PAIRING COLORED SOCKS AND FOLLOWING A RED SERPENTINE WITH SOUNDS OF MUSICAL INSTRUMENTS

Guido Bologna[1], Benoît Deville[2], Thierry Pun[2]

[1]UNIVERSITY OF APPLIED SCIENCE
Rue de la prairie 4, 1202 Geneva, Switzerland
guido.bologna@hesge.ch

[2]COMPUTER SCIENCE DEPARTMENT, UNIVERSITY OF GENEVA
Route de Drize 7, 1227 Carouge, Switzerland
{benoit.deville, thierry.pun}@cui.unige.ch

ABSTRACT

The See ColOr interface transforms a small portion of a colored video image into sound sources represented by spatialized musical instruments. This interface aims at providing visually impaired people with a capability of perception of the environment. As a first step of this on-going project, the purpose is to verify the hypothesis that it is possible to use sounds from musical instruments to replace color. Compared to state of the art devices, a quality of the See ColOr interface is that it allows the user to receive a feed-back auditory signal from the environment and its colors, promptly. Two experiments based on a head mounted camera have been performed. The first experiment pertaining to object manipulation is based on the pairing of colored socks, while the second experiment is related to outdoor navigation with the goal of following a colored serpentine. The "socks" experiment demonstrated that seven blindfolded individuals were able to accurately match pairs of colored socks. The same participants successfully followed a red serpentine for more than 80 meters.

1. INTRODUCTION

See ColOr (Seeing Colors with an Orchestra) is an ongoing project aiming at providing visually impaired individuals with a non-invasive mobility aid that use the auditory pathway to represent in real-time frontal image scenes. In the See ColOr project, general targeted applications are the search for items of particular interest for blind users, the manipulation of objects and the navigation in an unknown environment.

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 [1]. Note that with this special cane, it is possible to perceive the environment by listening to a sound coding the distance.

"TheVoice" is another experimental vision substitution system that uses auditory feedback. An image is represented by 64 columns of 64 pixels [2]. Every image is processed from left to right and each column is listened to for about 15 ms. Specifically, every pixel gray level in a column is represented by a sinusoidal wave 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 [3]. 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 [4]. 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.

Our See ColOr interface encodes colored pixels by musical instrument sounds, in order to emphasize colored entities of the environment [5][6]. The basic idea is to represent a pixel as a directional sound source with depth estimated by stereo-vision. Finally, each emitted sound is assigned to a musical instrument, depending on the color of the pixel.

In previous work of the See ColOr project [5][6], we performed several experiments with six blindfolded persons who were trained to associate colors with musical instruments. The participants were asked to identify major components of static pictures presented on a special paper lying on a tactile tablet 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. Overall, the results showed that learning all color-instrument associations in only one training session of 30 minutes is almost impossible for non musicians. However, color was helpful for the interpretation of image scenes, as it lessened ambiguity. As a consequence, several individuals participating in the experiments were able to identify several major components of images. As an example, if a large region "sounded" cyan at the top of the picture it was likely to be the sky. Finally, all experiment participants were successful when asked to find a pure red door in a picture representing a churchyard with trees, grass and a house.

In this work the first purpose is to verify the hypothesis that it is possible to manipulate and to match colored objects with an auditory feed-back represented by sounds of musical instruments. The second purpose is to validate that navigation in an outdoor environment can be performed with the help of the sound related to a colored line. We introduce two experiments (a related video is available on http://129.194.70.56/see_color_demos); for the first, the goal of seven blindfolded individuals is to pair colored socks by

pointing a head mounted camera and by listening to the generated sounds. In the second experiment the same participants are asked to point the camera toward a red serpentine and to follow it for more than 80 meters. Results demonstrate that matching colors or following a path with the use of a perceptual language, such as that represented by instrument sounds can be successfully accomplished. In the following sections we present the auditory color encoding, the See ColOr interface, the aural color conversion, the experiments, followed by the conclusion.

