

Review on Image Sonification: A Non-visual Scene Representation

Rajib Sarkar

National Institute of Technology
Rourkela
Rourkela – 769 008, Orissa, India
sarkar.rajbrciit@gmail.com

Sambit Bakshi

National Institute of Technology
Rourkela,
Rourkela – 769 008, Orissa, India
sambitbakshi@gmail.com,

Pankaj K Sa

National Institute of Technology
Rourkela,
Rourkela – 769 008, Orissa, India
pankajksa@gmail.com

Abstract: With the advent of image and video representation of visual scenes in digital computer, subsequent necessity of vision-substitution representation of a given image is felt. The medium for non-visual representation of an image is chosen to be sound due to well developed auditory sensing ability of human beings and wide availability of cheap audio hardware. Visionary information of an image can be conveyed to blind and partially sighted persons through auditory representation of the image within some of the known limitations of human hearing system. The research regarding image sonification has mostly evolved through last three decades. The paper also discusses in brief about the reverse mapping, termed as sound visualization. This survey approaches to summarize the methodologies and issues of the implemented and unimplemented experimental systems developed for subjective sonification of image scenes and let researchers accumulate knowledge about the previous direction of researches in this domain.

Keywords: Image representation, Sonification, Stereo vision, Non-visual image representation, Auditory image.

I. INTRODUCTION

Data conversion between the visual and audio domain has been an active research area due to a growing interest of researchers into non-visual techniques such as sound, touch and smell to represent information. The same information can be represented by a sound or tactile display. These forms of representing a data are more appealing to human beings and they serve the purpose of facilitating communication or interpretation. Sonification is defined as the sound (auditory) representation of non-speech data to convey information [1]. More specifically, sonification is the transformation of any data relation into perceived relation in an acoustic signal. The capability of human auditory system of discerning subtle changes in acoustic patterns makes sonification a useful method of data representation. Research proves that auditory display is superior to its visual equivalent [2]. By its very nature, sonification is interdisciplinary. The conversion of image to auditory signal is coined as image sonification. It is perhaps

the most admired form of sonification due to its versatile applicability in real life. As an image can be visualized as a two dimensional pixel-space each pixel having a discrete value, hence an image can also be represented as a three dimensional matrix having three indices as: x-position, y-position, and intensity-value. The mathematical nature of the process of image sonification is the conversion of a data from static two-dimensional domain to a one dimensional time domain. The helical coordinate system [3] is used as a framework by many researchers to analyze two-dimensional data with methods for one-dimensional space.

II. PIONEER RESEARCHES

Initial attempts of image-to-sound mapping failed to bring out useful results. The major bottleneck was the failure of the researchers to follow the principles of psychoacoustic in implementing image-to-sound conversion methods. Therefore there has been a strong foundation of research that has subsequently taken place regarding intensity, frequency, and temporal discrimination of static audible sounds as discussed in [4]-[6]. The determinants of pitch and loudness, the effect of masking [7], and human auditory localization abilities [8] are also explicitly researched and understood. Recently, researchers have begun extending their investigations towards the perception of more complex, dynamic auditory patterns in speech and music, which is of primitive relevance for sonification research [9]-[11]. Further researches revealed that a one-to-one mapping (invertible) from image to sound can ensure the preservation of visual information [12]. For image sonification, the strategy adopted by the researchers in one of the earliest experimental system for auditory representation of image is to sweep a line across the image and converting the pixels falling on the sweeping line into sound [12]. Related works to this can be found in [13]-[16].

Other strategies (viz. query-based method and tour-based method) have been adopted for vector and object-based images. Authors in [17] have proposed a novel path-based model for sonification.

