



Audio Engineering Society Convention Paper

Presented at the 119th Convention
2005 October 7–10 New York, New York USA

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Frequency-Based Coloring of the Waveform Display to Facilitate Audio Editing and Retrieval

Stephen V. Rice

The University of Mississippi, University, MS, 38677, USA
Comparisons Corporation, Grass Valley, CA, 95945, USA
rice@cs.olemiss.edu

ABSTRACT

The audio waveform display provides the visual focus in audio-editing systems yet sounds are difficult to see in the display. Using a new technique, the display is colored to represent the frequency content to make sounds more visible. This requires extraction of frequency information from the audio signal and an appropriate mapping of this information to the color space. Ideally, the coloring is independent of recording level, and similar sounds are represented by similar colors. Audio-editing systems are enhanced by the improved user interface. Audio-retrieval systems can present colored waveform displays as visual “thumbnails” in a list of sound search results.

1. INTRODUCTION

A user interface that is “WYSIWYG” (What You See Is What You Get) has been a hallmark of document-editing systems for more than 20 years. Audio-editing systems have not attained this status due to the difficulty in making sounds visible.

The focal point of user interaction in audio-editing systems is the audio waveform display, the well-known graph of the amplitude of an audio signal over time. The user may pan and zoom the display. To pan, the user scrolls the display horizontally to view a particular time interval. A shorter or longer time interval may be viewed by zooming in or out. Editing operations are

specified with respect to the waveform display: the user clicks on the display to select a position or region in an audio track and then applies editing operations to the selection. A multi-track audio-editing system presents one waveform display for each track of a recording; for example, a 24-track recording appears as 24 parallel waveform displays.

The frequency content of the audio signal is not evident in the waveform display unless zoomed in to a brief time interval less than one second in duration. However, it is common to zoom out to see many seconds or even minutes of the waveform at once. The display then indicates only the relative loudness of sounds. Consequently, it is difficult to locate sounds in the display.

Users exert mental effort to learn and recall the shapes of the amplitude envelopes of sounds because these are the only visual landmarks in the display. Unfortunately, these shapes are altered by zooming in or out. Users repeatedly audition segments to learn the association between what they see (shapes in the waveform display) and what they hear (sounds of the audio recording). Frequency information *is* apparent in a spectrogram but it consumes precious screen space and is unwieldy if one is required for each track.

This article presents the technique of color coding the waveform display to convey the frequency content of an audio signal. This technique was invented by the author and patented [1]. The coloring communicates a wealth of information to the user and makes the audio-editing process faster, easier, more accurate, and more enjoyable. It also facilitates audio retrieval.

2. TECHNIQUE

A sequence of digital audio samples is partitioned into brief consecutive intervals (for example, 50 milliseconds for each interval) and frequency information is extracted from each interval using any of a number of methods, such as FFT, linear prediction, or wavelets. A “signature” is created for each interval to hold a description of the frequency content of the interval. It is preferable for a signature to represent the relative strengths of frequencies while being invariant with respect to scaling and polarity. That is, amplification or attenuation of the digital audio samples (multiplication by a nonzero constant) does not alter the signature.

A function is needed for mapping each signature to a color. There is an infinite number of possible mappings; however, it is advantageous to utilize a mapping in which sounds perceived as similar by human listeners are mapped to similar colors. Within the waveform display, each of the time intervals is painted by the color derived from the signature representing the time interval. The colors appear as vertical stripes in the display.

In the author’s standard implementation, a 16-byte signature is mapped to a 24-bit color in the RGB color space. The mapping from sound to color is designed so that similar sounds are mapped to similar colors. Red has a connotation of alarm and so does a woman’s scream, so the red component is increased in colors

assigned to high-frequency sounds. Low-frequency sounds are given dark colors so they both look and sound ominous. Middle to high frequencies are shaded green, while low to mid-range frequencies are shaded blue. “Noisy” sounds, such as white noise, receive a shade of grey. When distinct sounds are played together, the louder sound has proportionally greater influence on the color assignment.

