Identifying Major Components of Pictures by Audio Encoding of Colours

Conference Paper · June 2007

DOI: 10.1007/978-3-540-73055-2_10 · Source: DBLP

Guido Bologna
University of Applied Sciences and Arts Western Switzerland
79 PUBLICATIONS **1,320** CITATIONS

Identifying Major Components of Pictures by Audio Encoding of Colours

Guido Bologna[1], Benoît Deville[2], Thierry Pun[2], and Michel Vinckenbosch[1]

[1] Laboratoire d'Informatique Industrielle, University of Applied Science HES-SO
Rue de la Prairie 4, 1202 Geneva, Switzerland
Guido.Bologna@hesge.ch Michel.Vinckenbosch@hesge.ch

[2] Computer Science Center, University of Geneva
Rue Général Dufour 24, 1211 Geneva, Switzerland
Benoit.Deville@cui.unige.ch Thierry.Pun@cui.unige.ch

Abstract. The goal of the See ColOr project is to achieve a non-invasive mobility aid for blind users that will use the auditory pathway to represent in real-time frontal image scenes. More particularly, we have developed a prototype which transforms HSL coloured pixels into spatialized classical instrument sounds lasting for 300 ms. Hue is sonified by the timbre of a musical instrument, saturation is one of four possible notes, and luminosity is represented by bass when luminosity is rather dark and singing voice when it is relatively bright. Our first experiments are devoted to static images on the computer screen. Six participants with their eyes covered by a dark tissue were trained to associate colours with musical instruments and then asked to determine on several pictures, objects with specific shapes and colours. In order to simplify the protocol of experiments, we used a tactile tablet, which took the place of the camera. Overall, experiment participants found that colour was helpful for the interpretation of image scenes.

1 Introduction

In this work we present *See ColOr* (**See**ing **Col**ours with an **Or**chestra), which is a multi-disciplinary project at the cross-road of Computer Vision, Audio Processing and Pattern Recognition. The long term goal is to achieve a non-invasive mobility aid for blind users that will use the auditory pathway to represent in real-time frontal image scenes. Ideally, our targeted system will allow visually impaired or blind subjects having already seen to build coherent mental images of their environment. Typical coloured objects (signposts, mailboxes, bus stops, cars, buildings, sky, trees, etc.) will be represented by sound sources in a three-dimensional sound space that will reflect the spatial position of the objects. Targeted applications are the search for objects that are of particular use for blind users, the manipulation of objects and the navigation in an unknown environment.

Sound spatialisation is the principle which consists of virtually creating a three-dimensional auditive environment, where sound sources can be positioned all around the listener. These environments can be simulated by means of loudspeakers or headphones. Among the precursors in the field, Ruff and Perret led a series of experiments on the space perception of auditive patterns [13]. Patterns were transmitted through a 10x10 matrix of loudspeakers separated by 10 cm and located at a distance of 30 cm from the listener. Patterns were represented on the auditory display by sinusoidal waves on the corresponding loudspeakers. The experiments showed that 42% of the participants identified 6 simple geometrical patterns correctly (segment of lines, squares, etc). However, orientation was much more difficult to determine precisely. Other experiments carried out later by Lakatos taught that subjects were able to recognise with 60-90% accuracy ten alphanumeric characters [10].

Our See ColOr prototype for visual substitution presents a novelty compared to systems presented in the literature [9], [12], [4], [5] and [8]. More particularly, we propose the encoding of colours by musical instrument sounds, in order to emphasize coloured objects and textures that will contribute to build consistent mental images of the environment. Note also that at the perceptual level, colour is helpful to group the pixels of a mono-coloured object into a coherent entity. Think for instance when one looks on the ground and it “sounds” green, it will be very likely to be grass. The key idea behind See ColOr is to represent a pixel of an image as a sound source located at a particular azimuth and elevation angle. Depth is also an important parameter that we estimate by triangulation using stereo-vision. Finally, each emitted sound is assigned to a musical instrument, depending on the colour of the pixel.

In this work the purpose is to investigate whether individuals are able to learn associations between colours and musical instrument sounds and also to find out whether colour is beneficial to experiment participants. To the best of our knowledge this is the first study in the context of visual substitution for real time navigation in which colour is supplied to the user as musical instrument sounds.

