Interfaces for multi-sensor systems for navigation for the blind

Penny Probert, David Lee and Gordon Kao

Robotics Research Group, Department of Engineering Science, University of Oxford,
19 Parks Road, Oxford, OX1 3PJ, ENGLAND
Tel: +44-1865-273926 Fax: +44-1865-273908

pjp@robots.ox.ac.uk, david@robots.ox.ac.uk

ABSTRACT

This paper describes work towards a multi-sensor system to assist blind people to move around an urban or indoor environment. The communication of navigation information is constrained by both the type of information imparted and, in many ways more crucially, by the type of information which real sensors can extract. In this paper we describe the use of ultrasound to provide two types of information: first the low level directional information which may be provided currently by a guide dog and then information on features to allow the blind person to locate and direct themselves within the environment. A directional system with a simple vibrator interface is described and its performance assessed by a group of users. Finally we discuss the possibility of feature extraction for navigation by recognition and give some results using frequency modulated sonar.

Keywords: ultrasound, mobility aids, blindness, human interface

1. BACKGROUND AND STRUCTURE OF PAPER

Vision is by far the most important sense for most people, and trends in virtual reality and computer interfaces reflect this emphasis. Sighted people appreciate detailed images and are able to abstract relevant data from images with ease, and the challenge in virtual reality is not in abstracting the data but providing enough detail at sufficient speed. In contrast, the difficulty in presenting information to blind people is reducing information: non-visual channels are less direct and the available sensory channels have much lower bandwidth.

In this paper we address the problem of abstracting and presenting information for navigation. A number of sensors to assist mobility have been on the market for some time, normally using ultrasound as the sensing modality. However they have not met with widespread success. We see two reasons for this. One is that the information is often presented as a continuous, or near continuous, tone, usually through headphones, whereas blind people need their ears to pick up other signals from the environment. The work on the auditory representation of the environment is encouraging as it shows the possibilities of using sound as a rich information system, but does not overcome this problem [Kay95]. Another reason for failure is the restriction on the information which can be gathered. In the ideal world people would like a system which can pick out and convey information on landmarks such as bus stops, on pavement hazards such as parked cars and holes, and on general environmental features such as steps, and to pass on such information in a form so that the user can build up an internal map of his immediate location and probably relate to previous information. Unfortunately this is a very difficult goal to achieve, especially if significant data abstraction has to occur computationally to leave the audio channels free.

Another method of presenting information to the blind, which is more directly analagous to vision in providing spatial awareness, is through a tactile interface. There has been considerable research over the last two to three decades on the provision of an array which might act as a form of vision for blind people and some work suggests that certain functions, for example to do with grasping objects, could be stimulated by this method [Bach-y-Rita95]. However it seems clear that a fully visual sense cannot be conveyed, especially to older people, partly because of the quality of tactile array which can be made and partly because of the problems of interpretation by the brain. Work on the provision of tactile maps has similar problems because of technological problems in making tactile arrays: precision, dynamic range and cost are all major limitations [Frick96]. In addition the tactile sense gets tired easily, and for a large category of the blind, those whose blindness is associated with diabetes, it is often degraded.

There must therefore be a balance between the extent to which information should be abstracted before presentation to the user, which depends on the type of navigational information needed and on the sensing modalities available. Computer vision can provide huge amounts of data but abstraction is difficult and the complexity is unnecessary for simple navigational commands. In this paper we extract two levels of navigational information, and discuss how to present information for each. We describe the progress so far using ultrasound sensing, a compact and low cost technology.

2. INFORMATION PROVISION FOR NAVIGATION

We can view navigation as one of two tasks: a low level task in which someone simply wants directions to follow a safe route along a known path (which we call micro-navigation) , and a more information based approach in which the user wants to move him or herself through the environment by recognising and navigating from landmarks. Current mobility aids fulfil mainly the first of these. Guide dogs direct a person along a path through pressure on a handle and convey no higher level information on context. The long cane can provide more information since as well as warning of hazards it can be used to probe the immediate environment, both through touch and through the echo from a tapping sound. However the information it gathers is restricted and neither aid provides information on anything but the very close environment.

