Audio Engineering Society

Convention Paper

Presented at the 120th Convention
2006 May 20–23 Paris, France

Acoustic Rendering for Color Information

Ludovico Ausiello[1], Emanuele Cecchetelli[1], Massimo Ferri[1], Nicoletta Caramelli[2]

[1] *ARCES University of Bologna, Bologna, via Toffano 2, 40125, Italy*

[2] *Psychology Dep. University of Bologna, Bologna, v.le B.Pichat 5, 40127, Italy*

Correspondence should be addressed to Ludovico Ausiello (`lausiello@arces.unibo.it`)

ABSTRACT

The Espacio Acustico Virtual (**EAV**) is a portable device that acoustically represents visual environmental scenes by rendering objects with the sound of virtual rain drops. Here, an improvement of this device is presented, which adds color to the information conveyed. Two different mappings of color into sound were implemented. **Georama** is a geometric coding based on **R**ed **G**reen **B**lue vectors, while **Colorama** is an associative coding based on the Hue and Saturation model of color space. An experiment was run on both sighted and blind participants in order to assess which of these coding is the most user friendly. The results showed that participants learnt to discriminate colors through sounds better when trained with Georama than with Colorama.

1. INTRODUCTION

Participants in the **EAV** [1] [2] project have realized a portable device that allows blind people to explore both in-and out-door spaces through an acoustic representation of the scene. This device represents the physical organization of the surrounding space locating a set of virtual sonorous sources, like rain drops [3], on objects. It consists of conventional eyeglasses linked to a comfortable-to-carry palmtop by a cable. Inside the glasses there are two miniature camcorders that provide the processor with the images necessary to compute the distance between the camcorder and the objects in the scene and to transform them into the sounds that give rise to the virtual scene. By means of convolution with previously measured head related transfer functions [4] [5], sounds are properly located in the virtual scene and can be heard in the two earphones that are part of the glasses [6] [7] [8]. Since it is geometrically manageable, the space, recognized by means of stereovision, is digitized until it reaches the form of physical blocks ("stereopixels"). This leads to a system of space polar coordinates, the origin of which is the head of the person, in which only the spherical cap immediately in front of the person is presented with image resolution depending on the limited number of the "stereopixels".

Adding color information to this system, offers a more sophisticated tool to blind people for better discerning objects and solving possible ambiguities. The suggestion is to dinamically change the timbre of the virtual rain drops by encoding colors into sounds and sound variations. Two different approaches have been developed and implemented. The first, **Colorama**, associated a 64 colors palette taken from the **Hunter** coding of color space [9] to another one consisting of 64 percussive sounds. The second, **Georama**, mapped the three basic vectors of color space, **R**ed **G**reen and **B**lue [10], into three primary timbres. Then, the amplitude of each vector was transposed into the volume of its associated timbre and mixed. The recognition of transposition of color into sound by people depends on different cognitive resources in the two approaches. In fact, the former approach, based on associative encoding that provides a mental table to be checked for, heavily depends on memory resources. The latter, instead, based on a geometric encoding affording chromatic continuity requires on line processing of incoming information whitout any memory load.

2. EXPERIMENT

Hypothesis

The aim of this research was to assess in a pilot experiment which one of the two aforementioned types of encoding is more user friendly and, thus, worth of implementation in the **EAV** system in order to enhance its possibilities. Accordingly, the hypothesis was that, after the same training, the participants trained with Colorama and those trained with Georama should differ in their mastering of the coding systems as a consequence of the different cognitive elaboration required by the two systems.

2.1. Method

2.1.1. Participants

Twenty participants, with normal hearing ability, volunteered their participation in the experiment. 10 of them were normally sighted people (mean age 27ys) and 10 were blind people (mean age 45ys). As knowledge of both music and color could bias the results, musical knoledge was balanced in the two groups of participants, while the normally sighted participants were checked for special interest in color.