2 Audio Encoding

This section illustrates audio encoding without 3D sound spatialization. Colour systems are defined by three distinct variables. For instance, the RGB cube is an additive colour model defined by mixing red, green and blue channels. We used the eight colours defined on the vertex of the RGB cube (red, green, blue, yellow, cyan, purple, black and white). In practice a pixel in the RGB cube was approximated with the colour corresponding to the nearest vertex. Our eight colours were played on two octaves : Do, Sol, Si, Re, Mi, Fa, La, Do. Note that each colour is both associated with an instrument and a unique note [2]. An important drawback of this model was that similar colours at the human perceptual level could result considerably further on the RGB cube and thus generated perceptually distant instrument sounds. Therefore, after preliminary experiments associating colours and instrument sounds we decided to discard the RGB model.

The second colour system we studied for audio encoding was HSV. The first variable represents hue from red to purple (red, orange, yellow, green, cyan, blue, purple), the second one is saturation, which represents the purity of the related colour and the third variable represents luminosity. HSV is a non-linear deformation of the RGB cube; it is also much more intuitive and it mimics the painter way of thinking. Usually, the artist adjusts the purity of the colour, in order to create different nuances. We decided to render hue with instrument timbre, because it is well accepted in the musical community that the colour of music lives in the timbre of performing instruments. This association has been clearly done for centuries. For instance, think about the brilliant connotation of the Te Deum composed by Charpentier in the seventeenth century (the well known Eurovision jingle, before important sport events). Moreover, as sound frequency is a good perceptual feature, we decided to use it for the saturation variable. Finally, luminosity was represented by double bass when luminosity is rather dark and a singing voice when it is relatively bright.

The HSL colour system also called HLS or HSI is very similar to HSV. In practice, HSV is represented by a cone (the radial variable is H), while HSL is a symmetric double cone. Advantages of HSL are that it is symmetrical to lightness and darkness, which is not the case with HSV. In HSL, the Saturation component always goes from fully saturated colour to the equivalent gray (in HSV, with V at maximum, it goes from saturated color to white, which may be considered counterintuitive). The luminosity in HSL always spans the entire range from black through the chosen hue to white (in HSV, the V component only goes half that way, from black to the chosen hue). The symmetry of HSL represents an advantage with respect to HSV and is clearly more intuitive.

The audio encoding of hue corresponds to a process of quantification. As shown by table 1, the hue variable H is quantified for seven colours.

Hue value (H)	Instrument
red ($0 \leq H < 1/12$)	oboe
orange ($1/12 \leq H < 1/6$)	viola
yellow ($1/6 \leq H < 1/3$)	pizzicato violin
green ($1/3 \leq H < 1/2$)	flute
cyan ($1/2 \leq H < 2/3$)	trumpet
blue ($2/3 \leq H < 5/6$)	piano
purple ($5/6 \leq H < 1$)	saxophone

Table 1. Quantification of the hue variable by sounds of musical instruments.

More particularly, the audio representation h_h of a hue pixel value h is

$$h_h = g \cdot h_a + (1 - g) \cdot h_b \tag{1}$$

with g representing the gain defined by

$$g = \frac{h_b - H}{h_b - h_a} \tag{2}$$

with $h_a \leq H <\leq h_b$, and h_a, h_b representing two successive hue values among red, orange, yellow, green, cyan, blue, and purple (the successor of purple is red). In this way, the transition between two successive hues is smooth. For instance, when h is yellow then $h = h_a$, thus $g = 1$ and $(1 - g) = 0$; as a consequence, the resulting sound mix is only pizzicato violin. When h goes toward the hue value of green, which is the successor of yellow on the hue axis, the gain value g of the term h_a decreases, whereas the gain term of h_b $((1-g))$ increases, thus we progressively hear the flute appearing in the audio mix.

Once h_h has been determined, the second variable S of HSL corresponding to saturation is quantified into four possible notes, according to table 2.

Saturation (S)	Note
$0 \leq S < 0.25$	Do
$0.25 \leq S < 0.5$	Sol
$0.5 \leq S < 0.75$	Sib
$0.75 \leq S \leq 1$	Mi

Table 2. Quantification of saturation by musical instrument notes.

Luminosity denoted as L is the third variable of HSL. When luminosity is rather dark, h_h is additionally mixed with double bass using the four notes depicted in table 3, while table 4 illustrates the quantification of bright luminosity by a singing voice. Note that the audio mixing of the sounds representing hue and luminosity is very similar to that described in equation 1. In this way, when luminosity is close to zero and thus the perceived colour is black, we hear in the final audio mix the double bass without the hue component. Similarly, when luminosity is close to one, the perceived colour is

Luminosity (L)	Double Bass Note
$0 \leq L < 0.125$	Do
$0.125 \leq L < 0.25$	Sol
$0.25 \leq L < 0.375$	Sib
$0.375 \leq L \leq 0.5$	Mi

Table 3. Quantification of luminosity by double bass.