The work we describe in the next sections consists of two systems, one for micro navigation and the other for feature based guidance. For the former we draw from work on robot guidance using time of flight sonar, which has been shown to work reliably in a wide range of environments [Borenstein88]. These sonar devices simply return the range of the nearest obstacle in the path of an ultrasound beam. Ultrasound sensors are cheap, have a wide field of view and reasonably modest power requirements. Feature based guidance is more of a problem. There is no well developed system to draw on here; all available techniques either require very well structured environments or are very restrictive. A different type of sonar sensor, which has been used for blind people through an audio interface (various configurations have been marketed such as the Sonicguide [see RNIB94]), offers an interesting solution, and keeps the cost of the system low. This is the basis of the system we describe in section 4, but, in contrast to the sensor as sold, we reduce the quantity of information presented audibly by including some automatic data abstraction. Ultimately other sensors, such as vision, may be used as well to extend the range of features which can be observed. Computer vision is sometimes seen as the panacea for all these problems but even with full stereo vision it is hard to identify many features in an uncontrolled environment because of the problems of changes in lighting, poor calibration from camera movement and occlusion. The configuration of ultrasound we use has good immunity to noise and, because of the wide beam, can generally handle partial occlusion.

We make no attempt to recognise detailed man made landmarks such as bus stops or telephone kiosks. As well as the fact that there is a huge variety of these and it is hard to build a generic model, a much more sensible approach to providing location is through a talking beacon identifying the landmark (see, for example, [OPEN95]).

For micro-navigation, the user requires a stream of directional information to guide him along a path, and around obstructions. One simple way of doing this, which has proved successful for guide dogs is using a handle. However, as we had expected, and following other studies [Petrie95] potential users favour hands free operation so although the handle seems to have advantages in terms of reliability and ease of use this option was rejected. The channels remaining are tactile or auditory. As we discussed in the last section, each of these presents problems, and detailed information cannot be communicated through a tactile interface. Therefore we felt that there was no point in trying to provide a detailed local image, instead experimenting with a simple reactive interface, in which we simply provide instructions such as: “turn right, turn left, turn around”. We describe a system which communicates this information in the next section. The first implementation was simply through an audio interface, and, as predicted, it proved irritating in use. However for so few commands it is easy to implement a tactile interface with vibrators, which are similar to the buzzers in pagers. Experiments showed that for the cheaper vibrators the human skin can detect very little in the way of frequency or amplitude of vibration, so a simple on-off interface was used. In the next section we present results from users of a sonar belt using this simple interface, which has proved very successful.

The higher level of navigation is through feature recognition, maybe to assist in localisation in a known street or office environment. Information is likely to be required less frequently, possibly only on demand and so the auditory interface is acceptable and versatile. At present we are investigating the level at which to process information in this area: whether to attempt full pattern recognition computationally and communicate in words, or to present the information as a sound following a minimum of processing pattern (an option which may offer better feature recognition but at the expense of requiring more of the audio channel). We discuss these issues further in section 4.

3. MICRO NAVIGATION: FOLLOWING A ROUTE SAFELY

The micro navigation task is implemented using time of flight sonar sensors. These emit a brief burst of ultrasound at about 50 kHz and then measure the time until an echo is detected. Multiplying this time by the speed of sound gives the round-trip distance to the reflecting object, which in turn gives its range. The key properties of this sensor are:

- Range accuracy of about 1 cm when the object lies in the centre of the sensor's beam. Overestimates of up to 4 cm can occur when the object is to the side of the beam.
- Low angular precision. The effective width of the ultrasonic beam is often quoted as 30°, although certain types of object can be only be detected through angles as small as 18°.
- Only the first reflecting object within the beam is detected.

A 'sonar belt' was constructed with three time-of-flight transducers mounted on it, directed horizontally. The angle between adjacent transducers was set at 15°. This was found to be an effective compromise between the need to ensure that all obstacles will be detected and the desire to cover a wide area in front of the user. With this separation a region about 60 cm wide was covered 2 m in front of the user.

