PERCEIVING GRAPHICAL AND PICTORIAL INFORMATION VIA TOUCH AND HEARING

Pubudu Madhawa Silva, Thrasyvoulos N. Pappas

EECS Department, Northwestern University
Evanston, IL 60208, USA
{pubudu.silva, pappas}@eecs.northwestern.edu

Joshua Atkins, James E. West

ECE Department, Johns Hopkins University
Baltimore, MD 21218, USA
{joshatkins, jimwest}@jhu.edu

ABSTRACT

With plain text as the dominant medium of communication, Braille and text-to-speech have been quite successful in keeping the visually impaired community up to speed with rest of the world. However, with the ever increasing availability of the Internet and electronic media rich in graphical and pictorial information (for communication, commerce, entertainment, art, education), it has been hard for the visually impaired community to keep up. We propose a non-invasive system that can be used to convey graphical and pictorial information via touch and hearing. The main idea is that the user actively explores a two-dimensional layout (consisting of one or more objects) on a touch screen with the finger while listening to auditory feedback. We demonstrated the efficacy of the proposed approach in a range of tasks, from basic shape identification to perceiving a scene with several objects. The proposed approach is also expected to contribute to research in virtual reality, immersive environments, and medicine.

Index Terms— Auditory-tactile display, user interface, virtual cane

1. INTRODUCTION

With the ever increasing availability of the Internet and electronic media rich in graphical and pictorial information – for communication, commerce, entertainment, art, and education – it has been hard for the visually impaired (VI) community to keep up. This paper explores the use of two other senses, hearing and touch, to convey visual information to the VI.

The use of one or more functioning senses to convey information in another sense is defined as *sensory substitution (SS)*. There are two main types of SS: invasive methods and non-invasive methods. Invasive methods generally require surgery, e.g., sensory prosthesis. The cortical or retinal electrode matrix display is a popular invasive approach for visual substitution [1], while Braille is a non-invasive approach. Another SS approach that has proven to be quite effective in providing visual information and assisting visually impaired people with certain visual tasks is the use of a tongue display [2]. It consists of an array of electrodes that can apply different voltages to stimulate the tongue, which is the most sensitive tactile organ and has the highest spatial resolution. However, the majority of visually impaired people find such presentations – as well as the presentation of electrical and other tactile stimuli on other parts of the body (back, abdomen) – quite invasive, and prefer to scan/explore with the finger [3]. The focus of this paper is on this latter type of non-invasive methods for visual substitution (VS), using the finger. The main idea is that the user actively explores a two-dimensional layout consisting of one or more objects on a touch screen with the finger while listening to auditory feedback. In addition to their utility for the VI community, the proposed VS methods are expected to be of use in situations where vision cannot be used, e.g., for GPS navigation while driving, fire-fighter operations in thick smoke, and military missions conducted under the cover of darkness.

Out of the five senses, vision has the highest bandwidth followed by hearing, touch, taste, and smell. Gustatory (taste) and olfactory (smell) sensors suffer from remarkably slow recovery times and are also more prone to adaptation than others, making it hard to utilize them in VS. This leaves three alternatives for VS: solely by touch, solely by hearing, and by both touch and hearing.

The simplest navigational aid based on touch and hearing would be a long cane, which is used by the majority of VI community. It was shown that VI can acquire spatial abilities by using maps, which can also be used as a navigational aid, e.g., to plan the route before to walking [4]. Jacobson implemented an audio enabled map in a touch pad, which uses voice and natural sounds [5]. NOMAD (1988) [1], "Talking Tactile Maps" (1994) [1], and "Talking Tactile Tablet" [6] are tactile maps that play back an auditory label depending on the position touched. However, these systems are not well-suited to interactive applications. Parente *et al.* [7] have developed an audio-haptic map using spatial sounds in 3D. Several auditory counterparts of GUIs such as, Soundtrack by Edwards (1989), Karshmer and Oliver's system (1993), GUIB by Savidis and Stephanidis (1995), and Mercator Project by Mynatt (1997), are also available [5]. Meijer's imaging system named "vOICe," maps a 64x64 image with 16 gray levels to a sequence of tones [8, 9]. Another imaging system called "soundview" was developed by Doel [10], where the user explores a color image loaded to a tablet with a pointer; the color

of each pixel color is mapped to a sound in the tablet. Subjective experiments were conducted to measure the ability of "soundview" and "vOICe" [11] are discussed in Section 4.

The paper is organized as follows. The proposed approach is presented in Section 2 and the subjective experiments in Section 3. Experimental results are discussed in Section 4.

