From pixels to pitches: unveiling the world of color for the blind

Sofia Cavaco*, Michele Mengucci*, J. Tomás Henriques†‡, Nuno Correia* and Francisco Medeiros§

* *CITI, Departamento de Informática*
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa
2829-516 Caparica, Portugal

† *CESEM, Faculdade de Ciências Sociais e Humanas*
Universidade Nova de Lisboa
Av. Berna 26-C, 1069-061 Lisboa, Portugal

‡ *Buffalo State College*
1300 Elmwood Ave, Buffalo, NY 14221, USA

§*LabIO*
Rua Nova do Desterro 23 4-Esq., 1100 Lisboa, Portugal

Email: scavaco@fct.unl.pt, mengucci@gmail.com, henriqjt@buffalostate.edu, nmc@fct.unl.pt

***Abstract*—We propose a tool, SonarX, which converts color information from still images or video frames into sound. The tool converts the hue, saturation and value parameters into sound parameters that influence the perception of pitch, timbre and loudness. The goal of SonarX is to help visually impaired individuals to perceive characteristics of the environment that are usually not easily acquired without vision. The tool has been experimented by visually impaired individuals, who confirmed that it can be used to give them information about the range of colors present in the images, presence or absence of light sources as well the location and shape of the objects.**

***Keywords*-Audio visual systems, image color analysis, image representation, image sonification, sensory aids**

I. Introduction

This paper presents and validates a simple and robust digital tool that transforms, in real time, color and light information into sound. The main goal of this tool, which we call SonarX, is to give visually impaired people information about the light and colors that compose the surrounding environment. Nonetheless it also has applications for sighted people. For instance, it can be used by daltonic people to distinguish colors that they cannot distinguish (like traffic lights, or ripe/green fruit). It can also be used in chemistry laboratories to signal chemical reactions that result in color change. It can be further used as a complement to artistic exhibitions (architecture, sculpture, painting, performance) and for therapeutic purposes if the sounds are played in slow mode and result from the mapping of *pleasant* images into sound. The synesthetic effect of looking at a colorful image while listening continuously to the sound it creates can be very relaxing and therapeutic according to the combination of sound and color therapy principles.

A digital image is a composition of pixels, which are its discrete elementary units. SonarX converts digital images into sound by mapping the hue, saturation and value parameters of its pixels into audio parameters. The images can be video frames captured in real time with a camera or still images loaded into SonarX. When a single image or video frame is *played*, the result is a composition of elementary sounds that originate directly from the pixels values. Since these elementary sounds are a mapping from the pixels in the image, they carry information about the colors in the image, the amount of darkness/light, the location of these light sources, and the location and shape of the objects in the image.

When SonarX is used in real time with a camera, it generates a pulsating texture of sonic events. The scanning rate is user controlled and different scanning rates can be used for different purposes. For instance, a high rate can be used when fast information about the environment is necessary (for example, when moving in crowded urban areas) and a low rate can be used to obtain a more detailed description of the surrounding environment.

Other tools that convert image into sound have been developed and reported in the literature. The vOICe, which scans shapes, is a pioneer work in this area and it is quite evolved in terms of quality of software and distribution (there are PC, Iphone and Android applications available) [1], [2]. Nevertheless, quite surprisingly it does not take into account color detection and processing. It only detects and processes shapes segmented in a automatic and rough way. The scan speed is not easily user defined.

Kromophone is a tool based on color blobs detection and maps colors to sound locations [3]. It plays just one color at a time, translating it into three sounds simultaneously positioned in the auditive space. It processes video frames from a webcam and starts by mapping an average area around the center pixel. The red, green and blue (RGB)

 DOI: 10.1109/SeGAH.2013.6665305

parameters of this center blob are converted into sound in the following way: the intensity of red is mapped into a high-pitch noise in the right ear, the green is mapped into a middle-pitch noise in both ears, and the blue into a low-pitch noise in the left ear. The intensity of the white is then separated into distinct high-pitch sounds. The resulting sounds allow distinguishing colors.

Finally, it is also worth mentioning the Prosthesis Substituting Vision for Audition (PSVA) [4] and the Vibe [5]. PSVA translates visual patterns into sound, while the Vibe looks for virtual *sources* in the image and maps them into a mixture of sinusoidal waves.

The proposed tool, SonarX, is described in section II. That section describes how the tool maps information from the pixels into sound. Sections III and IV discuss two pilot studies done with visually impaired people. These studies aimed to evaluate if SonarX's sounds reliably deliver color, location and shape differentiation. In section V we discuss the results and finally section VI presents the conclusions and future work.

