

SONIFICATION OF COLOR AND DEPTH IN A MOBILITY AID FOR BLIND PEOPLE

Guido Bologna

Benoît Deville

Thierry Pun

University of Geneva,
Computer Science Department,
Route de Drize 7, 1227 Carouge,
Switzerland
Guido.Bologna@unige.ch

University of Geneva,
Computer Science Department,
Route de Drize 7, 1227
Carouge, Switzerland
Benoit.Deville@unige.ch

University of Geneva,
Computer Science Department,
Route de Drize 7, 1227
Carouge, Switzerland
Thierry.Pun@unige.ch

ABSTRACT

The *See Color* interface transforms a small portion of a colored video image into sound sources represented by spatialized musical instruments. Basically, the conversion of colors into sounds is achieved by quantization of the HSL color system. Our purpose is to provide visually impaired individuals with a capability of perception of the environment in real time. In this work the novelty is the simultaneous sonification of color and depth, depth being coded by sound rhythm. Our sonification model is illustrated by several experiments, such as: (1) detecting an open door in order to go out from the office; (2) walking in a hallway and looking for a blue cabinet; (3) walking in a hallway and looking for a red tee shirt; (4) moving outside and avoiding a parked car. Videos with sounds of experiments are available on <http://www.youtube.com/guidobologna>.

1. INTRODUCTION

This paper presents on-going work of the *See CoLoR* project, which aims at improving the perception and the mobility of blind individuals. In previous works we introduced the sonification of colors by means of instrument sounds, as well as experiments related to image comprehension, recognition of colored objects and mobility [1].

In this work the novelty is the use of depth in the sound code. Generally, distance to objects is a crucial parameter for an individual evolving in a given environment. In our experiments, depth is captured by a stereoscopic camera. We perform experiments for which a well trained blindfolded individual takes advantage of depth sonification in a number of situations like: (1) detecting an open door in order to go in and out; (2) walking in a hallway and looking for a blue cabinet; (3) walking in a hallway and looking for a red tee shirt; (4) moving outside and avoiding a parked car. The resultant videos are available on <http://www.youtube.com/guidobologna>. In the following sections, section 2 describes mobility aids without the sonification of color, section 3 presents several works that proposed color sonification, section 4 summarizes our previous experiments, section 5 explains our model of color and depth sonification, section 6 illustrates several experiments, followed by the conclusion.

2. MOBILITY AIDS WITHOUT COLOR SONIFICATION

Several authors proposed special devices for visual substitution by the auditory pathway in the context of real time navigation. The “K Sonar-Cane” combines a cane and a torch with ultrasounds [2]. Note that with this special cane, it is possible to perceive the environment by listening to a sound coding depth.

“TheVoice” is another experimental vision substitution system that uses auditory feedback. An image is represented by 64 columns of 64 pixels [3]. Every image is processed from left to right and each column is listened to for about 15 ms. In particular, every pixel gray level in a column is represented by a sinusoidal wave sound with a distinct frequency. High frequencies are at the top of the column and low frequencies are at the bottom.

Capelle et al. proposed the implementation of a crude model of the primary visual system [4]. The implemented device provides two resolution levels corresponding to an artificial central retina and an artificial peripheral retina, as in the real visual system. The auditory representation of an image is similar to that used in “TheVoice” with distinct sinusoidal waves for each pixel in a column and each column being presented sequentially to the listener.

Gonzalez-Mora et al. developed a prototype using the spatialisation of sound in the three dimensional space [5]. The sound is perceived as coming from somewhere in front of the user by means of head related transfer functions (HRTFs). The first device they achieved was capable of producing a virtual acoustic space of 17*9*8 gray level pixels covering a distance of up to 4.5 meters.

