Image to Audio Frequencies Modulation for Visually Impaired People

Thomas Gonnot, Matthew Mikuta and Jafar Saniie
Department of Electrical and Computer Engineering
Illinois Institute of Technology, Chicago, Illinois, USA

***Abstract*— This paper presents a method to convey images to visually impaired people. It takes the images from a camera, which could be on a smartphone or embedded into eyeglasses, and then converts them into a set of modulated frequencies mixed together into one single audio melody, that is then played back to the user. Although it is very difficult for a normal user to makes sense of all the information that can be conveyed using this method, the hope is that with training visually impaired people will be able to interpret the data as a rudimentary vision system.**

I. Introduction

An estimated 285 million people are visually impaired worldwide according to the World Health Organization. Of those 285 million, 39 million are blind and 246 million have low vision. Approximately 90% of those who are visually impaired live in low-income settings [1]. In addition, the visually impaired may struggle to find work [2], pay for the cost of visual aids, and medical costs for any injuries [3].

Most of the research on technologies to assist the visually impaired, such as a cane with a proximity sensor or an advanced computer vision system [4], work by helping the user to grasp more information about the environment than their remaining senses can [5]. These technologies also focus on being universally usable and requiring no form of training. Therefore, the complex visual systems are built to acquire as much data as possible, determine the relevant information, and then feed it back to the user using either audio or sensory feedback. Doing so requires a lot of computing power, often resulting in a device that is costly or power consuming, and still might leave out relevant information to the user [6].

Another approach to this problem uses the human brain to make decisions on what is important by allowing the raw images to be "seen" by the visually impaired people. The idea is that the brain has been proven to be able to adapt to new stimuli in order to compensate for loss of vision [7], by increasing the sensitivity to sounds for example, and so would be able to adapt to a new form of input through the ears. The drawback is that the users must train themselves, and therefore their brain, to interpret the audio signals. The hope is that the brain would be able to execute the equivalent of a time-frequency representation, and reroute it through the visual cortex of the visually impaired person [8] [9]. Then the users would be able to "see" their environment and choose how to interact with it. Other technologies, such as Optical Character Recognition (OCR) and Text-to-Speech algorithms, can help with the user interpret texts and other stimuli that could be hard to convey with sufficient accuracy using this method.

The first part of this paper debates the feasibility of implementing such an algorithm while investigating the optimal frequency range and spacing for the human ear and audio systems. The second part describes the image to sound conversion in detail, and the third part shows experimental results. Finally, the last section discusses the implementation of this algorithm in the context of a visually impaired user.

II. Human Hearing

Although the frequency range of human hearing is typically from 20Hz to 20KHz, it doesn't react linearly to the frequency. The ears of any individual tend to better perceive lower frequencies, and is also affected by its age. With time, the higher frequencies get increasingly harder to perceive to the point of not being audible at all. For men over 40 for example, it is common to observe a 5 to 10 dB hearing loss for upper frequencies. This decay is a problem for a lot of elderlies and hearing aids are commonly used to compensate for the lost sensitivity. In the context of this application however, the system can be tuned to each user by increasing the amplitude of the frequencies that are perceived as lower, and make the perceived amplitude flat across the spectrum used by the device.

Figure 1 shows a representation of the detection threshold of several individuals across the entire spectrum of human audible frequencies. As we can see, there is a definite increase in sensitivity – decrease of the pressure level required to be detected – around 2kHz to 4kHz.

Another critical factor for this application is the ability to discriminate two different frequencies, usually referred to as the pitch resolution or frequency discrimination [11]. The goal is to select the minimum possible frequency perceptible frequency

978-1-5090-4767-3/17/$31.00 ©2017 IEEE

difference to allow the maximum of different frequencies, and therefore increase the number of lines allowed for the image.

Finally, in order to be properly interpreted by the users, the generated sound needs to change with a rate slow enough to be perceptible. In our project, this is determined by the time allocated for each column of the processed image.

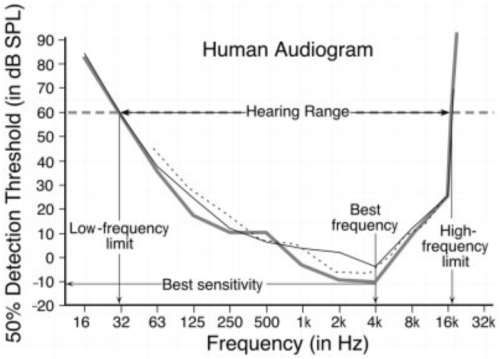

Figure 1. Human Audiogram [10]

III. Image to Sound Algorithm

The algorithm consists of 5 different operations. First the image is converted into grayscale. The resulting image is then converted to a binary map using an adaptive thresholding algorithm. The aim of these operations is to simplify the image as much as possible before conversion to sounds. The degree of simplification required on the experience of the user, and can be reduced for the advanced user to reveal more information.

