

HEARING IMAGES: INTERACTIVE SONIFICATION INTERFACE FOR IMAGES

Charles O'Neill and Kia Ng
ICSrIM – University of Leeds
School of Music and School of Computing
Leeds LS2 9JT
UK
info@icsrim.org.uk
www.icsrim.org.uk

Abstract – The paper describes an ongoing research for a method of interactive sonification of 2D image data. The method utilizes an existing device (from computer games console), Nintendo's wiimote controller [11] to provide a means of interacting with the image to aid exploration, providing the user with sonic and haptic feedback. This paper presents a method for segmentation and analysis of regions within an image to produce global and local descriptors of the specified region.

INTRODUCTION

Sonification methods present information through using sound (particularly non-speech), so that users obtain an understanding of the data or processes under investigation by listening [6].

Factors such as end application and the nature of the data under scrutiny play a major role in determining the best mapping and synthesis methods for the sonification process. A particular challenge lays in representing data which is typically time independent (such as image) within a modality which cannot exist without time (audio). In some instances, the mapping in the time domain can be implicit. In instances where this is not the case, the time domain must be constructed elsewhere within the system.

The system described within the paper looks at the sonification of time independent 2D image data and, in particular, the construction of the time domain through mapping of user actions to feedback. This current body of research looks at the sonification of the irregular shapes that make up organic images. The algorithms presented perform the processes of defining regions within the image, extracting features from these regions and constructing usable parameters from which sonic and haptic feedback can be produced. The system is driven by the users actions are therefore continued feedback comes as a result of continued interaction. An emphasis on interaction for exploring images should induce benefits such as learnability and subtlety of use [13] coming from increased familiarity with the system.

Work presented within [22] looks at issues of mapping time independent data onto the time domain (i.e., image to sound). Two methods of playback have been defined as scanning (automatic sonification route) and probing (user controlled sonification route). Previous attempts have utilized both of these methods in many different forms. Particular examples include a probing method in which sound parameters are mapped onto a spatial domain [7], and automatic scan methods based upon the raster scan technique (pixel by pixel) [21] and a left to right method (column by column) [10].

Continuing efforts have been focused upon implementing interactivity as a focus for sonification system design. Saue [16] introduces the concept of walking through data sets in order to determine global, intermediate, local and point data features. Hellstrom *et al.* [4] implement the mouse as a virtual microphone to explore data spaces. Pauletto and Hunt [12] have created a toolkit for interacting with data sets in which the user uses the mouse to interact with a data space, navigating sonified data in real time. Continuous feedback or the creation of context between modes can also aid the user in discrimination of sound [1]. The implementation of the highest level of real time continuous interaction has been found to produce the most “pleasing, efficient and fastest method of analysing data” [13].

Work presented in [4] highlights qualities which can be gained from the user being tightly embedded within an interactive control loop, sighting increased levels of 'control intimacy', a quality seen in the manipulation of musical instruments. Interfaces with sonic feedback have been realized most commonly in the form of mouse [2, 3, and 12] with additional forms of the keyboard [15] and tablet [5]. Devices which offer a multimodal feedback are less common. A successful binding of auditory and haptic feedback will provide the user with a multimodal description of the data under scrutiny.

Work is presented in [20] in which mobile device alerts the user to messages through combinations of sonic and haptic feedback in order to approximate physical objects 'dropping' into the device. Features such as weight, material and size of the object can all be approximated through combination of feedback across the two modalities to convey message parameters such as size, urgency, etc. Understanding of real world objects through physical manipulation and sonic and haptic feedback is particularly prominent with acoustic instruments. An expert violinist is constantly gauging a number of parameters through feedback both haptically (string tension, bow tension etc.) and sonically (pitch, timbre etc.) and it is this feedback which enables the highest level of control over the instrument. The following points are produced in [4] as a guideline for human machine interface design based upon the acoustic instrument example:

- Physical interaction for sonic response.
- Increased learning times producing higher level of performance.
- Interface reacting to physical interaction in a well known way.

Sonification system design may benefit from the same design ethos to that of electronic musical instrument design. Learning how to manipulate the image data to extract different types of response will require the user to become familiar with an interactive system inducing a learning curve and the added benefits that come with it.

DESIGN AND DEVELOPMENT

This section provides a system overview and describes the overall system in terms of its modular components and the individual algorithms which comprise it. The methods for region selection, feature extraction and the construction of parameters for sonic and haptic feedback are then explained.

