SOUND-BASED IMAGE AND POSITON RECOGNITION SYSTEM ∼ SIPReS ∼

Shin'ichiro Uno, Yasuo Suzuki, Takashi Watanabe

Nihon Fukushi University
Faculty of Health Sciences
Department of Human Care Engineering
26-2 Higashi-Haemi-cho Handa,
Aichi 475-0012, Japan
uno@n-fukushi.ac.jp
yasuo-s@n-fukushi.ac.jp
wata-t@n-fukushi.ac.jp

Miku Matsumoto, Yan Wang

Shouei Denshi Kenkyusyo Ltd.,Co.
26-145 Icchouda Kutsukake-cho Toyoake,
Aichi 470-1101, Japan
hirate@shouei.co.jp
wang@shouei.co.jp

ABSTRACT

We developed software called SIPReS, which describes two-dimensional images with sound. With this system, visually-impaired people can tell the location of a certain point in an image just by hearing notes of frequency each assigned according to the brightness of the point a user touches on. It can run on Android smartphones and tablets. We conducted a small-scale experiment to see if a visually-impaired person can recognize images with SIPReS. In the experiment, the subject successfully recognized if there is an object or not. He also recognized the location information. The experiment suggests this application's potential as image recognition software.

1. INTRODUCTION

1.1. STEM and data sonification

One of the important study themes in the assistive technology field is how to guarantee visually-impaired people's right to access information. Science Technology Engineering and Mathematics, or STEM, has been held to be important in recent years[1]. Emphasis is put on STEM also in the education field for disabled people[2]. Actually, there are more opportunities in the STEM field for disabled people to be independent and to exercise their abilities by making use of information technology. That is why efforts for better STEM education are being made in many areas such as in the education system, environment improvement, or software/ hardware development.

One of those efforts has been made in the study of conveying data through the medium of sound. Dubus and Bresin have surveyed 179 papers on sonification for the past 20 years, and made a database of them. Many of those papers are about sonification in the scientific fields. One recent example in the field of astronomy is sound models of the eight solar planets made by Michael Quinton. He reported visually impaired people can tell each planet's feature by sound[3].

Our project, the Astronomical Data Sonification Project, aims to develop methods with which space science data can be understood by visually-impaired people without depending on the visual information. The project is promoted by two institutions: Uno Laboratory under the Faculty of Health Sciences at Nihon Fukushi University and the Institute of Space and Astronautical Science(ISAS) of the Japan Aerospace Exploration Agency(JAXA). In this project, we explored possibilities of sonification of space science data. We already accomplished sonification of X-ray pulsar data [4] and sonification of Kp indices indicating the strength of geomagnetic disturbance[5]. In addition, we made prototype software called splot and sonified general one-dimensional histogram [6], [7]. After all these achievements, sonification method of two-dimensional data of astronomical bodies has been anxiously awaited.

1.2. Tactile Image Recognition Environment for Visually Impaired People

Various efforts have been made to develop methods for visually impaired people to recognize images.

Examples of tactile image representation are a) braille embossing printer, b) tactile images or three-dimensional copies made on capsule paper, and c) pin display, all of which are designed to give tactile information to readers.

A braille embossing printer can draw figures with raised dots just as an ordinary braille printer does when it prints letters. One of its advantages is that it can print letters along with images. So it is very useful for books such as expository books or braille books. One of its disadvantages is that printing requires time as it prints on paper. Another disadvantage is portability. It's too heavy to move it around. Figure 1 is an example of a chart made with a braille printer.

Tactile images or three-dimensional copies are made on special paper called capsule paper. It has a layer of capsules which expand with heat. An image is expressed by elevation of the paper. More specifically, the process is two-fold: making a copy with an ordinary copier and heating the paper with a dedicated machine. Then, a three-dimensional paper will be printed out. Compared to the other types of tactile image representation methods, it is easier to make three-dimensional diagrams on paper. Three-dimensional diagrams on paper are relatively easily made. The dedicated ma-

http://doi.org/10.21785/icad2018.005

chine, however, has similar disadvantages as a braille printer. One of them is its size. It is bulky.