III. RELATED WORKS

In The vOICe system developed by W. D. Jones in [18], the image captured using a suitable single video camera is scanned in left-to-right direction for sound generation. The implementation in this research approached to convert upper portion of the image into high frequency tones and bottom portion into low frequency tones. Hence there was no discrimination between objects and backgrounds in the image. To serve the objective of forming sound depending on objects present in the image, captured image is first converted to 32×32 greyscale image and an image processing operation is done on the image for foreground and background segmentation in research carried out in NAVI [19]. The foreground is assigned high intensity values whereas the background is assigned low intensity values. The stereo sound obtained converting the image has amplitude directly proportional to intensity of image pixels, and the frequency of sound is inversely proportional to vertical orientation of pixels.

There are basically two independent stages that occur during the process of sonification [20]. First a transfer function is generated that details how the data values are changed to sound. Second, the auditory data generated by the first stage is rendered with appropriate pitch and frequency. The two stages applied in succession produce an audible sound.

IV. TRANSFER FUNCTION GENERATION

The first step follows the image-to-sound mapping model as discussed in [12]. In this approach an image is transformed into a time-multiplexed auditory representation to achieve high resolution as shown in Fig. 1 [12]. The resulting sound represents the input image up to a resolution of 64×64 pixels with 16 gray-tones per pixel.

The way of traversal of the input data is to be taken into consideration. The process of transfer function generation can further be subdivided into three general classes on the basis of method of traversing input data.

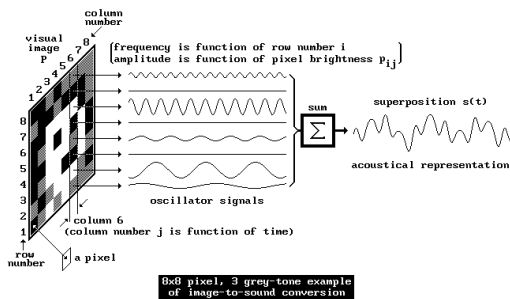


Fig. 1 Principles of the image-to-sound mapping

A. Query-based Method

Query based method is one where the system queries the user for pointing to a portion of information that needs to be sonified and the user uses a selector for continuously

selecting the information to be sonified, as proposed by authors in [21]. This method largely depends on human intervention throughout the process of sonification. Another example of a query based sonification is in [22] where sonification of heating schematics is presented.

B. Tour-based Method

Tour based method is one where the information is arranged into a predefined order for sonification. The tour consists of points of interests in the data set starting from a point and ending at the same point. A tour based method is suitable for sonification of a vector image, where the developer only needs to find a tour that visits the individual components that are of interest. Hence human intervention is not needed throughout the sonification process, but is needed only at the time of deciding the tour. However pre-deciding on a suitable tour for a given image is difficult to find.

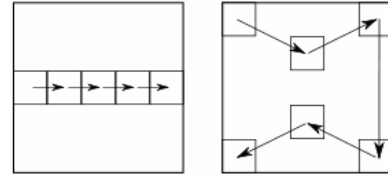


Fig. 2 Schematic arrangement of two possible path points

C. Path-based Method

A path based method is devised in [17] where the creation of path is based on a series of path points. This allows the path to be rescaled independent of the data. Devising different paths is important because the sonification on the same dataset along two different paths would be resulting in different sonified audio. Fig. 2 depicts two different paths as described in [17] which create different audios from sonification when the authors applied them onto the same dataset. Span defines the neighbourhood associated with every path point. In addition to the ability of specifying the span, the angle of interception between the span and the path can be specified by span angle. To optimize the length of the path, a span envelop is generated. Unlike other techniques, each path point can be placed anywhere within abstract space in a complete random order. There are two stages prescribed to generate a sonification from a path. The first stage deals with mapping the path onto the actual data set, which requires specifying how each span section maps onto the underlying data. The second stage involved generation of sound associated with each path point depending on the data values currently under the span and the transfer function.

V. RENDERING AUDITORY DATA

After an image is converted to its auditory representation, an issue of concern is how to render auditory data. The audio data to be rendered must conform to the principles of psychoacoustic, as well as should be reflecting subtle changes in different portions of input image. As the

area of Sonification is still considered to be in its infancy, current researches are progressing towards determining the best set of sound components to vary in dissimilar situations. There are several techniques for rendering auditory data as categorized below.

