This paper describes an ongoing research for a method for interactive sonification of 2D images. The method utilizes an existing device (from computer games console), Nintendo's wiimote controller [13] to provide a means of interacting with the image to aid exploration, providing the user with sonic and haptic feedback of the image under scrutiny. The system currently presents a method for the segmentation, extraction and feature analysis of the irregular shapes which comprise organic images.

1. 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 [8].

Factors such as application contexts and the nature of the data being sensed (e.g. tension, heat, movement, acceleration and visual data) play a major role in determining the best mapping and synthesis methods for the sonification process. 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. As a consequence, 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 [15] coming from increased familiarity with the system.

2. BACKGROUND
A number of design consideration need to be addressed in the design of interactive and sonification systems. This section looks at some of these issues which have been addressed in related work.

2.1. Mapping in the Time Domain
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.

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 [9] and automatic scan methods based upon the raster scan technique (pixel by pixel) [21] and a left to right scanning method (column by column) [12].

2.2. Interactivity as a Focus
Continuing efforts have been focused upon implementing interactivity as a focus for sonification system design. Saue [18] introduces the concept of walking through data sets in order to determine global, intermediate, local and point data features. Hellstrom et al. [2] implement the mouse as a virtual microphone to explore data spaces. Pauletto and Hunt [14] 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.

Furthering the implementation of interaction in sonification systems has led to the development of the newer concept of Model Based Sonification (MBS) [3]. MBS is based upon implementing interactivity, not just as the focus of the system, but as a fundamental part of the exploration and sonification method. Based upon real world interaction, the creation of a data driven interaction object mimics the interaction we see in the physical world, with sonic feedback only coming through physical excitation of the object under examination.
2.3. Feedback Loop: Audio and Haptic

Interfaces with sonic feedback have been realized most commonly in the form of mouse [4], keyboard [17] and tablet [7]. Devices which offer a multimodal feedback are less common.

Work presented in [5] highlights qualities which can be gained from the user being tightly embedded within an interactive control loop, sighted increased levels of 'control intimacy', a quality seen in the manipulation of musical instruments. The following points are produced in [5] 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.

Continuous feedback or the creation of context between modes can 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" [15].

3. DESIGN AND DEVELOPMENT

This section provides a modular design overview describing the layout and the functions of the modules which for the system design.

3.1. 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 [13], 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 [11], a visual programming environment which will be used for image processing and analysis, audio synthesis, and communication with the wiimote device. Max/MSP Jitter externals have been created in C++ to produce fast image processing and analysis modules which can be integrated within the same development environment as feedback modules.

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

3.2. Region Selection and Segmentation

The system allows the user to point and select regions within the image using algorithms based upon image segmentation techniques. Taking the individual pixel the user is pointing at, identifying regions with connected areas of similar pixel colour values (via flood fill algorithm [20]) creates a segmentation of the selected region. The region is analysed (through chain code algorithm [10]) to provide parameters to control feedback modules.

3.3. Region Analysis Modules

An intermediate step is implemented to define shape descriptors from the parameters gained from the segmentation algorithms. Two classes have been defined as follows:

3.3.1. Global Shape Descriptors

Global shape descriptors are parameters defining the characteristics of the entire selected region. The following parameters can be defined [16,19].

1. **Shape area** is defined as the number of pixels which comprise the flooded area of the shape.
2. **Shape perimeter length** can be defined as the number of adjacent steps + √2(diagonal steps) within the chain code.
3. **Shape complexity** is the perimeter length² / shape area.
4. **Form factor** is defined as (4.π.shape area) / perimeter length
5. **Aspect ratio** is defined as the width / height of a rectangular bounding box for the shape.
6. **Shape extent** is defined as the shape area / bounding box area.
7. **Center of gravity (x,y)** is defined as the mean average of coordinates included in representing the shape perimeter (for x and y separately).

### 3.3.2. **Local Shape Descriptors**

Using the position of the wiimote cursor within the selected region, a new class of local descriptors can be constructed as follows.

1. **Distance from Center of Gravity** is defined as the distance magnitude of the wiimote cursor from the shape center.
2. **Local perimeter section.** Variable size sections of the perimeter can be accessed based on wiimote cursor position. Parameters such as perimeter angularity and path can be defined based upon the chain code describing the section.
3. **Local perimeter points.** Points which comprise the perimeter can be sonified with respect to their distance, horizontal and vertical position relative to that of the wiimote cursor.

### 3.4. **Feedback Categorization**

Having produced two categories of input parameter, we can similarly categorize the type of feedback to be produced. We can define two sonic categories as follows.

1. **Low level sonic parameters** such as pitch, frequency components and dynamics. Typically, the sort of parameter control in standard electronic instrument design.
2. **High level musical parameters** which may control the tonality, rhythm, polyphony and melody of the output.

In addition to these, control parameters for the rumble feedback may include amplitude, frequency and rhythm of the motor.

### 4. **MAPPING TECHNIQUES**

The system currently implements a user configurable patching method for mapping input to feedback modules. 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. For example, in displaying global characteristics of the shape to the user, would high level parameters be more effective? Would low level parameters be more suited to providing localised data?

In looking at the potential benefits of the mapping of input data to feedback modules, the following considerations can be put forward towards the design of a default mapping strategy which will best facilitate the new user.

#### 4.1. **Instrument Design Strategy**

Mapping global shape descriptors to parameters usually governed by the natural properties of an instrument such as frequency components, amplitude envelope and frequency envelope will provide the user with an instrument ‘patch’ representing the input segment. The user can then drive the instrument with data produced from local parameter creation.

#### 4.2. **Providing Orientation within the Segment**

The following mappings have been designed to provide feedback regarding the locality of the wiimote cursor within the segment with respect to the gravitational center and points which make up the shape perimeter.

1. Increased frequency and intensity of the rumble feedback can be used to display closer proximity to the gravitational centre.
2. Timbre can be used to represent distance from the gravitational center. Treating the gravitational center as an anti-node and the perimeter as nodes (from vibrating systems [6]), the frequency response can range from a full sounding response in the middle of the segment to a ‘tinny’ frequency response nearer the perimeter.
3. The scalar properties of relative pitch (vertical) and panning (horizontal) are well understood and thus offer a good method for defining points (such as perimeter points and gravitational centre) relative to the cursor position.

### 5. **CONCLUSION**

The interface currently implements a method for user driven selection and parameter extraction of segments comprising an image which have been 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 localized feedback. Mapping considerations have been put forward based on the benefits of combining certain input and feedback modules.

Continuing work looks to developing the mapping considerations which have been addressed in section 4. Work in this area will point the system in the way of defining the most effective default setting for the new user. In addition, being able to offer effective control settings for user personalization is key in producing the most effective design.

Continuing work on the project includes the development of modules for colour and texture analysis, grouping and context of segments and development of new interaction and feedback methods to facilitate these modules.
6. REFERENCES