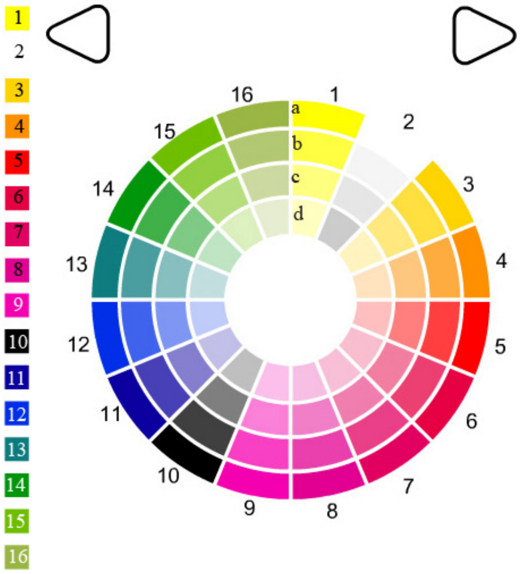

Fig. 1: Colorama: stimuli identification (numbers) and saturation degrees (letters)

2.1.2. Materials

Both Colorama and Georama run as browser plugins and were implemented using Shockwave Flash.

Colorama is based on the Hunter coding and maps a subset of Hue with a palette of 16 basic timbres. For every color there was an inner sub-palette with 4 different Saturation degrees (25%, 50%, 75% and 100%). In order to reproduce the loss of information emebedded in strongly de-saturated stimuli, the decreasing saturation was coded mixing an increasing percentage of white noise, keeping the energy of the stimuli unchanged. Figure 1 shows the architecture of the software with the colors and the buttons that produce the different timbres.

Georama is based on the **C.I.E.** coding and maps the **R**ed, **G**reen, **B**lue vectors with three primary timbres the volume of which represents the intensity of the color vector. In the software used in this experiment there were only 26 premixed stimuli instead of the continuous representation of the RGB gamut (with values from 0 to 255 for each vector) keeping the energy of the stimuli unchanged. Figure 2 illustrates the architecture of the software. Buttons on the edges of the triangle produce stimuli with only two primary timbres mixed together, buttons in the inner part of the triangle produce more complex stimuli obtained with all the three timbres.

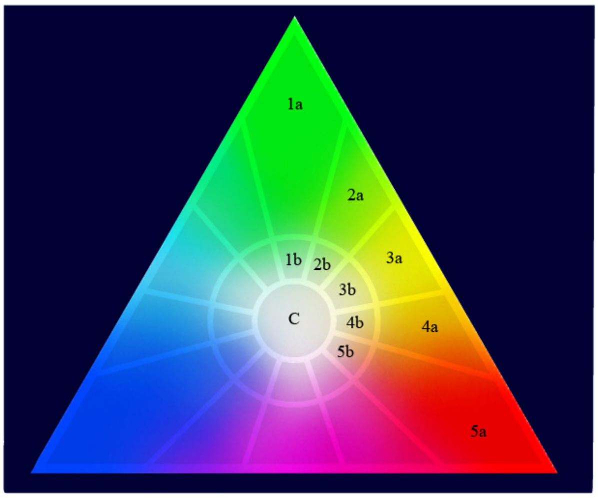

Fig. 2: Georama: stimuli ID (numbers) and saturation degrees (letters)

White is in the center of the triangle as it is obtained with all the three vectors mixed at maximum relative volume.

In both the softwares, i.e. **Colorama** and **Georama**, the same sets of stimuli to be presented were arranged. Furthermore, for each coding method two kind of sequences were prepared. The first one was supposed to be "easy", the second "difficult". The "easy" sequence contained a smaller number of sounds than the "difficult" one. Every sound sample was a PCM 16 bit 44.1 kHz, with a time duration of 500 ms and a normalized maximum SPL.

In **Colorama**, the "easy" sequences consisted of 8 contiguous colors, while the "difficult" ones consisted of 4 non-contiguos colors with all the 4 degrees of de-saturation where the letter "a" indicated 0%, "b" 25%, "c" 50% and "d" 75% of de-saturation.

In **Georama**, the "easy" sequences consisted of 12 colors of the outer crown (two timbres mixed), while the "difficult" sequences consisted of 15 colors (three timbres mixed) with white included. In this case the letter "a" stands for the stimuli taken from the outer crown and "b" for those taken from the internal one. White was labelled with "c", as the inner crown degenerates into a single element. Thus, the experimental materials can be summarized as follows:

- Colorama sequences:
  - 8 contiguous stimuli randomly presented ("easy" sequences)
  - 4 non-contiguous stimuli randomly presented at the 4 de-saturation degrees ("difficult" sequences)
- Georama sequences:
  - 12 contiguous stimuli randomly presented ("easy" sequences)
  - 15 stimuli taken from the two primary colors and white randomly presented ("difficult" sequences)

2.1.3. Procedure

The participants were tested individually and the experimental session consisted of two parts, the training and the test.

