

## INTERACTIVE SONIFICATION OF COLOR IMAGES ON MOBILE DEVICES FOR BLIND PERSONS - PRELIMINARY CONCEPTS AND FIRST TESTS

Andrzej Radecki, Michał Bujacz\*, Piotr Skulimowski, , Paweł Strumillo

Lodz University of Technology

Lodz, Poland

\*bujacz.m@p.lodz.pl

### ABSTRACT

The paper presents a proposed sonification algorithm for presenting images on mobile devices for blind children using an interactive auditory display. The sonification uses HSV color images and the software is being written for Android devices with touchscreens.

A brief discussion of the basic methods of sonication used in the presentation of images and the history of interactive sonification is presented in the review section.

The authors proposed several methods for interactive sonification in which the user indicates with touch the area of the image to sonify and depending on the image content and image processing filters a real-time additive synthesis of several sound buffers is prepared. The method presented in the paper sonifies just one pixel directly under the finger, using its HSV values as input to an additive synthesis transform.

The focus was to create a sonification algorithm that was real-time and not computationally complex, while allowing to clearly distinguish not only brightness, but also the saturation and hue components determining color.

### 1. INTRODUCTION

Attempts to utilize modern technologies to aid the blind using acoustic signals reach as far back as the XIX century with the Polish invention called Electroftalm [1] or the early XX century Optophone [2]. The term sonification itself has been formally defined in the 90s as "the use of non-speech audio to convey information" [3] or "data-dependent generation of sound in a way that reflects objective properties of the input data" [4]. An important organization responsible for discussion and dissemination of sonification-related research is the International Community for Auditory Display (ICAD, [www.icad.org](http://www.icad.org)).

Interactive sonification is a subtopic of sonification concerned with the creation of an interactive control loop with a human listener altering the properties of the synthesized audio [5][6].

Although sonification can be used for a variety of data sources [7], the authors' interest lies specifically in aiding visually disabled persons. Our previous efforts were focused on presentation of 3D scenes to the blind [8] and aiding in navigation [9], while the current project is aimed at interactive sonification of images on mobile devices.

#### 1.1. Sonification of images to the blind

Attempts to interactively sonify text or images reach as early as the 1920s. The first interactive sonification tool for the blind

was the Optophone that turned black and white print into chords that allowed to distinguish letters [2].

More extensive reviews of sonification and interactive sonification can be found in [10] and [11]. Probably the best known algorithm for non-interactive sonification for the blind is The vOICe (Meijer 1992)[12]. The algorithm is based on automated cyclic reading of pixel columns from a grayscale image and translating them directly into the momentary sound spectrum. It has recently become available as an app available for most mobile operating systems.

Due to the growing capability of mobile electronic devices to process images, touch and synthesize audio, the subject of interactive sonification for the blind has gained much interest in recent years [13][14][15].

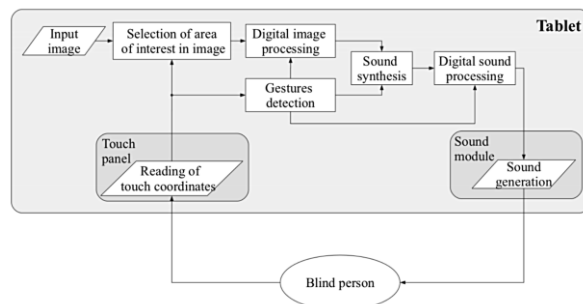


Figure 1. The block diagram of the proposed interactive sonification control loop.

### 2. PROPOSED INTERACTIVE SONIFICATION ALGORITHM FOR HSV IMAGES

The proposed method for image sonification aims to hand over full control over the sonification algorithms to the blind user. Real-time interaction is possible through the feedback loop that starts with the user's touch gestures, combines them with image information, synthesizes the sounds that the user hears and can decide to touch the image differently. This allows for fully interactive sonification, i.e. the blind user decides what region of the image to sonify and how to process the image information (e.g. by filtering). The block diagram of the interactive feedback loop is given in Figure 1.

