LA MACCHINA: REALTIME SONIFICATION OF A PAINTED CONVEYOR PAPER BELT

Alessandro Inguglia

Recipient.cc
Conservatorio "G.Verdi"
Milano, Italia
alessandro@recipient.cc

Sylviane Sapir

Conservatorio "G.Verdi"
Dept. of Music and New Technologies
Milano, Italia
sylviane.sapir@consmilano.it

ABSTRACT

This paper details a real-time sonification model named *"Scanline spectral sonification"*. It is based on additive synthesis, and was realized for the installation "La Macchina v0.6" (La Macchina). La Macchina was a project born from the latest collaboration between the artist 2501 and Recipient Collective. It is a kinetic / multimedia installation that aims to represent through sound the creative process of a pictorial work, all the while respecting its aesthetics and maintaining a strong synesthetic coherence between sounds and images. La Macchina is made from a long moving paper tape in a closed loop configuration which is activated by an electric motor via rollers. It becomes almost a kinetic canvas, ready to be painted on with paintbrushes and black ink. The image of the painting is continuously recorded by a camera, analyzed frame by frame in real-time and then sonified.

1. INTRODUCTION

This version of "La Macchina" is the last of a series of works born from the collaboration between the artist 2501 and Recipient Collective. The project originates from the artist's need to show his creative process as something flowing, rather than a static picture.

> *"This installation and body of work has developed through a progressive series of actions. My concept of painting is based on the continuity of experience, on flow rather than stillness, and it is for this reason that I am not going to show you a sequence of static, motionless slides, but something moving. Pictures and art pieces are static and indoor but they tell a story in motion and they are the result from outdoor processes."*

The first version of La Macchina was presented at Soze Gallery in Los Angeles for 2501's personal exhibition. It comprised two rollers fixed on a metallic grid, activated by an electric motor. This setup made it possible for a long paper tape to move in a closed loop. During the performance the artist used various customized paintbrushes to create patterns of lines and textures, until the tape ruptures. A second version was prototyped in 2015. Brand- new 3d-printed plastic bars were added to the structure. These bars permit the rollers to be anchored to the walls, and consequently allows for the installation to adapt better to a space. This version was presented during 2501's personal exhibition "Nomadic Experiment On The Brink of Disaster", at Wunderkammer in Rome. Another version was realized few months later following the same principle of adaptation to the given architectural space and to the environment. The dimensions of the installation were doubled to allow the public to interact with the paper tape using a set of custom paint brushes designed by the artist. The main intention was to trigger a collective pictorial act, to think about the role of the public in the context of so called neo-muralism art movement. In the most recent version of La Macchina a sonic feedback was introduced by means of using a purpose built sonification model.

This version has been presented for the first time at Movement Festival 2016 in Detroit. The unifying theme of this series of installations is concerned with the relationship between gesture and its graphical results, space and creative process. A sonic feedback was added for the first time in the last version described in this paper. It includes a camera, a computer with custom software, headphones and a video monitor (Fig. 1).

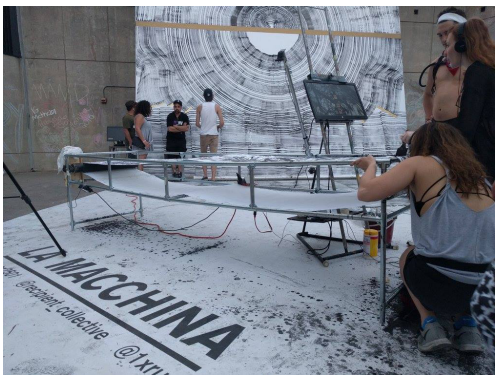

Figure 1: *La Macchina v0.6 at Movement Festival 2016*

The aesthetic and technical issues arising in the design process of an efficient sonification model (along with maintaining coherence with the sonic representation of the 2501's visual features) have proved to be complex. The artist's desire to portray the gestures as never statically depicted in order to remain in the flow of his movements and painted lines is evident. The scrolling movement of the paper tape suggests a temporal flow in which those same gestures are impressed. Another specific process of La Macchina is the closed loop, which allows for the progressive layering of visual materials. These aspects of the data can be transposed to a musical domain. Graphical materials (such as interweaving lines, texture and stipples) turn into their equivalent sonic materials, while the processes arising from the closed loop (repetition, accumulation) transform into musical generative processes. The model is based on a graphical representation of sound in the spectral domain. Visual elements painted on paper tape are used to form variable spectral sound shapes, which are then soni-

