Sonification of Range Information for 3-D Space Perception

Evangelos Milios, Bill Kapralos, Agnieszka Kopinska, and Sotirios Stergiopoulos

***Abstract*—We present a device that allows three-dimensional (3-D) space perception by sonification of range information obtained via a point laser range sensor. The laser range sensor is worn by a blindfolded user, who scans space by pointing the laser beam in different directions. The resulting stream of range measurements is then converted to an auditory signal whose frequency or amplitude varies with the range. Our device differs from existing navigation aids for the visually impaired. Such devices use sonar ranging whose primary purpose is to detect obstacles for navigation, a task to which sonar is well suited due to its wide beam width. In contrast, the purpose of our device is to allow users to perceive the details of 3-D space that surrounds them, a task to which sonar is ill suited, due to artifacts generated by multiple reflections and due to its limited range. Preliminary trials demonstrate that the user is able to easily and accurately detect corners and depth discontinuities and to perceive the size of the surrounding space.**

***Index Terms*—Audio systems, distance measurement, handicapped aids, laser applications, sound generation.**

I. Introduction

HELPING the visually impaired perceive the space around them with electronic travel aids (ETAs) relies on providing spatial feedback via nonvisual senses, primarily hearing and touch. Certain devices [7] have the ambitious goal of sonifying video images, but most are intended to facilitate navigation. Ultrasonic sensors have low cost and a wide beam width suitable for detecting obstacles in front of a visually impaired person, and have formed the basis of the majority of the devices [1], [2], [4], [5], [11]. One problem with sonar is that their resolution is not high enough for detecting dropoffs (e.g., steps down). Laser radar has been used as well. The Talking Cane is capable not only of obstacle detection, but also of detecting reflections from bar coded retroreflecting signs, which are converted into spoken messages [6]. The laser cane uses range measurement by triangulation, detects objects for two distinct elevations, and provides feedback with two different tones and vibrational rates. It also provides drop-off warning. Mobile robotics technology has been applied to ETAs. GuideCane is a mobile robot that guides the user to their destination, while detecting and avoiding both ground obstacles and overhanging objects [13]. Stereo vision-based obstacle detection has been exploited in [8], which combines disparity measurement for obstacle detection with motion estimation and the use of an inclinometer. The wheelchair pathfinder combines ultrasonic and laser radar sensors mounted on a wheelchair and offers forward, side, and step detection with audio and/or tactile feedback. Systems that integrate multiple technologies for travel aid have been designed. The MoBIC system [10] provides assistance in exploring maps and planning journeys, and in executing those plans by offering navigation and orientation assistance. A global positioning system (GPS) receiver is an integral part of this system, while local navigation information is expected to come from a long cane or guide dog. The issue of standardization and integration of ETAs has been studied in [3].

Unlike sonar, laser ranging allows for a greater range of operation. Its narrower beam and shorter wavelength combine to detect finer detail necessary for shape and pattern perception. A laser ranging device has provided range measurements for an autonomous robot [9].

Rather than object detection, for which sonar is better suited, we examine the sonification of infrared range measurements in order to perceive shapes and patterns (e.g., detection of corners, stairs, and depth discontinuities) in the user's environment. The ultimate goal of our work is to develop a device that will enable the perception of three-dimensional (3-D) space by the visually impaired user.

II. Sonification of Range Information

The range data sonification process involves a user holding or carrying on her head a point laser ranging device that produces a stream of measurements of the range to the object, at which the laser beam is pointing at any given instant. The user can scan the environment and thereby control the direction of the laser beam. Such "interactive" use of the laser range sensor allows the user to "inspect" surrounding space and perceive its spatial properties through perception of the stream of range measurements. Sonifying the stream of range measurements is a natural way for the user to perceive it.

Fig. 1 illustrates the range data sonification process. The users scan their environment and obtain eight range measurements per second (one measurement every 125 ms), with the laser range finder (LRF). Measurements are mapped to a particular feature, pitch (frequency of the audio signal), loudness (amplitude of the audio signal), or both in the audio domain. As pure tones are annoying to humans, we instead use musical instrument digital interface (MIDI) notes of different frequencies.