A pin display is a display to show what is on the computer display just like an ordinary braille display does. Figure 2 is an example of a pin display [8]. Its advantages are its easy connection with a personal computer and real-time display. On the other hand, the number of actuators is limited and its density is fixed. So the size of what can be shown on the display is limited and its granularity is coarse. In addition, it is too pricey, as it consists of special parts.

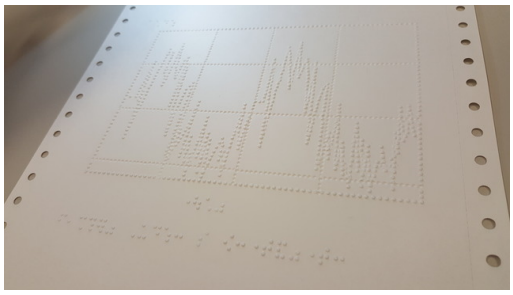


Figure 1: Example of an image printed with a braille embossing printer

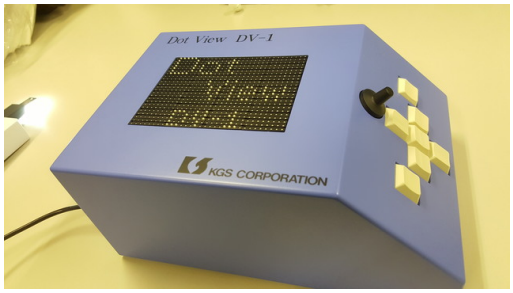


Figure 2: Example of a pin display

1.3. Auditory-Visual Sensory Substitution (AVSS) System

Many attempts have been made so far to convert two-dimensional image information into sound. This type of conversion is called AVSS, or Auditory-Visual Sensory Substitution System.

"The vOICe" is the first general-purpose AVSS[9], on which various researches have been conducted[10, and references there in]. "The vOICe" firstly cuts image information into vertically long strips, then allocates lower-part information of each strip to low frequency, higher part to high frequency, and brightness of each pixel to sound volume, and finally generates composite tone. It passes along two-dimensional information to a user by playing the composite tone along with its automatic sweep at a constant rate in a horizontal direction. The user can grasp a non-uniform image or space by the sweep.

Another system called GUESS creates two-dimensional auditory space and passes along the information of a shape to a user by moving the location of sound[11]. GUESS system has succeeded in recognition of a circle or a triangle. In addition, it conveys the location information with the help of a device such as a tablet.

Another system called sonicphoto allocates information of vertical section to sound pitch and that of horizontal section to time. It represents image by sound along with automatic scanning at a constant rate. The program sonicphoto is available on the Internet[12].

Another system called iSonic developed at the University of Maryland [13] is an interactive system and enables users to understand a choropleth map linked with data in a table. With operation using arrow keys or a touchpad, users can understand data such as statistics linked to a certain point on the map[14]. Using iSonic, Delogu *et al.* compared recognition rates of sonified maps with a keyboard or a touch pad. They showed potential of iSonic with touchpad operation, though its serviceability depends on the type of a target map and statistical data[15].

2. DEVELOPMENT

In this study, we tried to develop image description software using sound. By describing image information with sound on a small device, we aimed to achieve image recognition method with alacrity and portability, which were difficult to achieve with conventional methods. We set our eyes on smartphones and tablets to achieve our goal, since those devices have been rapidly spread among people. System features requirements are as follows:

- It can run on a smartphone and also on a tablet.
- It can recognize the coordinate of the point a user touches on the panel, and strike a note according to the brightness of the point.
- It can run camera application on the device, and take a picture and import it.
- It can read image files such as jpg, png, and pdf.
- It can read screenshot images.

We gave it a name Sound-based Image and Position Recognition System, or SIPReS. And its target operating system is Android. The development platform is Eclipse version 4.6.2, development language is Java version 8, and the current version of SIPReS is 1.0.

