Interactive Web-based Image Sonification for the Blind

Torsten Wörtwein
uncwt@student.kit.edu

Karin E. Mueller
karin.e.mueller@kit.edu

Boris Schauerte
schauerte@kit.edu

Rainer Stiefelhagen
rainer.stiefelhagen@kit.edu

Karlsruhe Institute of Technology
Vincenz-Prießnitz-Str. 3
76131 Karlsruhe

ABSTRACT

In this demonstration, we show a web-based sonification platform that allows blind users to interactively experience various information using two nowadays widespread technologies: modern web browsers that implement high-level JavaScript APIs and touch-sensitive displays. This way, blind users can easily access information such as, for example, maps or graphs. Our current prototype provides various 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.

Categories and Subject Descriptors

K.4.2 [**Computers And Society**]: Social Issues—*Assistive technologies for persons with disabilities*; I.4.9 [**Image Processing And Computer Vision**]: Applications

General Terms

Human Factors, Experimentation

Keywords

Sonification; Interactive; Blind; Assistive Technology; Web; Mobile; Image Processing

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, see Fig. 1. We implemented the system based on Javascript and modern Web 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).

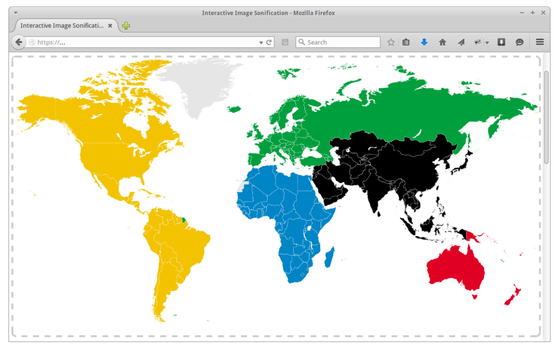


Figure 1: Our web-based user interface. The content of the area within the dotted border is sonified. New images can be loaded by drag-and-drop or a menu. The sonification method and parameters can be selected and changed during runtime by a menu. Our system works on tablets, smart phones, and PCs.

2. 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

ICMI '15, November 09-13, 2015, Seattle, WA, USA
ACM 978-1-4503-3912-4/15/11
http://dx.doi.org/10.1145/2818346.2823298.

target platforms. This has several advantages such as: (1.) Users do not have to install special software, since we rely on their readily available web browser. (2.) Although there are technological differences between web browsers themselves and web browsers across platforms such as 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 also 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 system is a website, it can be accessed from all around the world.

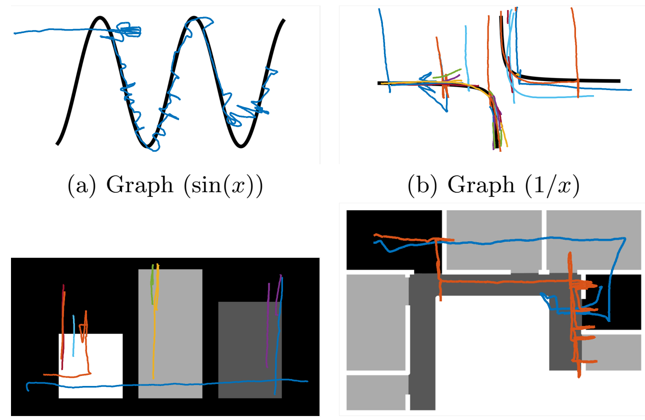
Our implementation relies on the W3C Web Audio API [3], which is encapsulated by the Timbre.js library [1] and allows us to synthesize arbitrary sounds at runtime. Currently, we only synthesize artificial sounds and do not use, for example, musical instruments for sonification. However, the current implementation of the Web Audio API has a major drawback for our application: It does not offer real-time guarantees, which often means that there are noticable delays between the mouse/finger movement and the sonification of the image content at the mouse/finger (cursor) location. In several preliminary studies, we noticed that this significantly irritates users – especially if unaware about the fact – and leads to problems when trying to identify shapes and details with swift finger movements.

Our sonification system follows a pipeline model (image pre-processing, image 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 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 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: (1.) To sonify the distance to the nearest wall when exploring floor maps and (2.) 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 [2].

Additionally, we capture click events to allow further user interaction. For example, we implemented a form of "sonar" that sonifies the distance (sampled in a pre-specified number of directions) from the current cursor position to the walls in a floor map. Moreover, we implemented the ability for text-to-speech in the browser and can also provide speech output to the user (*e.g.*, the name of the color at the current cursor position).

3. EXPERIMENTS

We have performed several experiments with various participants (see Fig. 2), 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:



(a) Graph ($\sin(x)$) (b) Graph ($1/x$) (c) Bar chart (d) Floor map

Figure 2: Examples that illustrate some experimental stimuli and tasks. The colored lines represent the finger movements of one participant.

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 very high 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 as the duration with which they used our system.

4. CONCLUSION

We present 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.

5. REFERENCES

[1] Mohayonao. Timbre.js. `http://mohayonao.github.io/timbre.js/`, 2015. Online; accessed 21 August 2015.

[2] S. S. Stevens, J. Volkmann, and E. B. Newman. A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8(3):185–190, 1937.

[3] W3C. Web Audio API. `http://www.w3.org/TR/webaudio/`, 2015. Online; accessed 21 August 2015.