Manuscript received October 28, 2000; revised September 10, 2001. The work was supported by a research grant from the National Science and Engineering Council of Canada (NSERC) and an NSERC Undergraduate Student Research Award to B. Kapralos. A summary of a preliminary version of the design without experiments on human subjects was presented at the Joint ICAD/ASA/EAA Workshop on Auditory Display, March 20, 1999, Berlin, Germany.

E. Milios is with the Faculty of Computer Science, Dalhousie University, Halifax, NS B3H 1W5, Canada (e-mail: eem@cs.dal.ca).

B. Kapralos and S. Stergiopoulos are with the Department of Computer Science, York University, Toronto, ON M3J 1P3, Canada.

A. Kopinska is with the Department of Psychology, York University, Toronto, ON M3J 1P3, Canada.

Digital Object Identifier 10.1109/TNSRE.2003.819645

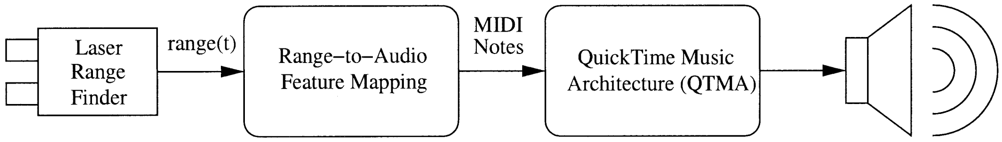


Fig. 1. Overview of sonification of range data.

The MIDI scale has 128 notes, and the mapping between MIDI note index m and frequency f in hertz is $f = f_0 2^{((m-60)/12)}$, where $f_0 = 261.625$ Hz, which corresponds to a middle "C." Loudness is adjusted by varying the "velocity" of the MIDI note (in MIDI terminology). Each MIDI note is output for 90 ms (e.g., there is a "small gap" of silence between successive notes), using one of the 128 instruments available with the quick-time music architecture (QTMA) software synthesizer.

Two modes of operation have been defined.

a) **Proportional mode**, where the audio feature is a nonlinear function of the instantaneous value of range. More precisely, the nonlinear function is a decaying exponential mapping.
b) **Derivative mode**, where the audio feature is a function of the temporal derivative of range. More precisely, the change between consecutive range measurements (an approximation to mapping the derivative of range as a function of time) are mapped to the audio domain.

In Section III, we will describe the mappings in detail.

III. Proportional Mode in Range-to-Audio Mapping

Since there are no general guidelines for sonifying data [14], several different range-to-audio (or change in range to audio) mappings were experimented with. For range-to-frequency mapping, the output produced with a decaying exponential mapping was found to be the most effective as it stressed range measurements close by, but allowed for a perceived change in frequency throughout the entire range.

Although the LRFs maximum range exceeds 100 m, the maximum range for the purpose of this study was restricted to 15 m with the proportional mode of operation. This restriction was placed to ensure the user is not overloaded with information and to restrict the range of notes output. For example, an increase in the maximum allowable range measurement from 15 m should also be followed by an increase in the range of frequencies used to ensure there is a perceived change in frequency throughout the entire range of allowable measurements.

At the other extreme, the minimum range measurement allowable was limited to $r_0 = 0.30$ m as it is used primarily for object detection/avoidance.

While in the proportional mode, a range measurement r is mapped to frequency using the relation

$$\text{Frequency} = kc^{-a(r-r_0)} \tag{1}$$

where $k = 4200$ Hz, $c = 2.718 = e$, $a = 0.25$ m^{-1}, $r_0 = 0.30$ m (the minimum range allowed).

The frequency value obtained is then mapped to the MIDI note with the closest frequency. Using this mapping allowed range measurements close to the user (e.g., small range measurements) to be mapped to higher frequencies thereby stressing their importance of objects nearby. The maximum frequency in this mapping (4200 Hz) corresponds to a range measurement of 0.30 m whereas the minimum frequency (106.46 Hz) corresponds to a range measurement of 15 m.

IV. Derivative Mode in Range-to-Audio Mapping

While in the derivative mode, rather than using the range measurements themselves, the change between consecutive range measurements (an approximation to mapping the derivative of range as a function of time) are mapped to the audio domain. Although such information may not necessarily be used to locate objects, it could be used to provide greater detail about the user's surroundings. For example, depth discontinuities will easily be detected, as there will be a sudden jump in the derivative between consecutive readings. At the other extreme, flat surfaces could also be detected as the derivative between consecutive range measurements will not change (or will change slightly).

