I J C T A, 9(11) 2016, pp. 5343-5350

FROM PIXELS TO PITCHES: UNDERSTAND THE IMAGE THROUGH SOUND

Raj Kumar[*] and Ajmer Singh[**]

Abstract: A sonification is totally depending on the perception of things. An image sonification is used to help the visually impaired people of the society. This paper proposed a "time-frequency mapping" algorithm which perceives the open shapes consists in a digital binary image. The goal of this algorithm is to assign each pixel values of the binary image with sinusoidal signals of changing frequencies. This helps visually impaired people to hear or taking the perception of the shapes present in the binary image.

Key Words: auditory display; probing; scanning; image sonification; spectrogram, open-shapes.

1. INTRODUCTION

Imagine listening to the flying bird and able to differentiate flying of different birds through sound. How brain wave seems to be sound like? How picture of galaxy sound like? How X-rays of variable frequencies sound like? How can sound be used for analyzes big data over a year? Answers to these entire questions are the domain of auditory display and sonification. Researchers examine how the human auditory system can be used for communicating and transmitting information. The goal of the auditory display system is to detect and finding out the changes in the structure of data or in the other words, any changes in the data would be detected by listening to the sound. So, sonification is the core component of the above discussion. It defined as the technique which converts any form of the data relation into the auditory relation. Also, define as the technique of translating sound in the response of data and interactions [1].

1.1 Sonification

When the data or information is conveyed to the human ears through sound is term as sonification. Recent definitions have focused more on the objective in terms of scientific method. Music, which is described as the entertainment purpose, needs new means of definition and expression. According to the researchers, that sonification is the new way to use the sound to express the data. So, the sonification is the transformation of the data relations to the auditory relation for the purpose of facilitating communication [1]. One example of sonification used in daily life is Earcons. Earcons are the sound used to convey the information to the user without actually see the things. For e.g. different ringtones applied to the different contact so that user by hearing the sound knows who is calling or bell sound used at the doorstep to indicating that there is someone at the

[*] School of Electrical and Electronics Engineering Lovely Professional University, Phagwara, Punjab
Email: raj.kumar2208@yahoo.in

[**] School of Electrical and Electronics Engineering Lovely Professional University, Phagwara, Punjab
ajmersinghbhinder@gmail.com

door. The subset of auditory display is sonification and it offers propagation between the information source and receiver (shown in figure 1)

Figure 1 General description of a communication system

Image Sonification

Imagine in future human can able to hear an image. An image is nothing but actually 2D static data and what if the human being is capable of detecting the intensity of the image in the presence or the absence of dark light through sound. For this, image sonification came into the picture. The principle dimension of the sound is time and all its parameter placed accordingly [1]. Now the challenge is an effective sonification of time independent images and their application to music.

Image sonification has software of vast majority uses the "time-frequency" approach in which an image acts as the spectrograph for a sound. As the image is drawn, but Audition is software allows the use of bitmap images and the vertical axis is represented by color as the intensity. MethSynth uses the image parameter i.e. color to represent the stereo position of the sound. The blind people live life on natural sound. They learn to neglect the useless sounds [2] and filter out those that are important. So, the image sonification is used to help them. "Mathtalk" is a program that translates a text-to-speech with non-verbal sound to make it easier to understand mathematical expression to the listener. "Audiograf" is a program that used to select a diagram with a finger on a touch-screen and generate a sound. In the situation where the need of moving the eyes is risky for the information such as driving in an emergency at that place, sound can be played an important role in acquiring the information.