2.1. Sonification

The system will strike a note of an assigned frequency according to the brightness of the point a user touches, and the note will continue as long as the user touches the point. Input images can be color or black and white. For black and white images we use the brightness value without any conversion. For color images, we convert RGB data to brightness information (Equation 1) [16] .

$$y_n = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (1)$$

Next, we derive the frequency value of a note it will strike based on the brightness information. Firstly, we read the minimal and maximum value of the brightness of an image and assign them to $y_{\min}$ and $y_{\max}$ respectively. Then we set y_n as the brightness of the pixel the user touches. y_n's frequency, f_n, is calculated using equation 2 below.

$$f_n = f_{\min} \times 2^{\left(\frac{(y_n - y_{\min})}{(y_{\max} - y_{\min})} \times \log_2 \frac{f_{\max}}{f_{\min}}\right)} \quad (2)$$

$f_{\min}$ and $f_{\max}$ are frequency values according to the brightness and can be set by users. Default value of $f_{\min}$ and $f_{\max}$ are 220 Hz and 1760 Hz respectively. y_n can be a) the brightness

value of a single pixel touched or b) the average brightness value of r layers of surrounding pixels from the point the user touches. The default value of r is 3, but it can be changed by users.

2.2. Application

After starting the application, a menu will come up for selection of an input source. A user can choose a file or take a picture for sonification here. According to the point the user touches, it strikes a note.

There is a menu icon in the upper right corner of the application. From the menu a user can choose an image or change settings of the application. It can also show the coordinate of the point touched, x and y, its brightness, y_n, and corresponding frequency, f_n, in the upper part. In Figure 3, we show the SIPReS's display.

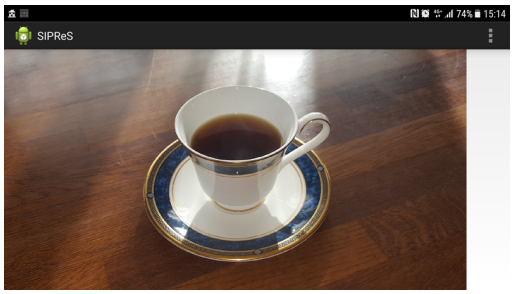


Figure 3: SIPReS's display. An image of a coffee cup is shown on it. There is a blank space due to the difference between the image's aspect ratio and the number of pixels of the display, and/or the height of the menu bar.

3. EXPERIMENT

We carried out a small-scale experiment to see if SIPReS enables readers to recognize images. The aim of the experiment at this stage is just to see if the application is useful or not. So the number of subjects is only one this time and used simple images. The subject is a 31-year old man blind from birth. We will experiment with a number of subjects using more complicated images from next time onwards.

Devices used in the experiment are a tablet, Galaxy Note Pro and smartphone, Galaxy Note 5, both of which were made by Samsung Electronics Co., Ltd. We installed SIPReS ver 1.0 to both devices and ran it with default settings. The specification of both devices is shown in table 1. Since images for the experiment were horizontally long, we used Galaxy Note 5 turned sideways.

3.1. Simple Images

Experiment 1 is the simplest. We had the subject look for the center of a white circle against the black background and touch the point. The white circle is gradational with the whitest center. The size of the picture is 2560×1440 pixels, and the radius of the circle is 300 pixels. The center is located 30% lower from the top and 30% right from the left, or at the coordinate of $(768, 432)$. In figure 4(a), we show the image used in the experiment.

The subject touched the tablet's display with his finger and based on the sound information gave us an oral answer on the location of the center, or how far, or the ratio on the whole surface,

Table 1: Device used for evaluation

	galaxy note pro	galaxy note 5
developer	samsung	samsung
display size (inch)	12.2	5.7
display resolution (pixel)	2560×1600	2560×1440
dimension (H/W/D, mm)	$204/295.6/7.95$	$153.2/76.1/7.6$
bezel (u/d/l/r, mm)	19/19/15/15	14/12/2/2
OS(android) version	5.0.2	7.0

The values of bezel are measured in the experiment. Other figures are extracted from the device catalogs.

the point is located from the upper left corner. When the subject said he found the point, we read the coordinates displayed on the SIPReS. Although we did not set any time limit, he reached the brightest point and gave an answer of a ratio of x and y in a matter of seconds.

Because of the frame of the device and the fringe parts of the application such as the menu bar, the center shows up at a little different location. As the frame size of a device is different from model to model and it is difficult to tactually tell the border between frame and display, we have decided not to offset this gap. The table 2 shows a) the location of the center in the picture, b) that in the whole device including its bezel, c) his answer on the ratio, and d) the real point he touched.