Irrespective of the user's scanning position, scanning at a faster rate will result in a more rapid change between range measurements as opposed to scanning at a slow speed.

To allow the user to discriminate between positive and negative range change, distinct signals were used for each case. Positive/negative change indicates that the current location being measured is farther away from/closer to the user relative to the last location.

Mapping parameters are as follows:

- maximum allowable change in range $(r_{\max})$: $\pm r_{\max} = \pm 2.3$ m;
- minimum allowable change in range $(r_{\min})$: $\pm r_{\max} = \pm 0.15$ m;
- total number of MIDI notes used (N): 70.

The mapping itself from the change in range Dr to a MIDI note, is a piecewise linear mapping Since MIDI notes come only in integer values, the result is rounded to the nearest integer.

$$\text{MIDI}(Dr) = \begin{cases} 25 + \frac{N}{2}\frac{r_{\max}-Dr}{r_{\max}-r_{\min}}, & r_{\min} < Dr < r_{\max} \\ 60, & 0 < Dr < r_{\min} \\ 69, & -r_{\min} < Dr < 0 \\ 69 + \frac{N}{2}\frac{|Dr|-r_{\min}}{r_{\max}-r_{\min}}, & -r_{\max} < Dr < -r_{\min} \end{cases} . \tag{2}$$

Several points should be made regarding this method.

1) Although a linear mapping was used between range and MIDI note, adjacent notes of the musical scale differ logarithmically. As a result, the change in range measurements to frequency did follow a logarithmic mapping.
2) Although several of the frequencies associated with positive changes in range measurements ("Moving Away")

TABLE I
SYSTEM PARAMETERS RELEVANT TO BOTH MAPPINGS

Range Readings per Second	8
LRF Accuracy	+/- 10cm (however, we have found this value is slightly more!!)
Headphone Volume	Adjustable
MIDI Note Duration	90ms - small gap of silence between consecutive sounds
MIDI Note Amplitude (Volume)	Constant value of 80 (max. allowable amplitude = 127)

are fairly low, and may sound displeasing when output as regular tones, generating the output using the piano instrument of the QTMA resulted in a pleasing sound

V. Testing the Accuracy and Precision of the Sonification of Range Measurements

With both modes, range measurements are mapped to frequency in the form of MIDI notes. The MIDI note was output using the QTMA software synthesizer. In particular, all notes were generated using instrument 1—piano. MIDI notes range from 1 to 128 and follow the scale of equal temperament in which every octave (a 2:1 change in frequency) is divided into 12 equal intervals allowing for the frequency of adjacent notes to differ by a factor of 1.05946 (twelfth root of two)—e.g., adjacent keys differ logarithmically.

Information relevant to both mappings is shown in Table I.

A. Subjects

Seven female and five male (total of 12) subjects participated in the study. The age of the subjects ranged from 18 to 50, with a mean of 32. Ten of the subjects had normal vision and were blindfolded during the entire experiment. Two of the subjects were visually impaired, in that they had their eyes removed in early childhood, two and three years of age, due to retino-blastoma. All the subjects were paid for their participation.

B. General Method

All the experiments were conducted in an empty spacious room (1460 $\times$ 420 cm). Each experimental session lasted for approximately 3 h for the normal vision subjects and approximately 5 h for the visually impaired subjects. At the beginning of the session the subjects received between 45 min–1 h of training on how to use the device. The training took approximately 2 h for the visually impaired subjects. To blindfold the subjects, a pair of "covered" ski goggles were used which prevented the subjects from viewing the experimental setup. The subjects sat on a chair that was pushed around by the experimenter and were holding the LRF unit with both their hands, as shown in Fig. 2. The distance between a "laser dot" created by the laser pointer attached to the LRF and the feature that subjects were instructed to detect (i.e., pointing error) was measured and recorded by the experimenter. The *accuracy* of pointing was obtained by taking the average of pointing errors for each subject on repetitive trials. The *precision* of pointing was measured by the standard deviation of the repetitive pointing errors. The data was further analyzed using the repeated measures ANOVA [12].