Luminosity (L)	Voice Note
$0.5 \leq L < 0.625$	Do
$0.625 \leq L < 0.75$	Sol
$0.75 \leq L < 0.875$	Sib
$0.875 \leq L \leq 1$	Mi

Table 4. Quantification of luminosity by a singing voice.

white and thus we hear the singing voice. Note that with luminosity at its half level, the final mix contains just the hue component.

3 Experiments

Our prototype is based on a sonified 17x9 sub-window pointed by the mouse on the screen which is sonified via a virtual Ambisonic 3D audio rendering system [3], [2], [6], [7], and [11]. In fact, the sound generated by a pixel is a monaural sound that is encoded into 9 Ambisonic channels; with parameters depending on azimuth and elevation angles. Then, the encoded Ambisonic signals are decoded for loudspeakers placed in a virtual cube layout. Finally, the physical sound is generated for headphones with the use of HRTF functions related to the directions of virtual loudspeakers. The HRTF functions we use, are those included in the CIPIC database [1]. The orchestra used for the sonification is that described in section 2. The maximal time latency for generating a 17x9 sonified subwindow is 80 ms with the use of Matlab on a Pentium 4 at 3.0 GHz.

The purpose of this study was to investigate whether individuals are able to learn associations between colours and musical instrument sounds. Several experiments have been carried out by participants having their eyes enclosed by a dark tissue, and listening to the sounds via headphones [12]. In order to simplify the experiments, we used the T3 tactile tablet from the Royal National College for the Blind[3] (UK). Essentially, this device allows to point on a picture with the finger and to obtain the coordinates of the contact point. Moreover, we put on the T3 tablet a special paper with images including detected edges represented by palpable roughness.

Six participants were trained to associate colours with musical instruments and then asked to determine on several pictures, objects with specific shapes and colours. Experiments involved a learning phase with the use of elementary pictures. At the end of the training phase, a small test for scoring the performance of the participants was achieved. On the 15 heard sounds, the average number of correct colours among the six participants was 8.1 (standard deviation : 3.4). It is worth noting that the best score was reached by a musician who found 13 correct answers. Afterwards, participants were asked to explore and identify the major components of the pictures shown in figure 1.

Regarding the children draw picture illustrated in figure 1, all participants interpreted the major colours as the sky the sea and the sun; clouds were more difficult to infer (two individuals); instead of ducks, all the subjects found an island with yellow sand or a ship.

For the picture depicted in figure 2 all participants interpreted the major colours as the sky and the sea; an individual said that the dolphin is a “jumping animal”, another said that it was a fish and the others determined a boat or a “round shape”; only a person found birds and no one was able to identify the small fish.

On the interpretation of real images, such as the picture shown in figure 3, four participants correctly identified the tree with the grass and the sky; a participant qualified the tree as a strange dark object and finally, the last individual inferred a nuclear explosion ! Concerning figure 4 , all subjects found major colours (blue and yellow); however

[3] http://www.talktab.org/

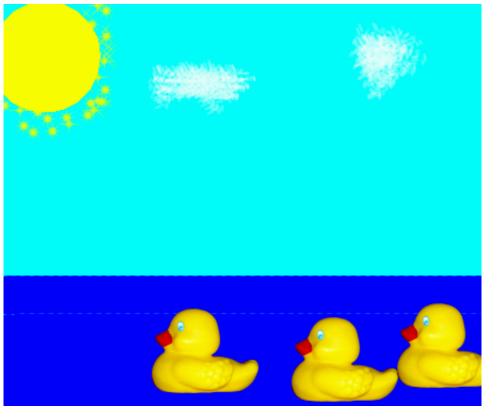

Fig. 1. The “Ducks” picture. Major components : sea (blue), sky (blue cyan), sun (yellow), clouds (white).

no one made the distinction between the sky and the sea. Moreover, no one identified the yellow cliff.

The last assignment was to find a red door in figure 5. All participants found one of the red doors in a time range between 4 and 9 minutes.

4 Discussion

The first experiment concerning the recognition of 15 colours corresponding to 15 sounds exhibited that correct answers were given in a little bit more than half of the times, on average. Therefore, roughly speaking our group gave correct answers for five colours out of nine. That is clearly better than black and white identification. Thus, this experiment demonstrated that learning all colours is possible, but difficult in a short training time. It is worth noting that learning Braille is also complicated and requires a long period of training. Accordingly, the training phase with musical instrument sounds should be repeated a reasonable number of sessions.