fied with an additive sound synthesis algorithm. The installation requires us to set a camera above the scrolling paper tape and to focus on the area which has just been painted by the artist. A single vertical scanline is set in the middle of the video canvas and represents the instantaneous spectrum of sound to be synthesized at that precise moment. The process comprises five main steps: image pre-processing, scanline data extraction, data analysis, the mapping process of these data and eventually; the sound synthesis. This paper will first outline a brief overview of similar works. It will then proceed by describing the sonification model named *"Scanline spectral sonification"* which has been specially designed for this installation. The last part of the paper will give some technical details about its implementation.

2. SIMILAR WORKS

2.1. First Experiments

Since the first years of the XXth century many artists and scientists have been deeply fascinated by the possibility of associating sounds and images, with newly arising technological means. An early device is the Optophone (1910) invented by Edmund Fournier D'Albe. It was designed to help visually impaired people to recognize typographic characters by converting a light intensity input to different sounds. In 1929 Fritz Winckel, a German acoustician, managed to visualize an audio signal on a cathode ray tube (CRT)[11]. The visual results of these experiments comprised figures which were similar to Chladni's patterns. Winckel also managed to receive a video analog signal on a radio [7]. It was one of the first documented attempts to generate audio signals from images. A different but more or less contemporary approach was based on sound-on-film techniques through analog optical technology. Since 1926 Russian artists like Arseny Avramov and Mikhail Tsekhanovsky started to investigate the possibility of synthesizing sounds by drawing directly onto film[11]. Avramov's first work was "Piatiletka"(1929). Over the same period similar works were developed also in Europe. Oskar Fischinger was a German animator and filmmaker based in Berlin. His "Sounding Ornaments" (1932) [4] were infact "decorations" directly drawn on the soundtrack of a film (Fig. 2). Norman Mac Laren, another famous Canadian animator and director, realized a series of similar experiments: "Boogie Doodle", "Dots", "Loops" "Stars and Stripes" (1940) are examples of such kind of short animations[1]. His technique was to directly draw on the motion picture film both figures and "sounds" with a pen, thus intending to create a strong correlation between sounds and images.

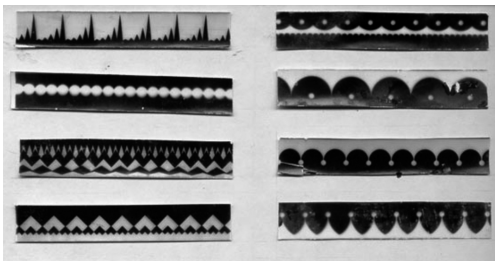

Figure 2: *Oskar Fischinger's "Sounding Ornaments[4]"*

A very well-known development harking from the first days of the digital era is notably that of Iannis Xenakis' s UPIC. It was a machine for music composition, designed and developed in Paris at CeMaMu at the end of 1970's, with the purpose of experimenting with new forms of notations. It could be defined as a "graphical composition system". The main interface in UPIC was an electromagnetic pen and a big interactive whiteboard where the user/composer was able to draw[6] (Fig. 3). All the resulting drawings on the whiteboard were recorded and visualized on a CRT monitor and possibly printed with a plotter. Graphics signs were then mapped to sound parameters following these principles. The system was based on a tree structure: the lower hierarchic graphical element was the "arch". A group of arches made up a "page", which can be considered as a sort of sonogram - but not necessarily, since it was possible to associate a specific meaning/function (waveshapes, envelopes, modulations, etc.) to drawn shapes. Eventually these pages could be grouped or layered. It was possible to "explore" the pages by moving a cursor, to give birth to the musical forms. The first system could work only in deferred time. In the second, faster and real-time version, the number of arches was limited to 4000 per page and 64 overlaying "voices". UPIC could also be defined as a sonification system as it converts graphical data to sounds by means of audio synthesis.

