Paweł Wołoszyn*, Eliasz Kańtoch*

An Acoustic Interface for Simple Binary Image Presentation**

1. Introduction

Using the most up to date studies, WHO estimates that the number of people with visual impairment is 285 million (65% of whom are aged over 50 years). Of these, 246 million have low vision (63% over 50) and 39 million are estimated to be blind (82% over 50) [1]. The top three causes of visual impairment are uncorrected refractive errors, cataract and glaucoma. About 90% of the world's visually impaired live in developing countries. There is about 20 thousand people who are blind in Poland according to Polish Blind People Association [2].

The bandwidth of a sense refers to the capacity of that sense to receive and perceive information. Studies show that vision is our highest bandwidth sense, followed by hearing and touch (tab. 1) [3]. As the highest bandwidth vision is arguably of the greatest importance among the senses, and therefore the hardest to do without. It is essential that our visual system is efficient because two-thirds of all information we receive is visual.

Table 1
Summary of information bandwidth for three senses [3]

Sense Modality	Limit bits/s
Skin	10^2
Ear	10^4
Eye	10^6

Blindness has a major negative influences on psychical well-being, creates barrier in human communication and limits cognitive skills. This barrier is particularly difficult to overcome in the information society, where image is the major communication medium.

* AGH University of Science and Technology, Faculty of Electrical Engineering, Automatics, Computer Science and Electronics, Department of Automatics, al. A. Mickiewicza 30, 30-059 Krakow, Poland

** This work was supported by AGH UST as a research project no. 11.11.120.612

On the other hand technological progress open new possibilities of improving the life of visually impaired people.

The aim of the research is to design, construct and implement an acoustic interface for simple binary image presentation. The paper is organized as follows. Section 2 looks at several vision substitution systems. Section 3 presents the system architecture. Section 4 presents results and discussion. Section 5 concludes the paper.

2. Vision substitution systems

There are several methods of presenting textual and graphical information for visually impaired people.

The most common is the Braille system that is widely used by blind people to read and write. The Braille system was based on a method of communication originally developed in response to Napoleon's demand for a code that soldiers could use to communicate silently and without light at night called night writing. Each Braille character, or cell, is made up of six dot positions, arranged in a rectangle containing two columns of three dots each[4].

One of the approaches evaluates a method for the automatic conversion of images from visual to tactile form. Tactile imaging is the process of turning a visual item, such as a picture, into a touchable raised version of the image. At first, image processing algorithms is applied to an image to produce a simplified version of the original. This caricaturized image is subsequently output in a raised tactile graphic form on microcapsule paper, suitable for display to a blind person. Graphics can also be applied to the microcapsule paper using ink pens, markers and other drawing implements. Once the image is applied to the microcapsule paper, it is inserted image side up into a heating machine, referred to as the tactile image enhance. When exposed to a heat source of 120–125 ^{o}C, portions of the paper that are printed in black expand. The microcapsules beneath the black lines of a diagram absorb more heat than the other microcapsules and expand in diameter, raising the drawing from the background [3].

There are many text based auditory interfaces which automatically read text document and text in the web pages. One big benefit of speech output is that users who cannot read Braille can use it. Reliable speech synthesizers are available for most computers, and the quality of speech is typically good.

Unfortunately, acoustically reading images comes with a lot of difficulties. There are attempts to elaborate algorithms and interfaces, which will facilitate reading images. System for acoustic identification of colors was described in paper [5]. It transforms a small portion of a colored video image into particular instrument sounds (ex. piano, viola etc.). It was shown that sounds from musical instruments provide an alternative way to vision for obtaining color information from the environment. Experiment was introduced in which several participants try to match pairs of colored socks by pointing a head mounted camera and by listening to the generated sounds. Experiment demonstrated that blindfolded individuals were able to match pair of socks with an accuracy of 94%.

Furthermore, human ability to perform complex analysis of acoustic signal should also be taken into account. Some researchers are experimenting with encoding image information into sound using time and frequency domain as analogues of two dimensions of the image. This approach is based on implicit assumption that human brain is analyzing sound in a manner resembling Fourier analysis. However the auditory system performs much more complex processing which involves separating sound sources, grouping signals into parallel streams and attaching some meaning to them during auditory scene analysis process [6]. It is very unlikely that such specialized system would suspend this advanced processing and allow low-level access to information encoded in frequency spectrum of synthesized sound.