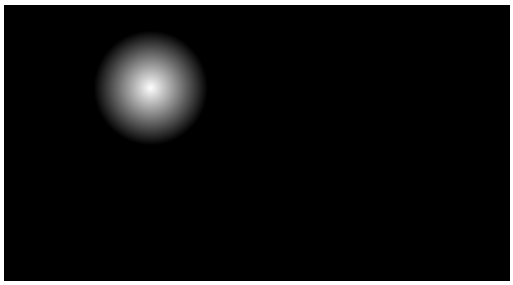
As the table 2 shows, the point he touched is almost correct with a margin of just 30 pixels or 2% of the size of the image. Though the point he touched is almost accurate, the margin of error of his oral answers is a little bigger. We should keep in mind that bigger margin of error is expected for location recognition due to the bezel of a device or the space for application's menu.

Table 2: The real location of the center of the circle and the subject's answer

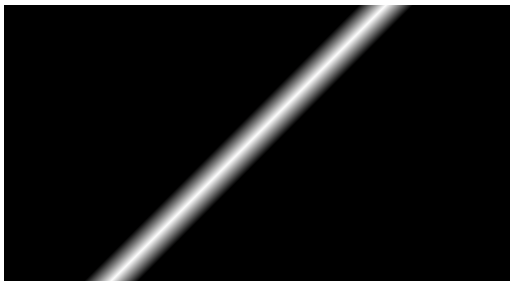
	galaxy note pro	galaxy note 5
Entire tablet (include bezel)		
from left	93mm(31.5%)	45mm(29.4%)
from top	80mm(39.2%)	32mm(42.0%)
Coordinates in the display (scaled)		
x,y(pixel)	(741,417)	(486,273)
Subject's answer (pointing)		
x,y	(775,418)	(491,290)
residual	(+34,+1)	(+5,+17)
(impressions including bezel)		
from left (%)	30%	30%
from top (%)	30%	30%

3.2. Experiment 2

We use an image of a slant line for the experiment 2 to see if the subject can trace the line. The line is gradational like the circle we used in experiment 1. The center in the cross direction is the whitest and the width is 200 pixels. It starts from the lower left corner, passes the center of the image, and ends in the upper right



(a) finding circle



(b) tracing line

Figure 4: Sample images for experiment 1 and 2. The image size is both 2560×1440 pixels. In the image (a) there is a circle with 300-pixel radius at the point 30% lower from the top and 30% right from the left. In the image (b) there is a line passing the center of the image at a slant angle of 45 degrees.

corner at a slant angle of 45 degrees. Figure 4(b) is the image we used.

All he knew before the experiment is there would be an image of a line. Then in the experiment, he gave us an oral answer on the location of the start points and end points along with the direction. He gave us answers like "the start point is 20% right from the lower left edge," and "the end point is 80% right from the upper left edge." With the prior information, he successfully trailed points which struck high notes. He told us, the clearer the border against the background, the easier he finds the line.

3.3. Experiment 3

We also carried out an experiment on SIPReS's another function, picture-taking function. One advantage of image recognition by sound is not limited to recognition of prepared images, but also of a shape in a picture taken on the spot. To make sure the function works as we intended, we experimented to see if the subject could tell the rough image of an object in a picture. We used a picture of a banana and an orange. Figure 5 is the picture we used.

After we presented the picture, the subject recognized there were two objects, one on the left was vertically long, and the other on the right was close to a circle. That said, he could not tell exactly what those objects were. After we told him there were a banana and an orange, he answered the left object was the banana. So if certain conditions are met, there is a possibility of providing shape information to the extent to which visually-impaired people can tell the difference between a round object and an elongated object.

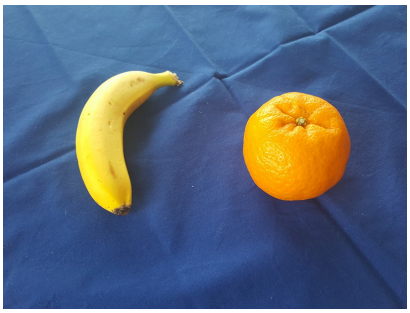


Figure 5: These fruits were used for a picture to see if the subject can recognize the shape of objects in it with SIPReS. We put the fruits on a blue cloth since the clearer the contrast between the object and the background is, more recognizable.

