Real-time Shape-based Sensory Substitution for Object Localization and Recognition

Hossein Pourghaemi, Torkan Gholamalizadeh, Ahmad Mhaish, Gökhan İnce and Damien Jade Duff
Faculty of Computer & Informatics Engineering
Istanbul Technical University, Turkey
Email: {anbardan16, torkan15, mhaish, gokhan.ince, djduff}@itu.edu.tr

***Abstract*—In this paper, we present a new approach to real-time tracking and sonification of 3D object shapes and test the ability of blindfolded participants to learn to locate and recognize objects using our system in a controlled physical environment. In our sonification and sensory substitution system, a depth camera accesses the 3D structure of objects in the form of point clouds and objects are presented to users as spatial audio in real time. We introduce a novel object tracking scheme, which allows the system to be used in the wild, and a method for sonification of objects which encodes the internal 3D contour of objects. We compare the new sonfication method with our previous object-outline based approach. We use an ABA/BAB experimental protocol variant to test the effect of learning during training and testing and to control for order effects with a small group of participants. Results show that our system allows object recognition and localization with short learning time and similar performance between the two sonification methods.**

***Keywords*–sensory substitution; sensory augmentation; point clouds; depth camera; sonification; object tracking.**

I. Introduction

Sensory substitution systems convert information from one sensory modality into another. They are potentially of great help for sense-impaired individuals, for example allowing individuals to locate and grasp objects [1], to navigate [2] or to appreciate visual patterns through sound [2][3]. They can also be used to answer questions about human perception [1].

We present a visual to auditory sensory substitution system, and we are concerned in particular with navigation, object recognition and object manipulation tasks. We would like users to be able to use the system to navigate with environmental awareness in unstructured environments, interpret novel objects from a distance, eat meals, and so forth.

The challenge in building such systems is in presenting sensory data to end users in a comfortable, learnable and understandable way. Until now, these systems have not reached a wide enough audience to supplant more basic but non-intrusive systems like the white cane or guide dog. We attempt to improve on these systems by using time-of-light and structured-light sensors which shortcut the problem of getting the depth of objects in a scene, and work by first extracting perceptual objects and spaces and then and generating meaningful sounds to allow their localization and recognition.

As a first step, we enabled users to localize single objects on a table top using simple spatial audio and tones to sonify direction and distance of objects [4]; next, we investigated different approaches to sonification of simulated 3D shapes [5]. An open research question is whether such sonification methods can be used in the real world; as such, in the current paper, instead of only sonifying simulated shapes, we test a new version of our system that enables localization and recognition of objects on the floor in an empty room. We allow

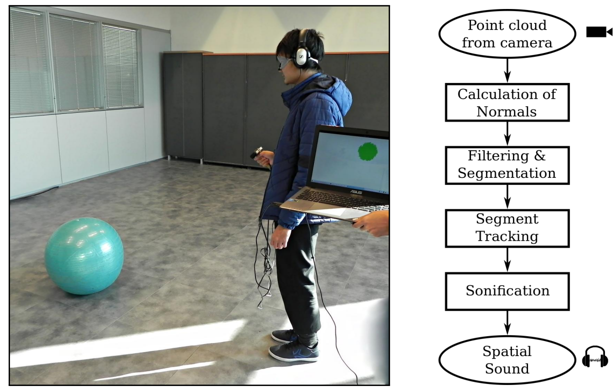

Figure 1. Left: Our system in use. Right: The data-flow pipeline of our sensory substitution system. Frames from the depth sensor are received at approximately $30Hz$ and are processed into sound.

users to move around freely and interact with objects actively using our system. The system in use during testing can be seen on the left of Figure 1.

In order to enable more sophisticated, longer lasting sounds to allow better discrimination of object parts, we build on previously developed object/part segmentation and background noise reduction techniques [4], implementing a scheme for consistently tracking object parts from frame to frame. We also realize a proto-object concept based on segmenting, tracking and sonifying multiple parts of objects separately so that users can understand the shape of objects as the assembly of their sonified parts. Finally, we improve the sonification approach to sonify not just the external contours (outlines) of objects, as previously done [5], but to completely sonify all internal contours of the object using a direct encoding of the full visible 3D shape of the object as sound waves. A summary of the flow of data for each frame in our real-time system can be seen on the right of Figure 1. We evaluated the performance of our system as it was being learnt and used, using a variant of an ABA/BAB test to capture the effect of learning, order effects, and to compare our different sonification approaches on a small sample of ten participants.