2. PROPOSED APPROACH AND IMPLEMENTATIONS

This paper proposes a new approach for conveying pictorial and graphical information (graphs, diagrams, charts, maps, photos, or video) via acoustic signals as the user actively scans a touch screen with the finger. The touch screen is partitioned into invisible regions, each with a particular sound field. Each region represents an object, part of an object, background, or other element of a visual scene. As the user scans the screen with the finger, she/he is listening to auditory feedback, played back via stereo headphones, corresponding to the finger's location on the touch screen. Our goal is to enable the user to build up a mental picture of a 2-D (or even 3-D) scene or environment by actively exploring the acoustic scene through the use of touch.

The concept was implemented in several configurations, using an *Apple iPad* touch screen. The identification of objects and their geometrical shapes is central and basic for almost all VS tasks such as perceiving maps, sketches, graphs, and images, as well as navigation in a real or virtual environment. It is thus important to find an efficient, intuitive, and practical algorithm for rendering geometrical shapes, using the proposed approach. Towards this goal, we tried four different configurations, which we describe below. These configurations assume one object at a time on the touch screen. However, the same algorithms, with minor adjustments, can be used to represent multiple objects in the screen. For the sake of simplicity, in the following discussions, we will assume that only one object is presented at a time.

2.1. Configuration 1: Object Shape Identification with Two Constant Sounds

There are two basic ways to represent a shape, visually or otherwise: as a line drawing or as a solid shape (filled interior). Thompson *et al.* [12] found that solidly represented objects (interior filled with embossed tactile textures) are easier to recognize than raised line drawings. Hence, we selected the solid shape representation approach. The touch screen is divided into two regions; one inside the object and one outside. One sound is played when the subject's scanning finger is inside the object and the other when finger is outside.

2.2. Configuration 2: Object Shape Identification with Three Constant Sounds

In this configuration, in addition to the solid shape representation, we include a narrow strip around the border of the objects, so that subjects can trace the object edges. Thus, the screen has three segments: the inner segment that represents the object, a narrow strip around the border, and the outer segment that represents the background. The strip is 50 pixels wide on the *iPad* touch screen, which has a resolution of 132 pixels per inch. One of three sounds is uniquely mapped to each segment. The sound played back at any moment, is determined by the position of the scanning finger.

2.3. Configuration 3: Object Identification with Tremolo

In order to facilitate the tracing of the object, we used a varying tremolo signal to convey proximity information in the acoustic feedback. The idea is that when the subject is approaching the border between background and object, she/he can get a sense of whether the finger is moving in the desired direction. Tremolo is a sound effect that is popular among musicians and can be described theoretically as a form of a low-frequency amplitude modulation. We use tremolos with two depth values, one inside and one outside the object. There is also a border region (strip) within each segment (object and background). The tremolo rate is constant within each segment, except when the finger enters the border region, where it increases as the finger approaches the border. In other words, when subject's finger is inside the border strip, they get a sense of how far the finger is to the middle line of the strip by the rate of the tremolo.

2.4. Configuration 4: Object Identification with HRTF

In this configuration, directionality is introduced to the border and outside sounds with Head Related Transfer Function (HRTF) of the user to guide the scanning finger. The sound is now played back via stereo head phones and it is tailored with the best match with general HRTFs by calibrating the subject with known directional sound. The touch screen is divided into three segments (object, background, and border region), each with its own unique sound. When the finger is inside the object, the sound is constant. When the finger scans the background (outside) segment, a 2D virtual acoustic scene is formed. In the 2D plane of the touch screen, the virtual listener (the subject) is assumed to be in the position of the scanning finger, facing north. The sound source (which emits the assigned unique sound for the object) is assumed to be located inside the object at the point nearest to the finger. To render this virtual acoustic scene, we used a special KEMAR Head Related Impulse Response (HRIR) signal from the CIPIC database [13]. To implement the sound directionality, the plane of the touch screen, with the virtual listener at the center was uniformly divided in to 24 pie slices of 15^o each. The source position with respect to the listener was then determined and the source was assigned to one of the 24 pie slices. Then the sound wave was convolved with the corresponding HRIR. The resultant wave was played back via stereo headphones; the volume of the playback was inversely proportional to the distance between the listener and the source. When scanning the border segment, the same as-

sumptions held for the virtual listener as in the background segment, but the sound source was placed in the direction that the user needs to follow in order to keep tracking the border clockwise. In other words, when the user is inside the border segment she/he should move the finger in the direction from which the sound is coming in order to continue tracing the border.