4. DISCUSSION

This study indicates possibilities for visually-impaired people to roughly recognize images using SIPReS on a tablet or on a smartphone. We believe the possibilities of location recognition without depending on visual information but with the combinatory technology of two-dimensional location information on the touch panel and information of different tones is very meaningful.

Having said that, we have not collected enough amount of data for verification yet. So we need to conduct larger scale of proving tests and assess its effectivity. As to the types of images, we may need to use more complicated ones such as a line of a rectangle or a star, or a lump of shapes such as a function or a chart. In addition, we may need to ask more subjects to join the experiment.

The current version of SIPReS 1.0 does not support video sonification. If it does, it is expected that users would recognize the change of image by the change of sound with the finger staying at the same spot. This is a challenge we should address from now.

4.1. Positioning

As mentioned above, the subject successfully located the center of the circle in experiment 1. He firstly moved his finger just like scanning the whole area, then looked for the center after recognizing the sound difference. He told us it was easier on a bigger tablet than on a smaller smartphone.

The gradation of the circle turned out to be very helpful. The subject understood the higher tone means brighter spot. With SIPReS which strikes a tone in real time, he looked for points with a higher tone and finally spotted the point. By making use of this mechanism such as changing gradation pitch, visually-impaired people might be able to recognize more complicated images such as a pointed cone, a cone with a bigger bottom, or hemisphere.

As we saw in experiment 2, however, it is easier to reach the point with the highest tone if the border between the line and the background is clear. In a gradational image, the frequency of the tone changes even by a tiny location change. This leads us to think it is easier to trace a line with the same tone without gradation. Line recognition could possibly be utilized for map reading or guideboard.

4.2. Sound Parameters

The default frequency range of SIPReS is from 220Hz to 1760Hz, three octaves across the standard pitch of 440Hz. We chose the range because it sits in the middle of frequencies that can be heard by humans and is expected to easier to distinguish.

Having said that, among those who have multiple difficulties such as auditory and visionary, there are people who cannot hear sound of specific frequency range. And the range they cannot hear varies from person to person. Therefore the output sound range is modifiable and can be set by users.

In the current version of SIPReS, output sound is formed by sinusoidal function. We had other sound form options such as triangle or box. However we deployed a simple sinusoidal function only, as our main aim, at least for now, is correspondence between the touched location and sound.

We think, however, allocation of multiple sound forms is effective to some users. Furthermore, there is a possibility to express subtle difference of colors by correlating each RGB parameter to specific sound form, which may lead to expression of more complicated information. Sound form does not have to be limited to mathematical formulas. There is also a possibility to express color and a shape with a chord, or even a discordance, by incorporating sound form of musical instruments.

4.3. Taking Picture on the Spot and Its Recognition

Experiment 3 indicates possibilities of recognition of a picture taken on the spot as long as it has a simple image on it. That said, the application's current ability is limited to recognition whether there is an object or not. We cannot tell how minutely people can recognize images. We need to start to think how we can evaluate it first. The subject made a suggestion that the application might be used when visually-impaired people look for something like a key they dropped inadvertently.

Let me talk about an example of image recognition requests here. The Astronomical Data Sonification Project is getting a request from a visually-impaired space enthusiast. He wants to join an astronomical observatory meeting and sense the waxing and waning of the moon. The moon is in stark contrast to the night sky, and the difference in shape between the crescent moon and the full moon is expected to be easily recognized. We take this as one challenge we should address.

4.4. Application Possibility of SIPReS

We developed SIPReS as one image recognition method for visually-impaired people. But it could be used for other purposes other than image recognition.

We tried to have SIPReS play music as one possible application. Figure 6 is a score of simple music. Each part's brightness is set so that its corresponding frequency can strike a tone of the musical scale. To clarify the maximum and the minimal there are two parts in the upper right corner showing 0% brightness and in the lower left corner showing 100% brightness. Simple music is played, while a user moves their finger in a set direction.