3. COLOR SONIFICATION

3.1. State of the art

Recently, the research domain of color sonification has started to grow [6], [7], [8]. A number of authors defined sound/color associations with respect to the HSL color system. HSL (Hue, Saturation, Luminosity) is a symmetric double cone symmetrical to lightness and darkness. HSL mimics the painter way of thinking with the use of a painter tablet for adjusting the

purity of colors. The H 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 color and the third variable represents luminosity. The H , S , and L variables are defined between 0 and 1.

Doel defined color/sound associations based on the HSL color system [6]. In this sonification model, sound depends on the color of the image at a particular location, as well as the speed of the pointer motion. Sound generation is achieved by subtractive synthesis. Specifically, the sound for grayscale colors is produced by filtering a white noise source with a low pass filter with a cutoff frequency that depends on the brightness. Color is added by a second filter, which is parameterized by hue and saturation.

Rossi et al. presented the “Col.diesis” project [7]. Here the basic idea is to associate colors to a melody played by an instrument. For a given color, darker colors are produced by lower pitch frequencies. Interestingly, based on the statistics of more than 700 people, they produced a table, which summarizes how individuals associate colors to musical instruments. It turned out that the mapping is: yellow for vibraphone or flute; green for flute; orange for banjo or marimba; purple for cello or organ; blue for piano, trumpet or clarinet; red for guitar or electric guitar.

Capalbo and Glenney introduced the “KromoPhone” [8]. Their prototype can be used either in RGB mode or HSL mode. Using HSL, hue is sonified by sinusoidal sound pitch, saturation is associated to sound panning and luminosity is related to sound volume. The authors stated that only those individuals with perfect pitch perform well. In RGB mode the mapping of colors to sounds are defined by pan, pitch and volume. For instance, the gray scale from black to white is panned to the centre, with black being associated to the lowest pitch sound. Blue and yellow are mapped to the left, with blue being associated to lower pitch than yellow. Similarly, green and red are related to sounds listened to the right. Finally, the intensity of each color is mapped to the volume of the sound it produces.

In one of their experiments, Capalbo and Glenney illustrated that the use of color information in a recognition task outperformed the performance of “TheVoice” (cf. section 2) [8]. Specifically, the purpose was to pick certain fruits and vegetables known to correlate with certain colors. One of the results was that none of the three subjects trained with “TheVoice” could identify any of the fruit, either by the shape contours or luminance.

3.2. Color sonification in See CoOr

Relative to the HSL color system, we represent the Hue variable by instrument timbre, because it is well accepted in the musical community that the color of music lives in the timbre of performing instruments. Moreover, learning to associate instrument timbres to colors is easier than learning to associate for instance, pitch frequencies. The saturation variable S representing the degree of purity of hue is rendered by sound pitch, while luminosity is represented by double bass when it is rather dark and a singing voice when it is relatively bright.

With respect to the hue variable, the corresponding musical instruments are based on an empirical choice:

1. oboe for red ($0 \leq H < 1/12$);
2. viola for orange ($1/12 \leq H < 1/6$);
3. pizzicato violin for yellow ($1/6 \leq H < 1/3$);
4. flute for green ($1/3 \leq H < 1/2$);
5. trumpet for cyan ($1/2 \leq H < 2/3$);
6. piano for blue ($2/3 \leq H < 5/6$);
7. saxophone for purple ($5/6 \leq H \leq 1$).

Note that for a sonified pixel, when the hue variable is exactly between two predefined hues, such as for instance between yellow and green, the resulting sound instrument mix is an equal proportion of the two corresponding instruments. More generally, hue values are rendered by two sound timbres whose gain depends on the proximity of the two closest hues.

The audio representation h_h of a hue pixel value h is

$$h_h = g \cdot h_a + (1 - g) \cdot h_b \quad (1)$$

with g representing the gain defined by

$$g = \frac{h_b - H}{h_b - h_a} \quad (2)$$

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.

