the bar chart sonification (3.8 vs 3.7) and the floor map understanding (3.1 vs 2.7).

Compared to the tactile graphics, we still observe a substantial gap in the users' quantitative task performances and subjective ratings. The tactile graphics still perform best in terms of task completion times (9.5, 28.8, and 50.0 s for graphs, bar charts, and floor maps, respectively) and *ASQ* scores (5.0, 4.8, and 4.3).

Overall, all users rate the intuitiveness of our mobile sonification system as being very high, *i.e.* a 4.0 on the *SUS* scale of 1 to 5. Furthermore, even though our sonifications differ substantially, the whole system is perceived as being consistent (*SUS* score of 4.33).

5 Conclusion

We have tested several tasks with various sonification methods and six blind users. We included tactile graphics as a baseline in our evaluation.

All our newly introduced sonifications reduced the time that the participants require to interpret mathematical graphs, bar charts, and floor maps. Unfortunately, we observed that lower completion times do not automatically correlate with a higher user satisfaction. Tactile graphics are still preferred over sonification by all but one of our participants. However, the fact that we already have users that prefer our sonification over tactile graphics is extremely promising and encouraging, because we should not forget that our web-based sonification is in an early technological stage and still suffers from technical challenges such as, e.g., the lack of real-time capabilities in modern browsers (see [21]).

Although we use vibration to notify users when they leave the sonified image area, we have not yet used vibration as a distinct information channel, but we plan to do this as part of our future work.

Acknowledgments. We thank Giuseppe Melfi for his help during the experimental evaluation and Gerhard Jaworek for being our voluntary blind user for alpha tests.

References

1. Banf, M., Blanz, V.: A modular computer vision sonification model for the visually impaired. In: Proceedings of the International Conference of Auditory Display (2012)
2. Banf, M., Blanz, V.: Sonification of images for the visually impaired using a multi-level approach. In: Proceedings of the Augmented Human International Conference (2013)
3. Bearman, N., Brown, E.: Who's sonifying data and how are they ng it? A comparison of ICAD and other venues since 2009. In: Proceedings of the International Conference on Auditory Display (2012)
4. Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R.C., Miller, R., Tatarowicz, A., White, B., White, S., et al.: VizWiz: nearly real-time answers to visual questions. In: Proceedings of the Annual ACM Symposium on User Interface Software and Technology. pp. 333–342. ACM (2010)
5. Bigham, J.P., Jayant, C., Miller, A., White, B., Yeh, T.: VizWiz: locateit-enabling blind people to locate objects in their environment. In: Proceedings of the Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 65–72 (2010)
6. Blüher, J.: About visorapps (2015). <http://en.visorapps.com/about-visorapps/>
7. Brooke, J.: SUS - a quick and dirty usability scale. Usability Eval. Ind. **189**, 4–7 (1996)
8. Cavaco, S., Henriques, J.T., Mengucci, M., Correia, N., Medeiros, F.: Color sonification for the visually impaired. *Procedia Technol.* **9**, 1048–1057 (2013)
9. Grond, F., Droßard, T., Hermann, T.: SonicFunction: experiments with a function browser for the visually impaired. In: Proceedings of the International Conference on Auditory Display. Georgia Institute of Technology (2010)
10. Hermann, T., Hunt, A.: An introduction to interactive sonification. *IEEE Multimedia* **12**, 20–24 (2005)
11. Hermann, T., Hunt, A., Neuhoff, J.G.: *The Sonification Handbook*. Logos Verlag, Berlin (2011)

12. Heuten, W., Henze, N., Boll, S.: Interactive exploration of city maps with auditory torches. In: CHI Extended Abstracts on Human Factors in Computing Systems (2007)
13. Klatzky, R., Giudice, N., Bennett, C., Loomis, J.: Touch-screen technology for the dynamic display of 2D spatial information without vision: promise and progress. *Multisensory Res.* **27**(5–6), 359–378 (2014)
14. Lewis, J.R.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *Int. J. Hum. Comput. Interact.* **7**(1), 57–78 (1995)
15. Martinez, M., Constantinescu, A., Schauerte, B., Koester, D., Stiefelhagen, R.: Cognitive evaluation of haptic and audio feedback in short range navigation tasks. In: Miesenberger, K., Fels, D., Archambault, D., Peñáz, P., Zagler, W. (eds.) ICCHP 2014, Part II. LNCS, vol. 8548, pp. 128–135. Springer, Heidelberg (2014)
16. Meijer, P.B.: An experimental system for auditory image representations. *Trans. Biomed. Eng.* **39**(2), 112–121 (1992)
17. Ribeiro, F., Florêncio, D., Chou, P.A., Zhang, Z.: Auditory augmented reality: object sonification for the visually impaired. In: Proceedings of the International Workshop on Multimedia Signal Processing (2012)
18. Rose, D.: Braille is spreading but who's using it? (2013). <http://www.bbc.com/news/magazine-16984742>. bBCNews
19. Sanchez, J.: Recognizing shapes and gestures using sound as feedback. In: CHI Extended Abstracts on Human Factors in Computing Systems (2010)
20. Schauerte, B., Koester, D., Martinez, M., Stiefelhagen, R.: Way to Go! Detecting Open Areas Ahead of a Walking Person. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014 Workshops, Part III. LNCS, vol. 8927, pp. 349–360. Springer, Heidelberg (2015)
21. Schauerte, B., Wörtwein, T., Stiefelhagen, R.: A web-based platform for interactive image sonification. In: Accessible Interaction for Visually Impaired People (2015)
22. Stevens, S.S., Volkman, J., Newman, E.B.: A scale for the measurement of the psychological magnitude pitch. *J. Acoust. Soc. Am.* **8**(3), 185–190 (1937)
23. Su, J., Rosenzweig, A., Goel, A., de Lara, E., Truong, K.N.: Timbremap: enabling the visually-impaired to use maps on touch-enabled devices. In: Proceedings of the International Conference on Human Computer Interaction with Mobile Devices and Services (2010)
24. Taibbi, M., Bernareggi, C., Gerino, A., Ahmetovic, D., Mascetti, S.: AudioFunctions: eyes-free exploration of mathematical functions on tablets. In: Miesenberger, K., Fels, D., Archambault, D., Peñáz, P., Zagler, W. (eds.) ICCHP 2014, Part I. LNCS, vol. 8547, pp. 537–544. Springer, Heidelberg (2014)
25. Yoshida, T., Kitani, K.M., Koike, H., Belongie, S., Schlei, K.: EdgeSonic: image feature sonification for the visually impaired. In: Proceedings of the Augmented Human International Conference (2011)