2. AURAL COLOR CONVERSION

The HSL (Hue, Saturation, Luminosity) color system 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. 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:

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

Note that for a given pixel of the sonified row, 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 $h_a \le H \le h_b$, 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. Do for ($0 \le S < 0.25$);
2. Sol for ($0.25 \le S < 0.5$);
3. Si flat for ($0.5 \le S < 0.75$);
4. Mi for ($0.75 \le S \le 1$);

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 (Do, Sol, Si flat, and Mi) 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.

3. SEE COLOR INTERFACE

We use a stereoscopic color camera denoted STH-MDCS2 (SRI International: http://www.videredesign.com/) or a Logitech Webcam Notebook Pro. An algorithm for depth calculation based on epipolar geometry is embedded within the stereoscopic camera, however in this work depth is not taken into account. The resolution of images is 320x240 pixels with a maximum frame rate of 30 images per second.

The See ColOr interface features two different modes, denoted “photographic” and “perceptual”. The “photographic” interactive mode consists in giving a rough sketch of the image scene, which is summarized to the user with the list of the largest homogeneous regions. Specifically, the size of the picture is decreased by a factor of ten, subsequently pixel color values are averaged by 3x3 blocks and then labels are associated to pixels with respect to the seven main colors (cf. previous section) with the addition of black and white. An arbitrary number of the largest colored areas are specified to the user as a sequence of sounds representing musical instruments. Specifically, for each colored region of a picture the See ColOr interface provides a user with a spatial sound corresponding to the average color of the region and a number between one and ten representing the second coordinate of the area centroid (the first coordinate is included in the 2D spatialization of the instrument sound).

Contrarily to the previous approach, the perceptual mode reacts in real-time. In practice, images are not processed and a row of 25 pixels in the middle part of the picture is sonified. We take into account a single row, as the encoding of several rows would need the use of 3D spatialization instead of simple 2D spatializazion. It is well known that rendering elevation is much more complicated than lateralization [7]. 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. Note that the 25 sounds corresponding to the 25 sonified pixels are played simultaneously. Moreover, lateralization 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 [8]. Measurements of the KEMAR

manikin [8] are those used by our See ColOr interface. All possible spatialized sounds (25*9*4 = 900) are pre-calculated and reside in memory. In practice, our main program for sonification is a mixer selecting appropriate spatialized sounds, with respect to the center of the video image.

In order to replicate 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.

4. EXPERIMENTS

In the experiments we use the perceptual mode of the See ColOr interface. The first experiment has been performed in a room, while the second has taken place in an outdoor environment with the use of a webcam able to quickly adapt to light changing conditions.

4.1. Pairing Colored Socks with See ColOr

The purpose is to verify the hypothesis that it is possible to manipulate and to match colored objects with an auditory feed-back represented by sounds of musical instruments. As it is difficult to learn the associations between colors and sounds in just one training session [5] [6], our participants are not asked to identify colors, but just to pair similarly colored socks.

In order to eliminate potential influence of other characteristics, such as haptic, smell and sound characteristics of sock pairs, the authors have organized the experiment that confirmed that blindfolded color pairing without help is random.

4.1.1. Training Phase

The experiments are performed by seven blindfolded adults. The training phase includes two main steps. First, we explain associations between colors and sounds in front of a laptop screen showing different static pictures. Specifically, we show the HSL system with seven main hues and several saturation varying pictures. We let our participants decide when they feel comfortable to switch to the second step aiming at learning to point the camera toward socks. With respect to each individual, Table 1 illustrates the time dedicated for the two training steps.

Participant	Static Training (mn)	Training with Socks (mn)
P1	12	12
P2	7	11
P3	18	15
P4	0	6
P5	0	24
P6	29	18
P7	16	16
Average	**11.7 ± 10.4**	**14.6 ± 5.7**

Table 1. Training time durations without socks and with a head mounted camera pointing at socks.

Note that since the camera is above the eyes, it is difficult for our experiment participants to point correctly the camera. Moreover, for particular angles of view the artificial light in the room can be reflected by the socks. On the other hand, if a sock is too close to the camera the captured color is dark. After the training phase, the test starts with socks that have not been observed previously.