The three transducers fire in sequence. To avoid crosstalk, enough time was left for an echo to have returned from an object at the maximum range of interest before the next sensor was fired. The shorter the imposed maximum range, the more frequently the sensors can be fired. A maximum range of 3 m was selected (considerably less than the theoretical maximum of about 10 m), making it possible to collect range data at approximately 50 Hz.

Two different versions of the sonar belt were constructed and tested. The first used a set of audible tones and the second used vibrating motors. These two versions are described in the following sections.

3.1 Audible Output

Section 1 described the possibility of analysing the sensor results before passing them to the user. As a first example of this approach, this version of the sonar belt suggested a safe direction of travel to the user, instead of simply reporting the approximate location of obstacles. A threshold value was selected (2 m in the experiments reported here) and the user was given turn instructions depending upon which, if any, of the range readings is less than the threshold. A simple rule base translated the sonar readings into commands.

The turn actions were given to the user through an earpiece. Three different tone frequencies were used for 'turn left', 'turn right' and 'stop'. The system was silent when the user could move straight ahead.

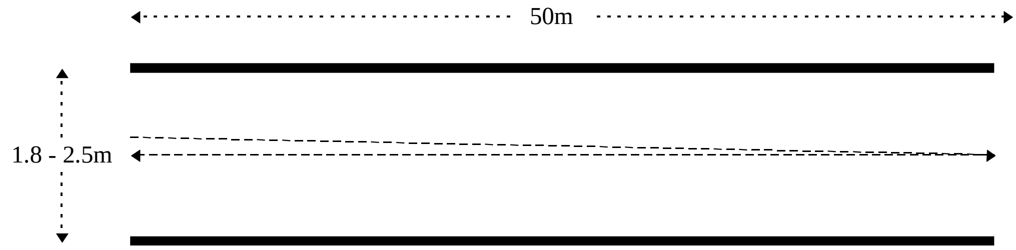


Figure 3-1. *Test run for the audio sonar belt*

To test the system, the user was required to walk blindfold along a test path which was bounded on both sides by walls, turn around at the end (when directed to do so by an assistant) and walk back to the starting point. The user's objective was to avoid collisions both with the walls and with other pedestrians using the path. The dimensions of the test path were as shown in Figure 3-1. To give a 'base case' for comparison, the same test was performed by a trained volunteer using a long cane.

Table 3-1 show the experimental results. A user of the belt was able to complete the test route at an average speed of 0.77 m/s, demonstrating the potential value of time-of-flight sensors in obstacle avoidance and path following. This speed is about 82% of the average speed that was achieved by the long cane user.

Most of the small number of collisions and sudden stops in the test results were caused by interactions with other pedestrians using the path. The narrowness of the path made it difficult to negotiate a clear path past the pedestrian. No meaningful comparison can be made with the number of collisions made by the long cane user because the collisions of the cane are the sensing mechanism itself.

Table 3-1. *Test results from audio sonar belt*

Test Run	Time (sec)	Speed (m/sec)	Collisions or Sudden Stops
1	145	0.69	1
2	130	0.77	2
3	141	0.71	0
4	135	0.74	0
5	154	0.65	0
6	111	0.90	0
7	132	0.76	1
8	129	0.78	0
9	110	0.91	0
10	119	0.84	1
Average	**130.6**	**0.77**	**0.4**

The following observations were made during these experiments:

- Listening to the environment was very helpful. It was, for example, easy to tell when the end of the walled pathway had been reached because the echoes from footsteps became much quieter. The presence of the earpiece and the turn signals could mask some of these audible clues.
- A fraction of a second was needed to convert the tone signal into a direction of travel.
- In some circumstances, the sensors are used for active examination of the environment. The user could, for example, make small body movements to determine the extent of an obstacle. Immediate sensor data, instead of turn signals, is more useful in this situation. The use of the rule base to convert the sonar signals into different turn actions appears unnecessary.
- The same range threshold is not appropriate in all circumstances. It would be helpful for the user to be able to select different thresholds in different environments.

With these points in mind, the tactile version of the belt was created and submitted to tests by blind and visually-impaired users. This device is described in the following section.

