HUE MUSIC – CREATING TIMBRAL SOUNDSCAPES FROM COLOURED PICTURES

Dave Payling, Stella Mills, Tim Howle (University of Hull)

Creative Technologies Research Group
Staffordshire University, Stafford, UK
d.payling@staffs.ac.uk

ABSTRACT

This paper discusses a technique to convert 2 dimensional images into music by associating hue values with timbres. The colour information in 2D still images was used to drive an 8 channel timbral audio mixer. 8 musical timbres were recorded to represent 8 hue values and these timbres were changed in amplitude dependent on the quantity of each hue in the image. A maximum of 64 blue 'pixel-blocks' for example resulted in a sound that was exclusively the timbre associated with blue. The technique is successful for enhancing some images but not others but after some familiarisation it is possible to distinguish the timbres and therefore to identify the colour composition of the picture through audio alone.

[Keywords: Colour, Timbre, Aesthetics, Pictures]

1. INTRODUCTION

Because both music and visual art are usually created with a view to producing an aesthetic appreciation from the audience or viewers it is not uncommon to believe there is an underlying link between both art forms. This 'desirable' aesthetic link has only been loosely investigated and then mainly in the visual realm with a strong research and artistic user group under the umbrella term of 'Visual Music' [1]. Visual Music looks at ways of conveying musical aesthetics with visual media such as paintings and video works. Fewer attempts have been made to represent visual aesthetics in a musical form and this is where this paper begins its investigation.

The investigation described here is a preliminary one to inform the creation of a system that will allow the generation of music automatically for moving images. Creating music automatically from images is possible in many ways and several software programs such as Matlab and MAX/MSP with Jitter allow this to be achieved quite simply. A greater problem is creating music that successfully represents or increases a viewer's appreciation of an image without any user or composer input and has a desirable and image enhancing aesthetic quality. Of course, each viewer is different and makes his or her own associations with the image but this work attempts to find a common ground where the music created for the picture will increase the viewer's appreciation of it in as many cases as possible. Conversely, it would be possible to compose a piece of music by traditional means that would guide the viewer into previously unconsidered associations with the picture much as a movie composer sets the scene with their own creative input. In this work, this approach is to be avoided as the intention is for the system to be used by non-composers with an appreciation of aesthetics but not necessarily any musical ability. Several implementations such as Metasynth's Image Synth editor [2] and the vOICe [3] have been used that take pixel lightness and vertical positional information to determine amplitude and pitch values respectively. Metasynth now also uses the colour information of pixels to determine the spatial positioning of oscillators in the stereo sound field. In both these implementations the image is scanned from left to right either at a fixed or user chosen rate. The result generally is a randomly changing timbre from which users can learn to perceive visual information through auditory means but has limited musical usefulness without careful programming. Both systems also use a type of raw additive synthesis with frequency components chosen by the image pixels. By starting with pre-recorded musical timbres as soundfiles the system devised in this paper could be thought of as a type of wavetable synthesis which can produce more interesting timbral soundscapes. Although this paper works exclusively with 2 dimensional still images it is envisaged that the approach that can be translated to moving images and automatic generation of music for film.

Section 2 looks at related work in this area followed by the main body of the paper with section 3 explaining the method of processing the image from colours to timbres. Section 4 shows how the image is scanned to focus on important areas of the image and section 5 discusses the method of timbral mixing after the quantity of hue values in the image have been identified. A series of rendered compositions are presented in section 6 with further work and conclusions given in sections 7 and 8.

2. RELATED WORK

The conversion of images to sound is currently a popular research topic both in computer music and sonification arenas but there is a long history connecting visual aesthetics to music. Notable contributors to the exploration of links between visual and musical aesthetics are Newton in 'Opticks' [4] and Kandinsky [5]. Newton's ideas were scientifically driven and the proposed technique of matching visual colour information to tonal music was basically a frequency division from the high frequency colour spectrum down to the relatively low frequency audio one. Kandinsky, coming from a more artistic viewpoint, believed that music was motivated by a purer aesthetic and compositional force than representational visual art and sought to put the music into pictures in his own canvases.