The rest of the paper is structured as follows. In Section II, we discuss related work. In Section III, we describe technical details, then experiments and results for localization and recognition problems in Sections IV and V, respectively. Finally, future work is discussed in Section VI.

II. Related Work

A variety of sensory substitution systems exist, including tactile-tactile, visual-tactile and visual-auditory; our work is a *visual-auditory sensory substitution system*. As a system that can create non-speech sound from data, ours may also be considered a *sonification system*.

The most well-known visual-auditory sensory substitution/sonification system is the vOICe [1][3] system, which scans gray-scale image snapshots acquired from video left to right over time, mapping vertical location and intensity of pixels to frequency and amplitude of sound waves. It is possible to use this system to sonify depth images; in contrast, our system is built around the metric 3D structure of the scene as embodied in its point cloud, spatial audio, and the principle of real-time responsiveness. With respect to shape-focused systems, Yoshida et al. [6] sonify 2D shapes on a touch screen by allowing users to explore edges in the image with a finger, producing a sound using the same scheme as with the vOICe, but local to the finger. When the user loses the edges, cues help them find their way back. In See ColOr [2], a depth camera is used, and different instruments (like piano, flute, trumpet) with different properties and in different combination are used to represent different hues.

In the augmented reality work of Shelley et al. [7], users can touch simulated 3D objects and manipulate them with a visual-haptic interface with sound generated from the 3D contours of the object. The local curvature and cross-sections of objects are transmitted as frequency over a carrier wave that is either sinusoidal, a cello wavetable or modally synthesized, and with a haptic force feedback component.

Another close work in conception is the real-time navigation aid of Dunai et al. [8], which uses a stereo system to calculate depth images. After tracking and segmentation, objects with high importance like cars, humans, buildings, animals, and also free space 5 to 15 meters in range, are determined, and the closest object and free space are sonified with a synthetic instrument. Frequency, binaural cues and other sound properties represent distance, direction, and speed. The system is designed primarily as a salient hazard detector.

In our previous work [5], we presented two different approaches to 3D shape sonification: a method based on object recognition techniques commonly used in cognitive robotics applications that first recognizes an object and then chooses an instrument to sonify it accordingly, and a method in which sound waves are directly generated based on objects' outlines in an image - no attempt is made to account for the internal shape of objects. Moreover, that system was tested using artificially generated 3D objects on the problem of object recognition. In contrast, although using a simple sonification scheme, mapping object size or distance to frequency, the earlier incarnation of our system [4] was tested in real scenarios, on the problem of object localization. Conversely, the current paper proposes a scheme for sonifying the interior shape of an object, and compares it to the outline contour approach implemented previously [5]; moreover, here we also introduce new object tracking capabilities to adapt the approaches to real physical scenarios on both tasks - object localization and recognition.

III. TECHNICAL DETAILS

Figure 1 shows the full data-flow of our system, from acquisition of a point cloud from the depth sensor to its sonification, and reflects at a high level our software architecture.

Our system was conceived as an application of the sonification of shape in the context of sensory substitution, to make use of depth cameras like the Asus Xtion used in the current experiments, which can access the 3D structure of most indoor environments directly, in the form of depth images and point clouds. Indeed, advances in stereo vision, structure from motion, and depth from monocular cues are rapidly making the use of point clouds cheaper and more accessible.

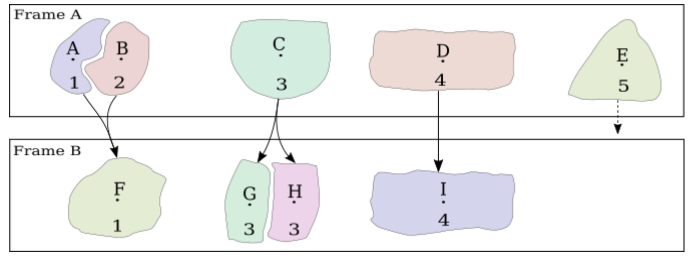

Figure 2. Example of the best association between segments in two consecutive frames (A and B). Letters show segment-ID, numbers show track-ID and black dots show the centroid of the segments. Each matched segment at frame t gets its track-ID number from its match at time $t-1$. If there is no match, it is given a new track-ID.