3.2 Tactile Output

This version of the sonar belt differs from the previous version in the following ways:

- Each transducer has an associated vibrating motor which is mounted on the inside of the belt, close to the transducer.
- Each motor vibrates constantly when the range measurement of the corresponding sensor is less than the threshold. No pre-processing of the range data is performed.
- The user can adjust the range threshold as necessary.

This system has recently been tested by 16 blind and visually-impaired people at the training centre of the Irish Guide Dogs Association in Cork. Of these people, thirteen used a guide dog and three a long cane, and fourteen made extensive use of these. They had the following profile:

1. Age

under 25	24-45	45-65	65+
0	9	6	1

2. Visual loss

Total	Perception of light only	Poor partial sight	Useful residual vision
10	2	1	3

Proc. 1st Euro. Conf. Disability, Virtual Reality & Assoc. Tech., Maidenhead, UK, 1996
©1996 ECDVRAT and University of Reading, UK ; ISBN 0 7049 1140 X

3. Home environment	City	Urban	Rural
	5	9	2

They were asked to negotiate the obstacle course used for training guide dogs and their handlers. The obstacle is delimited by a chain-link fence (which was found to be easily detectable by the sonar transducers, but is a problem for the long cane user as there is no ground plane obstruction), and consists of items such as a dustbin, a typical work in progress sign of the type on pavements, a bicycle and a horizontal barrier. Figure 3-2 shows the belt being tested by two mobility instructors at the training centre.

Figure 3-2. *Testing the tactile sonar belt*

The group of blind people underwent a day's training by the mobility instructors and were then asked their opinion of the belt. The functionality for path guidance was rated highly, with the vibrator operation deemed intuitive and well liked. The only negative comments were to do with the mechanical configuration, which 'did not allow for a sufficient range of stomach shapes!' and there were some problems owing to sonar 'blind spots'. These are easy to adjust and our next version of the belt should have all round coverage. However on the cautious side it must be borne in mind that this group were already highly mobile so the results cannot be taken as indicative of success over the whole range of the visually disabled. Indeed one problem with a tactile interface is the loss of touch sensation which accompanies diabetes, often a cause of blindness amongst the elderly, and for this group an audible or handle type of interface might be more appropriate.

4. COMMUNICATING FEATURES FOR NAVIGATION

For micro-navigation a simple interface seems adequate. The other demand for navigation is for the provision of information on features. The time of flight sonar sensors we have discussed throw away far too much information to be useful in feature recognition. In this section we describe the frequency modulated ultrasound sensor, which is proving useful for extracting certain types of feature information. We describe the sensor briefly, the information which it provides and then possible interfaces to the user.

4.1 Principle of Operation

The sensor consists of a separate transmitter and receiver. The frequency of the transmitter is modulated in a known pattern, and the frequency of the signal being transmitted is compared to that of the signal at the receiver, which was transmitted some time ago and has now returned after reflection. For the linear modulation which is shown in Figure 4.1, the difference frequency is proportional to the time of flight and therefore to range. This difference frequency is extracted automatically at the receiver.

4.2 Direct presentation of the signal

The difference signal is in the audible range for distances of up to a few metres. Sensors such as the Sonicguide, which work on the same principle, transmit the difference signal as a continuous tone to the user through an earpiece, thus presenting an auditory map of the environment. Different ranges appear as different pitches, and the loudness of the sound indicates how large a reflection occurred at that range. The user is provided with a rich source of information, with the possibility of distinguishing between single and multiple objects and of learning the ‘sound’ of particular feature shapes. However the main problem with the direct presentation of the beat signal is that it monopolises the user's sense of hearing, a vital source of information for a blind person. Interpretation of the signal also requires a lengthy training period before the signal can be well understood.