A. Audification

This method translates the data itself to amplitude values of the waveform. It is applicable if the data itself is a time series, e.g. data from a dynamic system like neural networks or seismic data. This sonification and its related forms are quite useful in monitoring large data sets in strongly compressed time. There is a need of scaling the pitch of the audio according to suitability of human perception. Tools for scaling the pitch or stretching the time-scale without affecting the pitch (one type of acoustic microscope) are useful to enhance this form of sonification.

B. Earcons (Auditory Icons)

This sonification is applied in sounds that are used in a metaphorical sense, so that the human effort to learn the display is decreased, e.g. filling a bottle with water produces a commonly known sound of evolution, which might be applied to a 'state of download-meter' sonification. A lot of research in the field of sonification focuses on how to find or select suitable sounds for a GUI. However, this display is rather unsuitable for presentation of general types of data.

C. Model-based Approach

Human acoustic senses have evolved and optimized to listen to natural sounds and are furthermore specialized to draw information from it. So, the idea behind this approach is to build a Sonification interface, in which the sound is produced similar to the physical model and where the modes of interaction with the data takes the usual proceeding in inspecting objects into account.

D. Parameter Mapping

It is a kind of sonic scatter plot. For each data record, an acoustic event is generated whose properties are driven by the data values (which are mapped to the sound attributes). The columns of a data set may be mapped to pitch, duration, volume, onset of the tone, vibrato strength, vibrato speed, brightness, spectral evolution of the sound, roughness, the envelope of the sound, etc. The high number of different attributes allows flexibility of designing a high dimensional acoustic display.

Apart from the techniques discussed above, another novel approach towards rendering auditory data through stream-based approach is proposed by S. Barrass in [23]. However none of these techniques utilizes the color information of an image for sonification.

VI. COLOR IMAGE SONIFICATION

Color information of an image can be of added interest while sonifying an image. The color models commonly used

are RGB (Red-Green-Blue), HSL (Hue-Saturation-Lightness), HSV (Hue-Saturation-Value), CMYK (Cyan-Magenta-Yellow-Key) models. Out of these models, RGB model is best suitable to be considered as a basis for the transformation of colors into sound. Apart from the standard models, a non-standard color model, coined as semantic color model is proposed in [24] to be more convenient than RGB to serve the purpose of color-to-sound conversion. The color space of semantic color model is a proper subspace of the RGB model as described in [24].

Semantic color model uses those colors as primary colors which are named and appear frequently in the language. User is allowed to add other colors of interest if they need. Default primary colors for this model are: red, green, blue, yellow, orange, purple, brown and gray. This model uniquely expresses a color by means of only two primary colors. The motivation behind is that two sounds are easily distinguishable by human auditory system, and it is also easy for approximately describing any color as a combination of two primary colors.

The sounds assigned to the primary colors can be artificial sounds, musical sound or other sounds that are mutually well distinguishable. The proposed approach in [24] is tested by students and blind users at Faculty of Informatics, Masaryk University in Brno in cooperation with Teiresias Centre.

Color properties of an image are not yet much exploited to generate sound. The distinctness of sounds assigned to primary colors are an open research area to study. Different color models can also be proposed by new researchers that are further suitable for color to sound conversion.

VII. APPLICATIONS AND TOOLS

Sonification has been applied to numerous application domains and dataset, including statistical information such as stock market data [25]. Ample techniques of sonification are applied and appreciated due to its interdisciplinary nature. Sonification also marks its applicability in representation of network diagrams (such as central heating schematics)[22], algorithm presentation (e.g. sorting algorithms)[26], representing physical data (such as oil and gas well sonification) [27], and medical applications [28]. Image sonification finds its application in real life as it can aid a blind person to understand his surrounding through audio. A person perceives the change in surrounding through his vision. Likewise a blind person can understand a change in his surrounding listening to the change in frequency and pitch in the audio which is generated and streamed from successively taken image frames capturing the surrounding of the person. AudioGraf is a query based diagram reader developed on the basis of an audio-tactile representation for the visually impaired people [29]. Recent research in [30] depicts experimental set up of an electronic blind aid, coined as The optophone.