The visible spectrum ROYGBIV [3] could be listened by using the sound of inverse correspondence of frequency. Low light frequency has a high sound wave frequency to provide strength and high light frequency has a low sound wave frequency. The two types of time-mapping concept for image sonification i.e., scanning and probing [4]. When the data is in the order, in that refers the term scanning mapping is done for sonification. It refers as inverse spectrogram method, which is a popular method for image sonification. (shown in figure 2)

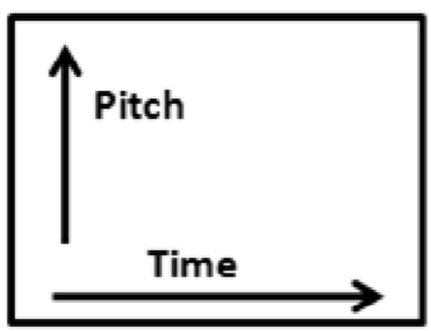


Figure 2: Inverse spectrogram scanning

In figure 2, the X-axis is mapped to the time, and at each time pitch is going to change or at each step of time sonification of corresponding vertical line. The fixed speed of scanning usually makes this method not dynamic. For dynamic scanning, another mapping introduce is probing where the speed and the "pointer" location can be varied by a user during sonification. In this method, there is a selection of pointer which can be a single point, a line/curve, an area, and set of distributed points. This method provides more flexibility for location/path or speed of sonification. The prototype for the conversion of image color information attributes to the sound parameters (pitch, timbre, volume, localization) [5]. A prototype used HSV model, having a digital image which is composed of discrete units of color or pixels. The author proposed a tool which converts color information into sound. For that purpose, Hue, Saturation, and Value (HSV) color model are used. Hue contains "pure color" information which mapped into the fundamental frequency of the sound. Saturation defines the deviation of color from grayscale which mapped into the spectral envelope of the sound wave to provide the perception of timbre. A value called intensity or brightness of the color which mapped into the volume of the sound. (shown in figure 3) There is the fourth parameter is pixel's abscissa used to define the location of the sound source by mapping the pixel's abscissa into azimuth.

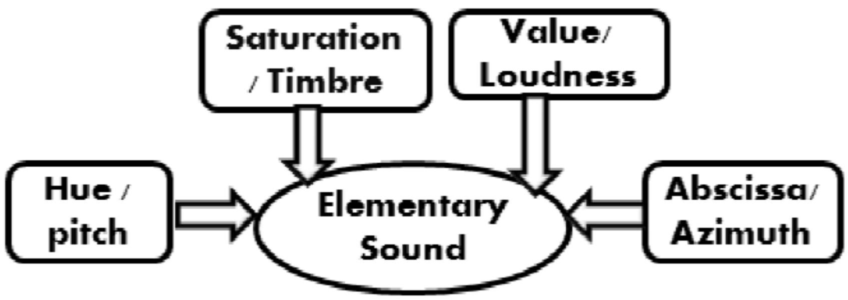


Figure 3: Elementary sound relationship between the image color models (HSV model) to sound parameter

The paper is been organized as follows: In section II, the proposed work algorithm used for the perception of open shapes in an image through melody sound is discussed. In section III, we discuss the result of the work and section IV presents the conclusion and future work.

2. PROPOSED WORK

Research and application require sound synthesis and manipulation algorithms in an auditory display that affords carefully controlled sonic results. Auditory display designer may use a wide variety of sound processing/ analysis/ synthesis algorithms that yield to the research in computer music, speech, acoustics and human audio perception.

Time-Frequency Mapping Algorithm

A digital image is composed of pixel values and it can be colored or grayscale digital image. Now algorithm is working in such a manner as consider a digital image of size "32X32", containing an open shape. Next task converts the color or grayscale image into binary image which consists open shapes like a straight vertical line as shown in figure 4(a), a diagonal line as shown in figure 4(b), a reverse diagonal line as shown in figure 4(c), a "greater than" symbol (">") as shown in figure 4(d) or a "smaller than" symbol ("<") as shown in figure 4(e).

(a) (b) (c) (d) (e)

Figure 4: Open Figure in binary image (a) a straight vertical line, (b) a diagonal line, (c) a reverse diagonal line, (d) a greater than symbol and (e) a smaller than symbol

Now mapping of a digital binary image to the sinusoidal matrix in such a way that frequency is changing increasingly for each pixel logarithmically in column wise and amplitude increase row-wise as shown in figure 5. An image is scanned horizontally first, assigning each pixel with some value of frequency and amplitude. And the method used to the scan the image known as raster scanning [7]. Now, select the fundamental frequency for this is 440 Hz. And increase the fundamental frequency logarithmically as shown in equation (1). By fixing the reference frequency f_R, obtain the equal-tempered frequency [6] f_L of interval L (L=0, 1, 2…, 11) as mentioned in equation (1).

$$f_L = f_R * 2^{L/12} \quad (1)$$

After that, an image is going to be played and producing melody sound. After hearing the sound, the perception of open shapes is made to determine which shape is present in a digital binary image.

