Sonification of Range Information for 3D Space Perception*

Evangelos Milios[1] Bill Kapralos[2] Agnieszka Kopinska [3]

Sotirios Stergiopoulos[2]

[1] Faculty of Computer Science, Dalhousie University, Halifax, Canada, B3H 1W5

[2] Department of Computer Science, York University, North York, Ontario, Canada, M3J 1P3

[3] Department of Psychology, York University, North York, Ontario, Canada M3J 1P3

October 23, 2000

Abstract

We present a device that allows 3-D Space Perception by sonification of range information obtained via a point laser range sensor. The laser range sensor is worn by the 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 3D 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

*Submitted to the IEEE Transactions of Rehabilitation Engineering, October, 2000. 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.

and accurately detect corners and depth discontinuities and to perceive the size of the surrounding space.

1 Introduction

The visually impaired rely on non-visual senses, primarily hearing, to help them locate and identify objects within their immediate and distant environment. Although all of the non-visual senses are able to convey information pertaining to an object (i.e. texture, temperature, and size), only the auditory system is capable of providing significant distance cues, which may be coded through properties such as intensity, echoes, reverberation, or Doppler shift. The Sonic Pathfinder [2] and SonicGuide [3] are examples of Electronic Travel Aids (ETAs) which sonify range information allowing object detection/avoidance by the visually impaired user. With both ETAs, range is mapped directly to pitch. In addition, both ETAs provide a directional cue. With the SonicGuide, sounds associated with objects to the left or right of the user will be louder in the respective ear. In the case of the SonicPathfinder, sounds associated with objects to the left or right of the user will only be heard in the respective ear.

To determine the distance to an object, most ETAs emit an ultrasonic sound and measure the time it takes for the echo to return. Due to the wide panoramic field of view, small range and reflection artifacts inherent with sonar, the use of sonar based devices is limited to localizing objects within an immediate and uncluttered environment [1].

They fail to provide sufficient resolution and artifact-free range data necessary for the perception of 3D space.

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 [4].

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 3D space by the visually impaired user.

Section 2 of the paper describes the device setting, including the laser range finder unit and the QuickTime Music Architecture that enables the generation of pleasant-sounding MIDI tones.

Sections 3 and 4 of the paper describes the mappings from range to audio. Proportional and derivative mode are the two major options we experimented with.

Section 5 describes experiments with human subjects. The precision and accuracy of detection of depth discontinuities, vertical and horizontal corners and object are measured.

Section 6 presents conclusions from our work and future possibilities.

2 The 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.

Figure 1 illustrates the range data sonification process. The users scan their environment and obtain eight range measurements per second (one measurement every 125ms), 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 MIDI (Musical Instrument Digital Interface) notes of different frequencies. The MIDI scale has 128 notes, and the mapping between MIDI note index m and frequency f in Hz is $f = f_0 2^{\frac{m-60}{12}}$, 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 90ms (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.

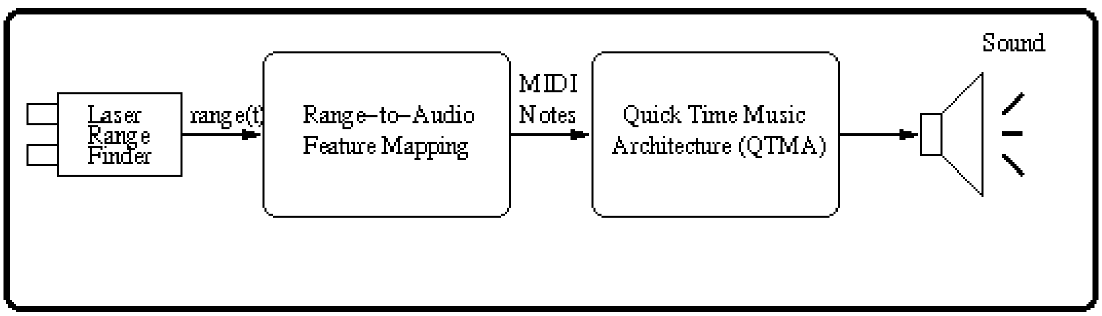

Figure 1: Overview of Sonification of Range Data.