II. The tool - SonarX

As mentioned beforehand, we developed a software tool, SonarX, that transforms color information into sound. The tool can process both still images and live videos: when a still image is loaded, the tool converts it into sound and plays it continuously until the user decides to stop. Alternatively the tool can receive the images from a camera and convert the video frames into sound.

In order to convert an image (be it a still image or video frame), the tool processes the pixels of the image. It maps the color Hue, Saturation and Value (HSV) attributes into sound parameters, and as a consequence, for each HSV triple, it produces one sound.

The hue value is mapped into the fundamental frequency of the generated sound (which is related to the sound's perceived pitch). This results in a sound that consists of a pure tone (i.e., a sinusoidal wave). The spectral envelope of this sound can be modified by the other parameters and as a consequence, the final waveform may be different from a pure tone. In addition, there is an inverse correspondence between the sound and color frequencies: when the color's frequency decreases from violet to red, the sound's pitch increases (by increasing its fundamental frequency) [6], [7].

The saturation is used to control the spectral envelope of the sound, which influences the perception of timbre. The shape of the waveform can vary from a sinusoid (which is obtained with the lowest saturation value) to a square wave with energy in the odd frequency partials (which is obtained with the highest saturation value).

The parameter value (ranging from 0 to 1) is mapped into intensity, which is related to the sound's perceived loudness. More specifically, the sound is multiplied by value, which affects its intensity.

Finally, the pixel's abscissa is used to define the location of the sound source, more specifically, its simulated position in the horizontal plane. This is done by mapping the pixel's abscissa into azimuth.

Similarly to what is done by other similar tools [1], [2], SonarX plays the image as it is being scanned and not all at once. The images are scanned from top to bottom with a user defined scanning rate. The rows are played in sequence (from top to bottom) and the sounds corresponding to one row are played simultaneously. As it will be seen later (sections III and IV) in this way the tool not only gives information about the colors present in the environment but also about the shapes and locations of the objects in the scene.

Also, instead of converting the HSV attributes of each pixel into a different sound, SonarX segments the image and produces only one sound for each short image segment. Since we process frames with a resolution of 360×480 pixels, each row is composed by 480 pixels. Playing 480 sounds at the same time would result in a very intricate sound texture, therefore the rows are divided into 12 segments. So for each row, SonarX makes an interpolation of the HSV values in each segment and generates a sound for each of those segments. These 12 sounds are played simultaneously. The segment's center abscissa is mapped into azimuth, which is assigned to one of 12 possible values (one for each segment of pixels).

For a more complete explanation of SonarX's technical details, please refer to [8].

III. Colors

The goal of the prototype is giving the visually impaired a way of knowing which colors compose the surrounding environment. Thus, the best way to validate the prototype is with a user study with visually impaired people. Therefore we devised and run a pilot study to evaluate if the sounds created by the prototype reliably deliver color differentiation. Succinctly, in this pilot study, after a short training period, subjects were presented several sounds and were asked to what colors they corresponded. Below we describe the study in more detail.

A. Stimuli

The visual samples were images that were loaded into SonarX. Each image consisted of only one color. We used the colors: red, orange, yellow, green, blue, purple and violet. After loading the images, the prototype transformed them into sounds as explained in section II. These sounds consisted of the audio samples that were presented to the subjects (the audio sample with the highest pitch corresponded to red and the audio sample with the lowest pitch corresponded to violet). The pitches interval corresponded to one musical octave and varied from 110 Hz to 220 Hz. Each subject was presented 20 sounds (since there were only seven visual samples, and consequently seven audio samples, we repeated each sample two or three times, to make the 20 trials).

B. Protocol

The same samples were used for all subjects but the order of presentation varied: to avoid having presentation order effects, we used four different orderings, and each ordering was used with two subjects. The test was performed on a computer. The subjects were informed that they were first going to hear reference sounds that had an association to specific colors, and that their job was to match new incoming sounds originated from still images, to their color correspondence. They were informed that the test would be done with headphones.

Before the actual test started there was a training stage: the visual samples mentioned in section III-A were loaded and the subjects heard the corresponding audio samples as often as they wished, in any order they wished. At this stage they were told the color of the images. The sounds were heard either through headphones or directly from the computer speakers (even though the formal instructions said that the subjects would use headphones, some subjects preferred not to use them). After a few minutes of training most subjects started their attempt to pick out the colors generated through the image-to-sound conversion. Upon each response they were given immediate feedback about the actual color of the sample. This was done to help improve their skill to perform the test. A few subjects used a different strategy during the training period by singling out and repeating over and over each of the colors from the reference sounds (probably to memorize them), then proceeded to realize the test.