A. Filtering and Segmentation

Our original aim was to segment the point-cloud scene acquired by the camera into proto-objects based on structural edges between relatively smooth surface parts; these proto-objects we hereafter call segments. Segmenting objects in this way provides proto-objects that are relatively simple in terms of the local structure of their surface and as such suitable for later mapping on to sound primitives. As this is the first time this problem has been approached in the literature, in the first instance we adapted a common segmenting trick used in tabletop robotics [4] - removing the table from the scene and finding connected segments in the remaining point cloud.

In the current work, we return to the proto-object concept and segment and sonify multiple proto-objects in real-time based on sharp changes in the orientation of surfaces in an object, anticipating that their collective sound should help identify the object from which they are made. However, we do continue to filter out large planes, typically constituting walls and floor, to enable users to focus on object understanding.

B. Segment Tracking

In previous work [4], we were able to sonify an object by segmenting each new frame and playing a short sound interpolated from the previous sound. Since frames are received approximately every $\frac{1}{30}s$, and each frame is treated independently, we could only play sounds with a structure lasting $\frac{1}{30}s$. Moreover, because we sonified one segment/object at a time, we did not face problems with segments being confused with each other. However, in the present work we use improved kinds of sonification, in which sounds can last for several seconds, and we sonify multiple segments. We need to keep track of the identity of correct segments over multiple frames. Our approach is to keep the segmentation part of our pipeline but to associate segments over time using combinatorial optimization.

Figure 2 shows an ideal output of the tracker's data association. Due to noise, the sensitivity of camera to some materials, and the sensitivity of the segmenter to thresholds, objects might be segmented in different ways over time and the order in which segments are output from segmenter can change arbitrary from frame to frame. As can be seen in Figure 2 one segment (e.g., C) can split (into G and H) and two

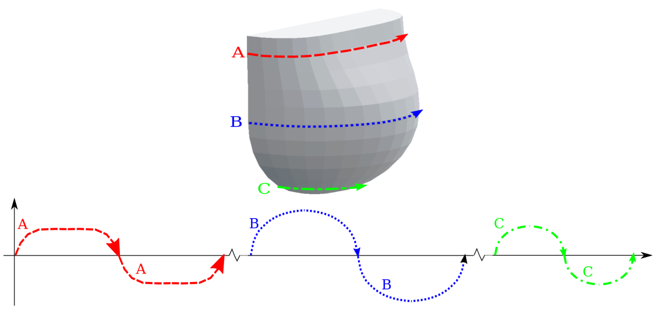

Figure 3. The proposed approach to shape sonification including internal contours. Top: The 3D object face, and selected superimposed internal contours from top to bottom (A,B,C). Bottom: corresponding waveforms; combined, an entire waveform can be constructed.

segments (e.g., A and B), can merge (into F). Our tracker handles merging and splitting by amalgamating close segments if the amalgamated segment matches a previous segment well. This is essential because appearing and disappearing segments lead to ambiguous and confusing sound.

Our data association approach evaluates each set of associations based on a cost function, which takes into account the distance between centroids and the first moments of segments. These features are simple, fast to compute and sufficient for the task at hand. After generating possible associations and evaluating them, the association with best cost is selected, track-IDs are allocated and segments are sent to the sonification subsystem.

C. Sonification

The sonification subsystem takes as input the segments extracted by the segmentation and their IDs as applied by the tracking. It keeps track of which segments (IDs) are new, which have previously been seen, and which have disappeared. The cleanest sound was produced not by adjusting the sound over time but instead reproducing the sound associated with the segment as seen in the frame in which the sound started playing. Once a sound is finished playing, the segment can be sonified again. As long as an object is visible, the 3D location of the sound is updated according to the present centroid of the object and the initially calculated waveform is fed incrementally to the sound rendering subsystem in such a way as to minimize delay and artifacts, by tracking the expected frame-rate. If an object stops being visible its sound is faded out over the period of a frame (roughly $\frac{1}{30}s$).