Many other models use a time-frequency approach which is in some ways similar to the one adopted for the realization of "La Macchina". Famous commercial software like MetaSynth or Adobe Audition can be good examples in this case. The basic idea is that of considering an image as a score which is progressively read from left to right. While Metasynth uses color data to move the sound on the stereo front, in Adobe Audition the same kind of information is directly mapped to the amplitude of resulting sounds. Another similar model is Meijer's [9]. It was developed as a medical aid for people with visual impairments. Similar to "La Macchina" it uses a camera which scans from left to right, transforming pixel positions on the vertical axis to frequencies, while the amplitude is directly proportional to the pixel brightness. In this case the data mapping is completely reversible. Put simply, the sound is generated from images, and from the resulting sound it remains possible to return to the original image as all the data is preserved in the process (involving no loss of information).

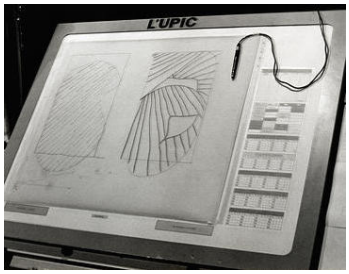

Figure 3: *UPIC whiteboard. (Centre Iannis Xenakis)*

2.2. Raster Scanning and other approaches

Another more modern approach is based on Raster Scanning. This technique consists of reading consecutive pixels with left-right and top-bottom ordering, row by row. The sampled data is used to directly generate the audio signal as a waveform. Pixel values are mapped to linear amplitude values between -1 and 1. In this particular case, time does not develop on the horizontal axis. The image is read at sample-rate, and consequently the resulting pitch

is influenced by the image dimensions (in that respect, the rastrogram is a very interesting approach to graphic representation of sound [13]).

More recent sonification models expect[10] the image to be pre–segmented in a particular order before being analyzed and sonified. Others specify various paths inside the image[2], or user-selected areas that are selectively sonified [7]. In this regard an interesting example is the method adopted by Vosis, an interactive image sonification application for multi-touch mobile devices, which allows one to control in real-time the sonification process of images through gestures[6].

A peculiar experience in the field of sonification is the case of Neil Harbisson's eyeborg, even if it is probably more closely related to color sonification. In 2004 the Irish musician and artist, affected by achromatopsia (a condition which imparts total color-blindness) decided to have an antenna permanently implanted to his head. The device allows him to perceive colors as micro-tonal variations. Each color frequency is mapped to the frequency of a single sine wave. Low-frequency colors are related to low pitched sounds, high-frequency colors to high pitched sounds. The model divides an octave in 360 microtones that are relative to specific degrees of the color wheel. The device is also connected to the Internet and only five chosen people are authorized to send pictures to the system. During a public demonstration, which was followed over live-streaming by thousands of people Harbisson could identify a selfie as the image of a human face. Neil Harbisson refers to his particular condition as sonocromatism / sonocromatopsia. He excludes the term synaesthesia because in that case the sound/color relation is generally subjective.

3. SONIFICATION MODEL

3.1. Methodological approach

The sonification model of La Macchina (Fig .5) is based on the interpretation of visual elements painted on a paper tape as graphic representations of sounds in the spectral domain. These visual elements will determine sound shapes (referred to as Smalley spectromorphologies [12]).