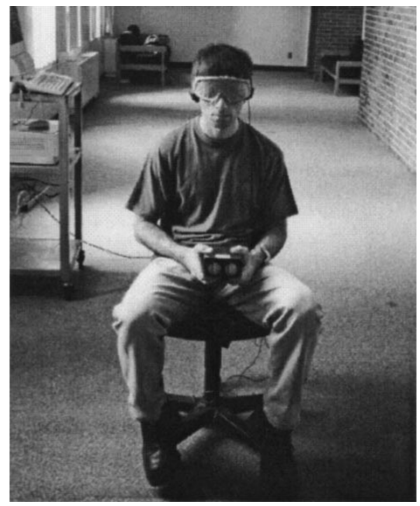


Fig. 2. Photograph of a subject during an experiment.

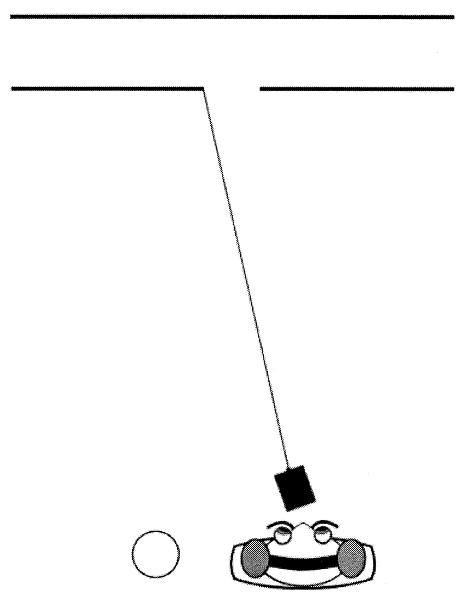


Fig. 3. Experiment 1—experimental setup. The subject positioned in front (30 cm to the left or right with respect to the center of a gap) of a vertical gap at a distance of 2, 6, and 12 m. The gap was 20 cm wide and 30 cm deep. The subjects were asked to indicate the left and right edge of the gap.

1) Experiment 1: Detection of Depth Discontinuities: As illustrated in Fig. 3, the subjects were positioned in front (30 cm

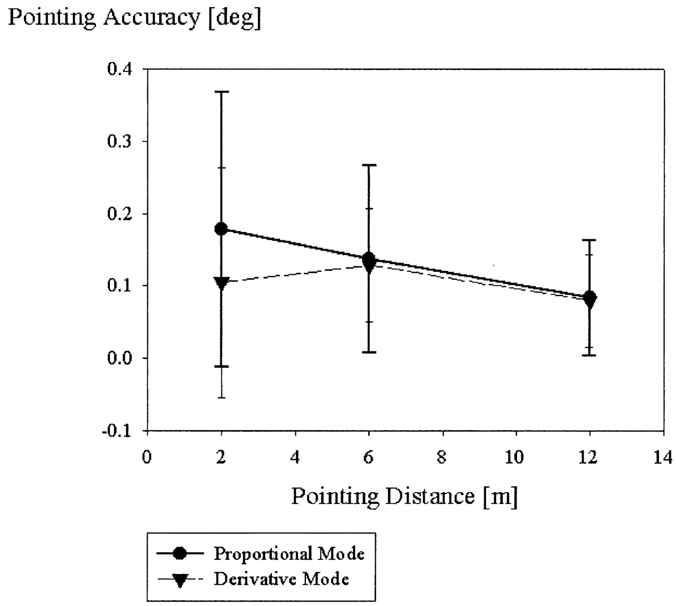


Fig. 4. Means and standard deviations of pointing accuracy when pointing while in the proportional (circles) or derivative (triangles) mode at the vertical gap located at 2, 6, and 12 m.