We use two approaches to generating the waveforms used in the sonification: "external" and "internal" contour.

1) External Contour: This approach [5] works on the 2D organized point cloud extracted from each segment. This is essentially a calibrated depth image containing a channel for each of x, y and z coordinates. Point clouds remain organized in a 2D array even after segmentation and tracking in order to maintain our high frame-rate. The organized point cloud for each segment is a cropped window around the segment with a mask defining the segment shape.

The external contour approach works by tracing the outline (external contour) of the object in the organized point cloud to create a carrier wave which is subsequently frequency and amplitude modulated; the modulation is done by scanning the object top-to-bottom and using the width of the object at each vertical location to modulate the carrier [5].

2) Internal Contour: Gholamalizadeh et al. [5] discovered that the lack of interior shape information (inside the external contour) was one drawback of the direct sonification method compared to the indirect (recognition-based) method. We attempt to rectify that by extracting information about both external and interior contours of the segment/object to be sonified. The new method is visualized in Figure 3. The object is scanned top to bottom, and for each row of points (essentially a row of pixels), the object depth at each point becomes a sample in a waveform. Interpolation is done to increase and decrease the frequency/speed of the wave. Thus, the frequency of the sound will depend on the width of the object as it is scanned top-bottom and the exact shape of the waveform produced will depend on the horizontal cross-sectional shape of the object.

Important caveats are attached to the current instantiation of this approach. Firstly, it was planned that depths in the object be produced relative to the average depth of the object. However, for an object with no significant internal contours, such as a flat surface facing the camera, this produces no sound. We could normalize the amplitude of the wave but this then vastly amplifies noise. So instead of relative depth of each point, we use the depth from the camera center to the point. However, the depth to the camera greatly overwhelms any other value and produces essentially a square wave no matter the shape of the object. Thus, in order to transmit internal contours to the user, in addition to the above-described scheme, we average multiple rows of the object and use the resulting averaged 1D array to do amplitude modulation of the object while the relevant part of the object is being sonified. This results in a consciously discernible oscillation in the wave that serves to encode cross-sectional shape.

IV. Training and Experiments

We investigated the ability of users to localize and recognize objects in a restricted indoor physical scenario, comparing internal and external contour shape sonification approaches. For this purpose, we designed an experiment involving both localization and recognition tasks, where each of 10 participants (mean age 25, 9:1 male/female ratio) participated in either an ABA or BAB experiment. Due to the experimental nature of the prototype, we only worked with sighted individuals. 50% of participants had experience with a previous iteration of the system. Half of participants used the system with internal contour sonification (A), followed by the external contour sonification (B), and again the internal contour sonification (A). The other half of participants used the external (B) then internal (A) then external contour based sonification (B).

The main idea behind conducting ABA/BAB experiments is to investigate the order effect of conditions [9]. Since we have limited time with each participant as well as a limited number of participants, we wanted to evaluate both conditions with each participant - moreover, random-sample based experiments require an order of magnitude greater sample size compared to matched experiments. We also want to analyze the effect of experience with the system on performance. Before conducting each of the three tests in the ABA or BAB sequence, an independent training session was conducted. In each test, participants are asked to first localize a single object in their near environment and then recognize it.

For the localization task, blindfolded users were asked to localize an object placed in one of six possible locations by walking and pointing. A map of the environment and possible

 ISBN: 978-1-61208-616-3

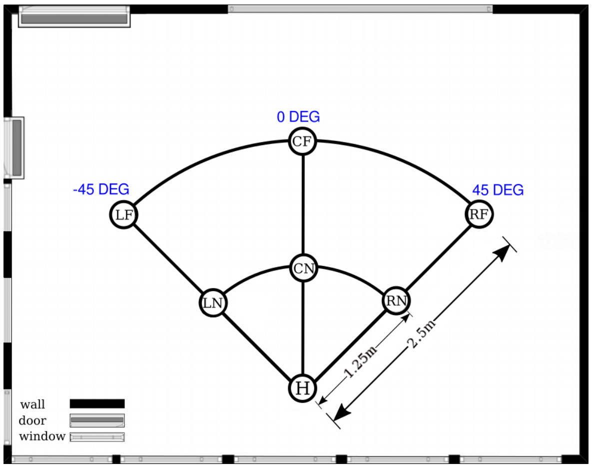

Figure 4. Map of the experiment room with six possible object positions. LN: left near, LF: left fat, CN: center near, CF: center far, RN: right near, RF: right far, H: Home location of the participant.

locations for objects is in Figure 4. After the localization task, users were asked to use the system to recognize the object.