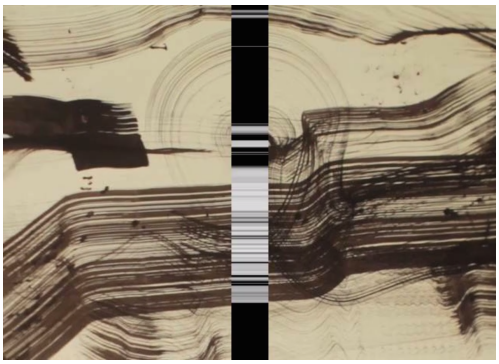

Figure 4: *A frame of the scanline sonification process*

This process considers the vertical axis of the tape as the frequency axis, and the color intensity of the pixels as the dynamics of the spectral components. As an arbitrary musical choice and in order to emphasise the sound shapes we introduced a process which varies the frequency mapping during the performance. In "La Macchina" the visual elements of the paper belt are captured by a camera and digitalized before being processed. Therefore, we work with a double time dimension which is defined by the scrolling speed of the paper-tape and by the frame rate of the video: substantially, a series of consecutive sonograms. To solve the problem of this timing ambiguity, we decided to use the data extracted from a central column of pixels (the scanline), to generate instantaneous spectra, and concatenate consecutive spectra in time, to form a sonogram (Fig. 4). The transition rate between consecutive spectra directly depends on the frame-rate and effectively affects the time-resolution of the sonification process. While the frequency resolution is determined by the number of pixels in the scanline (generally the height of the canvas in pixels), time resolution is simply the ratio between the speed of the paper tape, and the camera capture frame rate. In the current version of the model, the slide speed of the paper is about 2.5 cm/s, while the frame rate is 25 frames per second. This setup allows the system to run with a time-resolution about 1mm per frame. The choice of placing the scanline in the middle of the captured frame has been made empirically. Infact this frame is also displayed on a screen for the audience. We have experienced that setting the scanline next to the borders of the image did not create a good time synchronization between sounds and the new visual shapes which appeared on the right part of the screen. The analysis process of the scanline extracts color gradient values, by calculating the color intensity difference for each pixel in the scanline for a definite number of consecutive frames. Each pixel in the scanline represents a single component of the sound spectrum, which will then be activated whenever a sudden color variation occurs. In the overall flow of the installation gestures are transformed into signs (painted on paper), then into codified symbols (when the information is digitalized) and eventually into sounds. Somehow the sonic feedback will then influence a new gesture, as it has already been experienced during the live open sessions at Movement Festival in Detroit. Within this installation we have to deal with two types of feedback, a sonic feedback, and a graphic feedback due to the closed loop process.

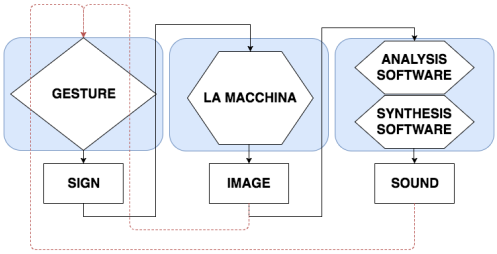

Figure 5: *Simple flow diagram of the model*

3.2. Software environment and developing tools

A first prototype was realized with Openframeworks, a set of C++ libraries for "creative coding" and then ported to Max/MSP in the last version. Infact Max/MSP has proved to be an efficient environment for the development of the sonification model and video analysis routines. It is a dataflow programming language which allows a rapid development of multimodal interactive applications with a deep focus on audio. Moreover, it supports GLSL, a shader scripting language, which can be useful to process the video on the GPU, leaving more resources for audio computing on the CPU.

