A perceptual interface for vision substitution in a color matching experiment

Bologna, Guido; Deville, Benoît; Vinckenbosch, Michel; Pun, Thierry

A Perceptual Interface for Vision Substitution in a Color Matching Experiment

G. Bologna, B. Deville, M. Vinckenbosch and T. Pun

***Abstract*— In the context of vision substitution by the auditory channel several systems have been introduced. One such system that is presented here, See ColOr, is a dedicated interface part of a mobility aid for visually impaired people. It transforms a small portion of a colored video image into spatialized instrument sounds. In this work the purpose is to verify the hypothesis that sounds from musical instruments provide an alternative way to vision for obtaining color information from the environment. We introduce an experiment in which several participants try to match pairs of colored socks by pointing a head mounted camera and by listening to the generated sounds. Our experiments demonstrated that blindfolded individuals were able to accurately match pairs of colored socks. The advantage 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. Our perceptual auditory coding of pixel values opens the opportunity to achieve more complicated experiments related to vision tasks, such as perceiving the environment by interpreting its colors.**

I. Introduction

Echolocation consists in perceiving the environment by generating sounds and then listening to the corresponding echoes. For instance, bats and dolphins have developed special skills, in order to perceive the environment by means of echo-location or sonar-location. Real neural networks in the brain can achieve such a difficult task of environment perception.

Visually impaired people use reverberated sounds, such as slapping of the fingers, or sounds from a cane, in order to become aware of the surroundings. 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 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 spatialization 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. 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]. For instance, when one looks above the horizon and it "sounds" blue using a piano sound, it will be very likely to be the sky. 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. In order to simplify the experiments, 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

Manuscript received November 30, 2007. This work was supported by the Hasler Foundation.

G. Bologna is with the University of Applied Studies, Geneva, Switzerland (phone: +41 22 5462831; e-mail: guido.bologna@hesge.ch).

B. Deville, is with the Computer Science Department, University of Geneva, Geneva, Switzerland (e-mail: benoit.deville@cui.unige.ch).

M. Vinckenbosch is with the University of Applied Studies, Geneva, Switzerland (e-mail: michel.vinckenbosch@hesge.ch).

T. Pun is with the Computer Science Department, University of Geneva, Geneva, Switzerland (e-mail: thierry.pun@cui.unige.ch).

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" green at the bottom of the picture it was likely to be grass. 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.

We have chosen to focus on auditory means for vision substitution, as for current navigation experiments this preference allows us to use conventional equipments (notebook and headphones) without the addition of special devices. Nevertheless, in the future we do not exclude to benefit from the combination of touch and audition.

In this work the purpose is to verify the hypothesis that sounds from musical instruments provide an alternative way to vision for obtaining color information from the environment. To this end, we introduce an experiment for which several participants try to match pairs of colored socks by pointing a head mounted camera and by listening to the generated sounds. The results demonstrate that matching colors with the use of a perceptual language, such as that represented by instrument sounds can be successfully accomplished. In the following sections we will present the models and methods behind the See ColOr interface, several experiments related to color matching, followed by the conclusion.

II. MODELS AND METHODS

A. Color Encoding

Many color systems are commonly used for computers and screen devices. An important drawback of the RGB model which represents three additive variables (red, green and blue) is that similar colors at the human perceptual level could result considerably further on the RGB cube and thus could generate in our interface perceptually distant instrument sounds. Therefore, we decided to use a much more intuitive color system, such as HSL (Hue, Saturation, Luminosity), which is a non-linear deformation of the RGB cube. HSL is a symmetric double cone symmetrical to lightness and darkness. The Saturation component always goes from fully saturated color to the equivalent gray. The Lightness in HSL always spans the entire range from black through the chosen hue to white.

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 for instance learning to associate absolute pitch frequencies to colors (which would require high level musician skills). 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:

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

This specific encoding of colors was chosen empirically. It is complicated to determine which instrument would be best suited to encode a particular hue. Generally, the different instruments should be equally well distinguishable, but it is unclear on how to satisfy this constraint, because it depends on each individual and on learning.

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. It would have been preferable to find a sound-representation which would preserve the structure of the color space, so that two colors that are visually similar would also appear auditorily similar. Unfortunately, the use of continuous sound frequencies coding the hue variable would involve very inharmonic and "frightening" sounds which would be unpleasant to our experimenters.