System Overview

The system comprises of 3 separate components as seen in Figure 1. A user input method in the form of Nintendo's *wiimote* [11], an input device containing a +/-3g 8-bit 3-axis accelerometer, 1024x768 infra-red camera capable of 4 point tracking and an additional 11 discrete buttons (including 4-way directional pad). The input device will be connected via Bluetooth to a PC running Max/MSP Jitter [9], a visual programming environment which will be used for image processing and analysis, audio synthesis, and communication with the wiimote device. Max/MSP externals have been created in C++ to produce fast image processing and analysis modules which can be integrated within the same development environment as the sonic and haptic feedback modules.

Audio feedback will be provided through stereo headphones and haptic feedback through communicating with the wiimote's built in force feedback motor.

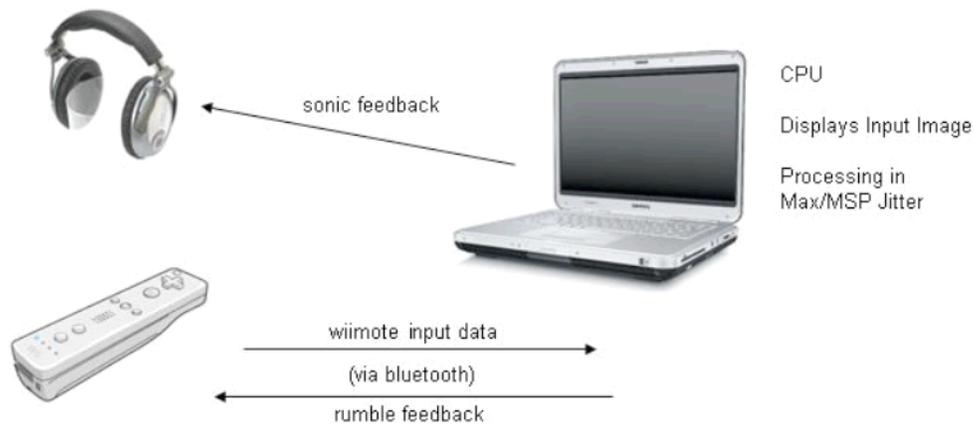


Figure 1. System Data Flow

Region Extraction

The system currently implements an area selection algorithm based upon image segmentation. Through identifying regions with connected areas of similar pixel colour values we can segment the image and thus analyze these sections separate from the image as a whole.

In the system, the wiimote is pointed at the computer monitor to scan over the image. The centre pixel is determined using the infra-red camera and point tracking hardware built into the wiimote. Depressing the A button will then initiate a procedure to segment the neighbouring area the user is pointing at based upon the single pixel being pointed at. The neighbourhood of pixels is defined using the following algorithm:

- A flood fill algorithm [18] is performed from the centre pixel coordinates on the original image.
- The flooded area is converted to a binary image with foreground representing the connected flooded areas.

-
- The image is resized to find a bounding box containing the selected region to be analysed for providing feedback to the user.
 - A chain code algorithm [8] is implemented to produce a set of integers denoting the movement taken from one pixel to the next in tracing the region perimeter:
 - A raster scan is performed to find the first instance of a flooded pixel.
 - From this position, a 4 neighbourhood chain code algorithm traces the path of the border clockwise around the shape perimeter.
 - In an instance where a connecting pixel is unavailable in the 4 neighbourhood setup, the remaining 4 diagonal neighbourhoods are checked (clockwise from bottom right) and the path is continued.
 - The chain code completing the shape perimeter is the chain code recorded.

Global Shape Describing Parameters

From the region extraction process we have obtained data describing the net area for the region and the region perimeter. In addition to these, we can derive a number of new parameters which can describe global shape descriptors for the region. An intermediate mapping step converts these raw acquired data sets into a set of parameters which aim to describe the shape as a set of more useable terms. These are obtained as follows [14, 17]:

- **Shape Perimeter:** $P(shape) = [adj + \sqrt{2}.diag]$
where *adj* refers to the number of adjacent movements *diag* to the number of diagonal movements performed within the chain code process.
- **Shape Complexity:** $shapeComplexity = shapePerimeter^2 / shapeArea$
- **Form Factor:** $formFactor = 4\pi.shapeArea / shapePerimeter$
- **Aspect Ratio:** $aspectRatio = shapeWidth / shapeHeight$
where *shapeWidth* and *shapeHeight* defined from the bounding box
- **Extent:** $extent = shapeArea / boundingBoxArea$
- **Centre of Gravity:** $gravityCentre = \frac{\sum_{0\ n} x_i + \sum_{0\ n} y_i}{chainCodeLength}$

where *n* is the series of x coordinates within the chain code path, *y* is the series of y coordinates within the chain code path and *chainCodeLength* is the number of steps taken in the chain code algorithm implementation.