There has been several image sonification software developed by the research organizations and software

vendors. *xSonify* [31] is a Java application for displaying numerical data as sound which is developed by Goddard Space Flight Center, National Aeronautics and Space Administration USA (NASA). *Sonification Sandbox v.3.0* is also a Java program that converts datasets to sounds. *Sandbox* is developed by GT Sonification Lab, School of Psychology, Georgia Institute of Technology. Examples also include *Metasynth* (by U&I Software), *MUSE*: A musical data sonification toolkit, and *Audiosculpt* (by IRCAM). All the tools mentioned are developed for specific applications.

VIII. SOUND VISUALIZATION

The domain of image sonification and sound visualization are complimentary to each other. Sound visualization refers to the inverse process of image sonification. Sound visualization maps an audio to its equivalent image representation. Rastogram is an approach for raster visualization of sound [32]. The importance of rastogram is in the fact that time scale modification and frequency shifting of a sound can be performed by resizing its rastogram. This facilitates analysis of sound with the help of rastogram.

IX. CONCLUSIONS

Though image sonification has been an active area of research interest for the last few decades, but there has hardly been a significant landmark outcome of the researches. Different methods has been tried out and tested for delivering well-perceivable sound from an image. The latest trend of research in the direction of image sonification is towards multi channel image data analysis using sonification. Image sonification opens a scope for processing an image by processing its equivalent sound. Likewise the research along the counter direction, i.e. sound visualization has marked its applicability for aiding the deaf people. Further practical implementation towards image sonification and sound visualization awaits evaluation of a system through experiments carried out with blind and deaf persons.

ACKNOWLEDGMENT

The authors would like to express their gratitude to all the co-researchers of Image Processing Laboratory of Department of Computer Science and Engineering of National Institute of Technology Rourkela for their active and hidden cooperation towards manifestation of this research.

REFERENCES

- [1] G. Kramer et al., "Sonification report: Status of the field and research agenda," National Science Foundation, 1999. [Online]. Available: <http://www.icad.org/websiteV2.0/References/nsf.html>
- [2] J. Tzelgov, R. Srebro, A. Henik, and A. Kushlevsky, "Radiation detection by ear and by eye," *Human Factors*, vol. 29(1), pp. 87-98, 1987.
- [3] J. Claerbout, "Multidimensional recursive filters via a helix," *Geophysics*, vol. 63, pp. 1532-1541, 1998.
- [4] W. M. Hartmann, *Sounds, Signals, and Sensation: Modern Acoustics and Signal Processing*, 1st ed., New York: Springer-Verlag, 1997.