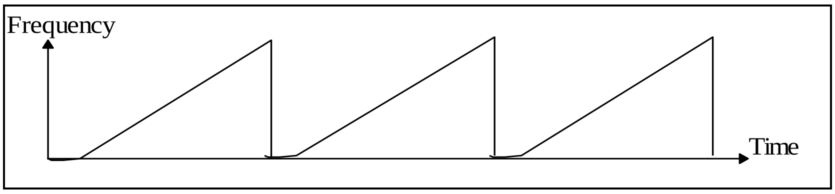


Figure 4.1: *the variation of frequency with time in the transmitter. Three cycles are shown. The frequency varies between 45kHz and 90kHz over a time of 160msec, and there is a blanking time of 24msec between the cycles. The blanking period is made long enough to avoid ambiguity between cycles and so determines the maximum range which can be measured.*

Nevertheless in spite of the disadvantages a small group of blind people use the Sonicguide, which consists of a central transmitter and two receivers built into spectacles, to great effect. Dedicated users report extraordinary perceptive powers with this technology: for example a highly mobile user reported recently to one of the authors that he could detect the difference between conifers and deciduous trees; others have reported being able to detect whether roses were in flower or not. However for most people the device remain inaccessible at least partly because of the interface.

4.3 Extraction of range information

The problem then is in how to extract and present information without overloading the auditory channels. A simple way to do this might be simply to provide information on demand; however for a direct interface it is likely that quite a long burst of the signal might be required. An additional problem, even for experienced and dedicated users, is that small features can get lost in the signal; for example the user is unlikely to pick up a pole in front of a large wall, since the reflection from the pole is swamped.

The range information in the signal at the receiver is in its frequency content, and the relationship between a signal in time and in frequency is given by the Fourier transform. From the Fourier transform it is easy to derive the power spectral density, which shows the amount of energy recovered at each range. For a single well defined surface, for example, the range image (which is equivalent to the power spectral density) shows a clear peak at a single range (Figure 4-2)

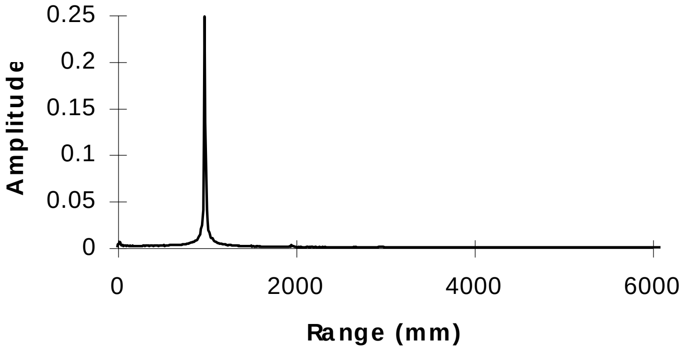
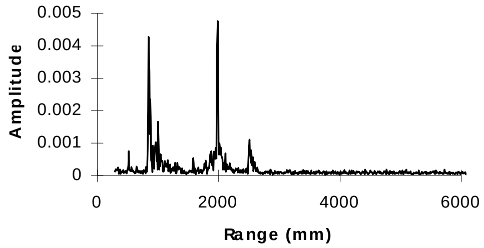


Figure 4-2. *A range image of the sensor pointed towards a wall*

When the sensor is pointing at a surface at normal incidence, the texture of the surface is indicated by the width of the peak. The height of the peak depends on the amount of energy reflected back to the receiver. More complex features are shown below. Note the multiple reflections from the car (from the front of the bonnet, from the windscreen/bonnet join and, the smaller furthest peak, probably from the top of the windscreen). The steps show a characteristic periodic structure.

Proc. 1st Euro. Conf. Disability, Virtual Reality & Assoc. Tech., Maidenhead, UK, 1996
©1996 ECDVRAT and University of Reading, UK ; ISBN 0 7049 1140 X



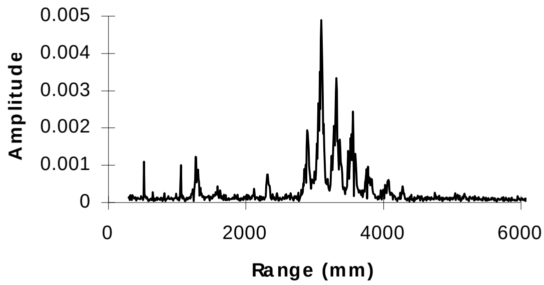


Figure 4-3. *Range maps from (a) the front of a car and (b) a flight of stone steps*

The visual shape of these gives us some ideas for how landmarks may be detected, and the information represented. For example it is easy to pick out some simple features from these range maps: periodicity, the positioning of multiple objects and an indication of texture. Unfortunately, of course, the visual interface is not accessible to the blind (and if it were the sensor would not be needed anyway, except in the dark!) so another method of presentation must be determined. A tactile array which reproduces the pattern is one possibility, but as we said earlier there remain significant technical difficulties in producing arrays with sufficient dynamic range to handle the problem of a high reflectivity background (such as the pole/wall example).