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 the next section, we will describe the mappings in detail.

3 Proportional Mode in Range-to-Audio Mapping

Since there are no general guidelines for sonifying data [6], 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 LRF's maximum range exceeds 100m, the maximum range for the purpose of this work was restricted to 15m 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 15m 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 0.3m as it is used primarily for object detection/avoidance.

While in the proportional mode, range measurements are mapped to frequency using the mapping shown below:

$$Frequency = kc^{-(rangeMeasurement - min)a} \tag{1}$$

where $k = 4200Hz$, $c = 2.718 = e$, $a = 0.25m^{-1}$, $min = 0.30m$

The $Frequency$ value obtained above 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 (4200Hz) corresponds to a range measurement of 0.3m whereas the minimum frequency (106.46Hz) corresponds to a range measurement of 15m.

4 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).

As another example, figure 2 below illustrates how this sonification could be used to detect corners. The change between consecutive range measurements (beginning at the point labeled 'Scanning Position'), of Wall X, will continue to increase until they reach a maximum value at the corner. Then while scanning Wall Y, (beginning from the corner), depending on the user's position relative to Wall X, the change between consecutive measurements will follow another pattern. In diagram A where the user's position is close to wall X, the change in consecutive range measurements will drop significantly within a small region, then level off and once again begin to increase. As shown in figure 2B, when farther away from Wall X, the change between range measurements will drop steadily while scanning Wall Y after encountering the corner. As a result, the sonification produced with such a mapping will result in a definite pattern. For example, in figure 2B while scanning wall X towards the corner, the frequency of adjacent notes will increase until they reach a maximum at the corner and then decrease. The user may interpret this information as the presence of a corner.

Irrespective of the user's scanning position, scanning at a faster rate will result in a more drastic 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.

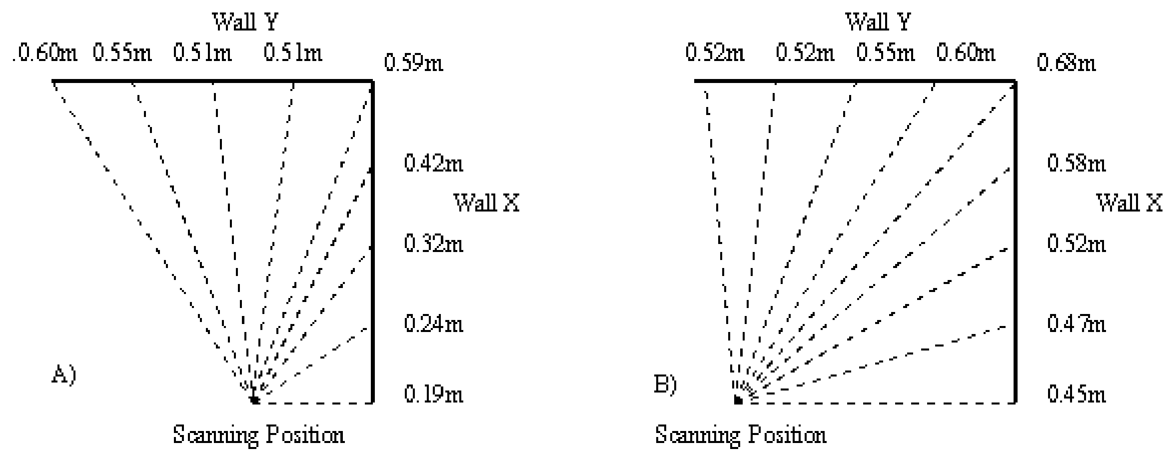

Figure 2: The detection of Corners using the Derivative mode.