- [5] B. C. J. Moore, *Handbook of Perception and Cognition*, 6th ed., New York: Academic Press, 1995.
- [6] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 4th ed., Orlando, FL: Academic Press, 1997.
- [7] W. L. Gulick, G. A. Gescheider, and R. D. Frisina, *Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics*, 2nd ed., Oxford University Press, 1989.
- [8] J. Blauert, and J. S. Allen, *Spatial Hearing: The Psychophysics of Human Sound Localization*, Cambridge, MA: MIT Press, 1996.
- [9] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, 1st ed., Cambridge, MA: MIT Press, 1990.
- [10] S. Handel, *Listening: An Introduction to the Perception of Auditory Events*, Cambridge, MA: MIT Press, 1989.
- [11] S. McAdams, and E. Bigand, *Thinking in Sound: The Cognitive Psychology of Human Audition*, Oxford: Clarendon Press, 1993.
- [12] P. B. L. Meijer, "An experimental system for auditory image representations," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 2, pp. 112-121, Feb. 1992.
- [13] S. Ferguson and S. D. Ferguson, "High resolution vision prosthesis systems: Research after 15 years," *Journal of Visual Impairment Blindness*, vol. 80, no. 1, pp. 523-527, Jan. 1986.
- [14] R. M. Fish, "Visual substitution systems: Control and information processing considerations," *The New Outlook for the Blind*, vol. 69, pp. 300-304, Sept. 1975.
- [15] A. D. Heyes, "Human navigation by sound," *Physics in Technology*, vol. 14, pp. 68-76, 1983.
- [16] L. Kay, "Electronic aids for blind persons: an interdisciplinary subject," in *Proc. IEE*, vol. 131, Pt. A, no. 7, 1984, p. 559-576.
- [17] K. M. Franklin, and J. C. Roberts, "A path based model for sonification," in *Proc. Eighth International Conference on Information Visualisation (IV'04)*, 2004, IEEE, p. 865-870.
- [18] S.Colton, A.Bundy, and T.Walsh, "On the notion of interestingness in automated mathematical discovery," *International Journal of Human Computer Studies*, vol. 53, no. 3, pp. 351-375, 2000.
- [19] G. Sainarayanan, "On Intelligent Image Processing Methodologies Applied to Navigation Assistance for Visually Impaired," Ph.D Thesis, Universiti Malaysia Sabah (UMS), 2002.
- [20] R.S.Tannen, "Breaking the sound barrier: Designing auditory displays for global usability," in *Proc. Human Factors & the Web*, Basking Ridge, USA, Jun. 1998.
- [21] S.Helle, G.Lepître, J.Marila, and P.Laine, "Menu Sonification in a Mobile Phone - a prototype study," in *Proc. International Conference on Auditory Display*, Espoo, Finland, Jul. 2001, p. 255-260.
- [22] D.J.Bennett, "Effects of Navigation and Position on Task when Presenting Diagrams to Blind People using Sound," in *Proc. Diagrams 2002*, Mark Hegarty, et al. Ed., LNAI 2317, 2002, Springer.
- [23] S. Barrass, "Developing the Practice and Theory of Stream-based Sonification," *SCAN Journal of Media Arts Culture*, vol. 6, no. 2, ISSN 1449-1818, Sept. 2009.
- [24] I. Kopecek, and R. Oslejsek, "Hybrid approach to sonification of color images," in *Proc. Third 2008 International Conference on Convergence and Hybrid Information Technology*, 2008, p. 722-727.
- [25] D. L. Mansur, M. M. Blattner, and K. I. Joy, "Sound-Graphs: a numerical data analysis method for the blind," *Journal of Medical Systems*, vol. 9, no. 3, pp.163-174, 1985.
- [26] M. H. Browns, and J. Hershberger, "Color and sound in algorithm animation," *IEEE Computer*, vol. 25, no.2, pp. 62-63, 1992.
- [27] S. Barrass, and B. Zehner, "Responsive Sonification of Well-logs," in *Proc. International Conference on Auditory Displays (ICAD) 2000*, p. 72-80, Atlanta, April 2000.
- [28] W.T. Fitch, and G. Kramer, "Sonifying the body electric: Superiority of an auditory over a visual display in a complex, multivariate system," in *Proc. the First International Conference on Auditory Display (ICAD)*, G. Kramer, Editor. 1992, Santa Fe Institute, Addison-Wesley: Reading, Massachusetts, p. 307-325, 1992

- [29] A. R. Kennel, "Audiograf: A diagram-reader for the blind," *ASSETS Vancouver*, British Columbia, Canada, 1996.
- [30] M. Capp, and P. Picton, "The optophone: An electronic blind aid," *Engineering Science and Education Journal*, vol. 9, no. 3, pp. 137-143, 2000.
- [31] xSonify, Goddard Space Flight Center, National Aeronautics and Space Administration USA (NASA) [Online]. Available: <http://spdf.gsfc.nasa.gov/research/sonification/sonification.html>
- [32] W. S. Yeo, and J. Berger, "Raster scanning: A new approach to image sonification, sound visualization, sound analysis and synthesis," in Proc. *the International Computer Music Conference*. New Orleans, LA, USA, 2006.