Six arbitrary objects with disparate shape, size and complexity were chosen as an initial challenge to the capability of our sonification method. The objects and output of segmenter for each object are illustrated in Figures 5 and 6, respectively.

Next, training and test protocols are described.

A. *Training sessions*

The training protocol was exactly same for both encoding approaches (internal and external). First, a short verbal explanation of the system, encoding approaches, and the rules for experiments were given by the experimenter; this took five minutes. Then, three objects, namely box, drawer and bucket were placed in locations LF, CF and RF and participants were allowed to familiarize themselves with the system with open eyes, look at the objects from different view points, and listen to the sounds on the headphones; this took 12 minutes. Then, five minutes of training were allocated to learning about the localization task, where users were instructed to stay in the home location, scan the environment from left to right and down to up, walk towards objects and localize them. During all training sessions, users were allowed to move the camera freely with their hand, which was more comfortable for all users. They could see the output of the tracker on the computer display (as with Figure 6) and were suggested to also listen to the sound with closed eyes. After the first set of objects, a second set of objects - plant, stool and ball - were provided in locations LF, CF and RF, respectively. After learning with both sets of objects, users were allowed to interact freely with all objects placed at locations that they requested, for up to six minutes. As such, the maximum allowed training time for each test was $5 + 12 + 5 + 12 + 5 + 6 = 45$ minutes. These durations were selected considering the trade-off between our unpaid participants' time and the need to test our system with experienced users. The amount of time used to train with the system in these experiments is low compared to that which would be expected if the system were to be used day-to-day.

Figure 5. Set of objects used in experiments.

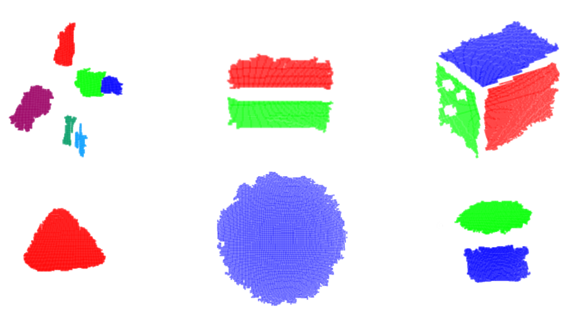

Figure 6. Segmented 3D images obtained from the objects in Figure 5. Top-left: plant, Top-middle: box, Top-right: drawer, Bottom-left: stool, Bottom-middle: ball, Bottom-right: bucket.

B. *Test sessions*

In test sessions, one randomly selected object was placed in a randomly selected location and blindfolded participants located at the home location were given three minutes to walk towards the object and localize it by pointing. Here, six trials with six different objects were conducted with each participant, plus an extra unrecorded trial with a random object, conducted to prevent the participant from guessing the object. Localization performance was evaluated as *precise*, *poor* or *failed* and participants were informed about the correctness of their answer. *Precise* localization meant the participant was able to point to the center of the object, *poor* localization meant the participant could point to the boundary of the object, and a *failed* localization meant the participant was unable to find the object. In the case of poor localization or failure, the experimenter helped the participant to precisely localize the object for the recognition phase of the test.

After localization of each object, participants were given two minutes to move the depth camera around the same object, listen to its sound and identify it. The experimenter then gave the correct answer, as participants were invariably curious considering their limited experience with the system during the training. The use of our ABA/BAB protocol controlled for the effect of experience during testing.

V. RESULTS

A. *Localization performance*

Figure 7 illustrates results of ABA and BAB experiments on the localization task for all participants. The internal approach with 73.3% precise localizations, 23.3% poor localizations and 3.3% failed localizations showed a similar performance to the external approach with 66.6% precise, 26.6% poor and 6.6% failed localizations (unpaired T-test, $N = 15$, $p = 0.31$). Localization performance for internal-external-internal and external-internal-external experiments increased by 22% and 30%, from the first to third test, respectively.