Mapping parameters are as follows:

- Maximum Allowable Change in Range (r_{max}): $\pm r_{max} = \pm 2.3m$.
- Minimum Allowable Change in Range (r_{min}): $\pm r_{max} = \pm 0.15m$.
- 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 (Figure 3). Since MIDI notes come only in integer values, the result is rounded to the nearest integer.

$$MIDI = \begin{cases} 25 + \frac{N}{2}\frac{r_{max}-Dr}{r_{max}-r_{min}} & 0.15m < Dr < 2.35m \\ 60 & 0 < Dr < 0.15m \\ 69 & -0.15m < Dr < 0 \\ 69 + \frac{N}{2}\frac{|Dr|-r_{min}}{r_{max}-r_{min}} & -2.3m < Dr < -0.15m \end{cases} \quad (2)$$

The discontinuities of the above mapping occur at:

- MIDI note 25 (frequency of 35.8Hz) corresponding to the max. change between consecutive measurements.
- MIDI note 60 (frequency of 261.625Hz) corresponding to the a change of range in the interval [0, +0.15] between consecutive measurements.
- $Dr = 0$ where the MIDI note flips between 60 and 69.
- MIDI note 69 (frequency of 594.36Hz) corresponding a change in range measurements in the interval [-0.15, 0].
- MIDI note 104 (frequency of 4340.52Hz) corresponding to the max. change between consecutive measurements

From the above we conclude that adjacent MIDI notes differ by 0.06143m $\simeq$ 6.1cm of change in range. Several points should be made regarding this method.

- 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.
- Although several of the frequencies associated with positive changes in range measurements ("Moving Away") 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

5 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 (Musical Instrument Digital Interface) notes. The MIDI note was output using the Quick Time Music Architecture

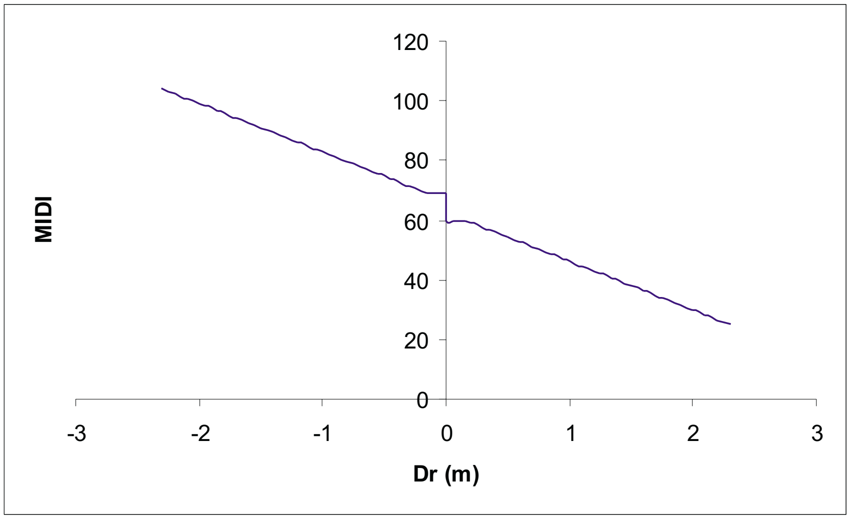

Figure 3: Graph of the piecewise linear mapping of the derivative of the range to MIDI notes.

(QTMA) software synthesizer. In particular, all notes were generated using instrument 1 - piano. Midi notes range from 1 - 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 1.

Range Readings per Second	8
LRF Accuracy	+/- 10cm (however, we have found this value is slightly more!!)
Headphone Volume	Adjustable
MIDI Note Duration	100ms - small gap of silence between consecutive sounds
MIDI Note Amplitude (Volume)	Constant value of 80 (max. allowable amplitude = 127)

Table 1: System parameters relevant to both mappings

5.1 Subjects

Seven females and five males (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. All the subjects were paid for their participation.