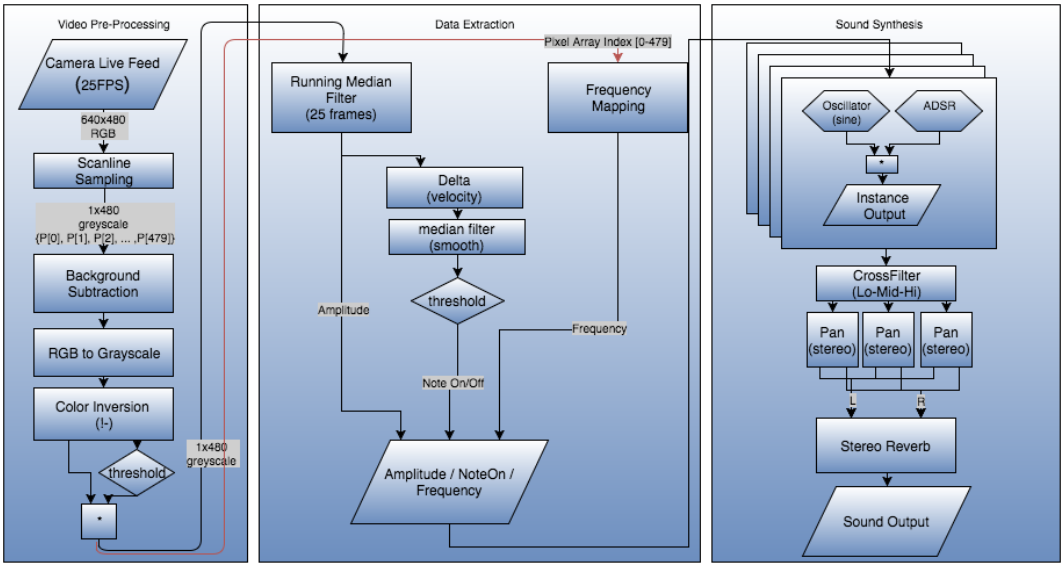

Figure 6: *A more detailed flow diagram of the sonification model*

4. REALIZATION

The developed Max/Msp application is made-up of three main functions: the scanline analysis and pixel parameters extraction, the mapping process of these parameters and the sound synthesis (Fig. 6). The video live feed is pre-processed, with background-subtraction techniques and other filtering processes. The video analysis is made on a single vertical array of pixels (scanline) with greyscale color format. It is based on the color variations of each pixel between consecutive frames as described below.

4.1. Image Pre-processing

The video capture frame rate is 25 FPS, and consequently the video analysis algorithm works at the same speed. For each frame (a 640 x 480, 8-bit RGB pixels matrix) a single central column of pixels is extracted and stored in a 480 elements array (the scanline). The video live feed is pre-processed with a color-subtraction algorithm. In reality, depending on the variety of possible differing lighting settings for the installation, the paper could never result as completely white. The RGB data is then converted to greyscale by computing the luma brightness of each pixel, where 0 corresponds to black and 1 to white. As the painted ink is black and the paper is white, it is convenient to invert the image color array. To ensure that little imperfections on the paper or light shades will not be accidentally sonified a threshold is set, such that only pixels above a certain value will be considered. The signal is then smoothed with a running median filter of the tenth order which is useful to remove noise. If the processed data were visualized as an image, it would appear as the original video greyscale picture with a blur-like effect. Then the slope of the brightness variation of each pixel (the tendency to shift towards white or black) is estimated and is used to control the parameters of the audio synthesis model, as explained in the next sections.

4.2. Audio Synthesis Model

In this version of the software the total number of oscillators is equal to the number of vertical pixels in the video frame. As the scanline pixel column is a 480 pixels array we get 480 oscillators, which remains a sustainable computational cost for a modern average CPU. Clearly the calculation could result extremely heavy, with large and more defined images. Nevertheless this problem could be easily solved by undersampling the image on the y-axis. As a first prototype, an inverse-FFT based model was developed, for a direct spectral sonification approach. Even if the outcome may not be uninteresting, the sounds were too noisy for the desired result, as we had predicted. For this reason we developed a more flexible model in terms of frequency and amplitude control which is based on additive synthesis, using sine oscillators with independent static frequencies and amplitude envelopes. Data relative to color variations of each pixel between consecutive frames are used to trigger and to control the ADSR amplitude envelope of each sinusoidal oscillators. Frequencies are non-linearly mapped along the y-axis of the canvas, and therefore arbitrarily quantized to chosen modal scales as detailed in the following section. Finally some pseudo-spatiality is added to the synthesized sounds by using amplitude panning. A cross-filter subdivides the audio signal in three main spectral bands (Low-Medium-High), which will be independently spatialized. The panning process uses constant-power function and slow sinusoidal movements of the above mentioned bands. To further enhance the feeling of spatiality the signal is then processed with a digital reverb (Gverb Max/MSP external by N. Wolek, based on Griesinger's reverb model).