#### 2.1. The concept of the sonic space and the proposed image to sound transform

The proposed synthesis method consists of combining  $s_n^m$  parallel sound buffers which make up the sonic space  $S$ . The  $S$

space has  $n$  dimensions and each of its elements is defined by the  $m$ -th buffer contents having unique timbres, the frequency a given buffer will be sampled with  $f_{nb}$  and its amplitude scaling  $A_n$ . This gives the  $\mathbf{S}$  space with  $s$  elements that can be described as follows:

$$\mathbf{s} = (s_1^1(f_{1b}, A_1), \dots, s_n^m(f_{nb}, A_n)) \quad (1)$$

The basic sonification scheme presented in the paper aims to transform the HSV color space to the sonic space  $\mathbf{S}$  according to the  $g_{HSV}$  function which is described as:

$$\begin{aligned} g_{HSV1} : HSV &\rightarrow \mathbf{S}, \quad g_{HSV1}(H, S, V) = \\ &= (s_1^1(f_b + H \cdot c_1) / (S \cdot c_2), V), \\ &s_2^1(f_b + H \cdot c_1, V), \\ &s_3^1(f_b + H \cdot c_1) \cdot (S \cdot c_2), V) \end{aligned} \quad (2)$$

where:  $H, S, V \in [0, 1]$ ,  $c_1, c_2$  – normalizing coefficients,  $f_b \in \mathbf{R}^+$  – audio buffer playback frequency

Transformation (2) assumes generation of three independent sounds of predefined timbres, spaced apart in frequency. The spacing is dependent on the color saturation  $S$ , while the hue  $H$  shifts the base frequency. The brightness  $V$  determines the loudness for all the components.

The HSV model shows high promise for the use in sonification of images for the blind [15]; however, the cyclic nature of the HSV model (Figure 2) could be problematic in translation to frequency, as a point of discontinuity would need to be introduced.



Figure 2. The hue value of the HSV color model is circular.

The proposed transform  $g_{HSV2}$  that is taking into consideration the cyclic nature of the  $H$  component is described as:

$$\begin{aligned} g_{HSV2} : HSV &\rightarrow \mathbf{S}, \quad g_{HSV2}(H, S, V) = \\ &(s_1^1(f_b / (1 + S \cdot c_{21}), V \cdot g_{amp}((H + c_{31}) \bmod 1)), \\ &s_2^1(f_b / (1 + S \cdot c_{22}), V \cdot g_{amp}((H + c_{32}) \bmod 1)), \\ &s_3^1(f_b, V \cdot g_{amp}((H + c_{33}) \bmod 1)), \\ &s_4^1(f_b \cdot (1 + S \cdot c_{22}), V \cdot g_{amp}((H + c_{34}) \bmod 1)), \\ &s_5^1(f_b \cdot (1 + S \cdot c_{21}), V \cdot g_{amp}((H + c_{35}) \bmod 1))) \end{aligned} \quad (3)$$

where  $c_{21}=1, c_{22}=0.5, c_{31}=0, c_{32}=0.2, c_{33}=0.4, c_{34}=0.6, c_{35}=0.8$  – are normalizing coefficients chosen so that a consonant chord is heard at full saturation and

$$g_{amp}(x) = \begin{cases} 2 \cdot x & \text{for } x \leq 0.5 \\ 2 \cdot (1 - x) & \text{for } x > 0.5 \end{cases} \quad (4)$$

The mod1 operation is used to wrap around from 1.0 to 0.0 by cutting off any integer part of the number. The nature of the equation (3) is different than (2). In (3) all sound buffers except for  $s_3$ , which contains the base frequency all others are scaled relative to, are played with frequency dependent only on the saturation component. The perception of the frequency change along with hue value was achieved by an amplitude envelope (5) which is dependent of  $H$  component and sound buffer frequency localization (described by  $c_{31}$  to  $c_{35}$  coefficients). In transform (3) the synthesized sound is the same for  $H$  component approaching both values of 0 and 1, which follows the circular nature of hue. Although the synthesized sound might be more complex than for transform (1), due to the amplitude restraints in function  $g_{amp}$ , the complexity is less audible.

One should also notice the influence of the saturation component  $S$ , which reduces the complexity of the sound, the closer it is to 0. This means the proposed sonification approach may make greyscale images more easily interpreted than full color ones.

## 2.2. Spectral properties of the synthesized sounds

The use of buffers containing pure tones was initially considered, but produced very monotonous and irritating sounds, that quickly tired the listener. Instead, a number of more aesthetically pleasing buffers were used that had constant timbres deemed subjectively pleasant by several experimenters. The spectrogram of the sound buffer  $sn1$  is shown in Figure 3.