4.1.2. Testing Phase

As shown by Figure 1 we use five pairs of socks having the following colors: black, green, low saturated yellow, blue and orange. Figure 2 illustrates an individual observing a blue sock, while the results obtained by our experimenters are summarized in Table 2. It is worth noting that the average number of paired sock is high. Participant P_4 made a mistake between yellow and orange socks.

Figure 1. The colored socks; from left to right : black, green, yellow, blue and orange.

Figure 2. An experiment participant scrutinizing a blue sock with the use of a head mounted camera.

Participant	Time (mn)	Success Rate (pairs)
P1	16	5
P2	4	5
P3	18	5
P4	6	3
P5	15	5
P6	11	5
P7	7	5
Average	**11.0 ± 5.5**	**4.7 ± 0.8**

Table 2. Testing time duration and success rate of the socks' experiment.

A normal sighted person can match five pairs of colored socks picked up from a plastic bag in 25 seconds, on average. A question arising is the influence of training on the time required to pair socks. In fact, one of the authors who is very well trained can perform this task in 2.2 minutes, which is almost twice faster than the best participant (4 mn).

4.2. Following a Colored Serpentine

Here the purpose is to verify the hypothesis that it is possible to use the See ColOr interface to follow a colored line or serpentine in an outdoor environment. Figure 3 illustrates an individual performing this task. For this experiment we retain the same seven individuals who carried out the experiment with colored socks. The camera here is the Logitech Quickcam Notebook Pro.

Figure 3. A blindfolded individual following a colored line with a head mounted webcam and a notebook carried in a shoulder pack.

4.2.1. Training Phase

The training phase lasts approximately ten minutes. A supervisor manages an experiment participant in front of the colored serpentine. The experimenter is asked to listen to the typical sonification pattern, which is red in the middle area (oboe) and gray in the left and right sides (double bass). The image/sound frequency is fixed to 4 Hz. For experienced users it would be possible to increase the frequency at the maximal implemented value of 11.1 Hz. Afterwards, we ask to the person performing the experiment to move the head from left to right and to become aware that the oboe sound shifts, as well as the moving head. Note that the supervisor wears a headphone and can listen to the sounds of the interface. Finally, the experimenter is asked to start to walk and to keep the oboe sound in the middle sonified region. Note that the training session is quite short. An individual has to learn to coordinate three components. The first is the oboe sound position (if any), the second is related to the awareness of the head orientation and the third is the alignment between the body and the head. Ideally, the head and the body should be aligned with the oboe sound in the middle.

4.2.2. Testing Phase

The purpose of the test is to go from a starting point *S* to a destination point *T*. The testing path is different from the training path. Several small portions of the main path *M* can be walked through three possible alternatives denoted as *A*, *B*, and *C*. The shortest path *M* has length of more than 80 meters. It is important to note that it is impossible to go from *S* to *T* by just moving straight ahead. In Table 3 are reported for each experiment participant the training time duration and the testing time duration, while Table 4 illustrates the followed length path and the average speed. All our experiment participants reached point *T* from point *S* and no-one was lost and asked to be helped.

One of the authors who knows very well See ColOr, went from *S* to *T* through the path *M*+*C* in 4.2 minutes, corresponding to a speed average of 1257 m/h. Therefore, "novice users" could potentially improve their average speed after several training sessions.

Participant	Training Time (mn)	Testing Time (mn)
P1	11	7.3
P2	10	7.1
P3	8	13.6
P4	9	8.5
P5	10	10.4
P6	10	9.7
P7	10	12.9
Average	**9.7 ± 0.9**	**9.9 ± 2.6**

Table 3. Training and testing time duration of blindfolded individuals following a red serpentine.

Participant	Path Length (m)	Speed Average (m/h)
P1	M+C = 88	723
P2	M = 84	710
P3	M+B = 110	485
P4	M+A = 93	656
P5	M = 84	484
P6	M+A+C = 97	600
P7	M+A+C = 97	451
Average	**93.3 ± 9.2**	**587.0 ± 114.1**

Table 4. Path length and speed average of blindfolded individuals following a red serpentine.