The audio representation h_h of a hue pixel value *h* is

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

with *g* representing the gain defined by

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

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:

- Do for ($0 \le S < 0.25$);
- Sol for ($0.25 \le S < 0.5$);
- Si flat for ($0.5 \le S < 0.75$);

- Mi 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 (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.

The pixel depth variable *D* is coded by sound duration. We quantify four depth levels; from one meter to three meters, every meter. This may be enough levels to navigate in a living room, but certainly not in a large city or an open space. Hence, in the future we may take into account to code depth beyond three meters by sound intensity. Currently, pixel depth farther than three meters is considered at infinity. The time duration of an instrument sound at infinity is 300 ms. Note that because of real time constraints it would be unfeasible to use longer sounds. Depth values can be undetermined, because of the limits of the depth determination algorithm in homogenous areas without texture. The current associations between pixel sound duration and pixel sound depth are:

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

B. Sound Spatialization

Our senses can be regarded as data channels possessing a maximal capacity measured in bits per seconds (bps). The bandwidth of the human auditory system is about 20 Khz, thus by virtue of the Nyquist theorem the maximal amount of information that could be transmitted through this channel without loss is 40 Kbps. Note that the visual channel has the largest capacity with about 1000 Kbps [7], while the sense of touch is less powerful than audition with about 0.1 Kbps [7]. Total vision substitution would be impossible to achieve, as the bandwidth of touch or audition would be inadequate. Therefore, it is important to make crucial choices in order to convey a small part of vision, in such a way that the reduced visual information rendered by another sense preserves useful meaning.

When a sound source is on the left side of a listener, a diffracted wavefront arrives to the right ear with some delay with respect to the left ear. Moreover, the head represents an obstacle of fixed size. As a result, it is possible to simulate lateralization, also denoted as two-dimensional auditory spatialization, with appropriate delays and difference of intensity between the two ears.

Inter-aural time delay (ITD) and inter-aural intensity difference (IID) are inadequate for reproducing the perception of elevation. In fact, the folds of the pinna cause echoes with minute time delays within a range of 0-0.3 ms [8] that cause the spectral content of the eardrum to differ significantly from that of the sound source. Strong spatial location effects are produced by convolving an instrument sound with Head-Related Impulse Response (HRIR), which not only varies in a complex way with azimuth, elevation, range, and frequency, but also varies significantly from person to person [9].

Generally, reproducing lateralization with uncustomized HRIRs is satisfactory, while the perception of elevation is poor. Since one of our long term goals is to produce a widely distributed prototype, thus involving standard HRIRs, we only reproduce spatial lateralization on a row of 25 points with the use of the CIPIC database. Note that the human auditory system can differentiate sound sources every five degrees in azimuth and it is usually difficult to distinguish sounds in the back from sounds in front. As a consequence, it would be really arduous to improve by a significant factor the resolution of the spatial information presented to humans.

In practice, each sound of the sonified row of 25 points corresponds to the convolution of an instrument sound with the corresponding HRIR filter. Measurements of the KEMAR manikin 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 video image.

C. See ColOr Interface Description

We use a stereoscopic color camera denoted STH-MDCS2 (SRI International: http://www.videredesign.com/). An algorithm for depth calculation based on epipolar geometry is embedded within the camera. 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”, respectively. The photographic interactive mode provides the user with a rough sketch of the image scene, which is summarized by a list of the largest homogeneous regions enumerated by a combination of voice and instrument sounds. The processing steps are the following :,

- the size of the picture is decreased by a factor of ten;
- pixel color values are averaged by a 3x3 window;
- pixels are labeled according to colors.

Subsequently, an arbitrary number of the largest colored areas (usually, eight areas), are enumerated. Specifically, for each colored region the photographic modality provides the user with :

- a spatial instrument sound corresponding to the

average color of the region including by sound spatialization the *x* coordinate of the area centroid;

- a number provided by voice representing the ratio of the area surface with respect to the picture surface;

- a number between one and ten representing the *y* coordinate of the area centroid.