Local Shape Exploration Parameters

The system aims to provide both a global representation of the selected region and also a local representation, describing smaller areas of within the region accessible

through direct interaction. Through exploring the region with the wiimote cursor the construction of local parameters provides the user with a means of further investigation of the region under scrutiny. The following methods have been implemented to produce local parameters.

- **Distance from centre of gravity.** The distance from the centre of gravity of the user position within the shape can form a new parameter for feedback. (Figure 2)

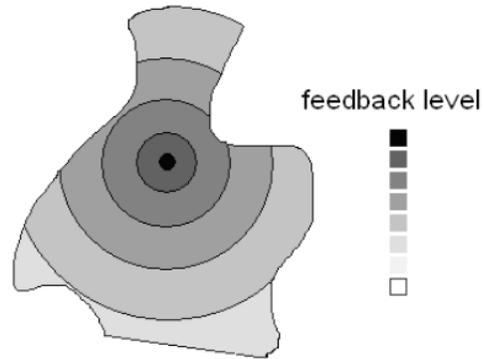


Figure 2. Distance Fields from Centre of Gravity

- **Local perimeter section.** Variable size sections of the perimeter can be accessed using the wiimote (Figure 3). The shape perimeter can be sonified with respect to angularity and path (from existing chain code segment) with a scan angle determinable by wiimote cursor position and tilt sensing capability (motion detection from accelerometers [19]).

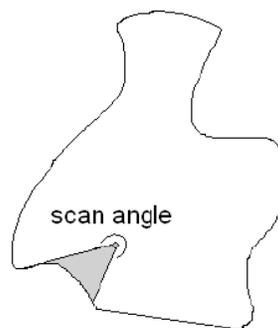


Figure 3. Scanning of Local Perimeter Section

- **Shape perimeter points.** Points which comprise the perimeter can be sonified with respect to their relative horizontal and vertical position and magnitude to that of the wiimote cursor. Figure 4 shows how these parameters are constructed.

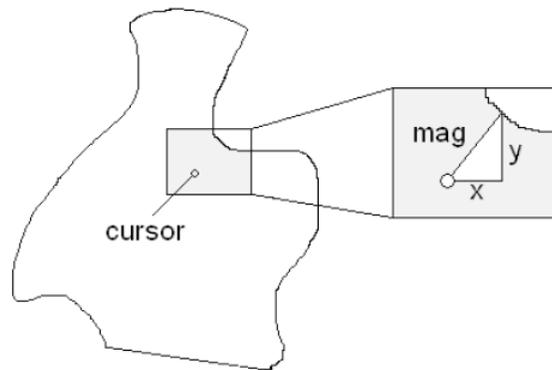


Figure 4. Determining Position of Relative Perimeter Points

User Configurable Mapping

The construction of global and local parameters allows for the categorization of input type. Similarly, we can group the feedback methods into the following categories:

- High level musical parameters such as melody, rhythm and harmony.
- Low level sonic parameters such as pitch, dynamics and frequency components.
- Amplitude and pulse frequency of wiimote rumble capability.

A user configurable mapping strategy allows the user to experiment with the perceptual effect of mapping input and output parameters singularly and also across categories types.

CONCLUSIONS AND FURTHER WORK

The interface currently implements a method for user driven selection and parameter extraction of regions comprising images loaded into the system. The regions are analysed to produce global descriptors from which we can build sonifications and haptic responses. Local descriptors are constructed through user interaction providing an additional category of parameters from which we can produce additional localised feedback.

Current work is focused on the investigation of mapping strategies in order to find effective ways of capitalizing on the segregation of input parameters we have obtained. Experimentation with the mapping of different parameter types to different sonic and musical descriptors across the two modalities should provide interesting and hopefully categorical results. Work in this area will point the system in the way of defining the most effective default setting for the new user, offering both generic settings and supporting user personalization.

Continuing work on the project can now include the development of modules for colour and texture analysis, grouping and context of segments and furthering of new interaction methods to facilitate these modules.