A good review of more current research in image to sound conversion is given by Giannakis [6]. His thesis on sound mosaics used empirical testing methods to inform the development of a visual mosaic sound synthesis program that successfully utilises saturation and lightness values mapped to pitch and amplitude respectively. The inclusion of texture was an important enhancement that allows the generation of different timbres mapped from the textural properties of images. This system is however a sound synthesis program that creates a sound from several mosaics rather than a direct interpretation of an existing picture.

Much of the work in making music from images has been undertaken to assist blind people and Alty [7] investigated various techniques for displaying graphical information using only music with the Audiograph system. Various scanning techniques were used to probe the graphs and shapes and the musical parameters were mapped from spatial co-ordinates to musical pitch and 2 timbres, organ and piano. A good level of success was achieved with blind users being able to recognise and manipulate various objects shape and size. A similar assistive technology was developed by Ausiello et al [8]. They demonstrated the use of red, green, blue (RGB) and hue, saturation, lightness (HSL) colour spaces for making auditory representations of colour information in an environmental visual scanning device known as the 'Espacio Acustico Virtual' (EAV or Acoustic Virtual Space). This device is to be used by sight impaired users to increase their environmental awareness and adds a colour dimension to the auditory landscape. Also Margounakis and Politis [9] demonstrated a technique using coloured 'bricks' and a 'colour wall' to control chromatic melodies based on the brick size, or contiguity of several pixels of the same colour, and hue values. The intention of Margounakis is to create an auditory display of pictures that will also involve the intervention of a composer to make the necessary adjustments to provide a satisfactory melody. Development of the Sonart application has seen Seung Yeo [10] [11] tackle many of the issues involved with scanning within images and choosing appropriate image to sound mapping parameters. A recent success in this area has produced a very intuitive raster scanning technique that can be used for sound synthesis and waveform display.

3. IMAGE PROCESSING

In the current investigation the add on library 'Jitter' along with Max/MSP was used for analysing and processing the image. In Jitter an image contains 4 planes of information representing the alpha (opacity), red, green and blue (ARGB) values for the 2D image matrix. Each plane consists of a 2 dimensional matrix storing the A, R, G or B values of each pixel. At the heart of the colour analysis part of the program for this investigation is a routine that breaks the image matrix into a series of 8x8 submatrices, identifies the average hue value of each submatrix and then quantises the average hue to the one nearest the list of hues given in table 1. Although the RGB colour space is used for much of the analysis work the hue values in HSL colour space are used for association with timbres because it is believed to be a more intuitive approach used by artists for choosing and mixing colours and allows extension to more colours at a later stage.

1. Black	2. Red	3. Yellow	4. Green
5. Cyan	6. Blue	7. Magenta	8. White

Table 1 *Hue Values used for Timbral Association*

The choice of these 8 hue values was made to allow the colour quantisation described in section 3.1 below to be enabled more effectively. As a simple example we can consider the image in figure 1, a simple half blue half yellow image.

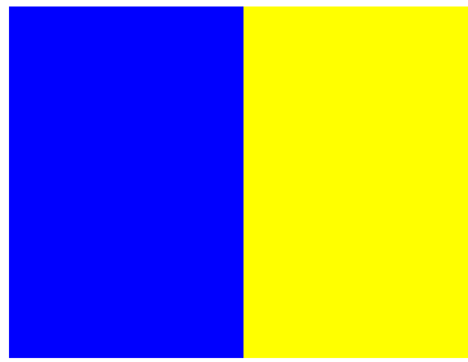

Figure 1. *Blue (left) Yellow (right) Picture*

This picture is loaded into a Jitter matrix and through a number of processes described later is then broken down into an 8 x 8 'chessboard' pattern of submatrices producing a grid as shown in figure 2. This display is made up from 64 coloured MAX panels whose colour is dynamically adjusted after the image is analysed.

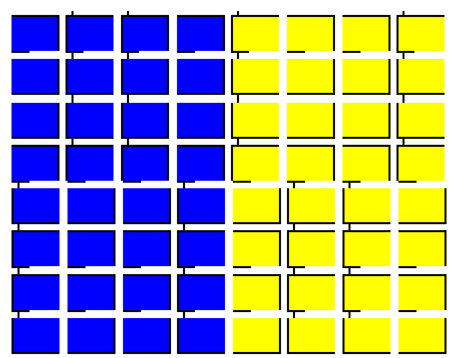

Figure 2. *Blue - Yellow Matrix*

These 64 coloured panels are for display purposes only and visually indicate that the downsampling process is working correctly. The panels are then counted and in this case produce values of 32 panels of blue colour and 32 panels of yellow. The discussion in section 5 will show how these numbers are then fed into the sound generation routine that will produce a sound which is a combination of the blue and yellow timbres at equal amplitude.