As an example, Figure 1 represents an original picture

Fig. 1. Original picture with sky tree and grass.

with sky, tree and grass, while Figures 2 and 3 show examples of green and cyan segmented areas. Note that we use false colors to represent several disconnected areas belonging to the same color label.

Contrarily to the previous approach, the perceptual mode reacts in real-time. In practice, we sonify a unique row of 25

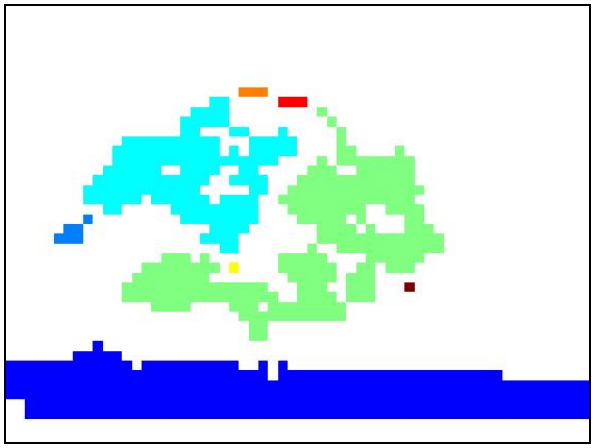


Fig. 2. Green areas of previous picture.

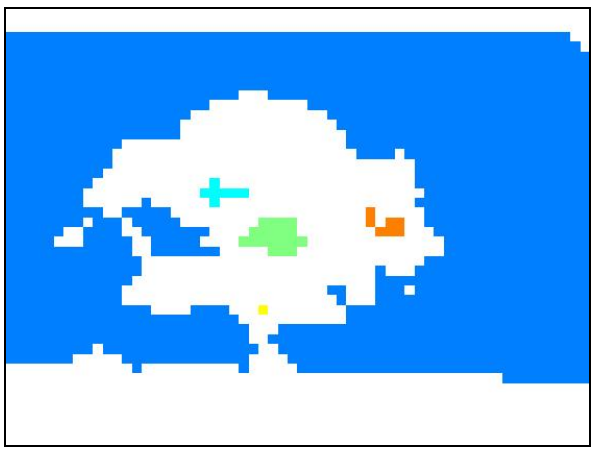


Fig. 3. Cyan areas of figure 1.

pixels at the centre 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 spatializazion. It is well known that rendering elevation is much more complicated than lateralization [8]. 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.

In order to replicate a crude model of the human visual system, pixels near the centre of the sonified row have high resolution, while pixels close to the left and right borders have low resolution. With this approach, a small periphery provides the user with local context and also permits to increase the angle of view. 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 omitted pixels.

[14 11 8 6 4 2 2 1 1 0 0 0 **0** 0 0 0 1 1 2 2 4 6 8 11 14]

The See ColOr perceptual mode is local, as a small region related to the centre of the picture is sonified. The photographic mode which is global complements the local mode. We think that it is important to have in the same system a local and a global modality, as with the human visual system.

III. Experiments

In the experiments we use the perceptual mode of the See ColOr interface. The purpose is to verify the hypothesis that sounds from musical instruments provide an alternative way to vision of obtaining color information from the environment. For the average participant it is difficult to learn the associations between colors and sounds in just one training session as shown by previous work [6]. However, our participants are not asked to identify colors, but just to pair off the same socks.

The experiments are performed by 10 blindfolded adults belonging to two groups. The first group of five persons has never used the See ColOr interface, while the second was involved in several experiments a year before these experiments [6], but has not reused the same interface in the meantime. The experiments present a training phase and four testing sessions, for which color is the only relevant parameter with varying hue, saturation and luminosity.