These color assignments are illustrated in Figure 1. An explosion in dark colors is followed by a woman’s scream highlighted in red. Then a siren with two alternating tones appears as an alternating sequence of green and blue bands. At the end, white noise is shown in grey. When sounds are mixed, their colors appear to be mixed. The siren’s “blue” sound commences just before the “green” sound has finished. In the brief interval when both sounds can be heard, the waveform is colored by a mixture of the blue and green shades.

The mapping from sound to color takes into account all of the frequency information, not just the pitch. This allows coloring of polyphony and inharmonic sounds, for which fundamental frequency cannot be determined. The influence of harmonics on the coloring can be seen in Figure 2. The same note (E4) is played by five musical instruments (bassoon, clarinet, English horn, trombone, and violin) but receives different colors according to the harmonics of the instruments. The striped patterns visible in the English horn and violin notes represent vibrato. Such subtle variations are apparent in the colored waveform display. A note bending sharp or flat is made evident by color changes.

Figure 3 shows the colored waveform display of the first two measures of Bach’s *Minuet* played by an oboe. Each note is assigned a unique color; however, subtle variations in the notes, for example at the attack and release points, are indicated by variations in color. In Figure 4, the sounds of bass drum, snare drum, and high-hat beats can be distinguished by their colors.

Figure 5 illustrates that zooming changes the shape of the amplitude envelopes, but the correspondence between color and sound is unchanged, providing a stable visual landmark. Because signatures are invariant with respect to scaling, the colors are unaffected by changes in recording level, as shown in Figure 6.

There are more than 16 million colors available in the 24-bit color space; however, the number of colors

discernible to the human eye is far less, perhaps as few as 100,000. The number of sounds represented by a signature is approximately 10^{30} , so there is no avoiding a many-to-one mapping from sound to color. Normally, the mapping assigns only similar sounds to a particular RGB color. However, due to the shortage of discernible colors, sounds dominated by very high frequencies (above 2 kHz) may be assigned colors that are also used for lower frequencies. This re-use of colors can be seen in bird songs for example.

Rather than map the entire sonic universe to the color space, each audio recording could be given a unique mapping of its sounds to the color space. While this would solve the color-shortage problem (because colors would be needed only for the sounds in the recording), it would require users to learn a different correspondence between sound and color for every recording and would make it impossible to compare the colored waveform displays of different recordings. By utilizing only one mapping from sound to color, users

learn the correspondence between sound and color and develop the ability to “read” audio. That is, they obtain an impression of how a recording will sound simply by looking at its colored waveform display.

In addition to their use in audio editing, colored waveform displays have proven to be valuable as visual “thumbnails” that represent recordings in a list of search results returned by an audio-retrieval system. The colored waveform display can help a user decide whether to audition a recording retrieved by the system. This feature is utilized effectively at FindSounds, the first Web search engine for sound effects (see the companion paper [2]). It is also used in the FindSounds Palette program [3], which is illustrated in Figure 7.

Colored waveform displays are also effective in representing audio tracks in a video-editing system. It is easier to synchronize a sound with a video frame when the sound is visible in the waveform display.

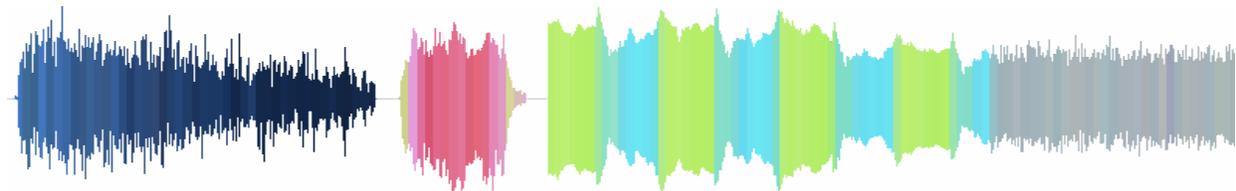


Figure 1 Sound effects;
from left to right: explosion (dark), scream (red), siren (alternating colors), and white noise (grey)