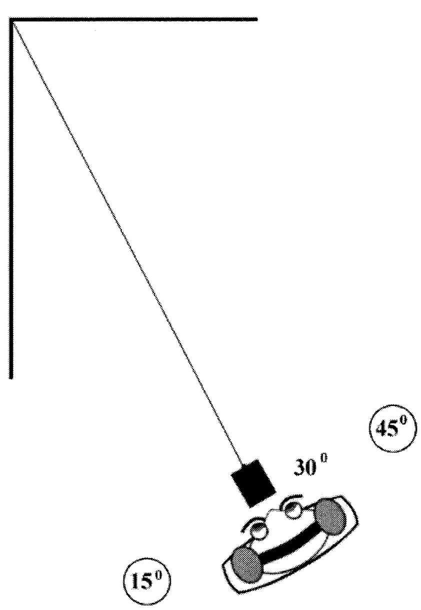


Fig. 5. Experiment 2—experimental setup. The subject positioned at a distance of three meters away from a corner between two walls. The task was performed from three different angular positions: 15°, 30°, and 45° with respect to the corner.

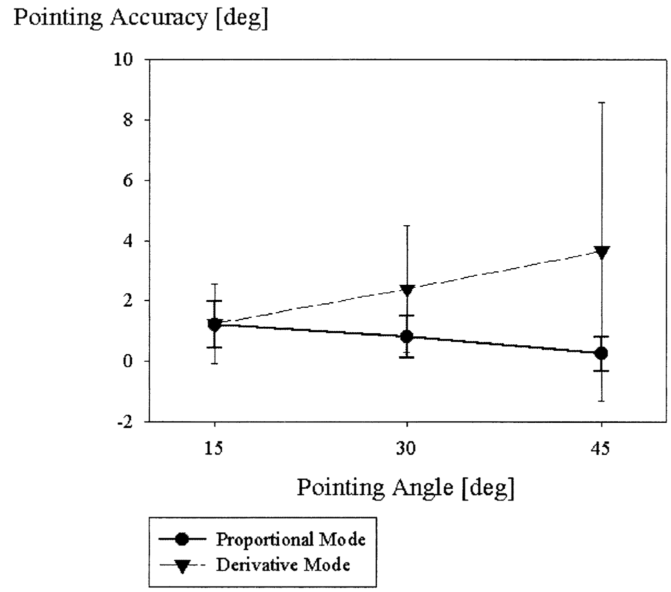


Fig. 6. Means and standard deviations of pointing accuracy when pointing while in the proportional (circles) or derivative (triangles) mode at the vertical corner.

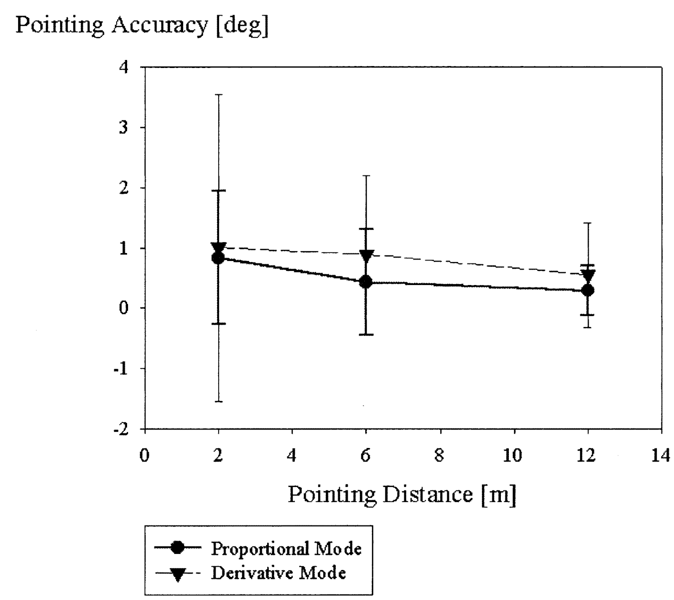


Fig. 7. Means and standard deviations of pointing accuracy when pointing while in the proportional (circles) or derivative (triangles) mode at the horizontal corner.

to the left or right with respect to the center of a gap) of a vertical gap at a distance of 2, 6, and 12 m. The gap was 20 cm wide and 30 cm deep. The subjects were asked to indicate the left and right edge of the gap. There were three repetitions for each edge at each of the three distances. Pointing in the direction of the gap was recorded as yielding a positive error and pointing away from the gap as negative.