As we experimented, we found it is a little bit more difficult than expected to move the finger straight and at the same pace. Though it takes some getting used to and complicated music cannot be played, there is a possibility that anybody with or without visual disability can play music and that even those who cannot read scores or have not played instruments can play music just by moving their finger on the tablet display.

Figure 6: Music score for SIPReS. By swiping from left to right at the same pace, one can hear songs played by SIPReS. By calculating brightness back from frequency, anybody can make music scores of a variety of songs. This example is the score of a song called Charmera, which was used by conventional ambulant vendors.

5. CONCLUSION

We developed SIPReS which describes a two-dimensional image with sound. SIPReS has possibilities that even a visually-impaired user can roughly recognize an image by hearing an assigned tone according to the brightness of the point a user touches. The subject, who is totally blind, successfully recognized whether there is an object or not along with the location information. The experiment indicates this software's practicability. We hope SIPReS can be applied to new methods of image recognition.

6. REFERENCES

[1] J. J. Kuenzi, "Science, technology, engineering, and mathematics (stem) education: Background, federal policy, and legislative action," *Congressional Research Service Reports*, 2008.

[2] J. Hwang and J. C. Taylor, "Stemming on stem: A stem education framework for students with disabilities," *Journal of Science Education for Students with Disabilities*, vol. 19 Iss.1 Article 4., pp. 39–49, 2016.

[3] M. Quinton, I. McGregor, and D. Benyon, "Sonifying the solar system," in *Proc. of the 22nd Int. Conf. of Auditory Display(ICAD)*, 2016, pp. 28–35.

[4] S. Uno, T. Kameyama, M. Horihata, N. Asano, K. Ebisawa, T. Tamura, I. Shinohara, O. Miyashita, A. Miura, K. Matsuzaki, H. Murakami, and F. Furusawa-Akimono, "The astronomical data sonification project: An overview of the project and its initial data output," *Journal of health sciences, Nihon Fukushi University*, vol. 10, pp. 1–9, March 2007.

[5] A. Miura, S. Uno, T. Kimura, and K. Ebisawa, "Visualization and sonification of space science data -for education and public relation-," *Journal of Space Science Informatics Japan*, vol. 1, pp. 13–22, May 2012.

[6] S. Uno, S. Sotoya, A. Miura, and K. Ebisawa, "The astronomical data sonification project 2: experimental production of sound-based data plot software splot," *Journal of health sciences, Nihon Fukushi University*, vol. 14, pp. 1–9, March 2011.

[7] S. Uno, S. Sotoya, A. Miura, and K. Ebisaw, "Current status of the astronomical data sonification project," *Journal of Space Science Informatics Japan*, vol. 1, pp. 7–11, May 2012.

[8] https://www.kgs-jpn.co.jp/index.php.

[9] P. B. L. Meijer, "An experimental system for auditory image representations," *IEEE Transactions on Biomedical Engineering*, vol. 39 Iss.2, pp. 112–121, 1992.

[10] https://www.seeingwithsound.com/literature.htm.

[11] H. M. Kamel, P. Roth, and R. R. Sinha, "Graphics and user' s exploration via simple sonics (guess): Providinginterrelational representation of objects in a non-visual environment," in *Proc. of the 2001 Int. Conf. on Auditory Display(ICAD)*, 2001, pp. 261–266.

[12] http://www.skytopia.com/software/sonicphoto.

[13] http://www.cs.umd.edu/hcil/audiomap/.

[14] H. Zhao, C. Plaisant, B. Shneiderman, and J. Lazar, "Data sonification for users with visual impairment: A case study with georeferenced data," *ACM Transactions on Computer Human Interaction*, vol. 15 Article 4., 2008.

[15] F. Delogu, M. O. Belardinelli, M. Palmiero, E. Pasqualotto, H. Zhao, C. Plaisant, and S. Federici, "Interactive sonification for blind people exploration of geo-referenced data: comparison between a keyboard-exploration and a haptic-exploration interfaces," vol. 7. Springer Berlin/Heidelberg, 2006, pp. 178–179.

[16] "Itu-r recommendation bt.470-6, conventional television systems," 1998.