A. Training Phase

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. Figure 4 shows a picture with several hues on the horizontal axis and several saturation levels varying vertically with constant luminosity fixed at 0.8. We let our participants decide when they feel

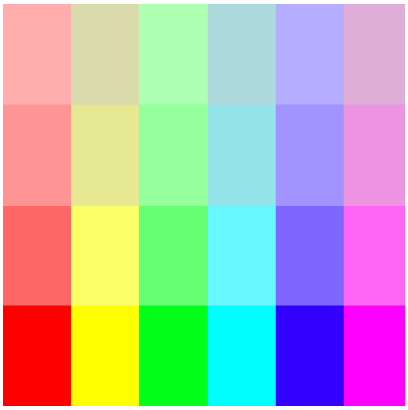


Fig. 4. A typical picture showed during the training phase of static pictures. Hue is represented horizontally, luminosity is constant (0.8) and saturation varies vertically (0.25; 0.5; 0.75; 1).

comfortable to switch to the second step. At this point, subjects wear the See ColOr device and learn to point the camera toward socks. With respect to each individual, Tables I and II illustrate the time dedicated to understand the HSL auditory encoding and the time duration for learning to point socks with the head mounted camera.

TABLE I
TRAINING TIME DURATION FOR THE FIRST GROUP

Participant	Static Training (mn)	Training with Socks (mn)
P1	17	14
P2	15	9
P3	11	8
P4	9	5
P5	12	12
Average	**12.8 (3.2)**	**9.6 (3.5)**

The second column indicates the time duration of the training phase with static pictures, while the third column relates to the learning with the head mounted camera pointing at several socks.

TABLE II
TRAINING TIME DURATION FOR THE SECOND GROUP

Participant	Static Learning (mn)	Learning with Socks (mn)
P6	8	5
P7	16	17
P8	3	13
P9	7	11
P10	18	15
Average	**10.4 (6.3)**	**12.2 (4.6)**

See Table I.

One of the difficulties is that several socks reflect the light in the room. Thus, an observer could be in trouble with the perception of the true color. As a consequence we advised our participants to learn to modify the angle between the camera and the socks. Note also that when a sock is manipulated, several parts of it could be hidden from the camera.

B. Testing Phase

Our experiment participants perform four different tests. In the first two series they need to pair five pairs of socks that are initially in a plastic bag. Socks have uniform colors. The goal is to pick up socks, then scrutinize them with the camera and match corresponding pairs. Figure 5 shows a participant observing a blue sock.

In the last two series of experiments the goal is to perform the same task without the See ColOr interface. In fact, we would like to demonstrate that just relying on touch involves random results.

Fig. 5. A blindfolded participant scrutinizing a blue sock with a stereoscopic camera mounted on his head.

1) First Series of Socks: We use five pairs of socks having the following colors:

- black
- green
- low saturated yellow
- blue
- orange

Figure 6 presents the socks, while Table III illustrates the prevalent sounds associated to their colors.

Fig. 6. The colored socks; from left to right : black, green, yellow, blue and orange.

Table IV illustrates the results for the first group of participants (the group of individuals who never used the See ColOr interface), while table V depicts those of the second group. It is worth noting that individuals of the

TABLE III
COLOR/SOUND ASSOCIATIONS OF THE FIRST FIVE PAIRS OF SOCKS

Socks	Hue and Saturation	Luminosity
Black	---	Bass (Do)
Green	Flute (Do-Sol)	Bass (Sol–Sib)
Yellow	Viola–Violin (Sol)	Bass (Do–Sol)
Blue	Piano–Trumpet (Sib)	Bass (Do–Sol)
Orange	Oboe–Viola (Sib-Mi)	Bass (Sol–Sib)

The second column indicates the instruments related to hue and saturation, while the last column depicts the instrument associated to the luminosity variable.

TABLE IV
FIRST SERIES OF SOCKS FOR THE FIRST GROUP OF PARTICIPANTS

Participant	Time (mn)	Success rate (pairs)
P1	12	5
P2	24	5
P3	7	5
P4	6	3
P5	16	5
Average	**13.0 (7.3)**	**4.6 (0.9)**

The second column indicates the number of minutes for achieving the task, while the last column indicates the numbers of correctly matched pairs of socks. Note that the standard deviation is given between parentheses.

TABLE V
FIRST SERIES OF SOCKS FOR THE SECOND GROUP OF PARTICIPANTS

Participant	Time (mn)	Success rate (pairs)
P6	5	5
P7	10	5
P8	4	5
P9	4	5
P10	18	5
Average	**8.2 (6.0)**	**5.0 (0.0)**

See Table IV.

second group are faster than those of the first group, on average and also that their success rate is higher. Participant P_4 made a mistake between blue socks, and yellow socks, because he was confused by piano representing blue and pizzicato violin representing yellow.