5.2 General Method

All the experiments were conducted in an empty, spacious room (1460cm x 420cm). Each experimental session lasted for approximately three hours for the normal vision subjects and approximately five hours for the visually impaired subjects. At the beginning of the session the subjects received between 45 minutes to an hour of training on how to use the device (the training took approximately 2 hours 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 set-up. 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 Figure 4. 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 standard deviation of the repetitive pointing errors was used as a measure of the precision. The data was further analyzed using the Repeated Measures ANOVA [5].

5.2.1 Experiment 1: Detection of Depth Discontinuities

As illustrated in figure 5, the subjects were positioned in front (30cm to the left or right with respect to the centre of a gap) of a vertical gap at a distance of two, six and 12 metres. The gap was 20cm wide and 30cm 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 positive and pointing away from the gap as negative.

The subject's accuracy for the different distances for the proportional mode ranged from -0.009^o to

Figure 4: Photograph of a subject during an experiment.

0.301^o and for the derivative mode from 0.025^o to 0.214^o. The precision ranged from $\pm 0.183^o$ to $\pm 0.549^o$ for the proportional mode and $\pm 0.212^o$ to $\pm 0.469^o$ for the derivative mode. For the proportional mode the overall accuracy was 0.133^o and precision $\pm 0.349^o$ and for the derivative mode 0.104 ± 0.319^o respectively. The accuracy and precision for the visually impaired subjects were -0.009 ± 0.549^o and -0.005 ± 0.304^o for the proportional mode and 0.085 ± 0.469^o and 0.056 ± 0.329^o for the derivative mode.

There were no differences in accuracy ($F_{1,360} = 1.179, NS$) and precision ($F_{1,72} = 0.014, NS$) between the two modes. Figure 6 illustrates accuracy and precision for different modes and distances across all the subjects.

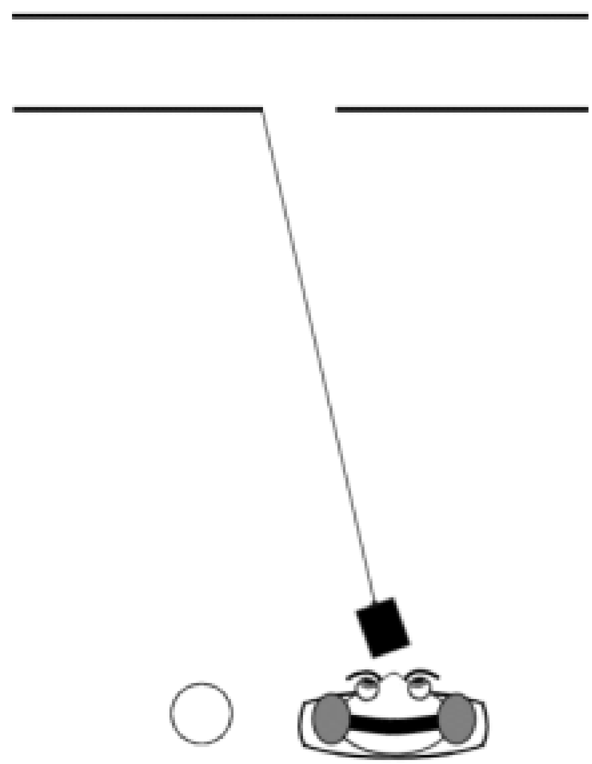

Figure 5: Experiment 1 - experimental set-up. The subject positioned in front (30cm to the left or right with respect to the centre of a gap) of a vertical gap at a distance of two, six and 12 metres. The gap was 20cm wide and 30cm deep. The subjects were asked to indicate the left and right edge of the gap.