Figures 4-6 show the audio spectrograms of several sample color bars. Sharp changes in color are transformed into sudden changes in the timbre of the sound (Figure 4), while smooth hue transitions are turned into smooth changes of sound (Figure 5), with a closed loop (i.e. the end of the bar produces the same sound as the start). When a color simultaneously changes its brightness and intensity (Figure 6), the sound increases in amplitude, while becoming spectrally less complex due to decreased saturation.

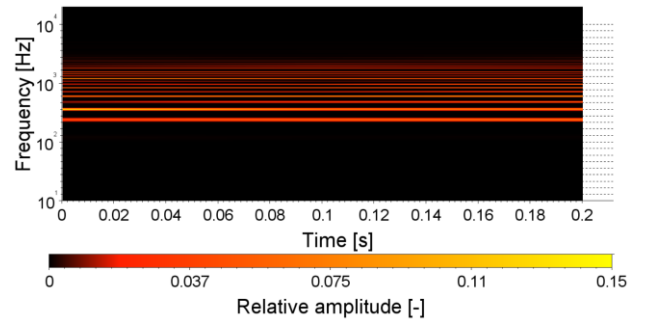


Figure 3. Spectrogram of the sound buffer used in the additive synthesis algorithm for sonification

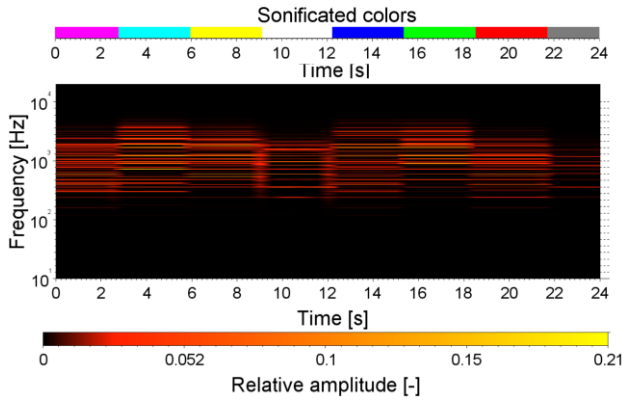


Figure 4. Spectrogram of the sonification of a discrete color sequence, each color resulting from a combination of three buffers scaled with amplitude and sampling frequency

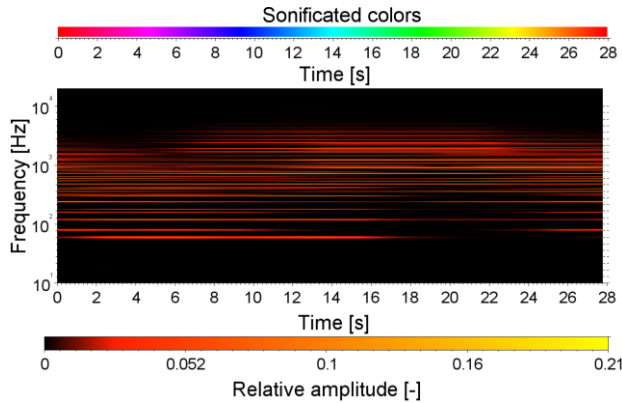


Figure 5. Spectrogram of the sonification of a continuous color sequence with a constant saturation value ( $S=1$ )

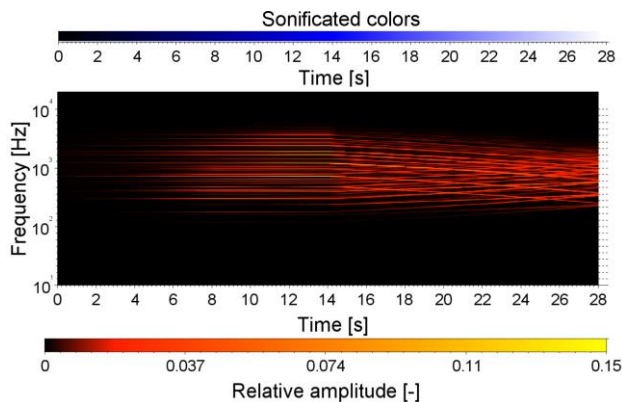


Figure 6. Spectrogram of the sonification of a blue color hue with a varying brightness and saturation

### 3. DATA COLLECTION TOOLS AND FIRST EXPERIMENTS

In order to verify the proposed sonification method and to judge the possibility of differentiating between the different HSV components, several software tools were prepared both for Windows and Android and a short experimental study was performed.