The most important limitation in acoustic presentation of images can be accounted to low information bandwidth of human auditory sense. An image is two dimensional concept requiring appropriate method of presentation which preserves spatial relations between image elements. Unfortunately human hearing sense is specialized in perceiving rather temporal than spatial relations between stimuli. Although people can discern the direction of sound under typical circumstances, this ability is far too limited to be considered as a mean of rendering spatial relations within an image.

3. System description and architecture

The perception-related obstacles mentioned above can be overcome by introducing an element of interaction into proposed interface. Instead of producing auditory rendering of entire image only a small focus area of it can be presented, for example one pixel at a time. Instead of ordering pixels into arbitrarily chosen sequence (like row-by-row scanning) the initiative of moving focus area can be given to the user. This ensures both that the amount of auditory data remains well within hearing capabilities and that spatial relations between image elements are not encoded in any way in the sound but inferred by the user from focus movements combined with aural feedback.

In this design the role of sound signals is narrowed to representing colour information of single pixel. This can be further limited to two-level monochrome images, in which case the sound is used only to indicate whether given pixel has foreground colour, as background colour is signalled by plain silence. No special training of recognizing different timbres or comparing musical tones and rhythms is required for the user beyond elementary ability of hearing sound.

A metal detector device can be considered as a simple metaphor of this approach. One can use such device not only to find hidden metallic object but also to obtain some information about its shape by slowly moving detector over an area and collecting its responses. Virtually the same procedure is needed for recognizing shapes on images presented in proposed interface.

System consists of microcontroller board based on the ATmega328, which is connected to PC computer via the USB, and 8 ohm piezoelectric element (Fig. 1).

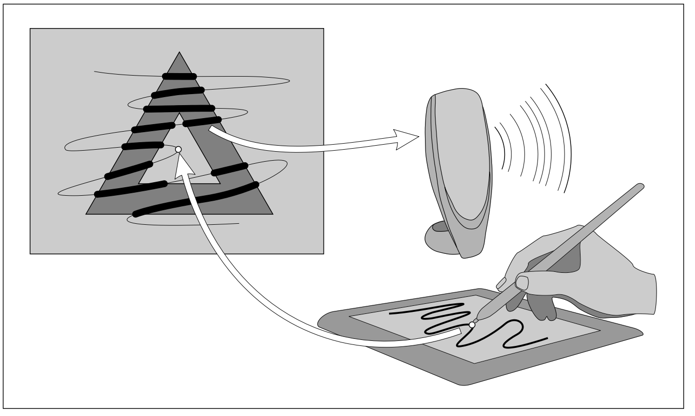

Fig. 1. System architecture

One terminal of the piezoelectric element is connected to digital pin through a 100 ohm resistor. The other terminal is connected to ground (Fig. 2).

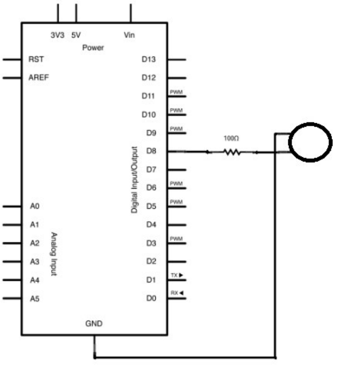

Fig. 2. Connection of piezoelectric element with microcontroller

Microcontroller was programmed to generate an alternating current through piezoelectric element what indicates reverse piezoelectric effect resulting in its contraction and expansion. Vibration of piezoelectric element generates acoustic sound waves.

The ATmega328 provides UART TTL (5 V) serial communication, which is available on digital pins (RX) and (TX). An ATmega8U2 on the board channels this serial communication over USB and appears as a virtual COM port to software on the computer. The ATmega8U2 firmware uses the standard USB COM drivers, and no external driver is needed. Simple transmission protocol which uses only one byte data packets was implemented in order to provide effective serial communication between microcontroller and computer software.

System is integrated with Java computer application which loads images into memory and tracks user pointing device. If user enters object area program automatically send information packet to microcontroller in order to generate audio feedback. Performed tests showed that the maximum delay is 40 ms. Signal frequency can be modified by user preferences.