The pitch of a selected instrument depends on the saturation value. We use four different saturation values by means of four different notes:

1. C for ($0 \leq S < 0.25$);
2. G for ($0.25 \leq S < 0.5$);
3. B flat for ($0.5 \leq S < 0.75$);
4. E for ($0.75 \leq S \leq 1$);

When the luminance L is rather dark (i.e. less than 0.5) we mix the sound resulting from the H and S variables with a double bass using four possible notes (C, G, B flat, and E), depending on luminance level. A singing voice with also four different pitches (the same used for the double bass) is used with bright luminance (i.e. luminance above 0.5). Moreover, if luminance is close to zero, the perceived color is black and we discard in the final audio mix the musical instruments corresponding to the H and S variables. Similarly, if luminance is close to one, thus the perceived color is white we only retain in the final mix a singing voice. Note that with luminance close to 0.5 the final mix has just the hue and saturation components.

The sonified part of a captured image is a row of 25 pixels in the central part of the picture. We take into account a single row, as the encoding of several rows would need the use of 3D spatialization, instead of simple 2D spatialization. It is well known that rendering elevation is much more complicated than lateralization [9]. On the other hand, in case of 3D

spatialization it is very likely that too many sound sources would be difficult to be analyzed by a common user.

Two-dimensional spatialization is achieved by the convolution of mono aural instrument sounds with filters encompassing typical lateral cues, such as interaural time delay and interaural intensity difference. In this work we reproduce spatial lateralization with the use of the CIPIC database [10].

4. OUR PREVIOUS EXPERIMENTS

In the first step of the See CoLoR project, we performed several experiments with six blindfolded persons who were trained to associate colors with musical instrument sounds [1]. As shown by figure 1, the participants were asked to identify major components of static pictures presented on a special paper lying on a T3 tactile tablet (<http://www.mcb.ac.uk/t3/index.htm>) representing pictures with embossed edges. When one touched the paper lying on the tablet, a small region below the finger was sonified and provided to the user. Color was helpful for the interpretation of image scenes, as it lessened ambiguity. As an example, if a large region “sounded” cyan at the top of the picture it was likely to be the sky. Finally, all participants to the experiments were successful when asked to find a bright red door in a picture representing a churchyard with trees, grass and a house.



Figure 1: Example of embossed picture on the T3 tactile tablet.

The work described in [11] introduced an experiment during which ten blindfolded individuals participants tried to match pairs of uniform colored socks by pointing a head mounted camera and by listening to the generated sounds. Figure 2 illustrates an experiment participant observing a blue socket. The results of this experiment demonstrated that matching similar colors through the use of a perceptual (auditory) language, such as that represented by instrument sounds can be successfully accomplished.



Figure 2: A blindfolded subject observing a blue socket.

In [12] the purpose was to validate the hypothesis that navigation in an outdoor environment can be performed by “listening” to a colored path. We introduced an experiment during which ten blindfolded participants and a blind person were asked to point the camera toward a red sinuous path painted on the ground and to follow it for more than 80 meters. Results demonstrated that following a sinuous colored path through the use of our auditory perceptual language was successful. A video entitled “The See CoLoR project” illustrates several experiments on <http://www.youtube.com/guidobologna>.



Figure 3: A blindfolded individual following a red sinuous path.

5. SONIFICATION OF COLOR AND DEPTH

We use a stereoscopic color camera denoted STH-MDCS2 (SRI International: <http://www.videredesign.com/>) and the “Bumblebee” (Point Grey: <http://www.ptgrey.com/>). An algorithm for depth calculation based on epipolar geometry is embedded within both the stereoscopic cameras. The resolution of images is 320x240 pixels with a maximum frame rate of 30 images per second.