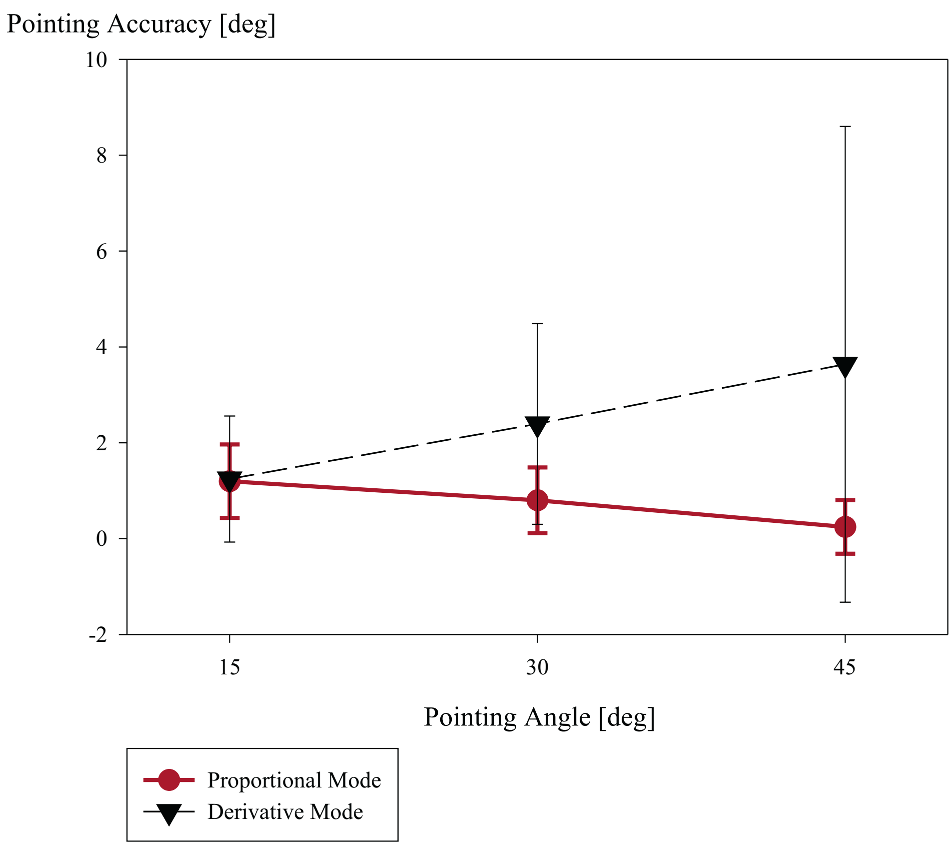

Figure 6: Means and standard deviations of pointing accuracy when pointing while in the proportional (grey circles) or derivative (black triangles) mode at the vertical gap located at two, six and 12 metres.

5.2.2 Experiment 2: Detection of Vertical and Horizontal Corners

To measure the subjects' ability to detect vertical corners, the subjects were positioned at a distance of three metres away from a corner between two walls. The task was performed from three different angular positions, i.e. 15, 30 and 45 degrees with respect to the corner (see figure 7). The subjects were never directly facing the corner; they were rotated by 15 - 20 degrees to the left or right with respect to the corner. There were five repetitions for each position tested.