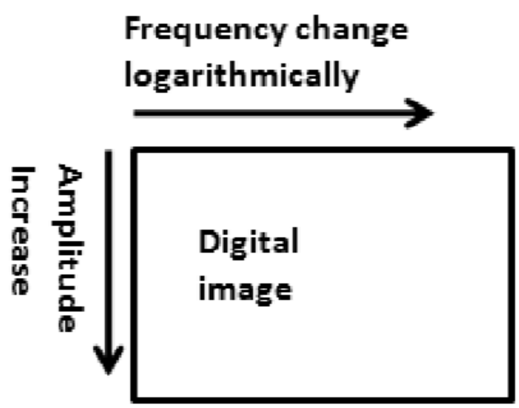


Figure 5: Time-Frequency Mapping

Pseudo-Code for Time-frequency mapping Algorithm

Step 1: Read the binary image of size"$32X32$". Each binary image contains only one open-shape.

Step 2: Take $t = 0.5$ (time is 0.5 second for each tone), $fs = 4400$ (Sampling frequency), $i = 0, j = 0$.

Step 3: Read the $(i, j)^{th}$ element of binary image. If the pixel value of binary image is 0, jump to step 5.

Step 4: Obtained the sinusoidal signal of frequency f_L mentioned in equation (1). Assigned that signal to matrix named as "B".

Step 5: If $j < 31$, increment $j = j + 1$ and go to step 3.

Step 6: If $i < 31$, increment $i = i + 1$ and go to step 3.

Step 7: A matrix named as "B" is obtained which convert into one column matrix named as "Array".

Step 8: Multiply matrix "Array" with Ramp function to increase the amplitude and formed another matrix named as "Array2".

Step 9: Play the matrix "Array2".

3. RESULT

The goal of this algorithm is giving the visually impaired people the perception of the shapes present in an image. A binary image consists of the two values 0 and 1 only. Digital binary images have an open shape and the perception of that shape is getting through the sound. An image is going to be scan horizontally.

Stimuli

The loaded samples were images. Each image consisted of only one open-shape. These open shapes are a straight vertical line, a diagonal line, a reverse diagonal line, a “greater than” symbol (“>”) and a “smaller than” symbol (“<”). After scanning the images, it transformed into sounds as explained in section III. The sound sample with the highest pitch is 2794 Hz and the lowest pitch is 440Hz. Each subject was presented 7 sounds and there were 5 shapes to examine.

Protocol

The order of the samples used is varied to avoid the guessing of the samples and used 3 different orders. The test was performed on a laptop and it would be done using headphones.

The training stage consists of audio samples of the two low and two high frequencies respectively. And also subjects heard the audio samples of the low-to-high frequencies and the high-to-low frequencies respectively with increasing amplitude. The variation of the frequencies in training stage is from 440 Hz to 880 Hz for increasing frequencies and 880 Hz to 440 Hz for decreasing frequencies logarithmically. After hearing the samples from the training stage number of times offered. The actual test started immediately.

Once the real test started, the feedback was not given. The instruction was given to the subjects that they have to choose their best choice whenever in doubt. Subjects allowed hearing the audio number of times they wished to answer.

Test Results

There were 20 subjects from the Lovely Professional University, participating in this study. There were 4 women and 16 men including students and faculties, with ages ranging between 18 and 30. None of them have musical knowledge but some of them have knowledge in identifying the low and high frequency.

