

# A Web-based Platform for Interactive Image Sonification

Boris Schauerte<sup>1</sup>, Torsten Wörtwein<sup>1</sup>, Rainer Stiefelhagen<sup>1,2</sup>

Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology<sup>1</sup>  
Study Center for the Visually Impaired, Karlsruhe Institute of Technology<sup>2</sup>

## Abstract

We present a web-based sonification platform that allows blind users to interactively experience a wide range of information using two nowadays widespread technologies: modern web browsers that implement high-level JavaScript APIs and touch-sensitive displays such as provided by, e.g., tablets. This way, blind users can easily access information such as, e.g., maps or graphs. Our current prototype provides a variety of different sonifications that can be switched depending on the image type and user preference. The prototype runs in Chrome and Firefox on PCs, smart phones, and tablets.

## 1 Introduction

The World Health Organisation estimates that 285 million visually impaired people live in the world of which 39 million are blind. The unfortunate fact that the majority of blind people lives in developing countries in combination with the aging global elderly population leads to a huge innovation pressure for affordable and intuitive tools that aid visually impaired people. With the decreasing costs of mobile computing power in the form of tablets and smart phones, computer-based assistive technologies become increasingly affordable and cheaper than traditional means of assistance. For example, blind users commonly access graphical information using tactile graphics printed by a tactile embosser, which cost between US\$1,500 and US\$150,000. For comparison, Android-powered tablets can be bought for less than US\$100 and modern web browsers are free, which makes mobile technologies interesting for cost-effective and easy-to-use assistive systems.

We present our prototype of a web-based image sonification platform. We implemented the system based on Javascript and modern Javascript APIs. This allows our system to operate on desktop PCs as well as mobile devices, with mouse and touchscreen support. We use image processing and computer vision techniques to (pre-)process the image and map the image content to sounds with support for interactive exploration of the image. The exact sonification method depends on the category of the image material and, of course, user

preferences. Our system already allows the sonification of various image categories (e.g., maps, graphs, charts, diagrams, and pictures of art) and can easily be extended to support further sonifications (e.g., different means to sonify low-level or task-specific information).



Figure 1 Our web-based sonification prototype. The image within the dashed-grey border is being sonified.

## 2 Related Work

Various assistive systems for blind people have been developed (Banf & Balz 2013; Su et al. 2010; Yoshida et al. 2011) that rely on a variety of different technologies such as, e.g., image processing (Yoshida et al. 2011) or crowd sourcing (Bigham et al. 2010). Turning information into auditory signals is called sonification (Hermann et al. 2011) and is a common mean to present visual information such as, e.g., maps to blind people (Su et al. 2010). Sonification systems may use different techniques of sound synthesis, which we can differentiate into two broad categories: Synthetic sounds and sounds composed with musical instruments (e.g., piano and guitar sounds) (Banf & Balz 2012; Banf & Balz 2013). As for sonification for visually impaired people, there also exist two broad categories when it comes to image sonification: Low-level sonification of arbitrary images (Banf & Balz 2012; Banf & Balz 2013; Cavaco et al. 2013), i.e. sonification of basic image properties like color (Cavaco et al. 2013) and edges (Yoshida et al. 2011), and task-specific sonification (Su et al. 2010). 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 (Meijer 1992) and systems that allow user interaction (Hermann & Hunt 2005), e.g. by sonifying the image area under the mouse cursor (Banf & Balz 2012; Yoshida et al. 2011). Web and mobile technologies have also been explored to assist blind people. Most prominently, Bigham et al. use the crowd to enable blind people to ask non-blind people for assistance in finding specific objects in camera images (Bigham et al. 2010).

### 3 System

One of our goals is to allow easy access to graphical information for blind users all over the world. For this purpose, we chose to implement our sonification system in Javascript with modern web browsers such as Firefox and Chrome as target platforms. This has several advantages: Users do not have to install software, assuming that nowadays a web browser is always available. Although there are technological differences between web browsers themselves and web browsers across platforms (e.g., Android and Windows), the sonification does not depend on the chosen platform, which means that blind users hear the same sounds regardless of whether they use a PC, tablet, or mobile phone. Both aspects mean that it is less problematic to use our system across different platforms, at work, or on another person's PC or tablet. Furthermore, since our sonification platform is a website, it can be accessed from all around the world.