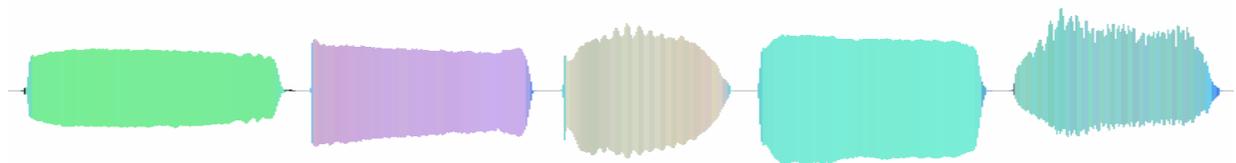


Figure 2 The note E4 played by five instruments;
from left to right: bassoon, clarinet, English horn, trombone, violin

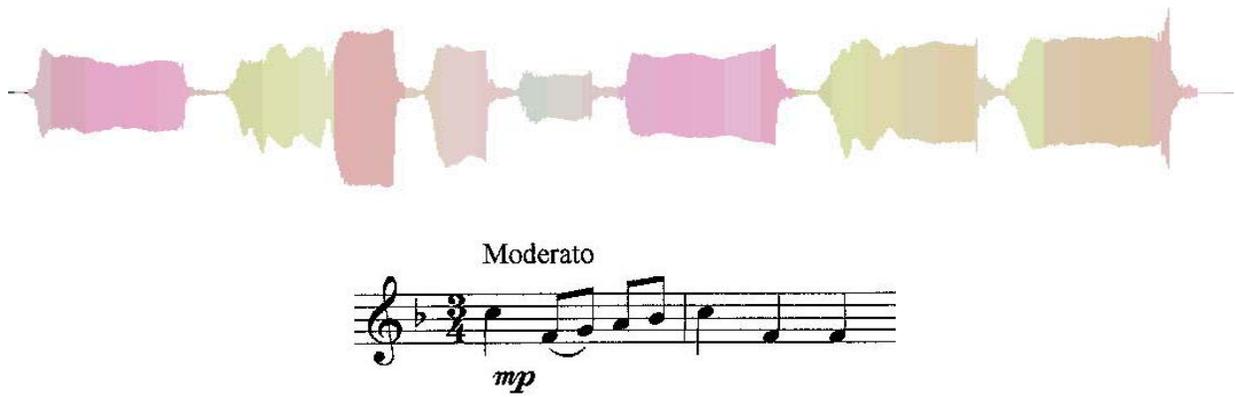


Figure 3 First two measures of Bach's *Minuet* played by an oboe

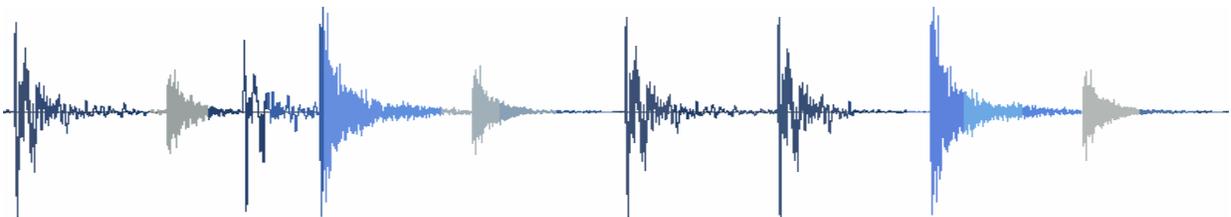


Figure 4 Drum loop: bass drum (dark), snare drum (blue), high hat (grey)