3.1 Hue Quantisation

The preceding image works without any problems since it only contains the hue values chosen for the experiment and listed in table 1. A more complicated problem is when the image contains other hue values not accounted for in the list of 8 hues given in the table as in a real world example. In this case the colours are quantised to the nearest hue value based on their closeness to the RGB values identified by Jitter. The down sampling into a series of 8x8 submatrices takes place in the same way. As an example of the process the photograph of some Manatee swimming in an aquarium in figure 3 is used as this produces a satisfactory final musical composition.

Figure 3. *Aquarium Photograph*

First the image is segmented into a series of 8x8 blocks or submatrices as before to allow colour analysis to be undertaken. The segmentation is done within Jitter to produce 8 rows and 8 columns. The segmentation process breaks the image into the desired number of subsections but displays the colour from the top left hand value of the submatrix. This is later resolved with the use of an analysis which takes values from the entire submatrix.

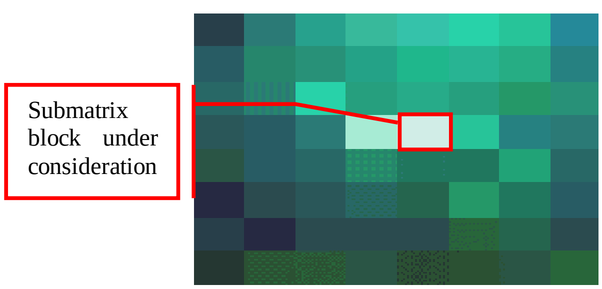

Figure 4. *Picture down-sampled into an 8x8 Matrix*

As shown in figure 4 it can be seen that the block containing the lightest shade of colour is in row 4 column 5; we will use this block for demonstration but the process is the same for all 64 blocks. This submatrix is analysed to determine the statistical mean value of the RGB values contained in it (the alpha channel is ignored in this implementation). The actual mean values were Red = 125, Green = 217 and Blue = 182. The maximum possible value for each plane is 255 with 128 being the halfway point.

RESULTANT HUE	RED value	GREEN value	BLUE value
Black	0	0	0
Red	255	0	0
Yellow	255	255	0
Green	0	255	0
Cyan	0	255	255
Blue	0	0	255
Magenta	255	0	255
White	255	255	255

Table 2. *Quantised Hue Values*

Anything above 127 is given a value of 255 and anything below 128 is given a value of 0 to give a series of quantised hue values shown in table 2.

The values from the block at row 4, column 5 would therefore be quantised to Red = 0 (actually 125 which is $<128 \therefore$ given a value of 0), Green = 255 (actually 217 which is $>127 \therefore$ given a value of 255), Blue = 255 (actually 182 which is $>127 \therefore$ given a value of 255). The combination of Red=0, Green=255 and Blue=255 gives a resultant hue of Cyan shown in table 2. The completed quantised matrix is displayed in figure 5.

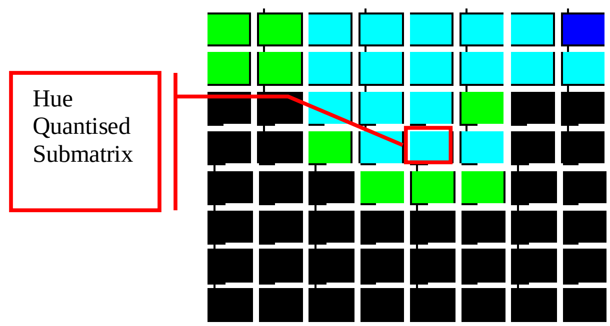

Figure 5. *8x8 Matrix of Quantised Hue Values*

3.2 Matrix Histogram

The 64 submatrix panels are counted to determine the quantity of blocks containing the same hue values. This is tabulated in a histogram graph (again only for display purposes). For this image the counted values are 36 black blocks, 10 green blocks, 1 blue block, 17 cyan blocks and the remaining colours are not represented. As before these histogram values are used to drive the timbral mixer described in section 5 and in this case produces a denser sound containing more timbres.