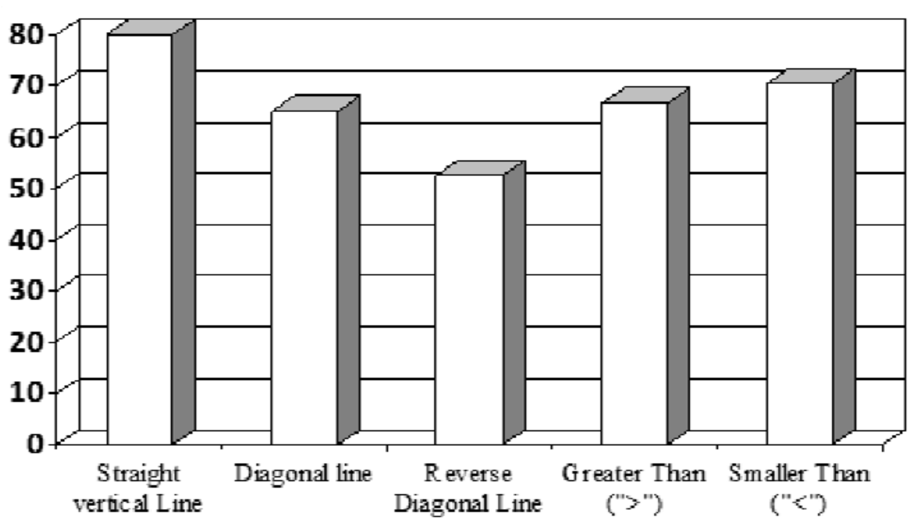


Figure 6: Samples classified correctly in percentage.

In figure 6 shows correctly samples classified in percentage, that is, the bar represents the percentage of classified samples correctly for each shape. For example, 80% of the straight vertical line audio samples were classified correctly shown by the first bar. The results vary from 52.5% for reverse diagonal line to 80 % for straight vertical line, with three shapes having average results above 60% (a diagonal, a “greater than” (“>”) and a “smaller than” (“<”).

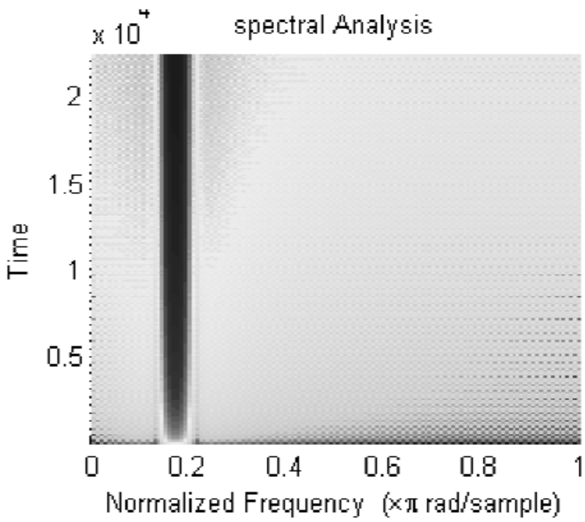
Now, if we compute the confusion matrix as shown in Table 1. When samples from the specific shape are classified wrongly, then mostly it classified as near or neighboring shape. For example, the first row of the matrix shows that the 5% of the subjects classified wrongly for a reverse diagonal line, the remaining 15% of the subjects classified wrongly for the diagonal line and the 80% of subjects correctly classified a straight vertical line. The second row of the matrix shows that 65% of the subjects classified correctly a diagonal line, the remaining 35% of the subjects classified as a straight vertical line (10%), a reverse diagonal line (20%) and “a greater than” symbol (“>”) (5%). The third row of the matrix shows that 52.50% of the subjects classified correctly a reverse diagonal line, 25% of the subjects classified wrongly either a straight line or a “smaller than” symbol (“<”), 17.50% of the s 1ubjects classified wrongly as a “greater than” symbol (“>”) and the remaining 5% of the subjects classified wrongly as a diagonal line. The fourth row of the matrix shows that 66.66% of the subjects classified correctly a “greater than” symbol (“>”), the 18.1 % of the subjects classified wrongly as a diagonal line, the 12.1% of the subjects classified wrongly as a “smaller than” symbol (“<”), and the remaining 3% of the subjects classified wrongly as a reverse diagonal line. The fifth row of the matrix shows that 70.37% of the subjects classified correctly, the remaining 29.63% of the subjects classified wrongly as a “greater than” (“>”) symbol (14.8%), a reverse diagonal line (3.7%), a straight line (3.7%) and a diagonal line (7.4%).

