SEE-THROUGH-SOUND

Bridging sensory worlds

Tomás Henriques, Buffalo State College NY, USA, (PI of Project*)
Mick Mengucci; F. Medeiros, Lab-IO Lisboa, Portugal
Sofia Cavaco; Nuno Correia, FCT-UNL Lisboa, Portugal

OBJECTIVE

Create a device that converts color information from live video and still images into sound, mapping color attributes to sound parameters such as pitch, timbre, volume and sound location.

The tool aims specifically at aiding the visually impaired to navigate through and assess the surrounding environment. It can also help anyone to sonically decode visual information, without directly observing or having access to its source.

Key words: Sensory substitution, image analysis, sound synthesis, human computer interface

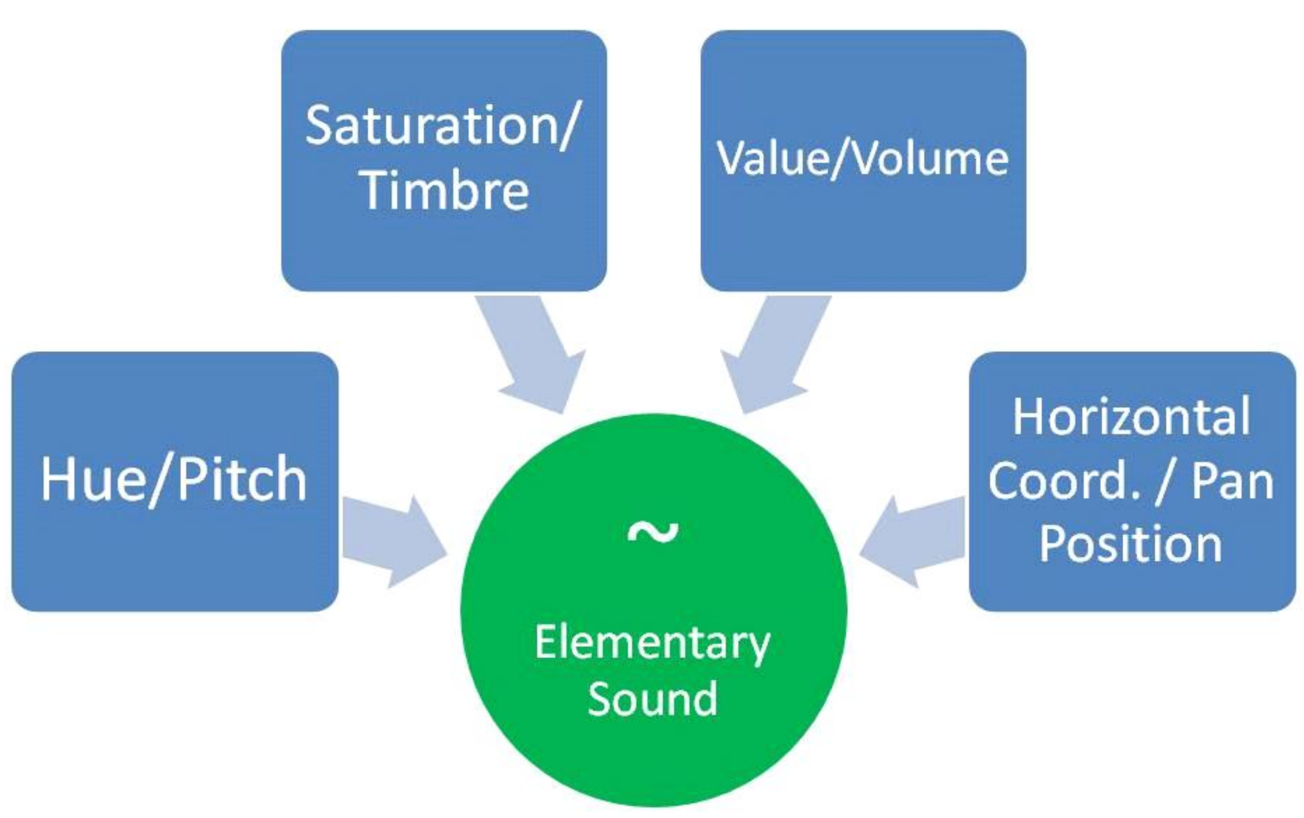

Fig. 1 - Pixels' HSV to sound mapping

APPROACH

A digital image is a composition of discrete elementary units of color/light, its pixels. Our research deals with the conversion of a digital image into its sonic print, where each pixel and pixel properties have an associated sound. Therefore, when a single image or video frame is "played", the result is a composition of elementary sounds originated directly by the pixels' values.

These sound events convey what the camera captures, giving information about the colors that surround the user, the amount of darkness/light, and the location of the light sources. The prototype can process either a still image or a live video feed, weaving a tapestry of sonic particles of varying complexity.