Our See ColOr prototype presents two sonification modes that render color and depth. The first replicates a crude model of the human visual system. Pixels near the center of the sonified row have high resolution, while pixels close to the left and right borders have low resolution. This is achieved by considering a sonification mask indicating the number of pixel values to skip. As shown below, starting from the middle point (in bold), the following vector of 25 points represents the number of skipped pixels:

[15 12 9 7 5 3 3 2 2 1 1 1 **1** 1 1 1 2 2 3 3 5 7 9 12 15]

In the first mode, depth is represented by sound duration. The mapping for depth D is given by :

- 90 ms for undetermined depth;
- 160 ms for $(0 \leq D < 1)$;
- 207 ms for $(1 \leq D < 2)$;
- 254 ms for $(2 \leq D < 3)$;
- 300 ms for $D > 3$

The second mode sonifies only a pixel of a particular area of 25 adjacent points in the middle of the image. Specifically, we first determine among these 25 points the greatest number of contiguous points labelled with the same hue. Then, we calculate the centroid of this area and the average depth. It is possible to have points of undetermined depth, especially in homogeneous areas like walls, for which the depth algorithm is unable to determine landmark points related to the calculation of the disparity between the left and right images. Points with undetermined depth are not considered in the average depth calculation. The final sonification presents only a spatialized sound source representing the average color and the average depth.

In the second mode, depth between one and four meters is sonified by sound duration (the same sonification scheme explained above), while after four meters the volume V starts to decrease by following a negative exponential function given by

$$f(V) = V * \exp(-k * D) \quad (3)$$

with k a positive small constant.

6. PRELIMINARY EXPERIMENTS

The experiments were performed by a very well trained blindfolded individual, who is very familiar with this color sonification model, but not with depth sonification. Although in the long term we will aim at complementing the white cane of blind people by a miniaturized version of our prototype, this person relied only on the See ColOr interface. The reason is that we wanted to be sure that our prototype represented the only sensing tool.

In the first video entitled “Going out from the office” and in the second video entitled “Going into the office” we aim at demonstrating that it is possible to perceive an open door and to pass through it. Figure 4 illustrates a picture of this experiment, which is performed with the second sonification mode. The

brown door is sonified by a viola and the rhythm is fast when the user is close to it. Note also that the user decided to move when slow sound rhythms or low volume sounds were discerned, indicating distant obstacles.



Figure 4: A blindfolded individual looking for an open door.

The third video entitled “Find a red tee shirt with sounds of musical instruments” illustrates the same individual in a successful search task. It is worth noting that here depth is often undetermined when the camera is pointed toward the floor or the white walls. Note also that the user trusted the depth information related to the trumpet sound representing the blue-cyan cabinets. The red tee shirt is sonified by oboe and when the user was close to it the rhythm frequency increased.



Figure 5: A blindfolded individual looking for a red tee shirt.

In the fourth video entitled “The blue cabinet” the user switched to the first sonification mode (with all 25 points sonified by color and depth). Here the goal was to find a blue cabinet sonified by a piano playing a medium pitched tone. This mode is more complex than the previous, since more than one color can be present in the current sonified frame. Here the distance to the floor is defined, as the floor is textured. Note also that the brown doors are sonified by viola sounds. From time to time, our experiment participant wished to ask to the computer the depth of the middle point of the sonified row. With the use of a mouse button the computer answered with a voice saying numbers in French. “One” means distance between zero and one meter; “two” means distance between one and two meters, etc. At the end of the video the user reached and indicated the cabinet.

In the last video entitled “Walking outside” the user walked outside. He switched again to the second mode with only a sonified sound. The sound of the ground is rendered by a singing voice or a double bass, depending on its gray level. Suddenly, the user found in his trajectory a parked car and he avoided it.



Figure 6: A blindfolded individual walking outside and avoiding parked cars.

After the experiments the blindfolded person was asked to give his impressions about the two different modes. The first impression is that the second mode (with the decreasing volume) is felt as “relaxing” compared to the first mode. The second mode is valuable in large areas (for instance, outside). Moreover, in some situations, it will be very useful to switch from the second mode to the first, as the first mode gives more precision and to some extent, peripheral view. A sonified compass could be also very useful, as it is very easy to lose orientation. Finally, while the first mode provides to some extent limited global information, a “global module” would be helpful in order to get a clear picture of the close environment geometry.