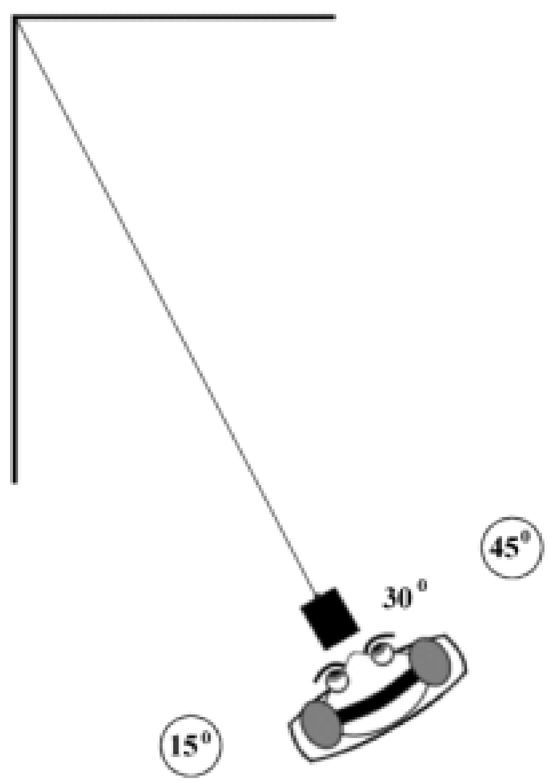

Figure 7: Experiment 2 - experimental set-up. The subject positioned at a distance of three metres away from a corner between two walls. The task was performed from three different angular positions: 15, 30 and 45 degrees with respect to the corner.

To measure the subjects' ability to detect horizontal corners, the subjects were positioned at two, six, and 12 metres 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.

Vertical Corner. The subject's accuracy for the different angles of pointing for the proportional mode ranged from -0.771^o to 2.017^o and for the derivative mode from -7.459^o to 8.517^o. The precision ranged from $\pm 0.136^o$ to $\pm 2.026^o$ for the proportional mode and $\pm 0.334^o$ to $\pm 10.566^o$ for the derivative mode. For the proportional mode the overall accuracy was 0.747^o and precision $\pm 1.371^o$ and for the derivative mode 2.424 ± 4.355^o. The accuracy and precision for the visually impaired subjects were 0.239 ± 0.565^o and 0.815 ± 1.461^o for the proportional mode and -2.607 ± 3.693^o and 3.596 ± 3.576 for the derivative mode.

Figure 8 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 ($F_{2,22} = 9.447, p < 0.001$).

Horizontal Corner. The subject's accuracy for the different distances for the proportional mode ranged from -2.223^o to 2.195^o and for the derivative mode from -5.191^o to 4.657^o. The precision ranged from $\pm 0.072^o$ to $\pm 3.225^o$ for the proportional mode and $\pm 0.212^o$ to $\pm 3.577^o$ for the derivative mode. For the proportional mode the overall accuracy was 0.515^o and precision $\pm 1.493^o$ and for the derivative mode 0.815 ± 2.266^o. The accuracy and precision for the visually impaired subjects were -0.560 ± 1.829^o and 0.791 ± 0.91^o for the proportional mode and 0.915 ± 1.064^o and 0.130 ± 1.519^o for the derivative mode.