Once the subjects started the actual test, no more feedback on their answers was given. Since this was a forced choice study, subjects were instructed to make their best guess whenever in doubt about the color of the sample. Even though some subjects felt they had to give an answer right away, subjects could take as long as they wished to answer and could hear the sample again before giving the answer.

C. Results

There were 8 visually impaired subjects (users or employees) from Biblioteca Nacional de Portugal, participating in this pilot study. There were 2 women and 6 men, with ages ranging between 26 and 63. Some participants were completely blind, others had a very high degree of blindness and 3 reported that they still could distinguish brightness from darkness or some colors under certain conditions. Most subjects reported having normal hearing but 2 reported that they suffered from (age related) hearing loss. There were 3 participants that reported having advanced knowledge of music, 1 that reported playing the guitar, while the others had only received music education in elementary school.

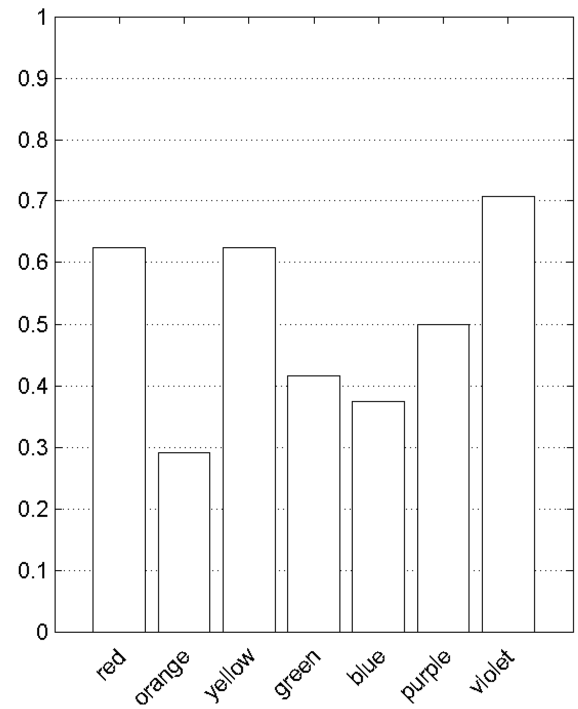


Figure 1. Percentage of correctly classified samples.

1) Colors and color neighborhoods: Fig. 1 shows the results for each color, that is, the ordinate represents the percentage of correctly classified samples for each color. For example, the first bar shows that 63% of the red audio samples (for all subjects) were correctly classified. The results vary from 29% for orange to 71% for violet, with three colors having average results below 50% (orange, green and blue).

While many of these percentages are quite low, if we inspect the confusion matrix (Fig. 2) we can see that when samples from a specific color are wrongly classified, most times they are classified as from a neighboring color, that is, a color with frequency close to the frequency of the sample. For instance, the first row of the confusion matrix shows that apart from one sample that was mistaken for green, the remaining wrongly classified samples were mistaken for orange, which is the closest color to red (in frequency). The same pattern repeats for most colors: violet samples are mostly mistaken for purple, purple samples are mostly mistaken for blue and violet, blue samples are mostly mistaken for purple, etc. Orange is an exception with actually more samples being classified as red than orange and three samples being mistaken for a frequency further away from the original frequency (and corresponding to blue).

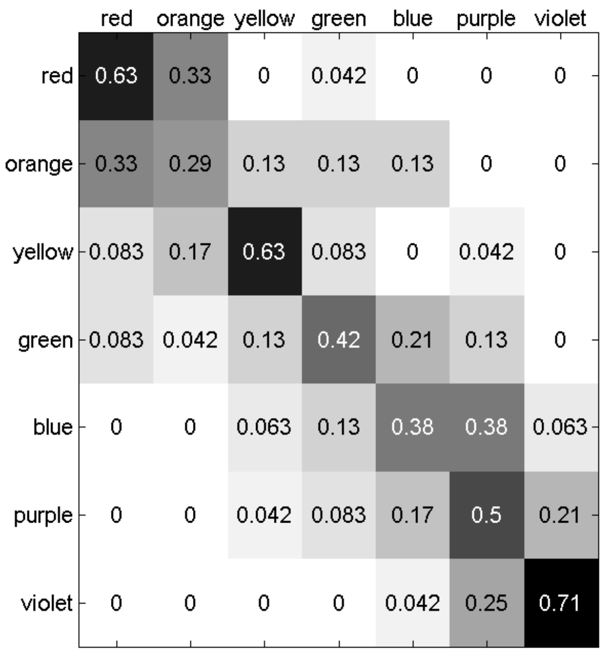


Figure 2. Confusion matrix.