A question arising concerns the influence of training on the time required to pair socks. One of the authors who is very well trained on the associations between colors and instrument sounds performed the matching task without error in 2.2 minutes, which is almost twice as fast than the best participant time duration. This person learned the color associations in about 20-30 sessions of half an hour.

2) Second Series of Socks: We replicate the previous experiment with another five pairs of socks. The question here is to determine whether individuals can perform the color matching task with additional socks that were not showed during the first experiment. Colors are:

- dark purple
- dark red
- light green
- white
- cyan

Figure 7 presents the socks, while Table VI illustrates the

Fig. 7. The colored socks; from left to right : dark purple, dark red, light green, white and cyan.

prevalent sound associations.

The second trial required subjects to make more subtle

TABLE VI
COLOR/SOUND ASSOCIATIONS OF THE SECOND FIVE PAIRS OF SOCKS

Socks	Hue and Saturation	Luminosity
Purple	Oboe–Sax (Do–Sol)	Bass (Do–Sol)
Red	Oboe (Sol)	Bass (Do–Sol)
Green	Flute–Violin (Do–Sol)	Singing Voice (Do–Sol)
White	---	Singing Voice (Mi)
Cyan	Trumpet (Sol–Sib)	Singing Voice (Do–Sol)

See Table III.

distinctions between colors. The main difficulty was related to the similarity of sounds between the dark purple and the dark red pairs. Table VII illustrates the results for the first group of participants, while Table VIII depicts those of the second group. Interestingly, with respect to the previous

TABLE VII
SECOND SERIES OF SOCKS FOR THE FIRST GROUP OF PARTICIPANTS

Participant	Time (mn)	Success rate (pairs)
P1	6	5
P2	10	5
P3	8	5
P4	9	5
P5	8	3
Average	**8.2 (1.5)**	**4.6 (0.9)**

The second column indicates the number of minutes for achieving the task, while the last column indicates the numbers of correctly matched pairs of socks.

experiment, individuals of the first group are substantially faster on average. On the other hand, members of the second group are also slightly faster, on average. Moreover, for both groups the standard deviation of the task time length was considerably reduced. Finally, the success rate was for both groups equal to 92%. Participants P_5 and P_6 made a mistake on the dark purple and the dark red sock's pairs,

TABLE VIII
SECOND SERIES OF SOCKS FOR THE SECOND GROUP OF PARTICIPANTS

Participant	Time (mn)	Success rate (pairs)
P6	5	3
P7	8	5
P8	6	5
P9	9	5
P10	9	5
Average	**7.4 (1.8)**	**4.6 (0.9)**

See previous table.

which involve quite similar sounds.

With this experiment we showed that another five colors can be matched with high accuracy, and also that individuals who never used the See ColOr interface before these experiments can perform the matching task faster, on average, as they become more accustomed. Finally, one of the authors who is very well trained performed the matching task without error in 2.5 minutes, thus twice faster than the best experiment participant (P_6).

Blindfolded individuals have chosen personal strategies, in order to match socks. For instance, color information enabled some participants to manipulate socks in ways that could not occur otherwise. Specifically, the desk represents the user's work-space (see Figure 5) where socks are positioned. Sometimes, a user forgets the colors and the positions of the socks lying on the desk, but with the camera it is possible to look over the desk and to determine the position of a particular sock. Subsequently, the observer can decide to grab the sock corresponding to a specific sound.

Finally, the variable resolution of the sonified row would involve similar results with constant resolution, since the majority of our participants learn to observe socks at a distance from the camera varying between 10 to 30 cm to the camera.

3) Experiments Without the See ColOr Interface: We must be sure that socks are not identifiable by touch. For instance we could think that according to the manufacturing process some of them could present recognizable features. As a consequence, participants were asked to match pairs of colored socks without the See ColOr interface. Table IX illustrates the results for the first group of socks, while Table X describes those of the second group.

In previous experiments over the two groups, for all socks the average matching accuracy was 94%. Without the See ColOr interface the overall matching accuracy was 16%. Thus, the See Color interface significantly helped blindfolded participants to identify pairs of socks. In both experiments without the See ColOr interface a statistical analysis shows that the results can be viewed as random.