The flow of the synthesized audio depends on the camera's scanning rate, which is user controlled. A high rate works best if fast information about the environment is necessary, as when moving in crowded urban areas; a low rate is more suitable if details of a scene are needed.

•See-Through-Sound is a research project funded by FCT, the Portuguese Foundation for Science and Technology.
•UTA-EXP/MAI/0025/2009

MAPPING COLOR TO SOUND

To convert image into sound we use the Hue, Saturation and Value (HSV) color model where *Hue* contains information about the "pure color" of the pixel. *Saturation,* measures the deviation of the color from gray and *Value* conveys information about the color's brightness.

The HSV attributes of the pixels are mapped into sound parameters, respectively as pitch, timbre and volume. A fourth parameter, the pixel's abscissa, was added to locate the sound within the horizontal plane (fig. 1). The pitch output of an HSV value can be assigned to user defined pitch ranges, from one to four octaves and with different frequency resolutions.

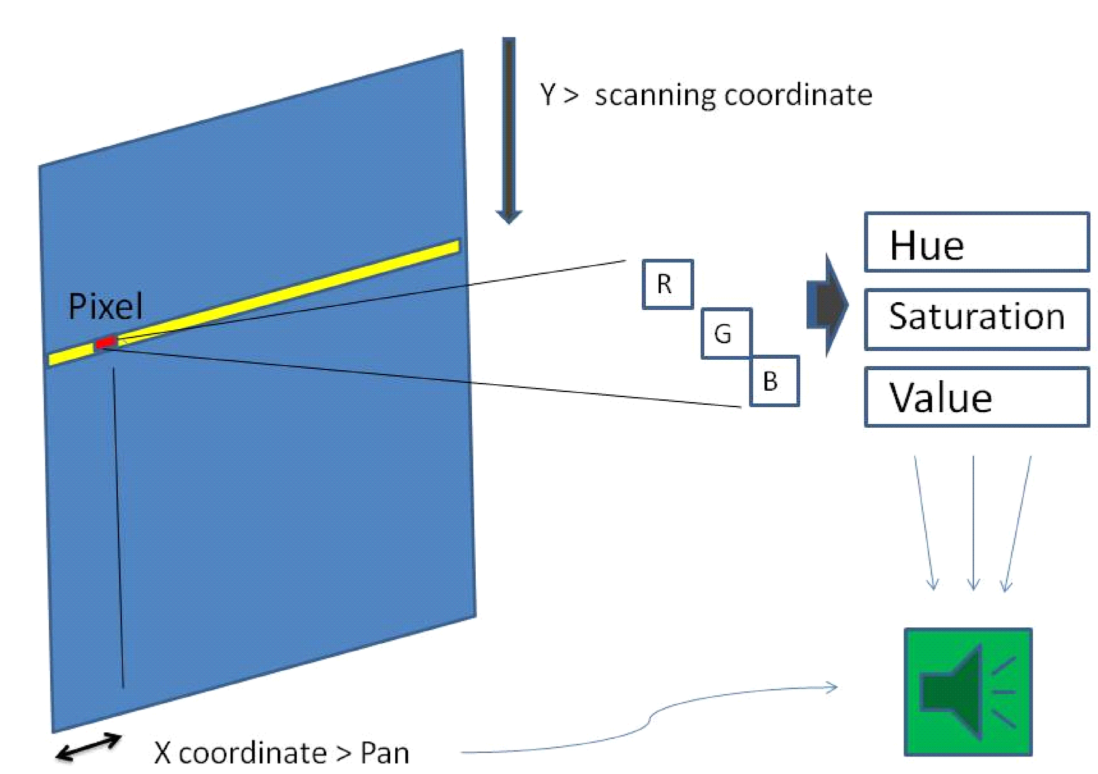

Fig. 2 - Interpolation of the HSV values

The prototype scans the images from top to bottom making an interpolation of the HSV values over 12 segments (columns) along the rows and generates a sound for each of the segments it plays. Images are played as they are being scanned (fig. 2).

PROTOTYPE

The software is implemented in Pure Data running a patch for video processing (using Pd's extension GEM) in tandem with a patch for sound synthesis. The two patches communicate through Open Sound Control. A Playstation PS3 Eye camera was used for testing. It captures standard video with frame rates of 60Hz at a 640x480 pixel resolution.

APPLICATIONS

The prototype enables people with visual disabilities to effectively interact with their environment: from matching clothes to wear, to locating door exits or interpret changing streetlights and signs, etc. Sighted individuals can use it as well to mine information about a scene, near or remote, that can't be directly observed. The audio generated easily conveys if the camera is pointing at a surface or at an object, or if an object has moved within the surveyed area. The prototype can thus be used as an intelligent motion detector, giving out unique details about the surrounding space.