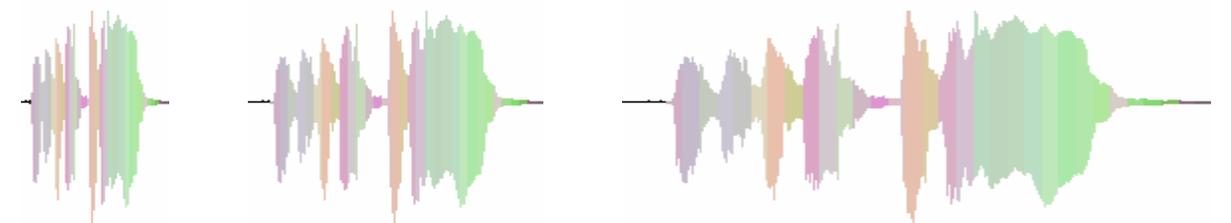


Figure 5 The same trumpet riff at three zoom levels

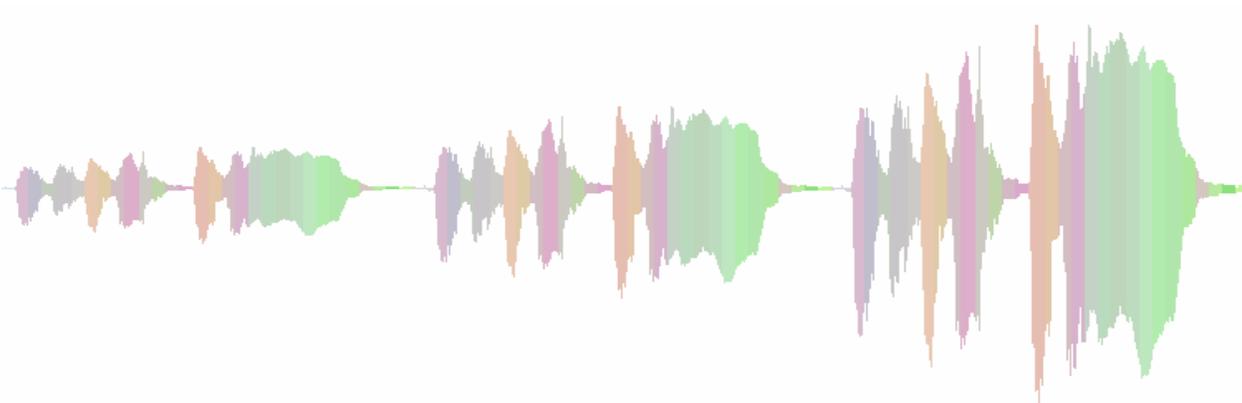


Figure 6 The same trumpet riff at three recordings levels

Waveform	File Name	Description
	Sirens.wav	toy siren
	fa-siren.wav	siren
	siren.wav	siren
	TrafficAlert...	siren
	thunder.wav	siren
	StreetYelp...	siren
	siren.wav	firehouse siren
	siren.wav	siren
	SIREN.WAV	siren
	Siren Loop ...	alarm/siren

Figure 7 Search results displayed by FindSounds Palette

3. CONCLUSION

The colored waveform display is a visually-attractive, computer-generated representation of audio. It facilitates audio editing by making sounds more visible. By coloring waveform displays to convey frequency information, an audio-editing system comes closer to the goal of WYSIWYG (What You See Is What You Get).

In the author's implementation, similar sounds are represented by similar colors, and changes in sound are indicated by changes in color. The display is computed efficiently and can be created in real time, as sound is recorded. This implementation is used by the FindSounds search engine [2] (www.FindSounds.com) and the FindSounds Palette program for audio editing and retrieval [3] (www.FindSounds.com/Palette.html). It is also available in the free Comparisonics Audio Player (www.Comparisonics.com/CSPlayer.html).

4. REFERENCES

- [1] Rice, S. V., and M. D. Patten, "Waveform Display Utilizing Frequency-Based Coloring and Navigation," U.S. Patent No. 6,184,898 (2001).
- [2] Rice, S. V., and S. M. Bailey, "A Web Search Engine for Sound Effects," presented at the AES 119th Convention, New York, New York, 2005 October 7-10.
- [3] Rice, S. V., and S. M. Bailey, "Searching for Sounds: A Demonstration of FindSounds.com and FindSounds Palette," *Proc. of the International Computer Music Conference*, 2004, pp. 215-218.