The main tool allows to synthesize a sound in real time depending on a touched pixel's HSV values and also log the path of a user's finger.

For the purpose of the study the experimenters trained themselves and one volunteer to discern two colors (white and red) and a brighter region (0.8 V value) from a "rainbow" background of a full H spectrum. This was done using the training images in Figure 7. All five participants were sighted, but blindfolded for the testing (though not the training). After a short training session, the task was to track a path with a finger along differently colored trajectories as shown in Fig 8, starting from one of the black regions and ending in the second one. Sample results of the test are shown in Figures 9 to 12. In each case the trajectory of the user's finger was tracked and the spectrogram of the sound was displayed. All the sample sounds can be found on the project website at <http://ztch.s.p.lodz.pl/~radecki/ISON2016/The> APK file for the image sonification software will be made available there as well.

As seen from the trajectories, the direction of the finger movement changes only after leaving the intended path. To provide additional information if the edge of a path was near, the paths were locally blurred and the task was repeated. The results are seen in Figures 12-14. The paths are evidently tracked more precisely, although at a cost of speed. Due to the increased complexity of the synthesized sounds, all subjects took more time to follow the tracks whenever they were blurred.

Due to the small number of participants and no quantitative data was gathered. The presented sonification scheme was just a proof-of-concept for a larger study planned in the near future.

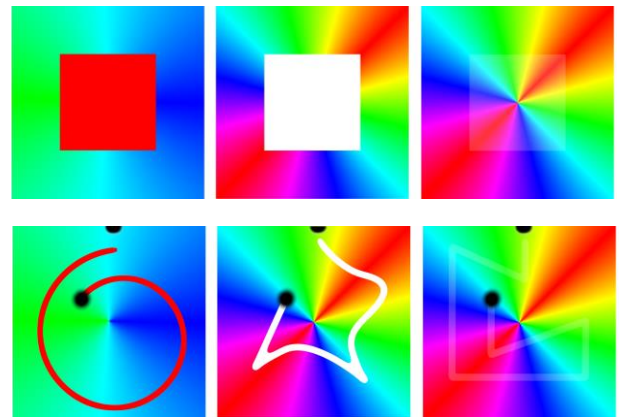


Figure 7. Training image set with a large central field corresponding to the path properties in the trajectory following tests (top) and the test images with the trajectories to be traced by finger using only sonification (bottom).

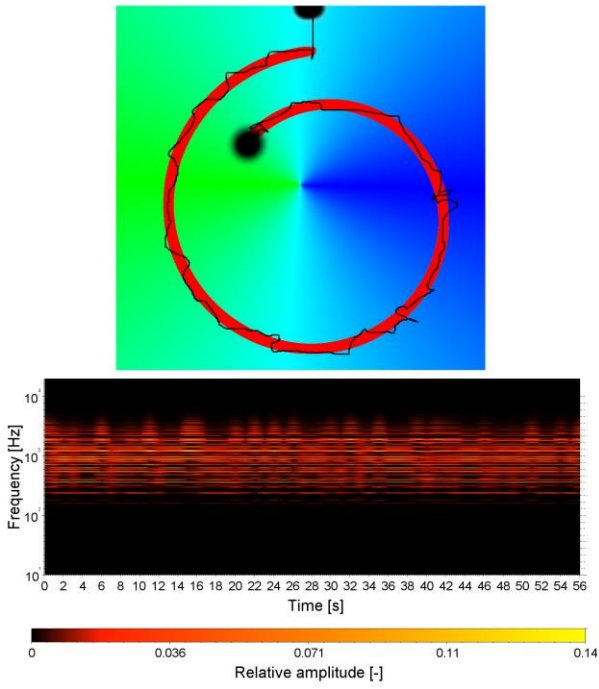


Figure 8. Tracking a path with minimum hue ( $H=0$ ).