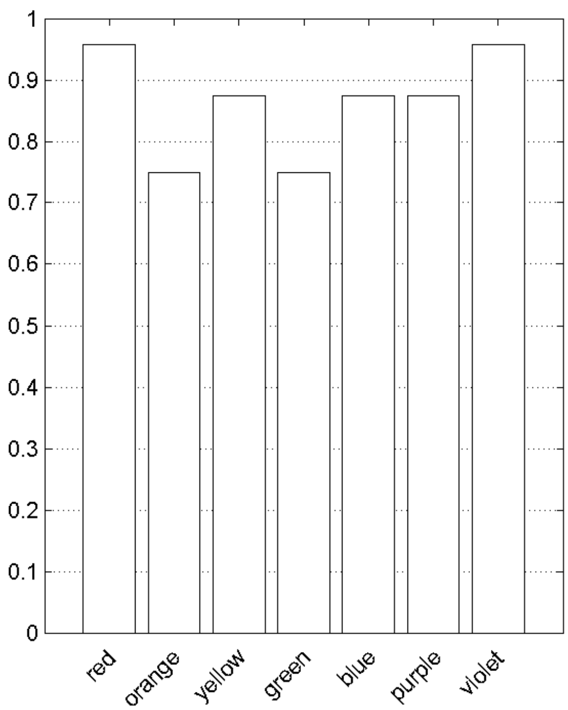


Figure 3. Percentage of correctly classified color neighborhoods.

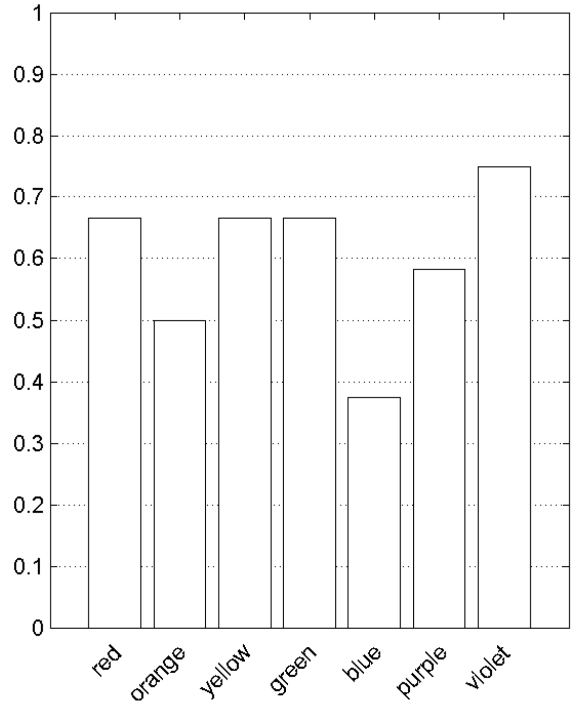


Figure 4. Percentage of correctly classified samples for subjects with music knowledge.

In fact, if we consider neighboring colors, the average classifications are considerably high, varying from 75% for orange and green and 96% for red and violet. Fig. 3 shows the average results for each color neighborhood, where color set X stands for the set of colors that includes color X, the color immediately before and the color imemdiately after X and where the color is one of the seven colors used in the study: $\{red, orange, yellow, green, blue, purple, violet\}$. More specifically, the sets are:
$red = \{red, orange\}$,
$orange = \{red, orange, yellow\}$,
$yellow = \{orange, yellow, green\}$,
$green = \{yellow, green, blue\}$,
$blue = \{green, blue, purple\}$,
$purple = \{blue, purple, violet\}$,
$violet = \{purple, violet\}$.

The high percentages shown in Fig. 3, confirm that if instead of considering a specific color, we consider the color and its immediate neighbors, SonarX's mapping from colors to sound is sufficient to inform visually impaired people about the range of colors being processed by the tool.