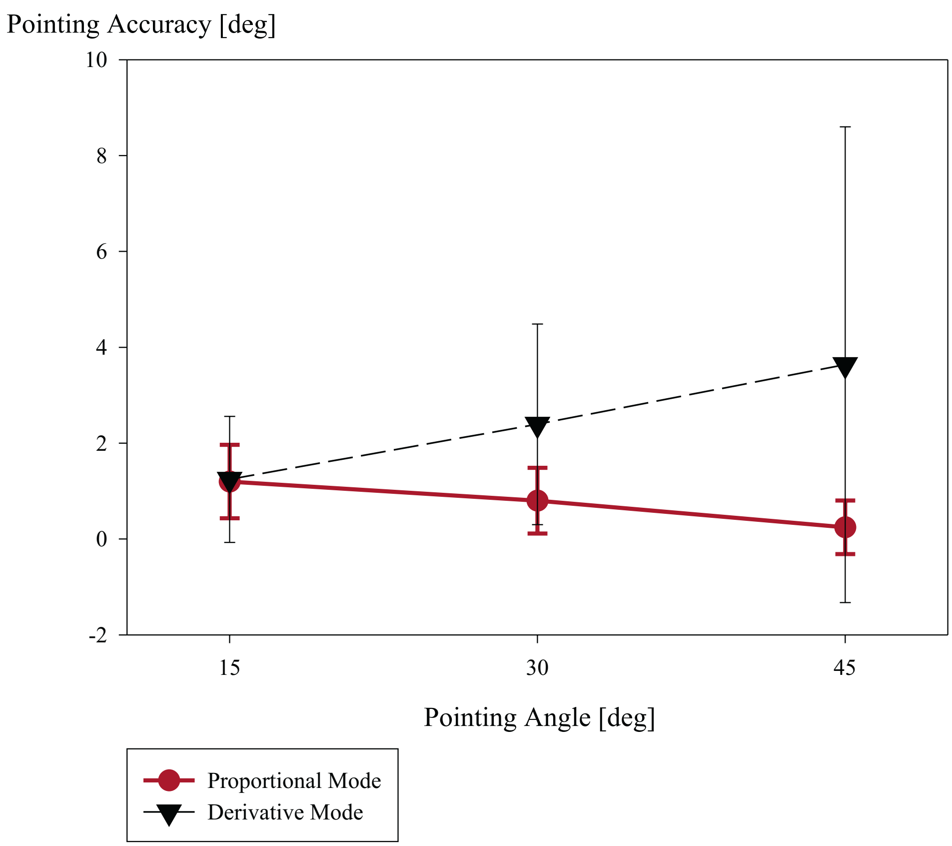

Figure 8: Means and standard deviations of pointing accuracy when pointing while in the proportional (grey circles) or derivative (black triangles) mode at the vertical corner.

Figure 9 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 ($F_{2,22} = 48.456, p < 0.0001$).

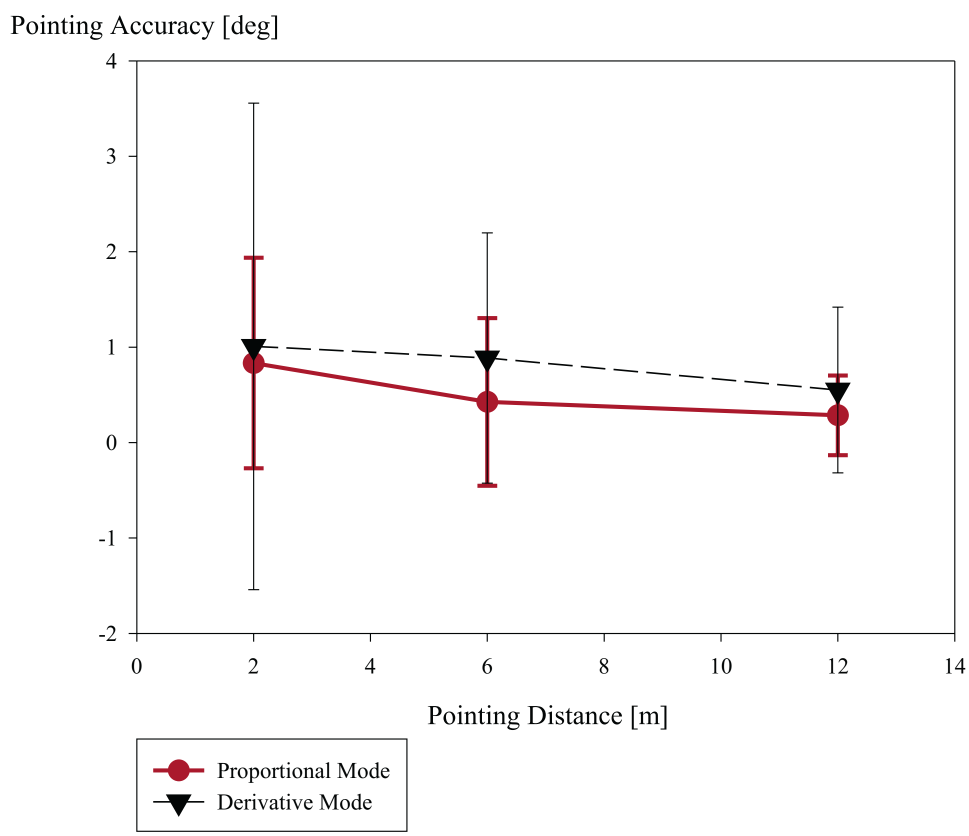

Figure 9: Means and standard deviations of pointing accuracy when pointing while in the proportional (grey circles) or derivative (black triangles) mode at the horizontal corner.