References

-
- [1] FERNSTROM, M and Brazil, E: Human-Computer Interaction Design based on Interactive Sonification – Hearing Actions or Instruments/Agents. *Proc. of the 2004 International Workshop on Interactive Sonification, Bielefeld University, Germany, 8th January 2004*
- [2] HELLSTROM, S. O and Winberg F: Qualitative aspects of the auditory direct manipulation: A case study of the towers of Hanoi. *Proc. of the 7th Int. Conf. on Auditory Display, 2001.*
- [3] HOLMES, J: Interacting with an information space using sound: Accuracy and patterns. *Proc. of the International Conference on Auditory Display, Limerick, Ireland, 2005.*
- [4] HUNT, A and Hermann, T: The importance of Interaction in Sonification. *Proc. of ICAD 04 – Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia, July 6-9, 2004.*
- [5] KILDAL, J and Brewster, S: Providing a size-independent overview of non-visual tables. *Proc. of the International Conference on Auditory Display, London, England, 2*
- [6] KRAMER, G. Ed: Auditory Display – Sonification, Audification, and Auditory Interfaces, Addison-Wesley, 1994.
- [7] LEE, Z., Berger, J and Yeo W. S: Mapping Sound to Image in Interactive Multimedia Art. Retrieved 15th January 2008 from:
<http://ccrma.stanford.edu/~zune/sources/papers/papers.files/ccrma2004.pdf>
- [8] LIU, Y. K and Zalik, B: An Efficient Chain Code with Huffman Coding – Pattern Recognition, Volume 38, Issue 4, April 2005, Pages 553-557.
- [9] Max/MSP Jitter. *Graphical Real-Time Programming Environment.*
<http://www.cycling74.com/products/maxmsp>
- [10] MEIJER, P. B. L.: Vision Technology for the Totally Blind. Retrieved 20th December 2007 from: <http://www.seeingwithsound.com/>
- [11] Nintendo. *Nintendo Wiimote.*
http://www.nintendo.co.uk/NOE/en_GB/systems/accessories_1243.html
- [12] PAULETTO, S and Hunt, A: A toolkit for interactive sonification. *Proc. of the 10th Int. Conf. on Auditory Display, 2004.*
- [13] PAULETTO, S and Hunt, A: Interacting with Sonifications: An Evaluation. *Proc. of the 13th International Conference on Auditory Display, Montreal, Canada, June 26-29, 2007.*
- [14] RUSS, J. C: The Image Processing Handbook, Third Ed. – CRC Press, 1999.
- [15] STOCKMAN, T., Hind, G and Frauenger, C: Interactive Sonification and Spreadsheets. *Proc of the International Conference on Auditory Display, Limerick, Ireland, 2005.*
- [16] SAUE, S: A model for interaction in exploratory sonification displays. *Proceedings of ICAD 2000*
- [17] TROUILLOT, X., Jourlin, M and Pinoli, J. C: Geometric Parameters Computation with Freeman Code. Retrieved 20th April 2008 from:
<http://icsxii.univ-st-etienne.fr/Pdfs/f72.pdf>
- [18] WEISFELD, S: Stack Based Flood Fill Algorithm. Retrieved 15th March 2008 from:
-

-
- <http://drowningintechnicaldebt.com/blogs/shawnweisfeld/archive/2006/12/04/Stack-Based-Flood-Fill-Algorithm.aspx>
- [19] Wiili.org Wii Linux. Motion Analysis. Retrieved 1st May 2008 from: http://www.wiili.com/index.php/Motion_analysis [22] YEO, W. S and Berger, J: A Framework for Designing Image Sonification Methods. *Proc of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*, pp 323 – 327. Limerick, Ireland, July 6 – 9, 2005.
- [20] J. Williamson, R. Murray-Smith, and S. Hughes, “Shoogle: Excitatory Multimodal Interaction on Mobile Devices.
- [21] YEO, W. S. and Berger, J: Application of Raster Scanning Method to Image Sonification, Sound Visualization, Sound Analysis and Synthesis. *Proc of the 9th Int. Conference on Digital Audio Effects. Montreal, Canada, September 18 – 20, 2006.*
- [22] YEO, W. S. and Berger, J: A Framework for Designing Image Sonification Methods”. *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*, pp 323 – 327. Limerick, Ireland, July 6 – 9, 2005.