The subject's accuracy (average pointing error) for the different distances for the proportional mode ranged from $-0.009°$ to $0.301°$ and for the derivative mode from $0.025°$ to $0.214°$. The precision (standard deviation of pointing error) ranged from $\pm 0.183°$ to $\pm 0.549°$ for the proportional mode and $\pm 0.212°$ to $\pm 0.469°$ for the derivative mode. For the proportional mode the overall accuracy was $0.133°$ and precision $\pm 0.349°$ and for the derivative mode $0.104 \pm 0.319°$, respectively. The accuracy and precision for the visually impaired subjects were $-0.009 \pm 0.549°$ and $-0.005 \pm 0.304°$ for the proportional mode and $0.085 \pm 0.469°$ and $0.056 \pm 0.329°$ for the derivative mode.

There were no statistically significant differences in accuracy ($F_{1,360} = 1.179$, NS) and precision ($F_{1,72} = 0.014$, NS) between the two modes. Fig. 4 illustrates accuracy and precision for different modes and distances across all the subjects.

2) Experiment 2: Detection of Vertical and Horizontal Corners: To measure the subjects' ability to detect vertical corners,

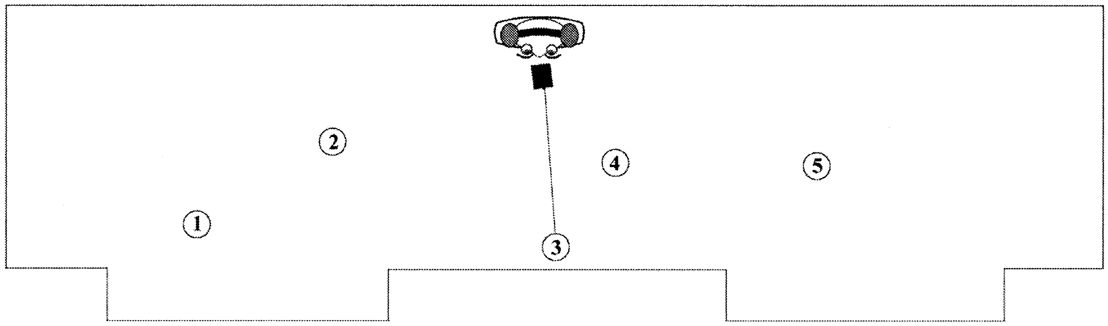


Fig. 8. Experiment 3—experimental setup. The subject scanned the surrounding space and pointed the LRF at the experimenter that was positioned at one of the five locations.

the subjects were positioned at a distance of three meters away from a corner between two walls. The task was performed from three different angular positions, i.e. 15°, 30°, and 45° with respect to the corner (see Fig. 5). The subjects were never directly facing the corner; they were rotated by 15°–20° to the left or right with respect to the corner. There were five repetitions for each position tested.

To measure the subjects' ability to detect horizontal corners, the subjects were positioned at 2, 6, and 12 m away from a wall and asked to point at the corner between the wall in front of them and the floor. There were five repetitions for each distance tested.

a) Vertical corner: The subject's accuracy for the different angles of pointing for the proportional mode ranged from −0.771° to 2.017° and for the derivative mode from −7.459° to 8.517°. The precision ranged from ±0.136° to ±2.026° for the proportional mode and ±0.334° to ±10.566° for the derivative mode. For the proportional mode the overall accuracy was 0.747° and precision ±1.371° and for the derivative mode 2.424 ± 4.355°. The accuracy and precision for the visually impaired subjects were 0.239 ± 0.565° and 0.815 ± 1.461° for the proportional mode and −2.607 ± 3.693° and 3.596 ± 3.576 for the derivative mode.

Fig. 6 illustrates the accuracy and precision for the different modes and angles of pointing across all the subjects. There was a difference in accuracy between the two modes ($F_{1,22} = 5.433$, $p < 0.05$) with pointing using the proportional mode being more accurate than pointing using the derivative mode. Also, pointing using the derivative mode was becoming progressively less accurate with an increase in the angle of pointing ($F_{2,22} = 4.330$, $p < 0.05$). Similarly, pointing using the proportional mode was more precise than pointing using the derivative mode ($F_{1,22} = 19.622$, $p < 0.001$), and precision of pointing using the derivative mode was becoming worse as the angle of pointing increased while precision of pointing using the proportional mode stayed the same ($F_{2,22} = 9.447$, $p < 0.001$).

b) Horizontal corner: The subject's accuracy for the different distances for the proportional mode ranged from −2.223° to 2.195° and for the derivative mode from −5.191° to 4.657°. The precision ranged from ±0.072° to ±3.225° for the proportional mode and ±0.212° to ±3.577° for the derivative mode. For the proportional mode the overall accuracy was 0.515° and precision ±1.493° and for the derivative mode 0.815 ± 2.266°. The accuracy and precision for the visually impaired subjects were −0.560 ± 1.829° and 0.791 ± 0.91° for the proportional mode and 0.915 ± 1.064° and 0.130 ± 1.519° for the derivative mode.