Table 1
A confusion matrix

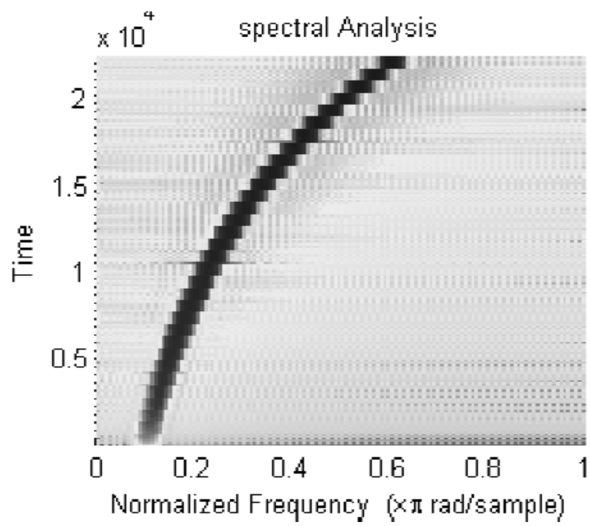
	Straight Line	*Diagonal Line*	*Reverse Diagonal Line*	*A Greater than (“>”)*	*A Smaller than (“<“)*
Straight Line		0.15	0.05	0	0
Diagonal Line	0.10		0.20	0.05	0
Reverse Diagonal Line	0.125	0.05		0.175	0.125
A Greater than (“>”)	0	0.181	0.030		0.121
A Smaller than (“<“)	0.037	0.074	0.037	0.148	

Spectrogram Result

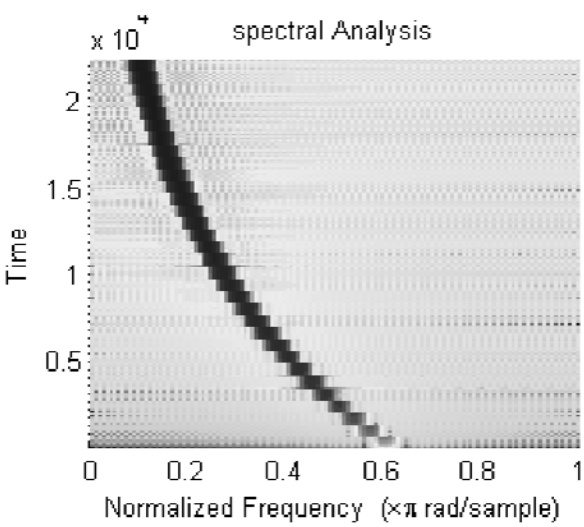
Spectral Analysis representation of the sound shows the frequencies variation with amplitude in results. Figure 7 represent the spectrogram of open figure mentioned in section III-A are a straight vertical line, diagonal line, reverse diagonal line, “greater than” symbol (“>”) and “smaller than” symbol (“<”). The spectral analysis of straight line shows the frequency cannot change but the amplitude increase with increasing in time. Yellow color changing from light to Dark shows the increasing in the amplitude. The spectral analysis of diagonal line shows the frequency increase logarithmically and amplitude increasing with time. For reverse diagonal line, shows the logarithmically decreasing in the frequency and increasing in the amplitude. For “greater than” symbol, first there is logarithmically increasing in frequency then decreasing but the amplitude is increasing continuously. But in “smaller than” symbol there is an inverse mapping of the greater than symbol but amplitude is increasing.