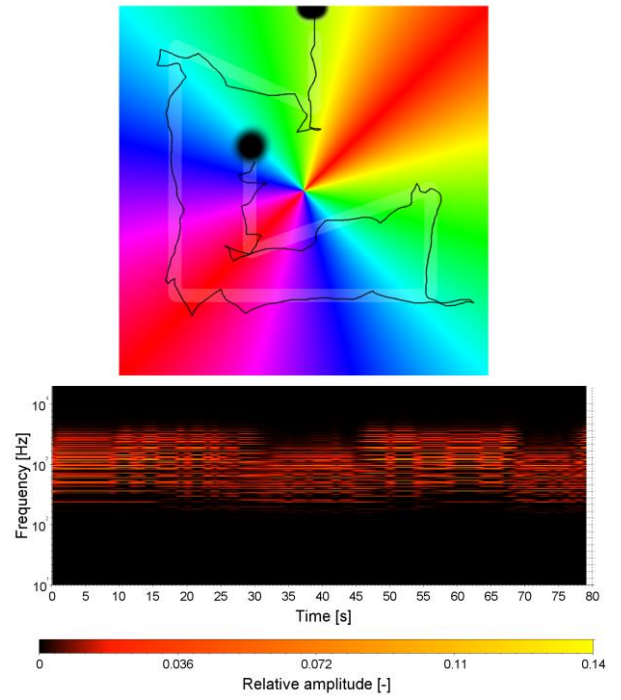


Figure 10. Tracking a path that has 20% less saturation than the background ( $S = 0.8$ ).

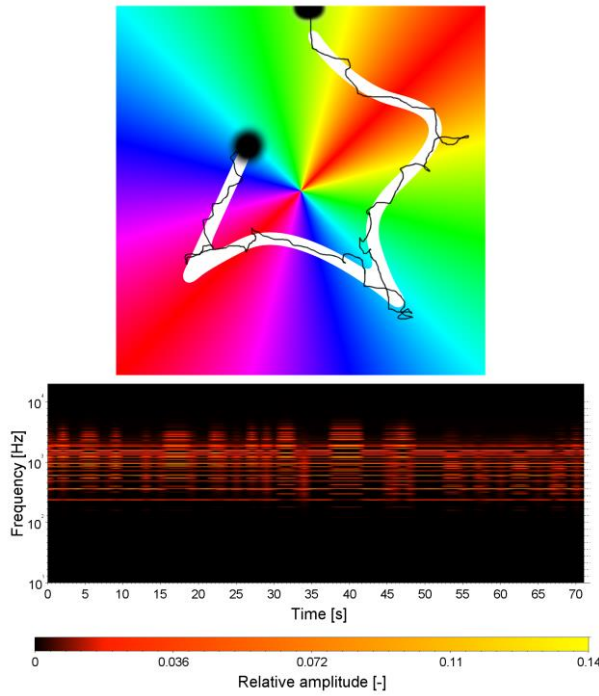


Figure 9. Tracking a path that has maximum brightness ( $V=1$ ).



Figure 11. Test images with blurred trajectories.

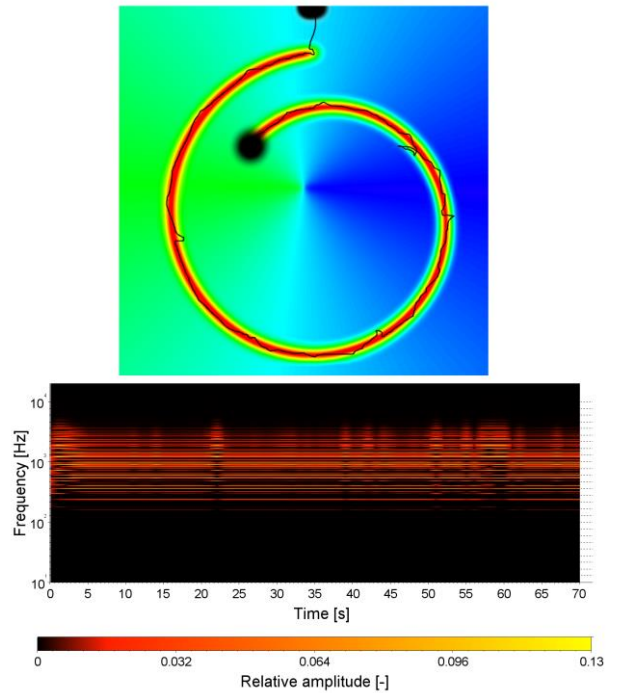


Figure 12. Tracking a blurred path with minimum hue ( $H=0$ ).