4.3. Data Mapping and Events Triggering

Two prototypes were first realized with raster scanning techniques and spectrographic sonification. The pixels values were directly mapped onto the amplitude of the single sample for the former, and onto the amplitude of a single FFT bin for the latter. As these "direct" mapping approaches were not satisfying our objectives we

chose instead to work on parametric data mapping. We decided to use data relative to color variations of a single pixel to control the parameters of a single sine oscillator: the frequency, the peak level and the duration of its amplitude envelope.

By calculating color difference in time, between consecutive frames, we obtained the color gradient(velocity) towards white or black, depending on the sign of the slope. For each pixel of the column, whenever a color gradient exceeds a threshold-value the amplitude envelope of the corresponding oscillator is activated and its peak value is determined by the instant intensity of the pixel color. When the color variation goes below another threshold-value the envelope is released. Furthermore, in order to avoid a too simple and predictable distribution of the frequencies along the vertical axis of the video (which may lead to poor musical results) the model provides a nonlinear mapping function for the frequency of the oscillators. Lower pixel positions match with low-pitched sounds, while high pixel positions match with high-pitched sounds.

The mapping depends on an arbitrary frequency range [50-17000 Hz] which has been quantized according to modal scales. In this version we used modal scales built on different degrees of the major scale. The mapping is based on a table-lookup algorithm, using the pixel number of the scanline (from 0 to 479) as an index to address an array of arbitrary frequency pitch values. For this installation we use 7-notes scales which are repeated over many octaves, in the limits of the audible frequency-range. We have seen that a number of around 63 pitch frequencies seemed to be appropriate for scales made-up of 7 elements (i.e. 9 octaves). Thus the total number of pitch frequencies stored in the array should depend on the number of notes used to generate the musical scale. Scales with large intervals between degrees have less notes thereby inducing a smaller pitch array.

As the number of pixels is mostly greater than the number of pitch frequencies we could not apply a one-to-one relationship between indexes and frequencies. In order to avoid a many-to-one mapping solution which would assign more pixels to a single frequency (thus yielding undesirable peaks of spectral energy) we decided to adopt the following strategy. The array would be addressed by applying a (kind of) quantization process on the index, but in order to diversify the frequencies and to enrich the overall spectrum each consecutive repetition of the same frequency would be substituted by an integer multiple of that frequency. This process generates the harmonic series of the base frequency, and whilst taking care not to exceed the maximum frequency of 17000Hz, it also guarantees no frequency repetitions thereby preserving the musical characteristics of the chosen scale. However "La Macchina" is not strictly tied to a specific scale or to the equal temperament. In fact it would be possible to manage the pitch system in many other different ways by providing any pitch frequency contents.

5. CONCLUSION

The outcome was positive since the first prototype, notably regarding synesthesia between brightness and sound intensity, lines and dynamics. Stipples and thin graphical elements relate to sounds with similar morphologies. Larger brushstrokes and interweaving lines produce real sonic textures. The closed loop of the paper belt which causes repetition, accumulation and layering of graphic elements is enhanced and also immediately perceived through the repetition, the accumulation and the densification of the sonic materials produced by the process of sonification. The dramatic visual effect is then accompanied by a corresponding increase in musical tension which both affects the painter and its gesture, definitively closing the loop.