4.3. Discussion

The reactivity of the See ColOr interface is important for tasks requiring real time constraints. The perceptual mode of the See ColOr interface provides the user with 25 points, simultaneously. Furthermore, using the perceptual language of musical instruments, the user receives sounds resulting from colors of the environment in 300 ms at most, which is clearly faster than a second, the typical time duration to convey a color name. Although our color encoding is quite natural, a drawback is that associations between colors and musical instruments should be learnt over several training sessions. Note however that learning Braille takes years.

With orally transmitted information every second, it would be unfeasible to follow a serpentine in real time, as the user feed-back would be too slow. Note that with the See ColOr spatialized row of 25 pixels at the center of the video image, the user can perceive the red serpentine in the left/center/right portions of the aural image. Therefore, when the red sound representation is lost the user can look for a red area by moving the head from left to right. As soon as the oboe sound appears, it is possible to shift the head and to put the red sonified area in the center of the aural representation. Then, the body can be aligned to the head and the user can start to walk. That would be really tricky with only one sonified point. In fact, with the use of 25 spatialized sounds, local context in the left and right sides is provided in addition to the central area, which makes the perception much more similar to our visual system.

This is our first experiment in an outdoor environment. The results are very encouraging because our experiment participants really perceived the red serpentine. Some of them said during the training phase: "Yes, I can see it". Moreover, this task was perceived as easier than the sock experiment. With further training sessions consisting in learning to anticipate turns by trying to distinguish more distant red patterns, it is very likely that the navigation average speed would be increased. As well as blindfolded experimenters, we are confident that blind individuals will successfully follow the red serpentine, because of their improved auditory sense.

5. CONCLUSION

We presented the current state of the See ColOr project, which provides the user with an auditory feedback of the colors of the environment. Inspired from the human visual system, this interface features a local and a global mode, the local mode giving real time feedback of the environment. Note that existing interfaces, such as "TheVoice" take into account the coding of gray levels.

With seven blindfolded participants we verified the hypothesis that it is possible to manipulate and to match colored objects, accurately. Overall, with only one training session, participants matched sock pairs with an accuracy of 94%. Moreover, it is very likely that with more training sessions the few mistakes that have been measured would disappear.

We started an experiment related to blind navigation with the help of a red serpentine. With the same experimenters we validated the hypothesis that with colors rendered by musical instruments and real time feed-back it is possible to follow a twisting path. To the best of our knowledge, experiments related to object manipulation and blind navigation based on sonification of colors have not been carried out.

In the future we will pursue the socks' and serpentine experiments, in order to increase our statistics. As well as blindfolded individuals, we are confident that blind persons will successfully follow the red serpentine, because of their improved auditory sense. Moreover, we will plan an experiment, for which depth represents an important parameter. Specifically, we could imagine the presence of obstacles and an experimenter should be able to estimate the distance separating him/her to an obstacle without touching it.

6. ACKNOWLEDGMENTS

The authors gratefully thank Mohammad Soleymani, Fedor Thönnessen, Stéphane Marchand-Maillet, Eniko Szekely, Bastien Francony and Marc Von Wyl for their valuable participation in the experiments and their precious comments related to the See ColOr interface. Moreover, we are grateful to the partners of the SIMILAR network of excellence for their collaboration. Finally, we express strong gratitude to the Hasler foundation for funding this very stimulating project.

7. REFERENCES

[1] 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.

[2] P.B.L. Meijer, "An experimental system for auditory image representations", *IEEE Trans. Bio. Eng.,* vol. 39, no. 2, pp. 112–121, 1992.

[3] 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.

[4] 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.

[5] G. Bologna, B. Deville, T. Pun, and M. Vinckenbosch, "Identifying major components of pictures by audio encoding of colors", in *Proc. IWINAC'07,* vol. 2, pp. 81–89, 2007.

[6] G. Bologna, B. Deville, T. Pun, and 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.

[7] R. Begault, *3-D sound for virtual reality and multimedia,* Boston A.P. Professional, ISBN: 0120847353, 1994.

[8] V.R Algazi, R.O. Duda, D.P. Thompson, and C. Avendano, "The CIPIC HRTF Database", *in Proc. WASPAA'01,* New Paltz, NY, 2001.