2) Music versus no music knowledge: As mentioned above there were 3 subjects that reported having some knowledge of music and 1 that plays the guitar. Fig. 4 shows

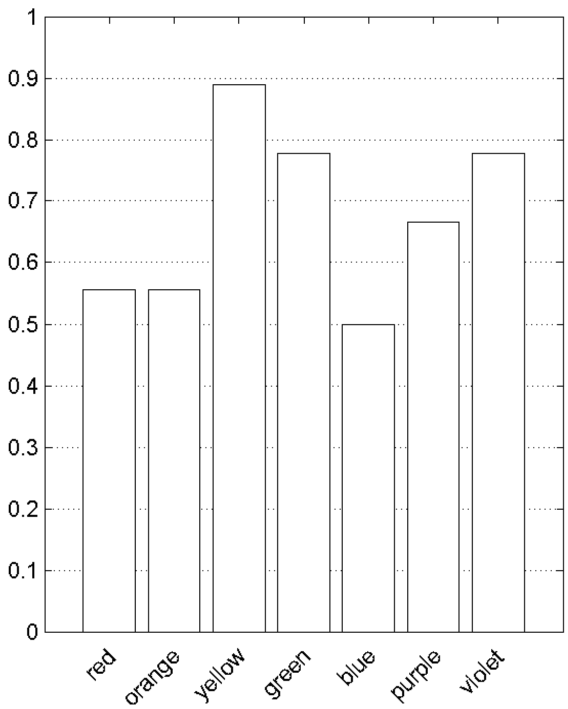


Figure 5. Percentage of correctly classified samples for subjects with music knowledge and no audition problem.

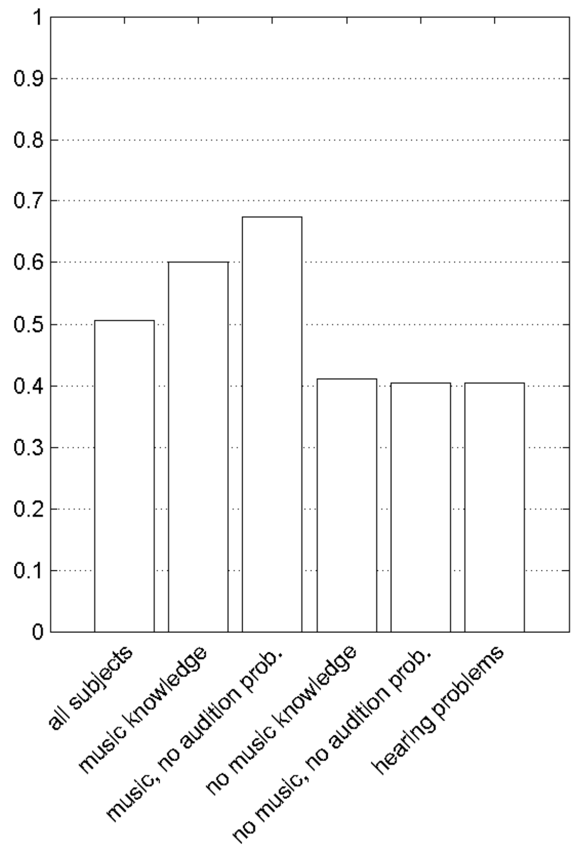


Figure 6. Percentage of correctly classified samples for different groups: all subjects (1st leftmost bar), subjects that reported having music knowledge (2nd bar), subjects with music knowledge and no hearing problems (3rd bar), subjects with no music knowledge (4th bar), subjects with no music knowledge and no hearing problems (5th bar) and subjects with hearing problems (6th bar).

the average classifications obtained for those 4 subjects. As can be observed by comparing figures 1 and 4, the results are better in this group than when we consider all subjects. There is an increase in the number of correctly classified samples, especially for orange and green. Nonetheless, we expected better performance from subjects with music training.

One of these four subjects, also reported having age related hearing problems, which had been confirmed by medical exams. In fact, this subject had the poorest results within this group. When we consider the results from this group without this subject, that is, when we consider the results from the three subjects with music knowledge and no audition problems, we see that the classification rates increase (Fig. 5). They are always greater or equal to 50% and reach 89% for yellow.

Fig. 6 shows the average results obtained with different groups of subjects, that is, the vertical axis represents the average number of times sounds are correctly classified within each group. The groups are: all 8 subjects, subjects with music knowledge, subjects with music knowledge and no audition problems, the 4 subjects with no music knowledge, the 3 subjects with no music knowledge and no hearing problems, and the 2 subjects who reported having hearing problems. One of the subjects with hearing problems had music knowledge and therefore this subject is also included in the second group.

As can be observed the highest percentage of correctly classified samples (67%) is obtained by people with music knowledge and no hearing problems. The lowest results (between 40% and 41%) are obtained by the group with hearing problems, the group with no music knowledge and the group with no music knowledge and no hearing problems. Amazingly, the results from the group with no music knowledge and no hearing problems are very close to those of the group with hearing problems, which suggests the training stage was not enough for people to learn the association between SonarX's sounds and colors.