In the "easy" **Colorama** sequences training, the participants had to listen to 4 sequences of 8 contiguous stimuli, which were identified by a number [11], e.g. from number 2 to number 10 and viceversa. The guided identification of a random sequence of 8 stimuli completed the training, with participant required to recognize the proper identification number of the colors. If participants were successful, the next stimulus was presented, otherwise the same stimulus was presented again till at ceiling performance was obtained.

In the "difficult" **Colorama** sequences training, the participants had to listen to 4 sequences of 4 non--contiguous stimuli ordered with increasing saturation and viceversa and to identify them using both the number and the letter. Then they had to identify 2 sequences of the same stimuli presented in random order.

In the test part of the experiment with the "easy" **Colorama** encoding, participants' identification of 4 randomly presented sequences of 8 stimuli was recorded in order to measure the distance between the stimulus and the identified sound according to a Manhattan norm. Thus, for example, if the participant identified the sound as number 5 when presented with stimulus number 2, then the **R**ecognition **D**istance (**RD**) was 3.

In the test part of the experiment with the "difficult" **Colorama** encoding, participants' identification of 4 random sequences of 16 stimuli with all the possible degrees of de-saturation was recorded. However, while the angular distance was computed as in the "easy" condition, the radial distance was computed assigning a distance equal to $0,25$ to each de-saturation degrees. Thus, for example, if the participant identified the sound as number 4c when presented with stimulus number 4a, then the **RD** was $0,5$.

In the "easy" **Georama** sequences training, the participants had to listen to 3 sequences of 12 contiguous numbered stimuli, which afterwards were randomly presented, and to identify them.

In the "difficult" **Georama** sequences training, the participant had to listen to 2 sequences of 15 non-contiguous stimuli from the outer crown of colors to white, and viceversa, and to identify them.

In the test part of the experiment with both "easy" and "difficult" **Georama** sequences the same measure of **R**ecognition **D**istance was calculated as with **Colorama** sequences.

The training sequences were balanced between sighted and blind participants for both **Colorama** and **Georama**.

2.1.4. Data Analysis and results

The **R**ecognition **D**istance obtained with the two different encodings were normalized (as the maximum measurable **RD** was 8 in Colorama and 6 in Georama). On the obtained identification measures an ANOVA was performed, the factor of which were the between factor group (sighted vs. blind participants), and the two within factors kind of encoding (Colorama vs. Georama) and condition (easy vs. difficult) thus obtaining a $2 \times 2 \times 2$ design.

Both the kind of encoding factor (Colorama vs. Georama) and the condition factor (easy vs. difficult) reached significance (respectively $F(2,72) = 6,879$, $Mse = 0,507$, $p < 0,0018$; $F(4,72) = 5,492$, $Mse = 0,507$, $p < 0,0006$). The factor group (sighted vs. blind participants), however, was not significant. This means that previous color experience did not bias the identification of colors when they were transposed into sounds. Instead, both the degree of difficulty and the different type of encoding influenced participants' performance. In fact, the "easy" condition in both Colorama and Georama was identified better than the "difficult" one (see Tab. 1).

Easy condition	0,860
Hard condition	0,997

Table 1: Mean normalized **RD** between stimuli and participants' identifications in the "easy" and "difficult" conditions

Furthermore (see Tab. 2) the overall performance with **Georama** software was better than that obtained with the **Colorama** one.

Colorama coding	1,028
Georama coding	0,830

Table 2: Mean normalized **RD** between stimuli and participants identifications with the Colorama and Georama

	Easy cond.	Hard cond.
Colorama coding	0,670	1,387
Georama coding	1,051	0,608

Table 3: Mean normalized **RD** between stimuli and participant identifications in both difficulty levels and coding systems

In the easy condition, the associative encoding of Colorama yielded a mean RD smaller than that yielded by the Georama encoding. In this condition the association of a palette of saturated colors to another consisting of single timbres was established more easily than in the Georama encoding. However, the more the resolution of the coding increased, thus requiring an increased sensitivity to the association between colors and sounds, the more participants' performance was impaired when compared to that obtained with Georama. In fact, the mean RD increased instead of decreasing due to the actual distances between the presented stimuli which were smaller than in the easy condition.