2.5. Configuration 5: Scene Perception by Virtual Cane

A number of disjoint objects are placed on the touch screen and a prerecorded tapping sound (of different materials) is assigned to each object. When the user's finger is inside an object, the tapping sound assigned to that object is played back; when the finger is in the background region, there is no sound. This can be thought of as a blind person exploring a scene, e.g., an outdoor scene outside her/his window or on the opposite side of the street, using a virtual cane – a very long cane in case of an outdoor scene – to tap on the objects. Information about the relative position of each object, the material it is made of, and some idea about its shape and size can be conveyed to the user in this configuration.

3. SUBJECTIVE EXPERIMENTS

Ten subjects took part in a series of experiments, in which they interacted with a touch screen (*Apple iPad*) and listened to auditory feedback. The average age of subjects was 31, ranging from 19 to 50; all reported normal or corrected vision and normal hearing. To prevent visual contact with the touch screen and the scanning finger, the screen was placed in a small box open in the front, so that the subject can put her/his hand inside to access the screen. The subject was seated in front of a table on which the box and the touch screen was placed, and was listening to sounds played back via stereo headphones (SENNHEISER HD595). The experiments were performed in a reasonably quiet room to avoid disturbances.

Subjective experiments were conducted for all five configurations. Before the beginning of the trials for a given configuration, the subject was given a written introduction about the experiment and a chance to ask questions. To familiarize with the system, the subject was first shown a training example, during which, the subject was able to see both the scanning finger and the shape on the touch screen. The subject was also asked to explore the training example under the box, in order to get used to the experimental procedure. There was no tight time limit for each experiment, but the actual time durations was recorded.

Each of first four configurations was tested with three shapes, a square, circle, and equilateral triangle. Each shape was centered in the touch screen and had approximately the same area in square pixels. The subjects didn't have any prior knowledge about the shapes they were going to be tested on, and the ground truth was not revealed until the end of the experiments. The sequence of the trials was randomized, both among configurations and subjects. The subjects were told that the shape they are going to be tested on in any given trial could be the same as that in a previous trial or a new one all together. At the end of each trial, the subject was first asked to draw the shape and then to name it. Subjects were then asked for comments.

The fifth configuration was tested with a scene consisting of three objects with a tapping sound of wood, glass, and metal assigned to each. At the completion of the experiment, the subjects were asked to write down the number of objects in the scene, to identify the material of each object, and to indicate their relative positions.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The results of the subjective experiments are presented in Table 1. The overall accuracy (averaged over all shapes, configurations and subjects) among all trial results was 74.2%. The percent accuracy for the three shapes in the four configurations, is significantly greater than what would be achieved by mere guessing. The results are strengthened by the fact that the subjects didn't have any prior knowledge about the shapes they were going to be tested on.

The average accuracies for each shape (across configurations and subjects for each shape) were: square 87.5%, circle 50.0%, triangle 85.0%. These figures clearly show that the subjects had more difficulty identifying the circle over the other two shapes. This indicates that the detection of curved edges is more difficult than that of straight edges. However, we should point out that the training example always had the same shape (cross) with only straight edges, which may have created the expectation of shapes with straight edges. In fact, when they were asked to draw the shape they experienced, two out of ten subjects attempted to approximate the circle with straight lines. Note that a 10% increase in accuracy for the second configuration over the first, justifies the addition of the narrow strip with distinct sound around the border. Since tracing the edge (of the shape) is easier in the second configuration than the first, the increase in performance may also be used to infer that the subjects preferred tracing the edge in identifying the shape. On the other hand, the addition of proximity feedback via tremolo didn't work out as expected, as can be seen by the drop in performance for Configuration 3, compared to Configuration 2. Perhaps the proximity information inside a relatively narrow border strip (50 pixels wide) was not helpful in carrying out the task. A more likely explanation, however, is that the tremolo signal is not optimal for this task. Indeed, some of the subject's comments, such as "inside/outside of shapes were not differentiable by assigned tremolos," "tremolo rate changes very fast within a small area," "tremolo rate changes were not noticeable," favor the latter argument. Finally, the superior performance of the fourth configuration, is due to the addition of spatial sounds. Yet, as some comments reveal, the addition of spatial sounds to the background was not of much use and it was the spatiality of the boarder strip sound which helped them. Since

	Configuration 1			Configuration 2			Configuration 3			Configuration 4		
	Square	Circle	Triangle	Square	Circle	Triangle	Square	Circle	Triangle	Square	Circle	Triangle
Accuracy	90 %	30 %	80 %	80 %	70 %	80 %	80 %	40 %	90 %	100 %	60 %	90 %
(overall)	66.7 %			76.7 %			70.0 %			83.3 %		

Table 1. Subjective Results

the shapes were always centered and occupied much of the screen, locating the shape in background (which was the intention behind adding spatiality to background sounds), might not be a challenging task. Spatial sounds in the border segment, on the other hand, are quite useful in guiding the finger in edge tracing and also provide clues about edge orientation. Having directionality as a guidance in tracing the edge, might have relieved the subject from the task of exploring and allowed her/him to focus more on identification, as Wijntjes *et al.* [14] explained.