We are looking at two types of auditory interface instead, balancing the possibilities of automatic feature extraction against information bandwidth. One takes information from the range map and uses pitch, volume and rhythm to code distance and peak amplitude. Since most people are more sensitive to differences in pitch than volume it may overcome the background problem. The second attempts to provide direct voice output describing the feature and information on its position. It depends on the success of an automatic feature extraction system to recognise certain categories of object. In the final section we look briefly at the possibilities in this area.

4.4 Extraction of feature description

Automatic feature extraction is not straightforward. As a start in this area we have developed algorithms to detect periodicity (for example steps) and texture. They use the statistical properties of the range map, and the Fourier transform to examine periodicity. In more detail there are three stages:

4.4.1. Determining the area of interest. Experiments on the range maps show that spurious peaks are introduced at close range (less than about 0.5m) from imperfections in the device and in the finite window size of the Fourier transform and so these are not included in the processing. To extract interesting features from the rest of the map a sliding threshold is used, which is based on the intensity integrated over the rest of the range map. Then only those parts of the image above the threshold are retained.

4.4.2. Distinguishing smooth surfaces from groups of objects and textured surfaces. The shape can most easily be characterised by the statistics of the first and second moments of the peaks in the area of interest: i.e. the mean and variance. Using the raw power spectrum however was found to be too sensitive to noise. A better method was to determine the autocorrelation function and then apply tests to examine its width. Figure 4-4 shows the results from several features, which can then be distinguished by a simple classifier.

Taking the autocorrelation function (ACF) provides a significant reduction in the effects of noise. However the method was found most robust if the power spectrum was reduced to a binary image before the autocorrelation (it also, of course, speeds up the algorithm). This is presumably because the thresholding (which was based on the mean intensity over the whole area of interest) acts as a crude low pass filter.

The rate of decay of the ACF is characteristic of the surface, with smooth single surfaces showing a fast rate of decay and periodic surfaces showing clear peaks in the function. Although at first sight the some raw power spectrum data appear as if there may be periodicity, it is clear from the autocorrelation function that such periodicity, if any, is very weak. The Fourier transform of the ACF can be used to picks out any periodicity (Figure 4-5); for non-periodic signals no energy lies above the adaptive threshold.

Because of the long wavelength of ultrasound (a few mm), the amount of energy reflected back to the receiver depends strongly on angle. When there is less energy, the peak spreads even for smooth surfaces because of the effects of noise. However work based on modelling following Kuc [Kuc92] suggest that this can be exploited to determine the angle of the reflecting surface; for example the angle of a wall. This again could be a useful indicator for navigation.

An automatic system for feature recognition based on these ideas, which communicates through a voice box, has been developed. The mean and variance of the ACFs is used to distinguish between smooth surfaces and more complex features. Early trials are promising but low signal levels (caused, for example, because a surface is viewed obliquely) cause errors. Thresholding the peaks on the Fourier transform of the ACF proved a good method of distinguishing periodicity. The user interface provides information on the class of object and its range using words. It would be nice to be able to project position through stereo sound so the user could just hear something shouting out its name, but although it is easy to simulate orientation the problems in projecting range realistically remain unsolved [Loomis95].