7. CONCLUSION

We presented the color and depth sonification model of the See ColOr mobility aid. A See ColOr prototype was tested by a well trained individual. He successfully (1) detected an open door in order to go in and out; (2) walked in a corridor with the purpose to find a blue cabinet; (3) moved in a hallway with the purpose to locate a red tee shirt; (4) walked outside and avoided a parked car. In the future, we would like to measure in a more systematic way whether the use of our prototype allows users to locate objects and to avoid obstacles of different sizes. Thus, we will perform experiments with more participants, in order to obtain more robust statistics.

8. REFERENCES

- [1] G. Bologna, B. Deville, T. Pun, M. Vinckenbosch, “Transforming 3D coloured pixels into musical instrument notes for vision substitution applications”, *J. of Image and Video Processing*, A. Caplier, T. Pun, D. Tzovaras, Guest Eds., Article ID 76204, 14 pages (Open access article), 2007.
- [2] L. Kay, “A sonar aid to enhance spatial perception of the blind: engineering design and evaluation”, *The Radio and Electronic Engineer*, vol. 44, pp. 605–627, 1974.
- [3] P.B.L. Meijer, “An experimental system for auditory image representations”, *IEEE Trans. Bio. Eng.*, vol. 39, no. 2, pp. 112–121, 1992.
- [4] C. Capelle, C. Trullemans, P. Arno, and C. Veraart, “A real time experimental prototype for enhancement of vision rehabilitation using auditory substitution”, *IEEE T. Bio-Med Eng.*, vol. 45, pp. 1279–1293, 1998.
- [5] J.L. Gonzalez-Mora, A. Rodriguez-Hernandez, L.F. Rodriguez-Ramos, L. Dfaz-Saco, and N. Sosa, “Development of a new space perception system for blind people, based on the creation of a virtual acoustic space”, in *Proc. IWANN’99*, pp. 321–330, 1999.
- [6] K. Doel, “Soundview: Sensing Color Images by Kinesthetic Audio”, *Proc. of the International Conference on Auditory Display*, Boston, MA, USA, July 6-9, 2003.
- [7] J. Rossi, F.J. Perales, J. Varona, M. Roca, «Col.diesis: transforming colour into melody and implementing the result in a colour sensor device», *Second International Conference in Visualisation*, Barcelona, Spain, July 15-17 2009.
- [8] Z. Capalbo, B. Glenney, “Hearing Color: Radical Plurastic Realism and SSDs”, *Fifth Asia-Pacific Computing and Philosophy conference (AP-CAP 2009)*, October 1-2 2009, Tokyo, Japan.
- [9] R. Begault, *3-D Sound for Virtual Reality and Multimedia*, Boston A.P. Professional, ISBN: 0120847353, 1994.
- [10] V.R. Algazi, R.O. Duda, D.P Thompson, C. Avendano, “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, 2001.
- [11] G. Bologna, B. Deville, M. Vinckenbosch, T. Pun, “A Perceptual Interface for Vision Substitution in a Color Matching Experiment”. In *Proc. Int. Joint Conf. Neural Networks, Part of IEEE World Congress on Computational Intelligence*, June 1-6, 2008, Hong Kong.
- [12] G. Bologna, B. Deville, T. Pun, “Blind navigation along a sinuous path by means of the See ColOr interface”. In *Proceedings of the 3rd international Work-Conference on the interplay between Natural and Artificial Computation: Part II: Bioinspired Applications in Artificial and Natural Computation* (Santiago de Compostela, Spain, June 22 - 26, 2009). J. Mira, J. M. Ferrández, J. R. Álvarez, F. Paz, and F. J. Toledo, Eds. Lecture Notes In Computer Science, vol. 5602. Springer-Verlag, Berlin, Heidelberg.