SYNTHETIC SYNESTHESIA: MIXING SOUND WITH COLOR

Kristinn R. Thórisson *Karen Donoghue*

The Media Laboratory
Massachusetts Institute of Technology
20 Ames Street, E15-411 Cambridge, MA 02139
kris@media.mit.edu

ABSTRACT
An interface is described that uses color and spatial relations to provide an intuitive interface for sound manipulation. A simple geometric shape, called the Geometric Sound Mixer (GSM), is used to mix sounds. Timbre is represented as color within the GSM; the relative loudness of these sound sources is represented visually by the color mixture. A dynamic representation of any sound mix can be viewed on the Mix Time Line, where relative moment-to-moment audio levels control the color mix and brightness as the sounds play in real time. Perceptually linear audio and color mixes are achieved using psychophysical functions. The result is an environment that allows for complex manipulations of sound in a highly simplified, structured environment.

KEYWORDS: Sound manipulation, color, perception, psychophysics, multi-media, user interface design.

INTRODUCTION
Sound is experienced temporally: once it goes away, the experience must be recreated from memory. It is often difficult to visualize the combination of several sound sources. Sound engineers must rely on vague visual cues such as fader positions and labels as their primary static representation of a sound mix. A synthesizer programmer often needs to control over a hundred parameters with only a few buttons and a small LCD screen to visualize the sound [4]. The rest is left to his overworked short-term memory. Computer displays can give a more detailed picture of the work environment, but most current interfaces for sound manipulation represent the mechanical basis of the sound generation or mix.

The complexity and transient nature of sound can put a serious load on a user's memory and cognitive functioning. The system described here is an attempt to alleviate some of this cognitive load by moving parts of the sonic functionality into the visual domain, providing an interface whose functions are derived psychophysically rather than physically. (For related work see e.g. [1].) It uses color and spatial dimensions in a way that transcribes what the user *hears* into what he *sees*, providing a permanent visual representation of the otherwise transient nature of sound. The full potential of this method will be apparent on computers that integrate digital sound processors and sound synthesis abilities, but lack an intuitive and concise way to access these technologies.

INTERFACE DESCRIPTION

Basic Units
The interface consists of three units: the Geometric Sound Mixer (GSM), a two-dimensional geometric shape used to define a color/sound-mix area (Figure 1A), the Mix Time Line (MTL), which displays the relative sound mix dynamically (Figure 1B), and menus for choosing colors and timbre (sound sources).

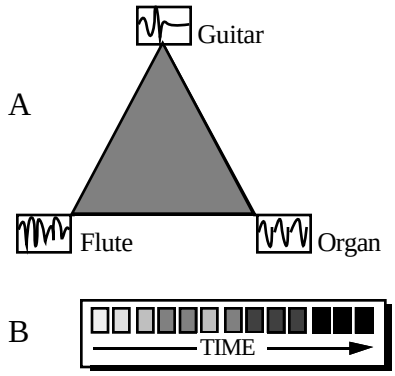

Figure 1. The Geometric Sound Mixer (A). The sound sources attached to its vertices (a triangle, in this case) are guitar, flute, and organ. (Waveforms could be synthesized or sampled.) The Mix Time Line (B) displays dynamically the mix of sound sources for any given point in the GSM as "snapshots" of color mixtures.

Example of Use
First, the user chooses the desired number of sound sources from the timbre menu, associating each of these with a vertex on the GSM. If the user picks three sounds, for example, the resulting GSM will be a triangle (Figure 1A). Colors are then chosen from the color menu and associated with each of the sound sources; the color of each sound source at a vertex blends linearly with the colors of its neighboring vertices. Any point on the triangle's two-dimensional plane represents a certain mixture of the three sound sources; the audio mix is directly reflected in the mixture of the colors. All possible combinations of the three sound sources are therefore instantly accessible to the

Published in: InterCHI Adjunct Proceedings '93, Conference on Human Factors in Computing Systems, Amsterdam, April 24-29, 65-66.

user's visual inspection: the color mix at any point within the area defined by the GSM's borders indicates the sound mixture at that point. By picking a point within the GSM (done by moving a cursor around within its border) the user can mix the sounds together in the desired proportions. When a chosen mix is played out, the resulting sound has the same perceptual combination of sound sources as the combination of colors at that point.

Because the interface perceptually matches sight and sound in a one-to-one relationship using psychophysical functions [2], the user can hear a sound mixture and see the exact same mixture expressed in color. The linear blend of the colors is *perceptually the same* as the audio mix. This has certain advantages: if the user wants to double the relative (perceptual) loudness of a sound source in one corner of the GSM, this can be done by halving the distance to that corner, or by finding a point with (perceptually) twice the amount of that sound's color. An equal amount of all sound sources can be found in the center of any GSM shape (point *t* in Figure 2).