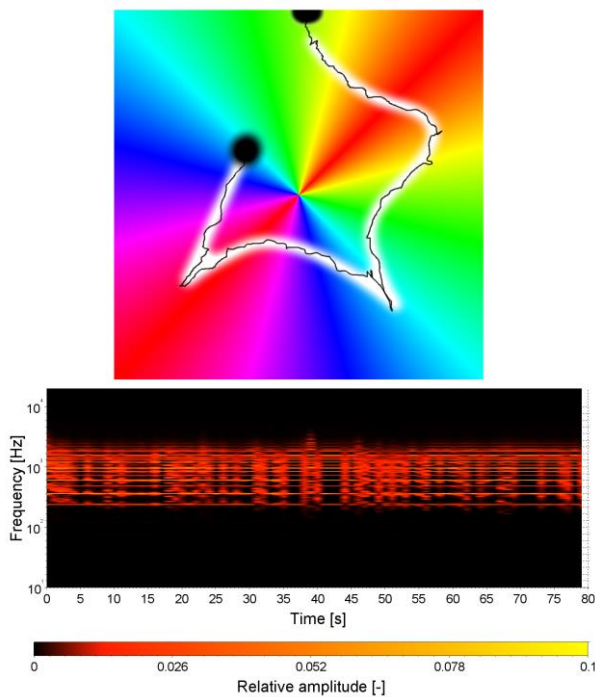


Figure 13. Tracking a blurred path that has maximum brightness ( $V=1$ ).

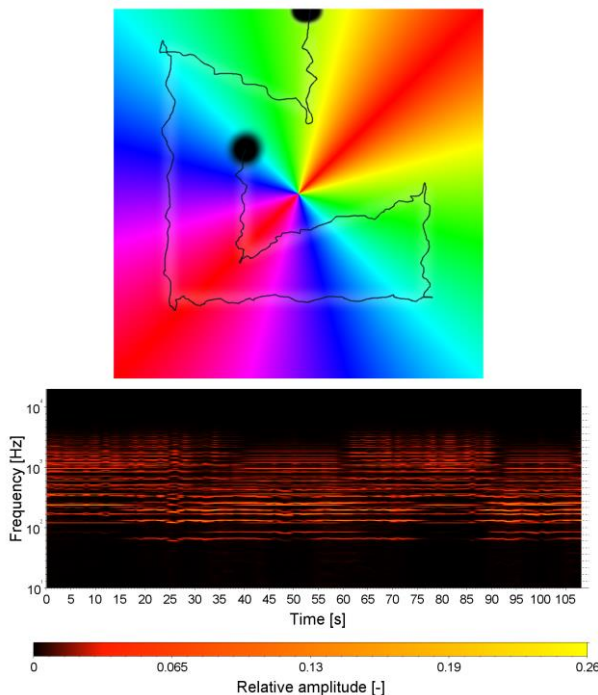


Figure 14. Tracking a blurred path that has 20% less saturation than the background ( $S = 0.8$ ).

#### 4. CURRENT AND FUTURE WORK

The presented sonification approach is the first of a number of approaches proposed in the project. The developed tools are intended to help conduct a larger study with a group of blind

school children, which will study if interactive sonification of images and maps could be added to their curriculum.

In the developed interactive sonification tool multiple image filters can be used to enhance an image, examples include color depth limitation, edge detection or Gaussian smoothing. Currently the image processing method is chosen manually, but an automatic adjustment basing on image content is being considered. Images loaded into the software can also have supplementary text information in xml files containing labels assigned to image regions, e.g. street names for maps, written descriptions of image content or instructions for educational tasks.

Two main categories of sonification approaches will be tested – cyclic, which creates looped waveforms with amplitude envelopes changing in short 1-2 s cycles, and non-cyclic, where the sound amplitude depends solely on the touch gestures as in the presented paper.

The sound synthesis will be performed by summing and amplitude modulation of up to 16 modifiable sound buffers, which can be read at different frequencies. This approach ensures smooth operation on most Android mobile devices. Additionally, Text-to-speech synthesis will be used for images with the supplementary xml text information.