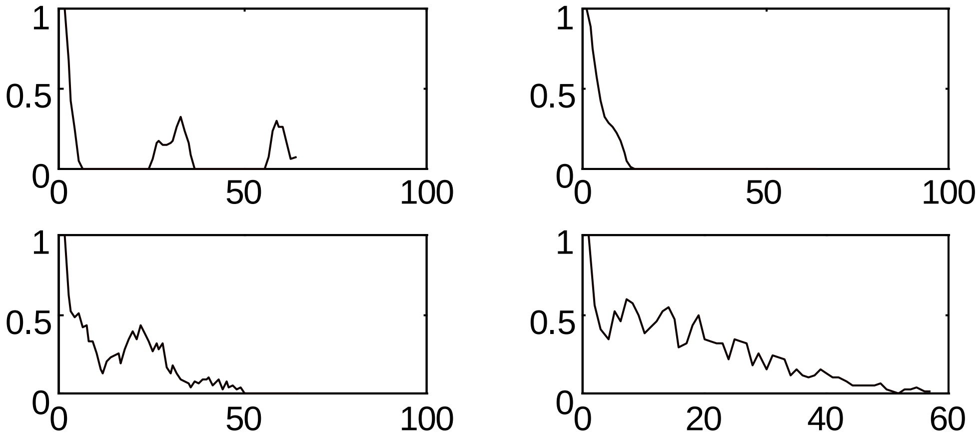


Figure 4-4. *The autocorrelation functions for steps (top right), branches (bottom right), a smooth wall (top left) and a cluster of rubbish bags (bottom right)*

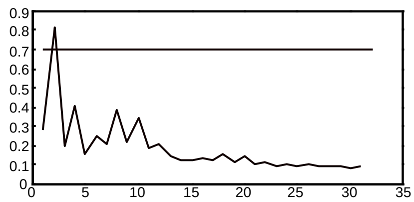


Figure 4-5. *Picking out periodicity: the FFT of the autocorrelation functions for the steps. Note the line marking the threshold, which depends on the average intensity in the image.*

5. CONCLUSIONS

In this paper we have described a two component system for navigation, which imparts directional information for local guidance through vibrators and higher level information on features through an audio interface. One of the ongoing themes in the project is how far information should be abstracted before it is presented to the user. People are far better than computers at recognising patterns, but the presentation of large amounts of raw data can overload the sensing systems available to the blind. The use of vibrators for the low level directional information has proved popular in user trials, but there remain problems of robustness, ease of use and cost if this method is ever going to have widespread acceptance. The early work on the frequency modulated sonar for providing feature information is encouraging but it is too early still to know whether a sufficient range of features can be included, either with the sonar alone or in conjunction with another sensor such as vision.

Acknowledgements: This work is funded by the European Union TIDE programme. One of the authors is supported by an EPSRC grant. We thank Simon Higgs of the Irish Guide Dogs Association for running the trials there and David Witt of Oxford University Department of Engineering Science for the frequency modulated sonar device.

6. REFERENCES

P. Bach-y-Rita and B. Hughes: " Tactile Vision Substitution: Some Instrumentation and Perceptual Considerations", in Electronic Spatial Sensing for the Blind, ed. D.H. Warren and E.R. Strelow, NATO ASI Series, Martinus Nijhoff 1985, pp125-140

J. Borenstein and Y. Koren, "Obstacle Avoidance with Ultrasonic Sensors", IEEE Trans. on Robotics and Automation, 42, 1988, pp213-218

O. Bozma and R. Kuc: "Characterising the Environment using Echo Energy, Duration and Range: the ENDURA Method", IEEE/RSJ International Conference on Intelligent Robots and Systems, 1992, pp813-820

J. Fricke "Tactile Interfaces to Graphical Information", Rank Prize Fund Symposium on Technology to Assist the Blind and Visually Impaired, Grasmere, UK, 1996

L. Kay: "Sensory Aids to Spatial Perception for Blind Persons: their Design and Evaluation" in Electronic Spatial Sensing for the Blind, ed. D.H. Warren and E.R. Strelow, NATO ASI Series, Martinus Nijhoff 1985, pp125-140

J. Loomis: "Personal Guidance System for Blind Persons", Conference on Orientation and Navigation Systems for Blind Persons, University of Hertfordshire, 1995

OPEN project: "Audio Interface", CC-TIDE Project TP 1182, 1995, issued by Royal National Instititute for the Blind, UK

H. Petrie: "User requirements for a GPS Travel Aid for Blind People", Conference on Orientation and Navigation Systems for Blind Persons, University of Hertfordshire, 1995

RNIB (Royal National Institute for the Blind): "Electronic Mobility Devices or Obstacle Detectors for Blind and Partially Sighted People", 1994