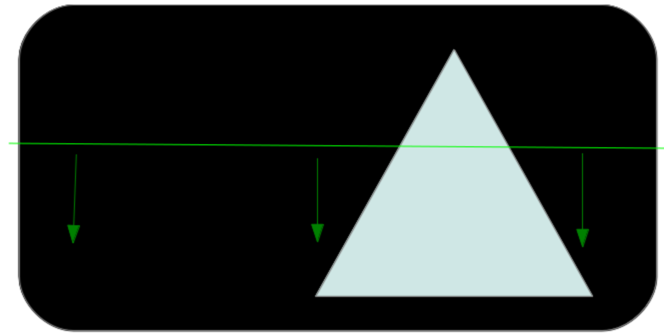
These results show that training matters. People that have music knowledge, and therefore, music training, can more easily distinguish the sounds from SonarX and make an association between the sounds and the colors (or between the sounds and names of colors, for those blind people who have never seen color). The training stage of the experiment was short. We are convinced that with more training and continuous use of the tool, the test's results would improve.

IV. Shapes and locations

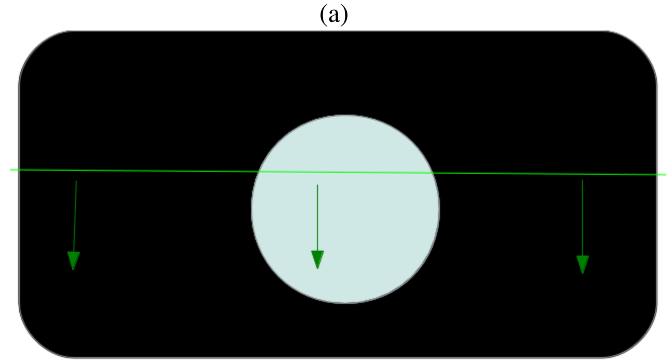
Besides the information about the colors, the sounds of SonarX also provide information about location and shape. As mentioned before, the azimuth of sound events (i.e., the azimuth of the simulated sound sources) is determined by the segments' center abscissae. In order to give the perception that a sound event originates from a specific location, we use interaural phase differences (the same sound is played in the left and right channel but the initial phase is different).

Since the images are scanned from top to bottom and the sounds of the rows are played in sequence (and not all rows at once), the sequence of these sounds gives information about the objects' shapes. The sequence of sounds will be heard as a single sound that remains constant, or with an increasing or decreasing intensity. For instance, a rectangle (with sides parallel to the screen) is characterized by a sound with constant intensity that is played while the rectangle is being scanned (downwards): the sound starts suddenly when the rectangle's top side is scanned, stays constant while the remaining rows of the rectangle are being scanned, and stops immediately after the bottom side has been scanned. On the other hand, a triangle is characterized by a sound with increasing or decreasing intensity (depending on the orientation of the triangle). For example, an isosceles triangle (Fig. 7.a) with the base parallel to the screen's bottom side, results in a sound that starts with just one elementary sound. In other words, assuming the background is not mapped into audible sound, there is just one pixel segment (the segment that contains the top vertex of the triangle) that is mapped into audible sound when the scan line first intersects this triangle. As the rows are being scanned (downwards), more and more pixel segments are mapped into sound. The sounds are all equal but as a result the loudness of the whole final sound increases. Another example is the circle (Fig. 7.b). In this case the sound's intensity continuously increases from the time the scan line first intersects the circle and until it reaches the center of the circle. Then it continuously decreases until it reaches the bottom of the circle.

In order to test these two functionalities, we run an experiment with three visually impaired volunteers, who already had participated on the first pilot study. This second experiment was performed in a more informal way, but for a longer time with each participant. This allowed to test the tool under many aspects and receive more feedback from the participants. This experiment was divided in three parts, to test colors, localization and shapes. Below we describe each of these three tests in more detail. Before the actual experiment started, the participants were allowed to experiment the tool with the camera in their hands.



(a)



(b)

Figure 7. Some images with shapes used in the second experiment. The horizontal line (scan line) and arrows illustrate the scan direction. The intersection of the scan line with the shape determines which pixel segments are *played*, that is, which of the 12 pixel segments are mapped into audible sound. (a) Triangle. The number of segments played increases until the scan line reaches the bottom of the triangle. (b) Circle. The number of segments played increases until the scan line intersects the center of the circle, then this number decreases.