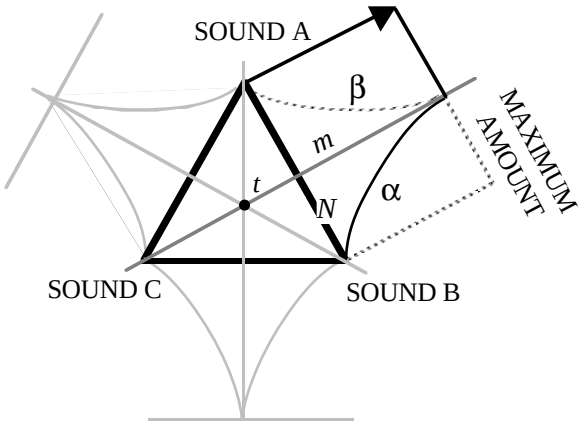

Figure 2. The relative amplitude of each sound falls off along the sides of the GSM according to function (1); curves α and β show how SOUND A and SOUND B fall off along their mutual side *N* of the GSM, respectively.

Technical Description

The method for transcribing amplitude into the spatial domain is based on psychophysical laws. Let's take the example of the triangle GSM shown in Figures 1 and 2. If we follow side *N* from SOUND A to SOUND B in Figure 2, we reach a line *m* perpendicular to *N*'s midpoint. Here the amplitude of SOUND A falls off according to the curve marked α. This is the function

$$I = \frac{L^{1.67}}{k} \qquad (1)$$

where I is the physical intensity, L is the apparent relative loudness, and *k* is a constant [2]. The same holds true for SOUND B in the opposite direction: it too starts falling off halfway to the other sound (marked β in Figure 2), making its volume zero in the opposite corner. The color mixture in the GSM is controlled by a lookup function to the Munsell color space on the hue (color) axis [3]. Since the Munsell space is perceptually derived, distances on each of the three axes (hue, brightness and saturation) are perceptually linear. Therefore, halfway between any two vertices in the GSM will be a color that is perceptually halfway between the two colors of the sounds at those vertices.

Since the sounds might not all have the same amplitude envelope, a dynamic display of any mix can be viewed on the Mix Time Line. Here, moment-to-moment audio levels control the color mix as the sounds play in real-time. Instead of controlling only hue, the dynamically changing amplitude of the sound sources is also transcribed to brightness by a lookup function to the Munsell color space [3]. In our current implementation, the brightness of the chosen hue represents maximum amplitude; black is silence.

Current Implementation

The current prototype has been developed on a Hewlett-Packard workstation, with access to synthesizers and a MIDI mixer via a MIDI interface. The user interacts with the system using a stylus.

DISCUSSION

The current implementation allows for cross-fades between sounds, which for example is useful in mixing live music (each vertex representing an instrument) or balancing levels between synthetic sound generators. However, the interface also lends itself to more complex manipulations such as interpolation between waveforms. Moving the interface and the sound generation to a common platform would make this, and many other options, possible. For example, sound effects could be introduced as a third dimension in the system, expanding the two-dimensional GSM plane to a three-dimensional shape. Other extensions include multiple GSMs in the same workspace, and an option for recording the movements of the pointer within the GSM, thus storing a dynamically changing mix of the sound sources. These could offer extremely powerful methods for creating new sounds, blurring the distinction between the interface and the instrument.

ACKNOWLEDGEMENTS

This work was done under the supervision of Muriel Cooper and David Small, and sponsored in part by Hewlett-Packard. The authors wish to thank Lisa Ezrol and Steve Librande for contributions. Special thanks to Richard Bolt.

REFERENCES

1. Abbado, A. (1988). *Perceptual Correspondences of Abstract Animation and Synthetic Sound.* Master's Thesis, Massachusetts Institute of Technology.
2. Coren, S. & Ward, L. M. (1989). *Sensation and Perception*. San Diego, CA: HBJ.
3. Foley, J, van Dam, A., Feiner, S., & Hughes, J. (1990). *Computer Graphics: Principles and Practice.* Reading, MA: Addison-Wesley.
4. *Synthesizer Programming* (1987). D. Milano (ed.). Hal Leonard Publishing Company.

Published in: InterCHI Adjunct Proceedings '93, Conference on Human Factors in Computing Systems, Amsterdam, April 24-29, 65-66.