5.3 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 figure 10.

When pointing using the derivative mode seven 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, eight 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.

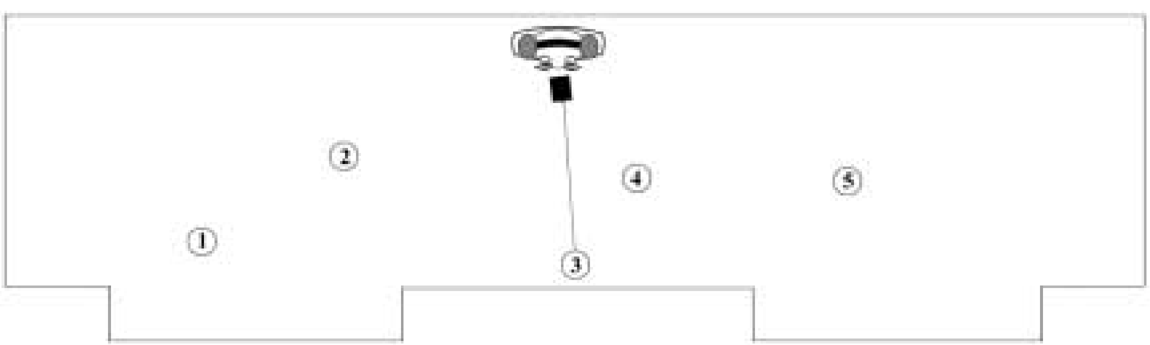

Figure 10: Experiment 3 - experimental set-up. The subject scanned the surrounding space and pointed the LRF at the experimenter that was positioned at one of the five locations.

6 Conclusions

The study suggests that the sonification of range measurements is quite effective in conveying information about spatial features. With both proportional and derivative mode the subjects were able accurately and precisely locate depth discontinuities as well as vertical and horizontal corners. The only exception to this, was pointing while in the derivative mode at the vertical corner from the angle of 45 degrees. 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.

Acknowledgements. The work was supported by a research grant from the National Science and Engineering Council of Canada (NSERC) and an NSERC summer research scholarship to B. Kapralos. We thank Prof. Patrick Dymond for suggesting the use of MIDI, Prof. Laurence Harris for suggesting the use of derivative mappings, and Robert Arrabito of DCIEM for his valuable comments and his encouragement. Greg Reid provided help with audio and signal processing facilities.

References

[1] Easton and D. Randolph. Inherent problems of attempts to apply sonar and vibrotactile sensory aid technology to the perceptual needs of the blind. *Optometry and Vision Science*, 69(1):3–14, 1992.

[2] A. Heyes. The sonic pathfinder: A new electronic travel aid. *Journal of Visual Impairment and Blindness*, 77:200–202, 1984.

[3] L. Kay. Acoustic coupling to the ears in binaural sensory aids. *Journal of Visual Impairment and Blindness*, 77:12–16, 1984.

[4] 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. *Robotics and Autonomous Systems*, 23(1-2):83–104, Oct. 1998.

[5] James Stevens. *Applied Multivariate Statistics for the Social Sciences (3rd ed.)*. Lawrence Erlbaum Associates, Inc., Mahway NJ, 1996.

[6] B. Walker and G. Kramer. Mappings and metaphors in auditory displays: An experimental assessment. In *Proc. of the Third International Conference on Auditory Display*, Palo Alto, California, Nov. 4-6 1996.

List of Figures