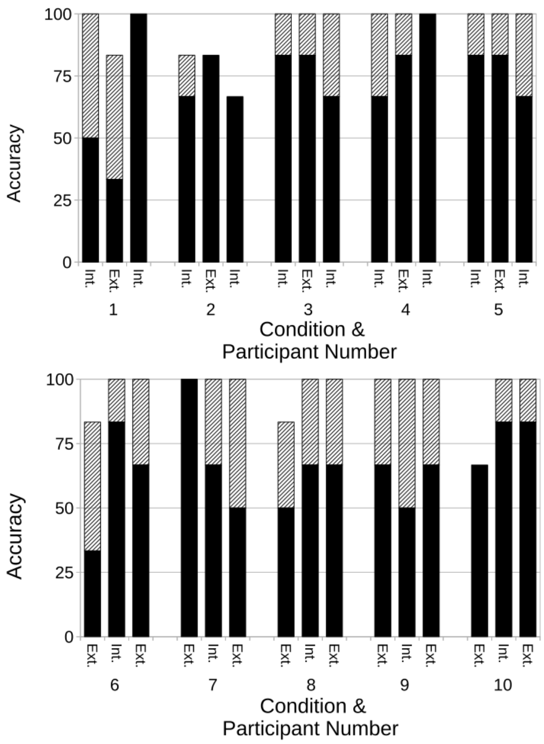

Figure 7. Results of ABA and BAB experiments for localization performance of all participants. Top: internal-external-internal order. Bottom: external-internal-external order. The black regions specify percentage of precise localizations and the shaded regions specify poor localizations.

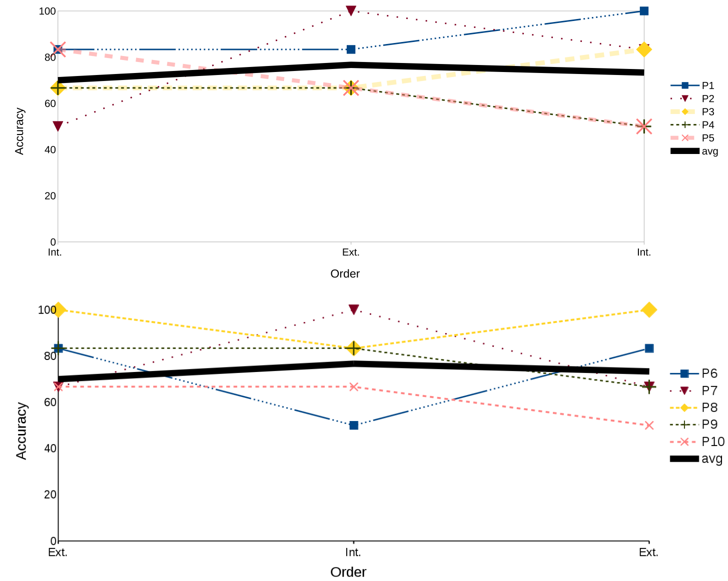

Figure 8. Results of ABA and BAB experiments for recognition performance of all participants (Pn: participant number n). Top: internal-external-internal order. Bottom: external-internal-external order.

For some objects, like plant, stool and bucket, the system had difficulty in detecting the objects when in distant locations like LF, CF or RF and when the user was at H. In these cases, participants tried to take a step forward and scan the environment again until they could find something. On the other hand, our system could detect the ball, drawer and box from the home position without a problem. For both approaches, participants reported that they had difficulty in localizing the drawer, because the large size of drawer created confusion. Therefore, the average localization time of drawer in both approaches was more than the average localization time for other objects. Moreover, among all experiments participants hit the object three times (1.6% of experiments); in two cases the object was the drawer and in one case bucket.

The system generated noisy sounds when the camera pointed at the walls and windows surrounding the test environment. In first training and then testing sessions, participants were confused when the system sonified segments related to part of the wall or window. Later, in the second and specially third tests, most participants showed that they can distinguish those noises, and they tried to change camera view to eliminate the noises and focus on the object.