The second experiment with children drawings demonstrated that the most important components of the pictures, such as the sky the sea and the sun were identified. Sometimes our participants were not completely sure, but with logical reasoning they inferred that if the top of the pictures is cyan and if the bottom is blue then the bottom is the sea and the top is the sky. Moreover, if something at the top of figure 1 is yellow and round shaped, this must be the sun. Another interesting observation is the difficulty to identify the three ducks. In fact, our common sense tells us that something yellow would be more likely to be the sand of an island or a yellow ship. Yellow ducks on the sea represent an unusual situation which is never considered by our participants.

The third experiment was performed with two real pictures. It is worth noting that figure 3 has three major components (sky, grass and tree), with a limited perspective view. Consequently, almost our participants gave a correct sketch of that picture. On the contrary, figure 4 presents noticeable perspective; as a result, the context of the picture was not determined by our six participants, although several individuals correctly identified the most important colours.

The fourth experiment consisted in finding one of the red doors of figure 5. All the people were successful, however the elapsed time was quite long. The first reason is that with A3 paper format on the T3 tablet, it takes a long time to explore the picture with a small sub-window of size 17x9 pixels. Moreover, the image scene is complicated with a high degree of perspective.

Five participants out of six said that colour was helpful for the interpretation of pictures. In fact, when one tries to identify a picture component, the presence of colour in the audio representation limits the number of possible interpretations. Finally, the experiments emphasised perspective as a major drawback for the understanding of two-dimensional figures.

5 Conclusion and Future Work

We presented the current state of the See ColOr project which aims at providing a mobility aid for visually impaired individuals. Because of real-time constraints, image simplification in our prototype was achieved by colour quantification of the HSL colour system translated into musical instrument sounds. With only a training session, the experiments on static pictures revealed that our participants were able to learn five out of nine principal colours, on average. Furthermore, colour was helpful for the interpretation of image scenes, as it lessened ambiguity. These experiments also demonstrated that the exploration time of pictures is quite long, probably because the sonified sub-window is small and should not expand too much for reasons related to the limits of human audio channel capacity. In the context of real time navigation, we foresee that a blind individual with a camera on her head would rely on image processing techniques, such as salient points determination, in order to quickly determine relevant parts of the environment.

Aknowledgements

We would like to thank the Hasler Foundation in Bern for financing the See ColOr research project under grant 1968. We also thank two internet sites[4] for making available the pictures used in the experiments.

References

1. Algazi, V.R., Duda, R.O., Thompson, D.P., Avendano (2001). The CIPIC HRTF Database. In IEEE Proc. Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House, (WASPAA'01), New Paltz, NY.

[4] http://www.freeimages.co.uk/galleries/buildings/country/index.htm
http://www.tuxpaint.org/gallery/

2. Bologna, G., Vinckenbosch, M. (2005). Eye Tracking in Coloured Image Scenes Represented by Ambisonic Fields of Musical Instrument Sounds. In Proc. IWINAC (1) 327–337.
3. Bamford, J.S. (1995). An Analysis of Ambisonic Sound Systems of First and Second Order. Master Thesis, Waterloo, Ontario, Canada.
4. Capelle, C., Trullemans, C., Arno, P., Veraart, C. (1998). A Real Time Experimental Prototype for Enhancement of Vision Rehabilitation Using Auditory Substitution. IEEE T. Bio-Med Eng., 45, 1279–1293.
5. Cronly-Dillon, J., Persaud, K., Gregory, R.P.F. (1999). The Perception of Visual Images Encoded in Musical Form: a Study in Cross-Modality Information. In Proc. Biological Sciences, 266, 2427–2433.
6. Daniel, J. (2000). Acoustic Field Representation, Application to the Transmission and the Reproduction of Complex Sound Environments in a Multimedia Context. PhD thesis, University of Paris 6.
7. Gerzon, M.A. (1977). Design of Ambisonic Decoders for Multispeaker Surround Sound. Journal of the Audio Engineering Society (Abstracts), 25, 1064.
8. Gonzalez-Mora, J.L., Rodriguez-Hernandez, A., Rodriguez-Ramos, L.F., Dfaz-Saco, L., Sosa, N. (1999). Development of a New Space Perception System for Blind People, Based on the Creation of a Virtual Acoustic Space. In Proc. IWANN, 321–330.
9. Kay, L. (1974). A Sonar Aid to Enhance Spatial Perception of the Blind: Engineering Design and Evaluation. The Radio and Electronic Engineer, 44, 605–627.
10. Lakatos, S. (1993) Recognition of Complex Auditory-Spatial Patterns. Perception, 22, 363–374.
11. Malham, D.G., Myatt A. (1995). 3-D Sound Spatialisation using Ambisonic Techniques. Computer Music Journal, 19 (4), 58–70.
12. Meijer, P.B.L. (1992). An Experimental System for Auditory Image Representations. IEEE Transactions on Biomedical Engineering, 39 (2), 112–121.
13. Ruff, R.M., Perret, E. (1976). Auditory Spatial Pattern Perception Aided by Visual Choices. Psychological Research, 38, 369–377.