Fig. 7 illustrates the accuracy and precision for different modes and distances of pointing across all the subjects. The accuracy of pointing did not differ between the two modes ($F_{1,22} = 1.203$, NS). Similarly, pointing using the proportional mode was more precise than pointing using the derivative mode ($F_{1,22} = 7.785$, $p < 0.05$), with precision of pointing using both the derivative and proportional mode becoming worse as the distance decreased, i.e., the angle of pointing increased ($F_{2,22} = 48.456$, $p < 0.0001$).

C. Experiment 3: Locating an Object in the Surrounding Space

Subjects were asked to scan the surrounding space and point the LRF at the experimenter that was positioned at one of the five locations as illustrated in Fig. 8.

When pointing using the derivative mode, 7 of the 12 subjects were accurate in finding the experimenter at all the locations. Four subjects located the experimenter at four out of the five locations. One subject failed to locate the experimenter at two out of the five locations. Five out of the six misses occurred when the experimenter was located at position one and one when the experimenter was located at position three. When pointing using the proportional mode, 8 of the 12 subjects were accurate in finding the experimenter at all the locations. Four subjects located the experimenter at four out of the five locations. All the misses occurred when the experimenter was located at position three.

VI. Conclusion

Our study suggests that the sonification of range measurements is a promising approach for conveying information about spatial features. With both proportional and derivative mode the subjects were able to accurately and precisely locate depth discontinuities as well as vertical and horizontal corners. The only exception was pointing while in the derivative mode at the vertical corner from the angle of 45°. Initially, the output from the proportional mode was reported as more comprehensible than the output from the derivative mode. However, after the experimental session most subjects expressed an opinion suggesting that the combination of both modes was more informative than either of the modes alone. This point is supported by the results of experiment three, where the derivative mode was superior in detection of an object closely located in front of a flat surface, while the proportional mode was more efficient in locating objects positioned in a more complex environment, with more spatial features in the background. Our data with visually impaired subjects is very preliminary. Proper clinical experiments with visually impaired subjects are required to investigate the extent to which this approach can be useful to members of that population, and to quantify the statistically significant differences, if any, in performance between visually impaired and blindfolded subjects. Future research should involve the comparison of this system with existing ETA technologies. Finally, the issue of information overload when exploring more complex environments needs to be addressed.

ACKNOWLEDGMENT

The authors would like to thank Prof. P. Dymond for suggesting the use of MIDI, Prof. L. Harris for suggesting the use of derivative mappings, and R. Arrabito of DCIEM for his valuable comments and his encouragement. G. Reid provided help with audio and signal processing facilities.

REFERENCES

[1] "The Trisensor (KASPA TM): Wideangle untrasonic head mounted mobility aid. SonicVisionN Ltd". [Online] http://www.sonicvision.co.nz/ (accessed on Sep. 12, 2001)

[2] "Mowat: A lightweight, hand-held, ultrasound mobility aid". Pulsedata Co., New Zealand. [Online] http://www.pulsedata.co.nz/ (not in production, accessed on Aug. 22, 2001)

[3] S.Simon Harper, "Standardising electronic travel aid interaction for visually impaired people," M.Phil. thesis, Department of Computation, Univ. of Manchester Inst. Science and Technology, Manchester, U.K., 1998.

[4] A. D. Heyes, "Sonic Pathfinder: A head mounted ultrasonic mobility device for outdoor use," Tech. Rep., http://www.sonicpathfinder.org/, 1994.

[5] Brytech Inc., "Sensory 6: A navigation aid for the blind," Tech. Rep., Ottawa, ON, Canada, http://www.brytech.com/, 1985–1995.

[6] S.Sten Lofving, The Talking Cane: A Blind People's Cane, With Built in Radar Functions, REAB Data AB, Sweden, 1994.