TABLE IX
MATCHING THE FIRST FIVE PAIRS OF SOCKS WITHOUT SEE COLOR

Participant	Time (mn)	Success Rate (pairs)
P1	2	0
P2	5	1
P3	5	1
P4	12	1
P5	3	0
P6	4	2
P7	2	0
P8	6	1
P9	2	0
P10	13	1
Average	**5.4 (4.0)**	**0.7 (0.7)**

The second column illustrates the number of minutes for achieving the task, while the last column indicates the numbers of correctly matched pairs of socks.

TABLE X
MATCHING THE SECOND FIVE PAIRS OF SOCKS WITHOUT SEE COLOR

Participant	Time (mn)	Success Rate (pairs)
P1	2	2
P2	1	0
P3	4	2
P4	8	1
P5	2	0
P6	3	2
P7	4	0
P8	8	1
P9	4	1
P10	7	0
Average	**4.3 (2.5)**	**0.9 (0.9)**

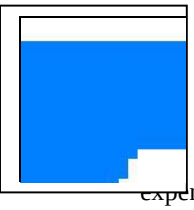
See Table IX.

IV. CONCLUSION

We presented the See ColOr interface, which has the characteristic to provide 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 colors of the environment. Ten blindfolded individuals validated the real time mode of this interface in a color matching task with the use of uniform colored socks. Overall, with only one short 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. When searching for items the color extensions would be only helpful if the object being looked for present a distinct color, which would be clearly different from the background. New experiments will be planned for the future. For instance, we will carry out an experiment with several blindfolded individuals aiming at following colored lines painted on the ground in a real environment.

ACKNOWLEDGMENT

The authors gratefully thank Isabelle Chapalay, Donn Morrison, Jean-Luc Falcone, Fokko Beekhof, Mohammad Soleymani, Joëlle Heldt, Guillaume Chanel and Fedor Thönnessen for their valuable participation in the



experiments. Finally, we are very grateful to the Hasler Foundation for funding this project.

REFERENCES

[1] L. Kay, “A Sonar Aid to Enhance Spatial Perception of the Blind: Engineering Design and Evaluation”, *The Radio and Electronic Engineer*, 44, pp. 605–627, 1974.

[2] P.B.L. Meijer, “An Experimental System for Auditory Image Representations”, *IEEE Transactions on Biomedical Engineering*, 39, vol. 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.*, 45, pp. 1279–1293, 1998.

[4] J.L. Gonzalez-Mora, A. Rodriguez-Hernandez, L.F. Rodriguez-Ramos, L. Dfaz-Saco, N. Sosa (1999), “Development of a New Space Perception System for Blind People, Based on the Creation of a Virtual Acoustic Space”, *In Proc. International Work-Conference on Artificial Neural Networks, IWANN*, pp. 321—330, 1999.

[5] G. Bologna, and M. Vinckenbosch, “Eye Tracking in Coloured Image Scenes represented by Ambisonic Fields of Musical Instrument Sounds”, *In Proc. Int. Work-Conference on the Interplay between Natural and Artificial Computation (IWINAC), June 2005, Las Palmas, Spain*, vol. 1, pp. 327-333.

[6] G. Bologna, B. Deville, T. Pun, and M. Vinckenbosch, “Transforming 3D Coloured Pixels into Musical Instrument Notes for Vision Substitution Applications”, *Eurasip J. of Image and Video Processing, Special Issue: Image and Video Processing for Disability*, A. Caplier, T. Pun, D. Tzovaras, Guest Eds., 2007, Article ID 76204, 14 pages (Open access article).

[7] T.P. Way, and K.E. Barner, “Automatic visual to tactile translation, part I: human factors, access methods and image manipulation”, *IEEE Trans. Rehabil. Eng*., 5 (1997), 81–94.

[8] R. Begault, *3-D Sound for Virtual Reality and Multimedia.* Boston A.P. Professional, ISBN: 0120847353, 1994.

[9] C.P. Brown, and R.O. Duda, “A Structural Model for Binaural Sound Synthesis”, *IEEE Trans. Speech and Audio Processing*, 6, vol.5, pp. 476-488, 1998.

[10] V.R. Algazi, R.O. Duda, D.P Thompson, and 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.