A. Colors

1) Description: While the pilot study described in section III used still images that were loaded into SonarX, in this test we used a video camera. The camera was facing a white wall and no *wall sound* was generated. SonarX has a *gray mode* option that can be turned on when we want to play the sounds corresponding to gray, or off otherwise. Since white is a particular tone of gray, with the gray mode option turned off, no sound for the wall was generated.

In order to facilitate training, SonarX has also the option of playing the sounds of the *reference* colors: the user can press a different key for each color and, as a result, the color is played along with a voice saying the name of the color. After the participants got familiar with the sounds of the reference colors, we placed an orange backpack, a yellow file and a blue book on a table (that was next to the white wall), and a red chair next to the table. These objects were used one at a time and in random groups. If they wished, the participants could hold the camera to better explore the locations of the objects or focus on a particular object.

2) Results: The participants could detect very easily the presence of objects when the camera was still and facing the wall. Sometimes they could not say the exact number of objects. They were able to successfully recognize the color of the objects, and when they failed, they picked a color in the neighborhood of the object's color (similarly to what happened in the first pilot study, see section III-C1). We observed that, the reference sounds were louder than the sounds that resulted from the analysis of real images (loaded photographs or video frames). This is due to the level of saturation of the reference colors, which is higher than the colors observed in the real images from this test. This can lead the users to fail the association between colors and sounds of those colors. For example, even though the reference red was very similar to the color of the chair, since the first was very loud and the second was soft, the participants were not sure about the color and reported it was "something between red and orange".

When the subjects tried to hold the camera with their hands, the results were very poor. They were not able to control the camera appropriately to give them the information they were looking for (for instance, they were not able to point the camera directly to the objects). Also they moved the camera very quickly. They did not wait for the sound of a frame to finish before moving to another direction, which resulted in confusing sounds. This shows that, while the tool can successfully be used when the camera is still, holding the camera can result in confusing sounds.

B. Localization

1) Description: This test focused on the location of sound events. We wanted to understand if the sounds generated provide information about the location of the objects in the images.

This test was done with still images loaded into SonarX. The visual samples consisted of a white filled circle on a black background. There were nine samples, each with the white circle at a different position: top left, top center, top right, center left, center of the image, center right, bottom left, bottom center, bottom right. There was no sound produced for the background: gray mode was on, but since the black value (from the HSV attributes) is zero, the sound that results from black has zero intensity and therefore is not audible.

As mentioned above, the azimuth of a sound event depends on the center abscissa of the pixels segment. That angle is used to determine the phase difference between the right and left channel signals. This phase difference gives the perception of (horizontal) location.

In order to give information about elevation, we use a click sound: when this sound is played, it means that the tool is starting to process the image. It is possible to infer the elevation of an object by paying attention to the time lap between the click and the sound event that corresponds to the object. The sound of an object at the top is played immediately after the click, while the sound of an object at the bottom is played immediately before the click (the image is scanned and processed continuously, that is, the tool scans and processes the image from top to bottom, and when it reaches the bottom of the image, it starts from the top again). The sound of an object in the center has approximately the same time lap after and before the click.

2) Results: All nine tested samples obtained 100% recognition rate (over all subjects). This success rate was obtained after just a few minutes of training. Note that each user chose the scan speed that suited him/her better. The results from this test show that SonarX's sounds can be used to successfully localize the objects in the images.

C. Shapes

1) Description: As mentioned before, the sounds produced by SonarX give information about the shape of the objects. This test investigated if the sounds give enough information to determine the exact shape of objects.

This test was also done with still images loaded into SonarX. Each image contained a white filled shape on a black background (Fig. 7). There was no sound played for the background. The shapes consisted of rectangles (including squares), triangles and circles. Before the test started, we explained how the sounds of different shapes are and illustrated with the audio samples that resulted from loading the shapes images into SonarX. The explanation had to be complemented with paper models of the shapes and making the participants feel with their hands the intersection between the scan line (the palm) and the shape. The participants were allowed to control the scanning velocity.

2) Results: This test also had 100% of recognition. After presenting the three shapes described above, we also used an extra sample with a star shape. The subjects were able to say it was none of the previously recognized shapes.

V. Discussion

The high percentages obtained for red and violet suggest that it is probably easier to distinguish the sounds at the extremes of the spectrum. This also suggests that, as mentioned by a couple the participants in the study, having a wider pitch interval between the different sounds, for instance by using pitches from more than one musical octaves, would be beneficial. One subject even mentioned that it would be preferable to have only two sounds per octave, which would imply using eight musical octaves. In response to this, we already modified SonarX to include more options related to the pitches interval.