The choice of pointing device used for moving image focus area requires further consideration. Standard computer mouse as the most commonly found device appears to be not suitable for the purpose of coordinating hand movement with auditory feedback. The location of mouse pointer is only vaguely related to location of hand, as the latter cannot be precisely defined. Depending on physical shape of mouse, placement of its sensors and anatomy of user's hand the exact coordinates recorded by mouse can refer to point lying either inside palm or somewhere between fingers. This point can even shift its location while hand becomes more stretched or contracted during movement, which is counterintuitive for the user and can be confusing.

This problem is also accompanied by operating system behaviour which alters mouse cursor movement in order to achieve greater accuracy for slower movement and shorter travel distance for faster moves. Variable pointer acceleration results in lack of repeatability of cursor placement. Even if user's hand moves twice to exactly the same location, the position of pointer can be different each time depending on movement path and speed.

For above reasons another pointing device was chosen: a graphic tablet. This kind of input device provides constant linear mapping from physical space to logical coordinates and employs comprehensible pointing method with the use of tangible stylus. It is thus easy for user to understand how to indicate specific point in image and how it is spatially related to other parts of image. Moreover the concept of screen pointer is not necessarily involved in explaining how the interface is intended to work. This is important especially for blind users who are not aware of typical visual interface features.

4. Results and discussion

A prototype implementation of proposed interface was used to perform several experiments with sighted people whose task was to recognize simple geometric shapes such as a circle, square, triangle or cross. Subjects were not able to see these images before and during the test. Each person was trying to recognize two random shapes, once while keeping eyes closed and second time while looking at the tablet and stylus. Recorded data included accuracy of user responses and time spent on recognizing images.

Although the interface is still in early development stage, collected results indicate that it is possible to recognize simple shapes with proposed method of interactive acoustic image presentation. However there are some problems which require further attention. One of them is the influence of ability to see the tablet device on shape recognition accuracy and speed. Persons allowed to look at tablet during test were giving better answers and spent less time on analyzing images. Seeing the stylus and its movement while memorizing auditory signals helped to develop mental map of sounding and silent areas. Of course this is impossible to achieve in case of blind users but instead of visual aid some kind of tactile clues could be used to facilitate association between sound and location.

Binary images contain only two colours and thus can be presented using one type of sound together with opposing silence. This approach however proved to be overly simplified because it contain no information about given point's neighbourhood. If the metaphor of metal detecting device is to be complete, the interface should provide some information about proximity of current pointer location to foreground image features. Only then the user could navigate the image without laboriously checking its every fragment in order to avoid omitting important details.

5. Conclusion

The paper proposes an acoustic interface for the blind and describes the prototype system which enables blind persons to read binary images. The system is low-cost and easy to install and use. In the future, we plan to design several touch sensors and integrate them with this system in order to provide user also with tactile feedback.

References

[1] World Health Organization, *Visual impairment and blindness*. http://www.who.int/mediacentre/factsheets/fs282/en (visited 9.05.2011).

[2] Polish Federation of the Blind, http://www.pzn.org.pl/ (visited 9.05.2011).

[3] Way T.P., Barner K.E., *Automatic visual to tactile translation. I. Human factors, access methods and image manipulation.* IEEE Transactions on Rehabilitation Engineering, 1997.

[4] Canadian Braille Authority/L'Autorité anadienne du Braile, *About Braille*. http://www.canadianbrailleauthority.ca/en/about_braille.php (visited 9.05.2011).

[5] Bologna G., Deville B., Vinckenbosch M., *A perceptual interface for vision substitution in a color matching experiment*. Neural Networks, 2008. IJCNN 2008 (IEEE World Congress on Computational Intelligence).

[6] Bregman A.S. (1993) Auditory scene analysis: *Listening in complex environments*. [in:] S.E. McAdams, and E. Bigand (Eds.) *Thinking in sound*. Oxford University Press, London 10–36.

[7] Kokjer K.J., *The information capacity of the human fingertip*. IEEE Trans. Syst., Man, Cybern., vol. SMC-17, No. 1, 1987.

[8] Hiroki Minagawa, Noboru Ohnishi, *Tactile-Audio Diagram for Blind Persons.* IEEE Transactions on Rehabilitation Engineering, vol. 4, No. 4, December 1996.

[9] Gomez J.D., Bologna G., Pun T. (2010), *Color-audio encoding interface for visual substitution: see color matlab-based demo*. Proceedings of 12th international ACM SIGACCESS conference on Computers and accessibility (ASSETS' 10). ACM, New York, USA, 245–246.