This software, more than a direct sonification system, could be defined as a generative process of events which musically controls an additive synthesizer. However it differs from the models used in commercial softwares like Adobe Audition or MetaSynth even if it partially shares with them a spectrographic approach. While in the first version of La Macchina (prior to the addition of the sonification system), the end of the process was due to the rupture of the paper tape, in this case it is produced by the servo-motor shutdown. At the moment there is no automatic interruption of the sonification process. The sound freezes on the last video frame. A process which smoothly interrupts the audio signal whenever the image is static could be easily introduced. Other future developments could include color data mapping, to associate RGB color variation with new parameters of the audio synthesis process, such as panning or frequency mapping functions. The model presented in this paper could also be used for the sonification of other looping mechanisms. An interesting application of the system could be the sonic enhancement of imperceptible imperfections on materials such as paper or porcelain.

La Macchina was presented for the first time at Movement Festival 2016 in Detroit (Fig. 7), where it was received with wide admiration amongst attendees. Experiments with non-painters and otherwise inexperienced people showed how the musical feedback of the installation influenced their drawings and how they were able to adapt their painting patterns to reach a significant musical result. For example many tried to draw stipples on the lower part of the paper tape, trying to generate sort of a bass drum; or repetitive patterns, to imitate the typical iterative structures of techno music. A second version of this installation has already been presented in Berlin. It was based on a paintable turning paper-disk. The substitution of the paper belt by a disk, the dimensions of the disk and its relatively high speed of rotation, compromise the time resolution and the efficiency of the sonification system. This confirms the importance of coherence between the sonification model and the artefact (or the data) to sonify it whilst designing an interactive audio installation.

Figure 7: *La Macchina at Movement 2016*

6. REFERENCES

[1] H. Beckerman, *Animation, The Whole Story*. Allworth Press, Feburary 2004, pp. 5152.

[2] K. M. Franklin and J. C. Roberts, "A path based model for sonification", in *Proc. Eighth International Conference on Information Visualisation (IV04)*, 2004, p. 865-870.

[3] T. Hermann, *Sonification for exploratory data analysis*. PhD thesis, Bielefeld University, Bielefeld, 2002.

[4] T. Hermann, A. Hunt, J. G. Neuhoff (Eds.), *The Sonification Handbook*. Logos, Bielefeld, 2011.

[5] T. Hermann and A. Hunt, "The Discipline of Interactive Sonification", in *Proc. Int.Workshop on Interactive Sonification (ISon 2004)*, Bielefeld, 2004.

[6] H. Lohnerand, "The UPIC System: A User's Report" in *Computer Music Journal*, 10(4) ,Winter 1986, pp. 42-49

[7] R. McGee, "VOSIS: a Multi-touch Image Sonification Interface", in *in Proc. New Interfaces for Musical Expression (NIME)*,2013.

[8] R. McGee, J. Dickinson and G. Legrady, "Voice Of Sisyphus: An Image Sonification Multimedia Installation", in *in Proc. of ICAD (ICAD)*, 2012.

[9] P. Meijer, "An Experimental System for Auditory Image Representations", in *in IEEE Transactions Biomedical Engineering*, vol. 39, pp. 112-121, 1992.

[10] R. Sarkar, S. Bakshi and P. K. Sa, "Review on Image Sonification: A Non-visual Scene Representation", in *Recent Advances in Information Technology (RAIT)* National Institute of Technology Rourkela, India, 2012.

[11] B. Schneider, "On Hearing Eyes and Seeing Ears: A Media Aesthetics of Relationships Between Sound and Image" in *See this Sound. Audiovisiology II, Essays. Histories and Theories of Audiovisual Media and Art*, Linz/Leipzig: Verlag der Buchhandlung Knig, 2011.

[12] D. Smalley , "Spectro-morphology and Structuring Processes" in *The language of electroacoustic music*, Springer, 1986.

[13] W. S. Yeo and J. Berger , "Raster Scanning: A New Approach to Image Sonification, Sound Visualization, Sound Analysis And Synthesis" in *Proc. International Computer Music Conference (ICMC)*, 2008.