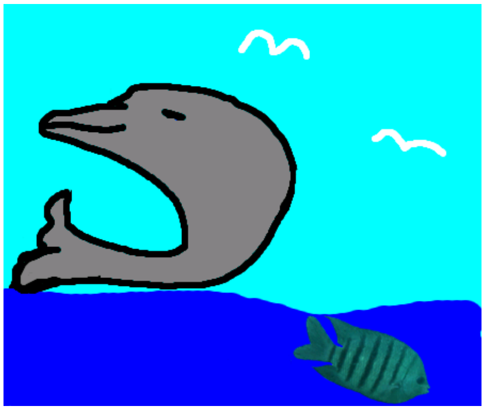

Fig. 2. The “Dolphin” picture. Major components : sea (blue), sky (blue cyan), dolphin (gray) and fish in the water (gray and cyan).

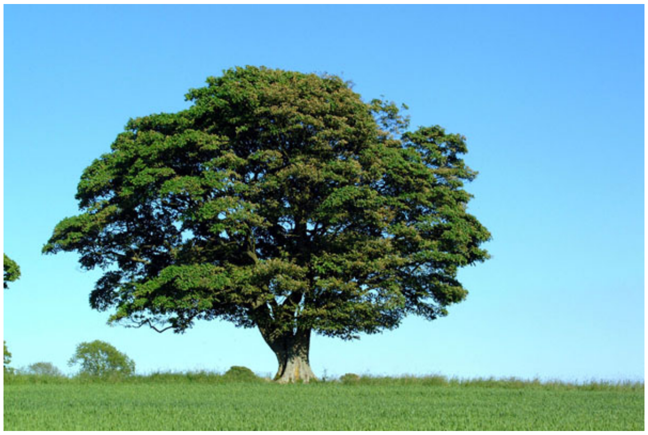

Fig. 3. The “Tree” picture. Major components : grass (green), sky (blue cyan) and tree (dark green).

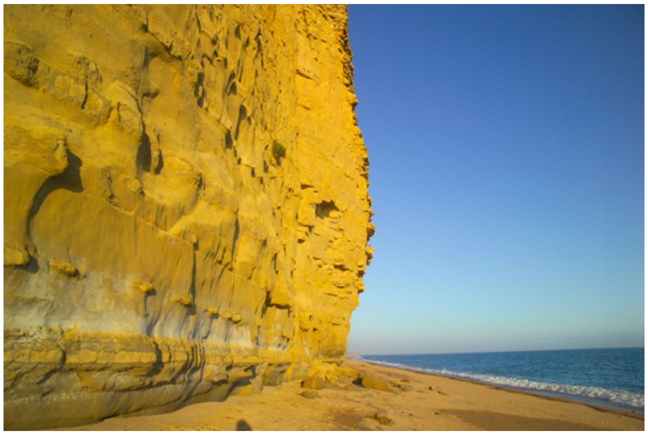

Fig. 4. The “Beach” picture. Major components : cliff (yellow), sky (blue and blue cyan) sea (blue) sand (bright-brown).

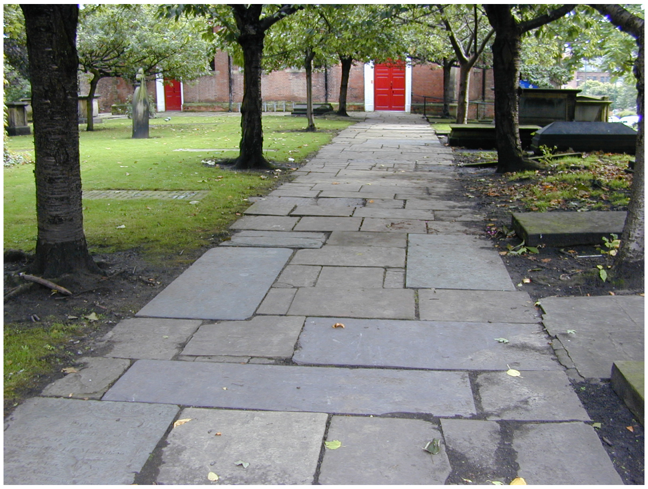

Fig. 5. The “churchyard” picture. The goal in the experiments is to find one of the red doors.