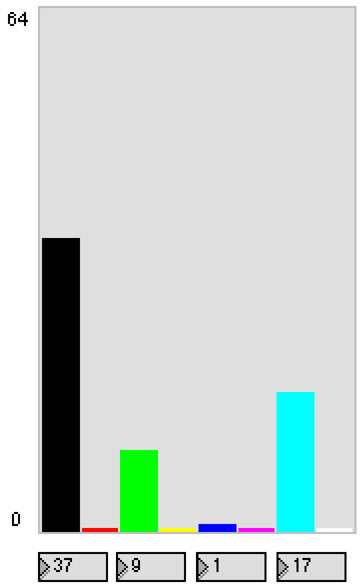

Figure 6. *Hue Histogram for the Aquarium picture. Colours left to right as in table 1.*

4. SCANNING AND FOCUS

It may seem that the effect of converting only the colour information to timbres conveys no sense of the composition of the picture but this is solved by the ability to zoom into a particular area in the image. The ability to scan through the pixels allows the timbre to change as it passes over different colours thus giving auditory clues as to the compositional nature of the picture.

A static approach would also produce a non-evolving timbral soundscape although some variety and 'movement' of the sound is produced because of the varying lengths of the soundfiles, the dynamic quality of the recordings and the way different timbres interact. Yeo [10] discussed several techniques that could be used for introducing a time base into a static image and to allow movement and focus between different areas and pixels in the picture. The technique used in this paper is to focus on a rectangular subset of the image initiated by a probing mouse click. This emulates the process of an observer choosing a focal point in the picture and looking at this area in more detail. The focus then returns to an overview of the picture after which the observer again focuses into the point of interest. The process of zooming in to the focal point and then moving back to an overview continues until a different focus point is chosen. Some high frequency jitter is included while the image is zoomed into an area as the eye refreshes its visual receptors by small high frequency movements as described by Gregory [12]. This excites the mix a little, adding a form of modulation to the produced music. The saccadic eye movement is emulated by the tracking from overview to zoomed localisation at a predetermined rate. This has a more gradual effect on how the sound evolves over time.

To initiate the zooming the user identifies a focus point in the image and clicks the mouse at the centre of the focus. The program then gradually repositions the display to the chosen co-ordinates and resizes the display around the rectangular focus.

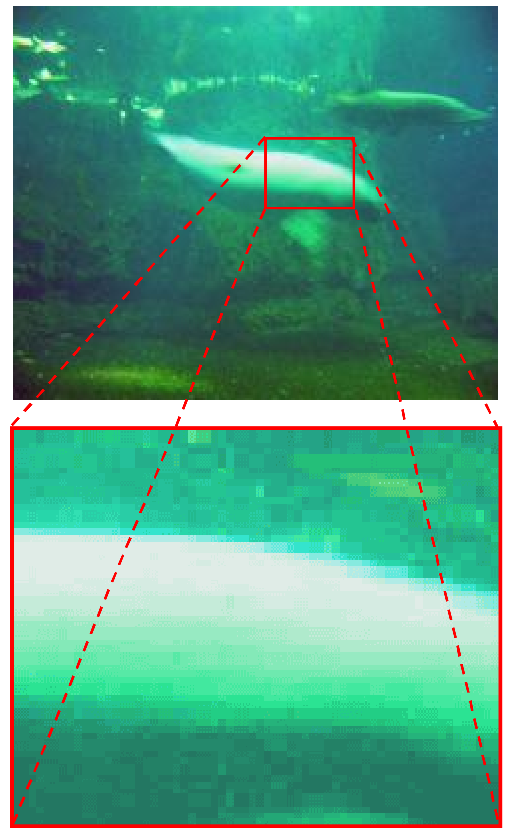

Figure 7. *Focussing on an image area*

For example in figure 7 the user has clicked into the light region on the back of the Manatee and the program enlarges this area with the clicked point being at the centre of the rescaled image. This action also has the effect of adjusting the relative amplitudes of the timbres thus creating a new, gradually evolving soundscape. This imparts a compositional aspect to the music which will now vary to a lesser or greater extent dependent on the concentration of hues in the zoomed area. This makes the music much more interesting than if a static image was converted to music.