B. Recognition performance

Results of ABA and BAB experiments for recognition performance of all participants can be seen in Figure 8. For the recognition task, results are independent of the order of conditions, with 4.7% improvement for internal-external-internal experiment and 8.3% deterioration for external-internal-external experiment, from the first to third test.

The internal contour approach, with an overall accuracy of 73%, did not show a significantly different performance from the external approach, with overall accuracy of 77% (unpaired T-test, $N = 15$, $p = 0.58$). Recognition accuracy showed individual differences. For example, one of the participants, who was a musician, obtained an accuracy of 100%, 83% and 100% on external, internal and external tests, respectively.

Participants, for both approaches, were encouraged to try to understand the logic of the encoding of object sounds. One of the strategies that participants were using for recognition was rotating the camera or going to other sides of the object to see it from different view points. These movements lead to different sounds for some objects due to the orientation-dependent top to bottom encoding in both approaches. It was observed that this cue helped users to distinguish between box and bucket in both approaches, and ball and stool in the external approach. Additionally, the other cue that was used heavily by users was the number of played sounds. By viewing some objects from different views, a different number of sounds could be heard. For example, by viewing a box from the front, just two sounds could be heard (one for the top segment and one for the side segment), whereas by viewing the same box from the side, three sounds could be heard (one for the top segment and two for side segments). On the other hand, for the ball and stool, just one sound could be heard. In the internal approach, some participants reported occasional interference of sounds of an object with multiple segments like box and bucket, making it hard for participants to distinguish number of segments. This led participants to confuse the stool with box and bucket. We attribute this effect to the square-wave nature of the internal contour encoding method we currently use.

For both approaches and for all objects, participants were able to successfully identify objects most of the time, as can be seen from the bold numbers in Table I. The two objects with the best recognition accuracy in both approaches were plant and ball, with 93% and 100% accuracy, respectively. The plant is distinctive because of a high number of segments

 ISBN: 978-1-61208-616-3

TABLE I. CONFUSION MATRICES FOR OBJECT RECOGNITION EXPERIMENTS. NUMBERS ARE PERCENTAGES ($N = 15$).

Internal approach		Response					
		plant	box	drawer	stool	ball	bucket
Actual	plant	$\mathbf{93\frac{1}{3}}$					$6\frac{2}{3}$
	box		$\mathbf{60}$	$6\frac{2}{3}$		$6\frac{2}{3}$	$26\frac{2}{3}$
	drawer		20	$\mathbf{66\frac{2}{3}}$			$13\frac{1}{3}$
	stool		$33\frac{1}{3}$		$\mathbf{40}$	$13\frac{1}{3}$	$13\frac{1}{3}$
	ball					$\mathbf{100}$	
	bucket		$6\frac{2}{3}$	$6\frac{2}{3}$			$\mathbf{86\frac{2}{3}}$

External approach		Response					
		plant	box	drawer	stool	ball	bucket
Actual	plant	$\mathbf{93\frac{1}{3}}$		$6\frac{2}{3}$			
	box		$\mathbf{40}$	$33\frac{1}{3}$	$6\frac{2}{3}$		20
	drawer		$6\frac{2}{3}$	$\mathbf{86\frac{2}{3}}$		$6\frac{2}{3}$	
	stool				$\mathbf{73\frac{1}{3}}$	$26\frac{2}{3}$	
	ball					$\mathbf{100}$	
	bucket	$6\frac{2}{3}$	$6\frac{2}{3}$	$6\frac{2}{3}$	$6\frac{2}{3}$		$\mathbf{73\frac{1}{3}}$

and the ball because of its size and recognizable contour. As expected, confusion is seen between objects similar in shape like box, drawer and bucket, the most confusions occurring between box and bucket, and also drawer and box in both approaches. Confusion between stool and ball also arises because of similarity of the contour of the two objects.

C. *Learning and testing duration*

Although we allocated a maximum of 45 minutes for training for each test, results shows that learning and using our system is possible in a shorter time. In both internal-external-internal and external-internal-external orders, the training time of participants decreased from the first test to the third test. In the internal-external-internal sequence ($N = 15$), the average training time for the first test was 23.5 minutes, which decreased by 47% for the third test; similarly, for the external-internal-external order ($N = 15$), time in the first test was 18.8 minutes, decreasing by 64% to the third test. Furthermore, the average time spent localizing and recognizing each object decreased from $181s$ by 22% and from $139s$ by 26%, from the first to third test. However, an unpaired T-test ($N = 15$, $p = 0.48$) verifies that there is no significant difference between internal and external contour in training and testing time.