One subject also referred that having an inverse association of colors to pitches would be more intuitive. In other words, this subject would prefer the correspondence of warm colors, namely red, to the lower pitches. Although this was only suggested by one subject, it would be interesting to confirm if inverting the current association of pitches to colors would result in higher classification rates.

As mentioned before, the amount of training is important. This was confirmed by some subjects that referred that they felt that with more training it would be easier to associate sounds to colors.

Most subjects reported that they used no conscious cues to recognize the sounds during the test. Only two subjects conscientiously used cues: one reported that he memorized the sound of red and the color order, while the other memorized the sound of green and then tried to guess if the trial sound had a higher or lower pitch than green. Curiously these were the only two subjects who did not remember or had never seen colors (one was born blind and the other got blind while still a baby). For these two subjects, colors are only a theoretic concept. The results of these two subjects together were higher than the remaining subjects: 60% against 48%. It is possible that since they do not remember what colors are, they felt the need to use a memorization technique, which in the end was more effective than the techniques used by the remaining subjects.

We believe that, at least for people who remember colors (people who were born with sight and lost their vision subsequently) the order of the colors is important, that is, it is important to have colors with neighbor frequencies to map to sounds with neighbor pitches. Even though we have not confirmed it, we believe that having a random association between colors and pitches would result in having lower accuracies, especially if we consider color neighborhoods and people who were not born blind.

VI. Conclusions and Future Work

We have presented a tool that converts color information, from the HSV parameters, into sound. There is a mapping between the HSV parameters and fundamental frequency, spectral envelope and intensity, which in turn give the perception of pitch, timbre and loudness. The location of the sound events depends on the location of the pixels. In addition we discussed two pilot studies that were performed with visually impaired subjects in order to validate the tool.

These tests showed that SonarX's sounds are sufficiently informative about color neighborhoods. That is when we consider color neighborhoods, such as $\{purple, violet\}$, instead of specific colors, the average classifications are high. These sounds are also sufficiently informative about simple shapes and location (azimuth and elevation).

As future work we plan to add other image processing techniques to the tool. These include edge detection, face detection and object tracking. The output of these processes will be mapped to more complex sounds as well as rhythmic patterns. While the tool presently runs on a small computer connected to a camera, we are working on a mobile application to run on a smartphone that will allow greater mobility for the user.

Acknowledgments

This research is made possible through the support of the Portuguese Foundation for Science and Technology (grant UTA-Exp/MAI/0025/2009) and the UT Austin Portugal Program in Digital Media.

The authors wish to thank all participants in the two pilot studies. Many thanks also to Carlos Ferreira (Biblioteca Nacional de Portugal) for his collaboration and help provided setting up the user studies.

References

[1] M. Auvray, S. Hanneton, and J.K. O'Regan, "Learning to perceive with a visuo-auditory substitution system: Localisation and object recognition with the vOICe," in *Perception*, 2007, vol. 36(3), pp. 416–430.

[2] C. Capelle, C. Trullemans, P. Arno, and C. Veraart, "A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution," in *IEEE Transactions Biomedical Engineering*, 1988, vol. 45, pp. 1279–1293.

[3] Z. Capalbo and B. Glenney, "Hearing color: Radical pluralistic realism and ssds," in *Proceedings of AP-CAP*, 2009, pp. 135–140.

[4] P. Bach-y-Rita and S.W. Kercel, "Sensory substitution and the human-machine interface," in *Trends in Cognitive Neuroscience*, 2003, vol. 7(12), pp. 541–546.

[5] M. Auvray, S. Hanneton., C. Lenay, and K. O'Regan, "There is something out there: distal attribution in sensory substitution, twenty years later," in *Journal of Integrative Neuroscience*, 2005, vol. 4, pp. 505–521.

[6] D.L. Datteri and J.N. Howard, "The sound of color," in *Proceedings of the International Conference on Music Perception and Cognition (ICMPC)*, 2008, pp. 767–771.

[7] I.C. Firth, "On the linkage of musical keys to colours," in *Speculations in Science and Technology*, 1981, vol. 4, pp. 501–508.

[8] M. Mengucci, J.T. Henriques, S. Cavaco, N. Correia, and F. Medeiros, "From color to sound: Assessing the surrounding environment," in *Proceedings of ARTECH 2012 - Conference on Digital Arts and New Media*, 2012, pp. 345–348.