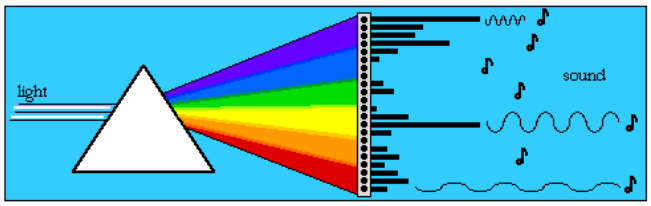
Artificial Synesthesia via Sonification: A Wearable Augmented Sensory System

Leonard N. Foner
MIT Media Lab
20 Ames St, E15-305
Cambridge, MA 02139
foner@media.mit.edu

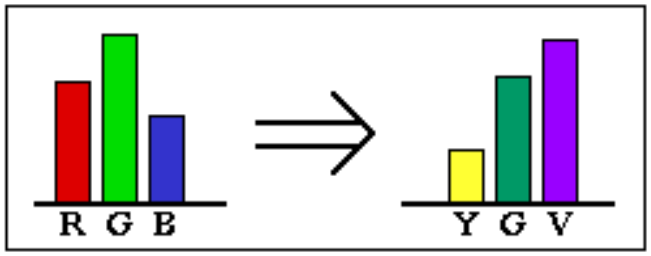


Abstract

A design for an implemented, prototype wearable artificial sensory system is presented which uses data sonification to compensate for normal limitations in the human visual system. The system gives insight into the complete visible-light spectra from objects being seen by the user; long-term wear and consequent training might lead to identification of various visually-indistinguishable materials based on the sounds of their spectra. A system design is presented, and some possible extensions to both the sonification and the sensor package are discussed.

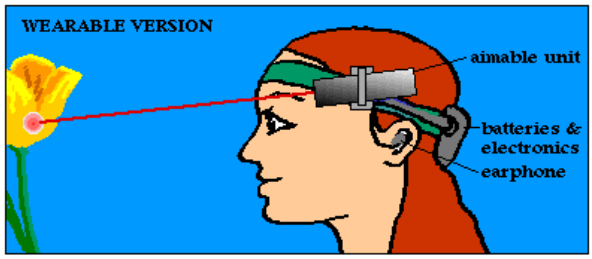
Introduction

Normal humans are trichomats [1], meaning that almost any three wavelengths of light, if their amplitudes are adjusted correctly, can be made to look like some particular color. (For example, this is the principle behind the red, green, and blue pigments in a CRT, and is the reason why people who lack one or more of the three color-vision systems are called "colorblind.") A sketch of this idea is below.



The system described here instead acts like a spectrometer, and, unlike the human eye, cannot be fooled by such a small number of wavelengths. Instead, it images in 128 wavelengths in the visible spectrum -- and, optionally, another 128 wavelengths in the near-infrared -- then uses sonification to make the resulting histogram of wavelength vs amplitude accessible to its user. It uses sonification, rather than a visual display, to keep the user's visual field uncluttered, and to enable a sort of artificial synesthesia.

The system is wearable. It sits on the side of the user's head, and images a patch of his or her environment about two degrees wide (the same width as the fovea), aimed in the direction that the user's head is pointed. (Full eye-tracking is too intrustive and too cumbersome to justify [6].) A prototype version of the system is pictured below.



Applications

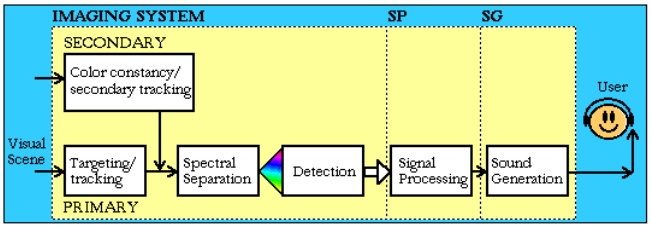
The system's wearability makes it possible to use the system most of the time, and to learn the mapping between materials commonly encountered and their sounds. This in turn may allow the user to identify what things are made of, or whether objects have undergone a change. (For example, imagine looking at a car and saying, "Well, it looks like metal, but sounds like painted plastic," or looking at one's lawn and saying, "Hey, the grass sounds funny today -- is it sick?")

In addition, there are several other applications, such as seeing through camouflage (the jeep which is painted like the jungle is *not* painted with the actual jungle), or scientific applications (certain skin cancers have unusual spectra compared to other types of moles, for example). In general, any time we have come up with different ways of looking at the world, be they different timescales, different media (light vs sound), or different electromagnetic wavelengths, we have learned something. It is hoped that routine use of this device may also reveal something unexpected.