Several tools were prepared for the logging of user interactions with the software, e.g. to track touch gestures and interactions with the interface. This will aid in documenting the experimental trials, once the software is tested on a larger group of blind children and adults.

The main goal of the trials will be to determine whether interactive sonification can be used to supplement or replace tactile materials used in classes for blind children.

To encourage reproduction of the research, the authors are considering preparing the experiments within the SonEX framework [16].

#### 5. CONCLUSIONS

The article presents an interactive image sonification approach using the HSV colour model. The main strengths of the proposed approach are:

- a simple synthesis algorithm (additive synthesis by combining a small number of audio buffers)
- transforming the circular nature of the hue component
- simplifying the synthesized sound with the decrease of the color complexity (smaller S value).

So far the model has shown potential usefulness of the continuous transform and of image blurring for smooth sound transitions. The prepared tools and the path tracking experiment are a proof of concept for a bigger scale test aimed at utilizing interactive sonification for the education of blind children.

#### 6. ACKNOWLEDGEMENTS

The work was supported by the National Science Centre of Poland under grant no 2015/17/B/ST7/03884 in years 2016-2018.

## 7. REFERENCES

- [1] M. Capp and P. Picton, "The optophone: an electronic blind aid" in *Eng. Science and Education Journal*, 2000, 9(3), pp.137-143.
- [2] M. Jameson,, "The Optophone: Its Beginning and Development", *Bulletin of prosthetics research*: 25–28, 1966
- [3] G. Kramer, "Some Organizing Principles for Representing Data With Sound", In G. Kramer (Ed.), *Auditory Display: Sonification, Audification and Auditory Interfaces*, SFI Studies in the Sciences of Complexity Proceedings (Vol. XVIII), Reading, MA: Addison-Wesley Publishing Company. 1994
- [4] T. Hermann, A. Hunt and J. Neuhoff, *Sonification Handbook*. Logos Publishing House, Berlin, 2000.
- [5] A. Hunt, T. Hermann, and S. Pauletto. "Interacting with sonification systems: closing the loop." *Information Visualisation*, 2004. IV 2004. Proceedings. Eighth International Conference on. IEEE, 2004.
- [6] N. Degara, A. Hunt, & T. Hermann, "Interactive Sonification" [Guest editors' introduction]. *IEEE MultiMedia*, 22(1), 20-23 2015
- [7] G. Dobus and R. Bresin, "Sonification of physical quantities throughout history: a meta-study of previous mapping strategies" in *Proceedings of the 17th International Conference on Auditory Display*, 2011.
- [8] M. Bujacz, P. Skulimowski, P. Strumiłło, "Naviton - a prototype mobility aid for auditory presentation of 3D scenes", *Journal of the Audio Engineering Society*, Vol. 60, No. 9, 2012 September, pp. 696-708
- [9] P. Baranski, P. Strumillo "Emphatic Trials of a Teleassistance System for the Visually Impaired" *Journal of Medical Imaging and Health Informatics* 5 (8), 1640-1651
- [10] R. Sarkar, S. Bakshi and P. K'Sa, "Review on Image Sonification: A Non-visual Scene Representation" in *Int. Conf. on Recent Advances in Information Technology*, 2012.
- [11] G. Dubus and R. Bresin, "Sonification of physical quantities throughout history", *The 17th International Conference on Auditory Display*, 2011.
- [12] Meijer, P. "An Experimental System for Auditory Image Representations" *IEEE Transactions on Biomedical Engineering*, 39, pp. 112-121, 1992
- [13] Matta, S.; Kumar, D. K.; Yu, X. & Burry, M. (2004), 'An approach for image sonification', *First International Symposium on Control, Communications and Signal Processing*, 431-434.
- [14] O'Neill, C. & Ng, K. (2008), 'Hearing Images - Interactive Sonification Interface for Images', *International Conference on Automated solutions for Cross Media Content and Multi-channel Distribution*.
- [15] S. Cavaco, "From pixels to pitches - Unveiling the world of color for the blind", *IEEE 2nd International Conference on Serious Games and Applications for Health*, 2013.
- [16] N. Degara, F. Nagel, T. Hermann, "Sonex: An Evaluation Exchange Framework For Reproducible Sonification". *Proc. of the 19th International Conference on Auditory Display (ICAD-2013)*. Lodz, 2013, pp 167-174