[7] P. Meijer, Voice: Vision Substitution Technology for the Blind, Philips Research Laboratories, Eindhoven, The Netherlands, 1968.

[8] N. Molton, S. Se, J. M. Brady, D. Lee, and P. Probert, "A stereo vision-based aid for the visually-impaired," *Image Vision Computing J.*, vol. 16, no. 4, pp. 251–263, 1998.

[9] S. B. Nickerson, P. Jasiobedzki, D. Wilkes, M. Jenkin, E. Milios, J. Tsotsos, A. Jepson, and O. N. Bains, "The ARK project: autonomous mobile robots for known industrial environments," *Robot. Auton. Syst.*, vol. 23, no. 1–2, pp. 83–104, Oct. 1998.

[10] H. Petrie, V. Johnson, T. Strothotte, A. Raab, R. Michel, L. Reichert, and A. Schalt, "MoBIC: an aid to increase the independent mobility of blind travellers," *Br. J. Vis. Impair.*, vol. 15, no. 2, 1996.

[11] G.Greg Phillips. Miniguide: Palm-held ultrasonic ranging device, with multiple modes (range sensititivy) and audio and tactile versions. GDP Research, Adelaide, South Australia. [Online]http://users.senet.com.au/~gphillip/ultra.htm

[12] J.James Stevens, *Applied Multivariate Statistics for the Social Sciences*, 3rd ed. Hillsdale, NJ: Erlbaum, 1996.

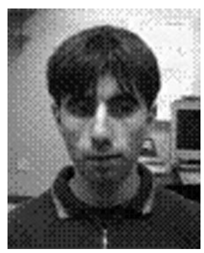
[13] I. Ulrich and J. Borenstein, "The guidecane—applying mobile robot technologies to assist the visually impaired," *IEEE Trans. Syst., Man, Cybern. A*, vol. 31, pp. 131–136, Mar. 2001.

[14] B. Walker and G. Kramer. Mappings and metaphors in auditory displays: an experimental assessment. presented at *Proc. 3rd Int. Conf. Auditory Display*. [Online]http://www.icad.org/websiteV2.0/Conferences/ICAD96/proc96/walker5.htm (accessed on Oct. 20, 2003)

Evangelos Milios (S'79–M'86–SM'98) received the B.S. degree in electrical engineering from the National Technical University of Athens, Athens, Greece, and the M.S. degree in electrical engineering, and the Ph.D. degree in computer science from the Massachusetts Institute of Technology, Cambridge.

After holding faculty positions at the University of Toronto, Toronto, ON, Canada, and York University, Toronto, he has been a Professor with the Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, since July 1998. He was Director of the Graduate Program from 1999 to 2002. His research interests include navigation and map construction in single- and multiagent robotics, networked information spaces, and Web information retrieval.

He was a member of the ACM Dissertation Award committee from 1990 to 1992 and a member of the AAAI/SIGART Doctoral Consortium Committee from 1997 to 2001.

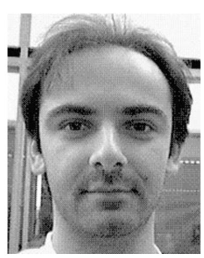


Bill Kapralos (S'01) received the B.Sc. and M.Sc. degrees in computer science from York University, Toronto, ON, Canada in 1999 and 2001, respectively. His M.Sc. research focused on the development of a video teleconferencing system combining both audio and video cues to automatically detect and track multiple speakers in a teleconferencing session. He is currently working toward the Ph.D. degree at York University.

His research interests include the development of a 3-D (spatial) audio system for York University's integrated virtual environment, a six-sided virtual environment completely enclosing its users.

Agnieszka Kopinska is currently working toward the Ph.D. degree at York University,Toronto, ON, Canada.

Her research interests include auditory-visual correspondence, in particular, how visual and auditory spatial localization is updated during eye, head and body movements, and how the brain handles the fact that sound travels slower than light.



Sotirios Stergiopoulos recieved the B.S. degree in computer science from York University, Toronto, ON, Canada.

He was previously with MacDonald Dettwiler and Array Systems Computing. He is currently an Application Developer for Palomino System Innovations, Inc., Toronto.