Regarding the geometric coding of Georama, although more demanding in the "easy" condition than Colorama, yielded better performances than

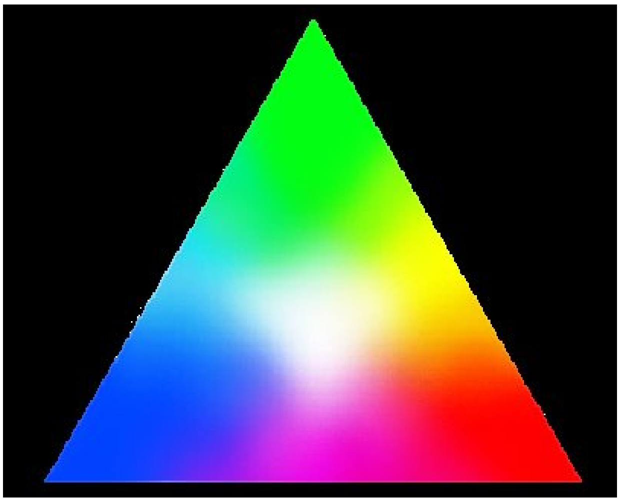

Fig. 3: PlayRGB: continuous coding software

Colorama with increasing resolution, thus showing a coherent effect in the two conditions (easy vs. difficult). In fact, during the test, mean RD improved approximately in 50% of the participants.

3. CONCLUSIONS AND DEVELOPMENTS

As the results of this experiment have shown that a geometric coding of colors into sounds is more effective, i.e. less cognitively demanding, than a simple associative coding, a continuous encoding software called PlayRGB was developed. This new software written in Visual Basic 5 is based on DirectX (Windows XP sound library). Its graphic interface allows people to load an image (Figure 3 shows the loaded image of the **R G B** Gamut) and to capture every color nuance by cliking on the points inside the triangle with the mouse. Then, the software, based on a fully continuous geometric encoding, reproduces the corresponding sound in real time. Further developments of this device will be aimed at obtaining personalized timbres in order to improve performance in color identification.

4. ACKNOWLEDGMENTS

We would like to thank Marco Flore of 21th Century Labs for help in developping the graphic interface of Colorama and Georama. This research was supported by ARCES, University of Bologna.

5. REFERENCES

[1] L. Diaz-Saco N. Sosa J. L. Gonzales Mora A. Rodriguez-Hernandez L. F. Rodriguez-Ramos, H. M. Chulani. Image and sound processing for the creation of a virtual acoustic space for the blind people, www.iac.es/project/eavi/.

[2] L. F. Rodriguez-Ramos L. Diaz-Saco N. Sosa J. L. Gonzalez-Mora, A. Rodriguez-Hernandez. Development of a new space perception system for blind people, based on the creation of a virtual acoustic space, www.iac.es/project/eavi/.

[3] John M. Hull. *Touching the Rock: An Experience of Blindness.* Vintage Books, June, 1992.

[4] William M. Hartmann. How we localize the sound. *Physics Today*, 52:24–29, November 1999.

[5] William M. Hartmann and Andrew Wittenberg. On the externalization of sound images. *J. Acoustical Society of America*, 99:3678–3688, June 1996.

[6] Frederic L. Wightman and Doris J. Kistler. Headphone simulation of free-field listening. I: Stimulus synthesis. *J. Acoustical Society of America*, 85:858–867, February 1989.

[7] Frederic L. Wightman and Doris J. Kistler. Headphone simulation of free-field listening. II: Psychophysical validation. *J. Acoustical Society of America*, 85:868–878, February 1989.

[8] Doris J. Kistler Frederic L. Wightmann Elizabeth M. Wenzel, Marianne Arruda. Localization using nonindividualized Head-Related Transfer Functions. *J. Acoustical Society of America*, 94:111–123, July 1993.

[9] C.I.E. The 1976 C.I.E. chromaticity diagram, www.cie.co.at/cie/.

[10] C.I.E. C.I.E. tristimulus values, www.cie.co.at/cie/.

[11] Jules Davidoff. Language and perceptual categorisation. *Trends in cognitive science*, 5:382–387, September 2001.