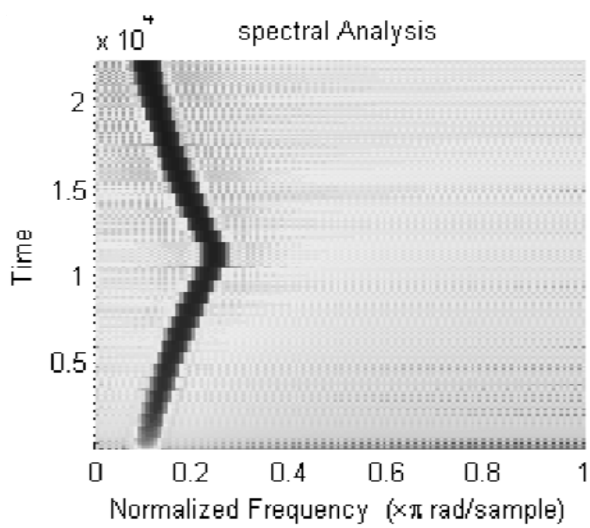
(a)



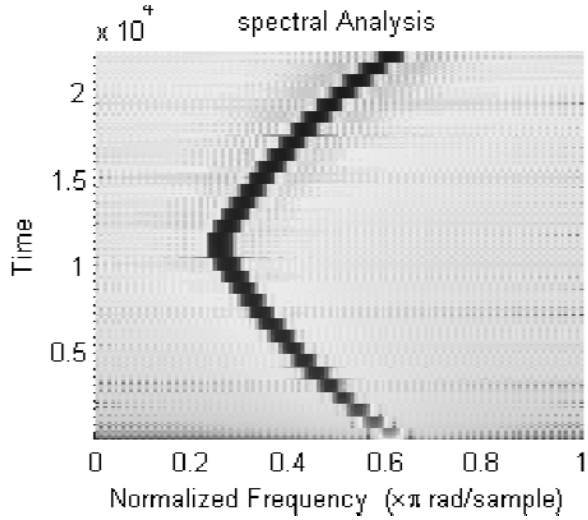
(b)



(c)



(d)



(e)

Figure 7: Spectral Analysis of (a) a straight vertical line, (b) a diagonal line, (c) a reverse diagonal line, (d) a greater than symbol and (e) a smaller than symbol

4. CONCLUSION AND FUTURE WORK

Sonification is the field in which so many applications have been implemented as the combination of multi-disciplinary teams. The sense of sight is very important sense which is cannot be substituted completely by any other sense. But trying to help the visually impaired people there is the requirement of the applications and this is the area where sonification excels. The perception of open shapes can be done using the algorithm mentioned in section III and their spectrogram results show that the perception could be achieved successfully.

Future work includes a number of open shapes in a single image and also the closed shapes such as a circle, triangle, square etc. Increasing the number of shapes in a single image includes localization of shapes in an image. So, that task of image sonification to help visually impaired people become more successful.

5. ACKNOWLEDGEMENT

We express our sincere gratitude towards all the teaching and non-teaching faculty members of Electronics and Communication Engineering department of Lovely Professional University for helping. This research would possible through the support of all the participants in the studies. Many thanks also to all for motivating to carry out this work.

Reference

[1] Thomas Hermann, Andy Hunt, John G. Neuhoff, The sonification handbook. 2011.

[2] Barrass S., Kramer G, "Using sonification," Multimedia systems. 1999, 7(1), 23-31.

[3] Datteri DL, Howard JN, "The sound of color" 2004, 767-771.

[4] Yeo WS, Berger J, "A framework for designing image sonification methods," Center for computer research in music and acoustic, department of music, Stanford University. Meeting, E., conference, I., & display, A. 2004, 323-327.

[5] Mengucci M, Henriques TJ, Cavaco S, Correia N, Medeiros F, "From color to sound: Assessing the surrounding environment," International conference on digital arts, 2012, 345-348.

[6] Garet h Loy, Musimathics, (n.d.) Vol. 1, 2006.

[7] Yeo, W. S., and Berger, J. (n.d.), "Raster Scanning : A New Approach to Image Sonification, Sound Visualization, Sound Analysis and Synthesis," in Proc. the International Computer Music Conference, New Orlean, LA, USA, 2006.