Our sonification implementation relies on the W3C's Web Audio API that allows us to synthesize arbitrary sounds at runtime. In our current implementation, we only synthesize artificial sounds and do not use, e.g., musical instruments for sonification.

In general, the system follows a pipeline model (image (pre-)processing, computer vision, map the information to sound, and sound (post-)processing), of which all elements are implemented in Javascript and run in the user's browser. First, the input image is pre-processed, which for example can include blurring (e.g., to reduce noise and abstract the image details) or thresholding as preprocessing step for a distance transformation. Then, the image is first processed using image processing and computer vision methods to precompute a map that encodes the auditory signal that we present to the user at a given cursor position. Here, the exact type of image processing depends on the image category (e.g., natural images or graphs) and the user preference (i.e., what does the user want to hear?). For example, we provide the distance transform as a basic sonification for two image categories: To sonify the distance to the nearest wall when exploring floor maps and to sonify the distance to the curve of a graph, which makes it simple to follow the curve with the finger. Then, we map image information onto frequencies (e.g., luminance range to a frequency range, or distance to curve to a frequency range with a specified cut-off level) and can combine various sources of information with different combination schemes (e.g., additive mix or overtones). The generated sound is then post-processed, for example, with a Hz to mel transformation.

Additionally, we can capture click events to allow further user interaction. For example, we implemented a form of "sonar" that sonifies the distances from the current cursor position to the walls in a floor map. Moreover, we implemented a text-to-speech interface in the browser and can provide speech output to the user (e.g., the name of the color at the current cursor position).

## 4 Experiments

We have performed several experiments with various participants, including born blind persons. So far, all users had normal hearing and experience with touchscreens. We currently focus our development and thus experiments on three tasks:

**Mathematical graphs:** Allow users to identify the underlying function of a presented graph.

**Charts:** Allow users to interpret the information in charts such as, e.g., bar charts.

**Floor maps:** Allow users to navigate on floor maps, e.g. to find the way through a corridor from one room to another on such a floor map.

It is interesting that even untrained users are able to correctly solve such graph identification, chart interpretation, and floor map navigation tasks with high very accuracy (consistently higher than 80%). We asked several users to repeatedly report aspects such as intuitiveness and ease of sound interpretation. Even after few uses, the users rated our system as being highly intuitive to use and the generated sonification as being easy to interpret. These already high ratings increased further the longer the users used our system.

## 5 Conclusion

We presented an interactive sonification platform based on modern web technologies and off-the-shelf hardware, which means that our system is just an URL away from blind users all around the world. In preliminary experiments, we have shown that our system is intuitive to use and allows to accomplish different tasks even with a minimal amount of training.

### Contact

Boris Schauerte, [boris.schauerte@kit.edu](mailto:boris.schauerte@kit.edu), Karlsruher Institut für Technologie, Vincenz-Prießnitz-Str. 3, 76131 Karlsruhe

## 6 References

- Banf, M. & Blanz, V. (2012). A modular computer vision sonification model for the visually impaired. In Proc. Int. Conf. of Auditory Display.
- Banf, M. & Blanz, V. (2013). Sonification of images for the visually impaired using a multi-level approach. In Proc. Augmented Human Int. Conf.. New York: ACM.
- Bigham, J. P., Jayant, C., et al. (2010). Vizwiz: Locateit-enabling blind people to locate objects in their environment. In Proc. Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE.
- Cavaco, S., Henriques, J. T., et al. (2013). Color sonification for the visually impaired. *Procedia Technology*, 9, 1048–1057.

- Hermann, T. & Hunt, A. (2005). An introduction to interactive sonification. *IEEE Multimedia*, 12
- Hermann, T., Hunt, A. & Neuhoff, J. G. (2011). *The sonification handbook*. Berlin: Logos Verlag.
- Meijer, P. B. (1992). An experimental system for auditory image representations. *Trans. Biomedical Engineering*, 39, 112–121.
- Su, J., Rosenzweig, A., et al. (2010). Timbremap: Enabling the visually-impaired to use maps on touch-enabled devices. In *Proc. Int. Conf. HCI with mobile devices and services*. New York: ACM.
- Yoshida, T., Kitani, K. M., et al. (2011). Edgesonic: image feature sonification for the visually impaired. In *Proc. Augmented Human Int. Conf.*. New York: ACM.