Another level of simplification is applied, by averaging blocks of pixels together. Therefore, an image of 1920x1080 pixels, resolution commonly available under the terms FullHD or 1080p, can be reduced to a more manageable resolution of 384x216 pixels by averaging blocks of 5x5. The result is a grayscale image, even if the input is a binary image.

For each row of this image, the algorithm then generates a sine wave. The frequency for each row is determined by the desired minimum and maximum frequencies, and the height of the image. The frequency increases linearly from the minimum frequency at the bottom of the image to the maximum frequency at the top of the image. The length of the sinewaves is defined by the desired pixel length, the width of the image and the number of audio samples per pixel. Once the sinewaves are generated, the time period dedicated to a certain column is then multiplied by the value of the pixel.

The last step consists in mixing all the sinewaves together to form one audio stream that can be played to the user. The mixing in this case is a simple sample-wise averaging of all the signals. Figure 2 shows a flowchart of the entire algorithm. The algorithm requires to set few parameters to fine-tune the audio response. The user can set the minimum and maximum frequencies, the size of the averaging blocks, and the time dedicated to each pixel column.

IV. Results

The algorithm was implemented in MATLAB, and was used on a set of test images for testing purposes. The audio generated was then fed to a spectrum analyzer program called Spectrum Lab which displays a waterfall representation of the audio.

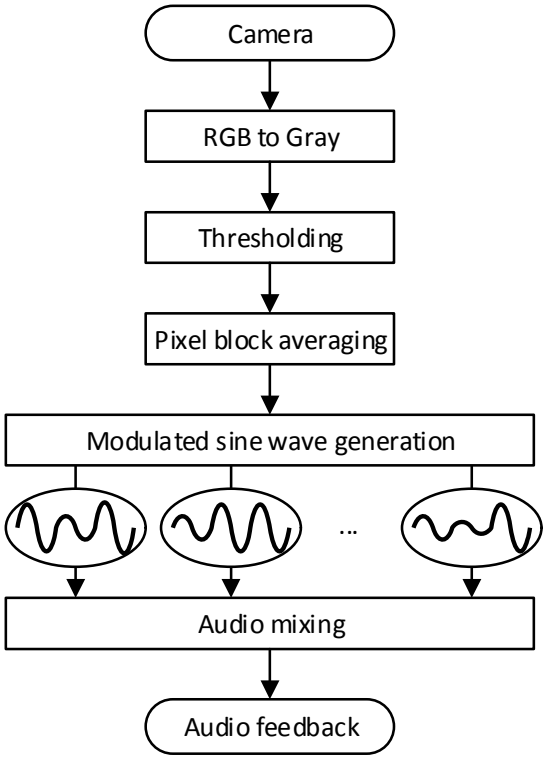

Figure 2. Image to sound algorithm flowchart

Figure 3 shows a pattern sample, with various geometries, that was processed by the algorithm. The selected parameters are frequencies within the range of 1kHz to 5kHz, 20ms per column, and averaging blocks of 5 pixels, and the input image has a resolution of 1722x824. The result is an audio stream of around 7 seconds in length, and shows in the time-frequency representation all the shapes expected.

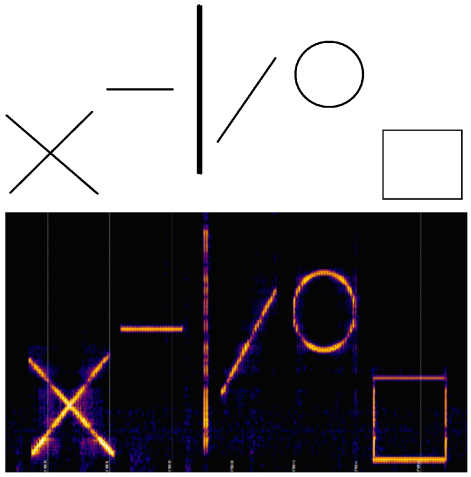

Figure 3. Test pattern and its resulting audio signal

Figure 4 shows a sample picture of a traffic sign. The aim is to know whether the text is readable. The assumption here is that what the spectrum analyzer software is showing us is what a trained user would be able to "see". In this case for instance, the stop sign can clearly be recognized. Figure 5 shows a picture from a street, containing a lot of information. Although no text is readable, we can recognize the two cars in the center of the image, and it gives some clues about the environment.

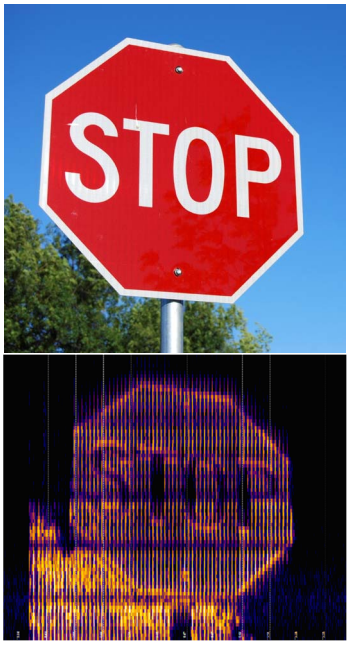

Figure 4. Sample picture of a traffic sign

Figure 5. Picture from a complex environment

V. Implementation for the Visually Impaired

One important objective in the development of a device for the visually impaired is to make it the least cumbersome as possible. The proposed algorithm being very simple, it is possible to run it on small devices, such as for example a phone. It can also be used as a complement of other solutions, in case crucial information of the environment can't be convey otherwise.