Future extensions of the system to a wider electromagnetic bandwidth (e.g., far-infrared and near-ultraviolet) promise to improve its utility substantially. Designs that incorporate unusual, nonhuman senses -- such as a polarimeter (allowing one to navigate like a honey bee, via the polarization of the blue sky [5]), an RF-field sensor, or a magnetometer -- possess similar promise.

Sonification of the Spectral Data

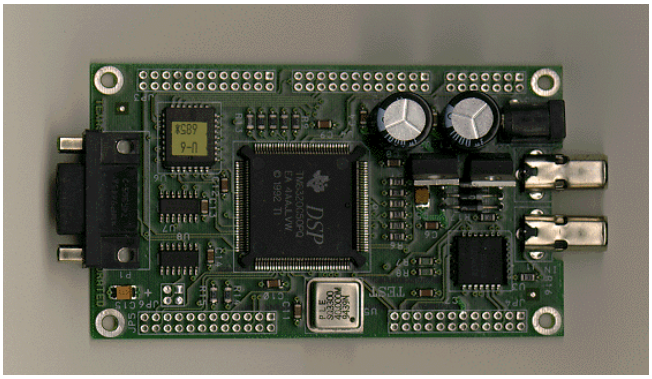
At the moment, the sonification strategy is quite simple: each visual wavelength imaged controls a corresponding audio wavelength. The data is scaled logarithmically to match human sensory characteristics and translated through various calibration tables. Finally, the resulting collection of sine waves at different frequencies and amplitudes is then simply summed and presented to the user via headphones. A highly conceptualized view of this system appears at the very start of this paper, in the diagram above the abstract; a block diagram of the actual system implemented appears below.



It is clear from the psychoacoustics literature [2][3][4], however, that there are many better strategies for presenting the data, and some of these methods will be incorporated into future designs. For example, incorporating timbre into the individual frequencies, and slightly staggering their relative attacks, should make it easier to differentiate various spectra. In addition, because people can often remember *where* a given sound came from better than they can remember the individual pitches that composed it, a more ambitious design, using a more powerful digital signal processor than is present in the current system, would use stereo output channels and would spatialize the resulting audio -- imagine wrapping a piano keyboard around your head. Other possible mechanisms include turning the continuous stream of audio into plucked notes (possibly plucked only when a significant change in visual input is detected), or using actual melodies to express some of the spectral information.

Implementation

The optical input to the system is a tiny spectrometer, consisting of a diffraction grating and a linear array of photodiodes, which detect the separated spectrum's amplitude at each wavelength. From there, the data is converted to digital form and used as the basis of a polyphonic synthesis algorithm, currently implemented via a TI 320C50 DSP, a 28 MIPS, fixed-point unit. (The DSP board, along with its A/D and D/A converters, and other support electronics, is shown at 75% scale below.)



The actual input to the spectrometer is head-mounted, so that it gathers light from a roughly foveal-sized (1-2 degrees of arc) region based on where the user's head is pointed. In addition, the system surrounds the active region with laser spots, which make the imaged area obvious, for use while the user is learning to use the system. Once the user learns what common surfaces sound like, the targeting spots may be turned off -- the user can tell where the device is aimed just by listening to it.

The system is undergoing continuous improvement and redesign. Sonification is a critical part of its construction, and experiments are ongoing in determining the best way to represent the sensory information.

Conclusions

An implemented system has been presented which is designed to be constantly and ubiquitously worn by its user, and can extend the sensory range of even normally sighted individuals by sonifying the complete visual spectra of objects in the visual field. Careful attention to system design yields a device small enough to be convenient, and hence used most of the time. This presents a rich set of sensory data, making creative sonification a challenging and rewarding task.

Many potential avenues for improvement remain to be explored, primary in two areas:

- Extending the sensory capabilities of the instrument, either into nonvisual electromagnetic bandwidths, or by adding different kinds of sensors entirely.
- Improving the sonification of the resulting sensor data.

Work is currently continuing on both these fronts. It is my hope that this could spawn a number of efforts for the routine sensory augmentation of ordinary humans with extraordinary senses, and further work into the social impact of wearable devices of many kinds.

References

[1] Cornsweet, Tom, *Visual Perception*, Harcourt Brace Jovanovich, 1970.

[2] Handel, Stephen, *Listening*, MIT Press, 1989.

[3] Kramer, Gregory, ed., *Auditory Display: Sonification, Audification, and Auditory Interfaces*, Addison-Wesley, 1994.

[4] Wenzel, Elizabeth, "Spatial Sounds and Sonification," *Auditory Display: Sonification, Audification, and Auditory Interfaces*, Gregory Kramer, ed., Addison-Wesley, 1994.

[5] Winston, Mark, *The Biology of the Honey Bee*, Harvard University Press, 1987.

[6] Young, Laurence, and Sheena, David, "Survey of Eye Movement Recording Methods," *Behavior Research Methods & Instrumentation 1975*, Volume 7, number 5, pp. 397-429.