5. AUDIO TIMBRE MIXER

The timbre of a sound has been discussed at length elsewhere [13] [14] but can be thought of as the characteristic quality or sound of an instrument that distinguishes it from others. The association of timbre with hue is a logical association to make when mapping HSL parameters to musical ones and was first suggested by Barrass [15]. It does have problems however, in that it can be seen as an oversimplification as timbre has many identifiable qualities whereas hue can be described by a single number. It is common however for musicians to refer to the timbre of instruments as their colour and in the past scores have been printed in different colours for different instruments as demonstrated by R. H. M. Bosanquet [16].

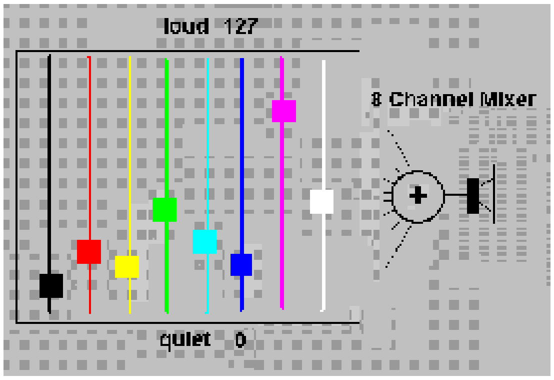

Figure 8. *Timbral 'Hue' Mixing. Colours left to right as in table 1.*

The 8 timbres (7 actual sounds with silence for white) are available on the website accompanying this paper [17]. Each of the sounds was recorded with the authors concentrating on a sound that 'resembled' the chosen colour based on the literature together with some compositional free license. For example, Kandinsky [5] writes *'it would be hard to find anyone who would try to express bright yellow in the bass notes, or dark lake in the treble'*. This implies a synaesthetic quality to sound and colour perception that many people have to some degree or another.

The synthesis programs were tweaked from a starting preset sound until the timbre matched, in the authors opinions, the colour being associated. As well as the synthetic sounds two 'found sound' samples, one of fire for red and another a breaking glass sound for magenta, were processed from existing copyright free recordings. All the sounds were recorded to last approximately 30 seconds with some variation in length between the audio files to ensure more variety when they were looped during playback. All the soundfiles were normalised in amplitude to ensure consistent sound levels.

The values generated from the colour matrix and identified in the histogram shown in figure 6 are doubled to give values between 0 and 127 (initially values between 0 and 63 are generated because of the 64 blocks in the segmented image but the audio output controller has double this resolution at 127). A value of 0 for a hue would mean an amplitude of zero for that sound i.e. the sound would not be heard. A value of 127 on any hue would mean that the associated timbre was played at full amplitude (and in this case no other sound would be heard).

The whole process can be envisaged as a 7 channel stereo audio mixer with white creating a reduction in overall volume where there is a prevalence of white in the image (an analogy to white space in an image containing no information apart from contextualizing the objects of interest). Figure 8 helps in visualising the process that occurs when mixing the timbres together. Each of the stereo samples is continuously looped and played back from disk with their amplitudes set by the histogram values. In figure 8 the timbre associated with magenta would be the loudest one in the mixed music. There would be a combined sound at the output of some quantity of each of the other colours. The mix levels will change dynamically as the image is scanned as described in section 4.

6. COMPOSITIONS

To demonstrate the process further some more pictures that have been rendered through the software to make music are presented below. All the audio files can be found on the website [17]. To highlight a potential problem with the system and discussed in section 7 the wedding picture in figure 9 generates similar music to the aquarium photograph in figure 3.

Figure 9. *Wedding Dance*

Black and green dominate this image and the result is a gentle sweeping wash soundtrack similar to the one already heard for the aquarium.

To demonstrate the effect of a quantity of white in an image figure 10 is a close up of a tree's bark. Although the predominant colour appears to be a general shade of brown, when run through the quantising process many of the brown shades reduce to white because of low values in the RGB planes. The overall result of this is a reduced amplitude piece of music with concentrations of black, red and yellow timbres.

Figure 10. *Tree Bark*

Finally we will consider a small group of flowers. The picture contains mainly yellow and green hues. These are translated to yellow and green timbres producing distinctively yellow and green music.