This algorithm can also make obstacle avoidance much easier for the visually impaired when combined with a depth camera. In this case, the intensity of the audio can represent how close the obstacles are, and even give a generic idea of the shape of the object. Figure 6 shows an example of that with the depth map of a motorcycle and the resulting audio spectrum.

Another approach would be to implement the algorithm directly on hardware, using for example an FPGA. This method has the advantage to be standalone, and could be made available as a single device, with its own battery, camera(s) and audio output.

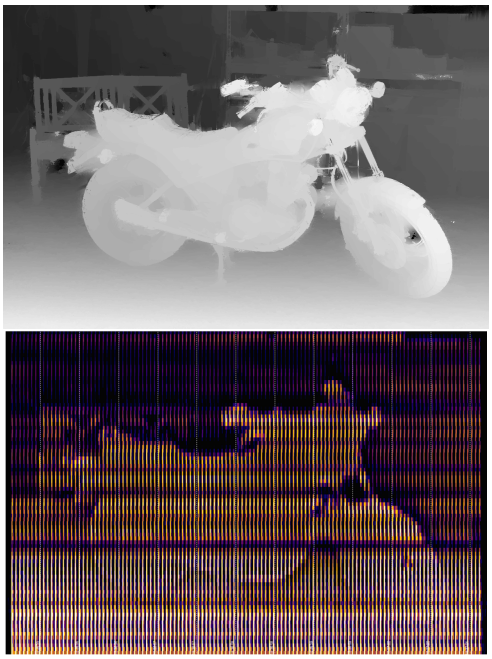

Figure 6. Example of depth map converted using the algorithm

VI. Conclusion

In this paper, we introduced an algorithm to help the visually impaired people being able to perceive their environment by converting images from a camera into a set of modulated frequencies mixed together. Preliminaries results shows that it is possible to convert images with sufficient resolution to recognize traffic signs, shapes, or even convey depth information for collision avoidance. Future work includes the optimization of the algorithm and its implementation on a mobile platform.

REFERENCES

[1] World Health Organization, "Visual impairment and blindness - Fact Sheet N°282," August 2014. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs282/en/. [Accessed January 2017].

[2] K. Jernigan, "BLINDNESS-DISCRIMINATION, HOSTILITY, AND PROGRESS," [Online]. Available: https://nfb.org/images/nfb/publications/articles/blindnessdiscrimination hostilityandprogress.html. [Accessed January 2017].

[3] Centers for Disease Control and Prevention, "Vision Health Initiative (VHI) - Data & Statistics - National Data," 30 September 2015. [Online]. Available: https://www.cdc.gov/visionhealth/data/national.htm. [Accessed January 2017].

[4] K. Gomez, "Multi-sensor navigation gadget for people who are blind," 19 July 2013. [Online]. Available: https://electronicsnews.com.au/multi-sensor-navigation-gadget-for-people-who-are-blind/.

[5] American Foundation for the Blind, "Technology Resources for People with Vision Loss," [Online]. Available: http://www.afb.org/info/living-with-vision-loss/using-technology/12. [Accessed January 2017].

[6] R. Manduchi, "Mobile Vision as Assistive Technology for the Blind: An Experimental Study," in *Computers Helping People with Special Needs: 13th International Conference, ICCHP 2012, Linz, Austria, July 11-13, 2012, Proceedings, Part II*, Springer Berlin Heidelberg, 2012, pp. 9-16.

[7] M. Bedny, H. Richardson and R. Saxe, ""Visual" Cortex Responds to Spoken Language in Blind Children," *Journal of Neuroscience,* vol. 35, no. 33, pp. 11674-11681, 2015.

[8] S. McAdams, "Spectral Fusion and the Creation of Auditory Images," in *Music, Mind, and Brain: The Neuropsychology of Music*, Springer US, 1982, pp. 279-298.

[9] S. McAdams, "Spectral Fusion, Spectral Parsing and the Formation of Auditory Images," 1984. [Online]. Available: https://ccrma.stanford.edu/files/papers/stanm22.pdf.

[10] R. S. Heffner, "Primate hearing from a mammalian perspective," *The Anatomical Record Part A Discoveries in Molecular Cellular and Evolutionary Biology,* vol. 281A, no. 1, pp. 1111-1122, 2004.

[11] C. Micheyl, P. R. Schrater and A. J. Oxenham, "Auditory Frequency and Intensity Discrimination Explained Using a Cortical Population Rate Code," *PLOS Computational Biology,* vol. 9, no. 11, pp. 1-7, 2013.