D. *Questionnaire*

After the experiment, users were given a three-question questionnaire asking them to rate on a scale of one to five the quality of sound in each approach and the usability of our system. Most participants reported that they found the generated sound of the external contour approach (average score of 3.3 out of 5) more pleasant than the sound of the internal approach (average 2.6 from 5), most likely due to the fact that the internal approach always generated roughly a square wave (see the description of the method above), though participants acknowledged the suitability of the internal approach for sonifying the curvature of objects like a ball. With respect to usability, the average score was 3.2 out of 5, along with feedback that users considered that the system has usefulness in the current simplified scenario but cannot guess how it might be used with more objects and in more unstructured scenarios.

VI. Conclusion

This is the first time that sensory substitution from 3D shape has been attempted. An internal contour based encoding method for sonification of 3D objects was presented. This method is able to encode depth and curvature of objects, and is compared with our previous external contour based approach. A data-association based tracking system was introduced to enable the use of these sonifications in real interactive physical scenarios.

An ABA/BAB experiment was performed to evaluate the ability of blindfolded participants in localization and recognition of different objects, and to investigate order effects and the effect of experience. Results show similar performance of users in localization and recognition of objects in both approaches and quick mastery of the system. The internal approach was expected to show a better performance in recognition due to encoding more information but in its current iteration suffers from an inability to represent a variety of contours and produce harmonious sounds at the level of carrier wave.

The next steps in this work are developing the internal encoding to produce more pleasant and informative sound and exploring more approaches for dealing with background noises to make our system work in more general scenarios.

Acknowledgment

This work is supported by the Scientific and Technological Research Council of Turkey (TÜBITAK), Projects 114E443 and 116E167.

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'," Perception, vol. 36, no. 3, 2007, pp. 416 – 430, ISSN: 1468-4233.

[2] J. D. Gomez Valencia, "A computer-vision based sensory substitution device for the visually impaired (See ColOr)," Ph.D. dissertation, University of Geneva, 2014.

[3] P. Meijer, "An experimental system for auditory image representations," IEEE Transactions on Biomedical Engineering, vol. 39, no. 2, 1992, pp. 112–121, ISSN: 1558-2531.

[4] A. Mhaish, T. Gholamalizadeh, G. İnce, and D.J. Duff, "Assessment of a Visual-to-Spatial Audio Sensory Substitution System," in Proc. IEEE Conf. on Signal Processing and Communications Applications, 2016, pp. 245–248, ISBN: 978-1-5090-1679-2.

[5] T. Gholamalizadeh, H. Pourghaemi, A. Mhaish, G. İnce, and D. J. Duff, "Sonification of 3d Object Shape for Sensory Substitution: An Empirical Exploration," in ACHI 2017, The Tenth International Conference on Advances in Computer-Human Interactions, Mar. 2017, pp. 18–24, ISBN: 978-1-61208-538-8.

[6] T. Yoshida, K. M. Kitani, H. Koike, S. Belongie, and K. Schlei, "EdgeSonic: image feature sonification for the visually impaired," in Proceedings of the 2nd Augmented Human International Conference. ACM, 2011, p. 11, ISBN: 978-1-4503-0426-9.

[7] S. Shelley et al., "Interactive Sonification of Curve Shape and Curvature Data," in Haptic and Audio Interaction Design. Springer, Berlin, Heidelberg, Sep. 2009, pp. 51–60, ISBN: 978-3-642-04076-4.

[8] L. Dunai, G. Fajarnes, V. Praderas, B. Garcia, and I. Lengua, "Real-time assistance prototype - A new navigation aid for blind people," in IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society, Glendale, Arizona, USA, Nov. 2010, pp. 1173–1178, ISBN: 978-1-4244-5225-5.

[9] C. Eilifsen and E. Arntzen, "Single-subject withdrawal designs in delayed matching-to-sample procedures," European Journal of Behavior Analysis, vol. 12, no. 1, 2011, pp. 157–172, ISSN: 1502-1149.