Figure 11. *Flowers*

7. FURTHER WORK

The current implementation has several limitations which can possibly be resolved by further investigation. First and probably most importantly, is the fact that pictures with similar colour values will produce the same music even though the pictures may contain very different compositions and content. This is partially resolved by the scanning process but it is envisaged that this can be further improved by a greater subdivision of the picture into more submatrices. Also it should be possible to identify important objects in the image which could be used to produce a strong timbral focus or identifiable musical composition itself. To further emphasise the important 'objects' in the image the background colours would be given less priority in the mix and could be set to 'fill out' the music with a background randomised pad or melodic sound. The image objects would therefore be given priority in the audio mixer and the background would have less emphasis but would still be important in putting the objects into context.

More colours and therefore timbres could also be used to generate a greater variety in combined timbres; alternatively the additional manipulation of the pitch of the timbre files via analysis of the colour saturation as described by Giannakis [6] would greatly enhance the variety of the generated music. This could also address the problem described above.

Several trials also need to be undertaken to generate qualitative data for gauging users' opinions on the success of the timbre-hue mapping and scanning algorithms.

8. CONCLUSIONS

This paper has shown that it is possible to enhance (and represent) a picture with automatic timbral mixing based on hue content of an image. The investigation provided most success with the aquarium picture but less so with other images. This is probably due to the nature of the slowly evolving timbres and the aquarium suggesting slow sweeping movements. Further refinements of the application will pay more attention to important objects within the picture and a greater subdivision of the picture into smaller submatrices will bring out finer detail that the images contain.

9. REFERENCES

[1] O. Mattis. (2005) *Visual Music: Synaesthesia in Art and Music Since 1900.* Thames & Hudson Ltd

[2] E. Spiegel. (2006) *"Metasynth 4.0. User Guide and Reference'"* Accessed 28/12/2006. http://uisoftware.com/DOCS_PUBLIC/MS4_Reference.pdf

[3] P. B. L. Meijer. (1992) "An experimental system for auditory image representations." *In IEEE Transactions on Biomedical Engineering* 39, 2: 112-121.

[4] I. Newton. (2003) *Opticks.* Prometheus Books, New York. ISBN 1591020956.

[5] W. Kandinsky. (1977) *Concerning the Spiritual in Art.* Dover Publications inc. ISBN 0486234118.

[6] K. Giannakis. (2001) "Sound Mosaics A Graphical User Interface For Sound Synthesis Based On Auditory-Visual Associations" PhD Thesis, Middlesex University School of Computing Science.

[7] J. L. Alty (1998) "Communicating Graphical Information to Blind Users Using Music: The role of context". *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* Los Angeles, United States. Pages: 574 – 581.

[8] L. Ausiello, E. Cecchetelli, M. Ferri, and N. Caramelli. (2006) "Acoustic Rendering for Color Information" *Presented at the 120th Convention 2006 May 20–23 Paris, France.*

[9] D. Margounakis and D. Politis. (2006) "Converting Images To Music Using Their Colour Properties" *Proceedings of the 12th International Conference on Auditory Display*. London, UK, June 20-23, 2006. ISBN 0-902-23820-5.

[10] W. S. Yeo and J. Berger. (2005) "A Framework For Designing Image Sonification Methods" *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*, Limerick, Ireland, July 6-9, 2005.

[11] W. S. Yeo and J. Berger. (2006) "Raster Scanning: A New Approach to Image Sonification,Sound Visualization, Sound Analysis And Synthesis" *Proceedings of the International Computer Music Conference*. New Orleans, LA, USA, 2006.

[12] R. L. Gregory. (1997) *Eye and Brain – The Psychology of Seeing.* Oxford: Oxford University Press.

[13] J. M. Grey. (1975) *Exploration of Musical Timbre.* PhD Thesis, Report No. STAN-M-2, Stanford University, Stanford CA.

[14] S. Bregman. (1990) *Auditory Scene Analysis.* Cambridge, MA: MIT Press.

[15] S. Barrass. (1997) *Auditory Information Design.* PhD Thesis, The Australian National University.

[16] P. Scholes. (1950) *The Oxford Companion to Music.* Eighth Edition. Oxford University Press, London. pp 181 - 189.

[17] D. Payling, S. Mills and T. Howle (2006) "Hue Music" Accessed 28/12/2006. url: http://www.staffs.ac.uk/~dp11/research/huemusic/huemusic.htm.