In Configuration 5, the accuracy of detecting the number of shapes in the scene was 100%. The subjects were able to locate the wooden object with 90% accuracy, the glass object with 80%, and metal object with 70%. Glass was confused with metal 10% of the time, and vice-versa 20% of the time as the sounds we used for the two were not easy to distinguish. In the future, we plan to use more distinguishable sounds, even if not as realistic.

We now compare our results with those reported in [11]. In "soundview" they used two sounds, one inside the shape and one in the background, as in Configuration 1. However, the sound played to the subject at a given time depended on both the location and the velocity of the pointer. In addition, they used six shapes (square, circle, and triangle, with and without a hole in the middle). In contrast to our experiment, they allowed subjects to have visual contact with the tablet and scanning pointer, thus using vision in shape identification – which is unrealistic for VI subjects. They used three different experimental procedures. In the first, the subjects didn't know the shapes they are going to be tested, and had to draw the shape after each trial. In the second, the participants were asked to chose the shape they perceived among 18 shapes. In the third, they had to pick one among the six possible shapes. The overall accuracy for the three experiments was 30.0%, 38.3% and 66.2%, respectively. They also tested "vOICe" with the third experimental procedure and got overall accuracy of 31.0%. With the exception of the number of shapes, we may say that their experiments were comparable or easier than ours. However, our results are clearly better.

In conclusion, we have proposed a new appoach for conveying graphical and pictorial information without utilizing vision, and proved its applicability in perceiving basic geometric shapes, significantly outperforming existing approaches. We are currently conducting experiments to apply the proposed approach to navigation, map perception, and imaging. We have also shown that a basic scene can be perceived and objects can be located, identified and distinguished using a "virtual cane." To further explore the shape of a selected object in finer resolution, we are exploring a zoomed-in mode that can be triggered by double-tapping inside the object. In the future, we will also consider about integrating GPS, accelerometer, camera and GIS enabled maps, with the presented approach.

5. REFERENCES

[1] P.B.L. Meijer, "Sensory substitution - vision substitution," http://www.seeingwithsound.com/sensub.htm, Oct. 2010.

[2] P. Bach-y-Rita, K. A. Kaczmarek, "Tongue placed tactile output device," US Patent # 6430450, Aug. 2002.

[3] K. Gourgey, "Personal communication," .

[4] S. Ungar, S. Blades, and C. Spencer, "The role of tactile maps in mobility training," *British J. Visual Impairment*, vol. 11, pp. 59–61, July 1993.

[5] D.R. Jacobson, "Navigating maps with little or no sight: Anovel audio-tactile approach," *Content Visualization and Intermedia Representations*, 1998.

[6] S. Landau, L. Wells, "Merging tactile sensory input and audio data by means of the talking tactile tablet," *EuroHaptics*. IEEE Computer Soc., 2003, pp. 414–418.

[7] P. Parente, G. Bishop, "Bats: The blind audio tactile mapping system," *Proc. ACM South Eastern Conf.*, 2003.

[8] P.B.L. Meijer, "An experimental system for auditory image representations," *IEEE Tr. Biomed. Eng.*, vol. 39, no. 2, pp. 112 –121, Feb. 1992.

[9] P.B.L Meijer, "See with your ears! - the voice," http://www.seeingwithsound.com, Oct. 2010.

[10] K. V. D. Doel, "Soundview: Sensing color images by kinesthetic audio," *Int. Conf. Auditory Display*. IEEE, 2003, pp. 303–306.

[11] K. V. D. Doel, *et al.*, "Geometric shape detection with soundview," *Int. Conf. Auditory Display*, 2004.

[12] L.J. Thompson, *et al.*, "The role of pictorial convention in haptic picture perception," *Perception*, vol. 32, pp. 887–893, 2003.

[13] V.R. Algazi, *et al.*, "The CIPIC HRTF database," 2001, pp. 99 – 102.

[14] M.W.A Wijntjes, *et al.*, "Look what I have felt: Unidentified haptic line drawings are identified after sketching," *Acta Psychologica*, vol. 128, no. 2, pp. 255 – 263, 2008.