Exploring Direct Downconversion of Ultrasound for Human Echolocation

Shane D. Pinder
School of Engineering
Auckland University of Technology
Auckland, New Zealand
shane.pinder@aut.ac.nz

T. Claire Davies
Systems Design Engineering
University of Waterloo
Waterloo, Canada
cdavies@uwaterloo.ca

ABSTRACT
Human-computer interfaces for mobility devices to aid visually impaired individuals are deficient in the ease of use of systems. This paper examines the techniques used by blind individuals to spatialize naturally in their environments and seeks to explore the possibility of providing the same information in a manner that may be otherwise more socially acceptable.

Author Keywords
Human echolocation, visually impaired, audio interfaces.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
How do humans see with sound? There are well-established physical mechanisms by which the sense of hearing can be used to determine the direction from which a sound arrives, and there has been considerable research in this field. However, we lack a true understanding of what constitutes an ideal sound stimulus for human beings in any particular situation. Consider an athlete performing in a team sport, where the urgency of communication from a team-mate must elicit an immediate response. How do we better influence the human response to sound? Will we, in time, reduce this qualitative evaluation of the effectiveness of sound stimuli to an exact science?

Human-machine interfaces are becoming increasingly dependent on the ability of the user to localize sound as a means to reduce cognitive load, thereby increasing task efficiency. Bimodal aircraft instruments designed to elicit a specific response to an emergency, for instance, provide the pilot with the benefit of distributing their attention across both visual and auditory senses. Some characteristics of sounds easily localized by humans have been suggested[6,12,14], such as wide spectral variation, though there exists the possibility that better-suited waveforms may be identified if we permit a wider set of parameters to be considered.

BACKGROUND
Understanding the methods people use to spatialize their auditory environment is becoming important for the enhancement of human-computer interfaces. People with visual impairments have learned to capitalize on information from their auditory environments to allow them to locomote effectively and efficiently[13]. We can learn from them how to better understand our environments.

This research follows from studies done with both blindfolded and visually impaired individuals to evaluate their methods of understanding their environments [9]. For instance, one study involved the use of the long white cane to detect obstacles in the environment and step over them. This study showed that the kinesthetic feedforward information provided by the cane did not allow sufficient information to accurately describe the height of obstacles and as such the participants overcompensated for the lack of information provided. Instead of stepping over the obstacle with only sufficient clearance for the toe of the lead and trail limb as a person who had full vision would do, the feet were both lifted much higher to ensure adequate clearance.

Since cane clearance did not adequately allow these individuals sufficient information to effectively clear obstacles, it was important to look at other devices that enabled visually impaired travellers to detect obstacles in their environment. As such, we examined secondary mobility devices. These are devices that can be used in addition to the cane to enable detection of environmental obstacles that will not otherwise be detected by the cane such as wall-mounted bookcases or pedestal displays. A sonar mobility aid known as the Sonic Pathfinder displays auditory information to the user about obstacles. This device was used to evaluate head clearance of individuals...
passing under obstacles. A similar problem was observed to that of the cane.

Individuals tended to detect obstacles in their path then crouched down to a position much lower than the obstacle to clear it. In this case, the head clearance increased relative to the obstacle. The higher the obstacle, the more clearance observed. This is the same phenomenon as before; feedforward information received from the auditory display of the Sonic Pathfinder did not provide enough information to just clear the obstacle as was evident in people who had full vision.

Next, we chose to evaluate the environment from the perspective of an ecological interface design [4]. This method involves a five-level hierarchy from the abstract to the concrete and defines the overall purpose of the system right down to the system components. All aspects of the system are able to be modelled using this method. This display was found to be effective at providing information to the participants, but required some training to enable understanding of environmental information.

HUMAN ECHOLOCATION

After finding that individuals have difficulty with information provided to them by both kinesthetic means and present auditory means, it became apparent that we should perhaps examine how people with visual impairments utilize sounds to evaluate their environments.

Prior to and including the first half of the twentieth century, blind individuals were taught how to locomote effectively using echolocation. Echolocation involves the use of reflected echoes to localise obstacles in a given environment. With the advent of the cane (and the introduction of guide dogs), this skill is no longer taught and is believed to be “socially unacceptable” drawing too much attention to the individual.

The basis for echolocation is that sound made by an observer is reflected off surfaces in the environment, shown schematically in the leftmost column of Figure 1. There are several factors that influence the echoes off obstacles including obstacle parameters, the sound source intensity and duration, the spatial relationship of the obstacle and the presence of background noise (Kish, 1995). If the individual is moving in a dynamic situation, Doppler shift, involving a change in frequency and thus pitch, can also be used for velocity estimation as one moves toward or away from a moving obstacle [1]. Perception of obstacles in the environment occurs at a very young age. Ashmead et al [3] found that congenitally blind children from 5-12 years of age could detect obstacles in their pathway with only the presence of auditory stimuli resulting from their own footsteps and acoustic environment.

We have drawn on this information and have begun a research path that may better enable individuals to use the auditory domain to understand their environments. The system developed for this purpose is based on radar systems and acts as an echolocation device but actually provides information that may otherwise be undetectable.

RADAR SYSTEM ARCHITECTURE

The rightmost column of Figure 1 identifies a generic block diagram of a common phased-array radar system. The process begins with the generation of a waveform, often a pulse analogous to the click used in human echolocation. This waveform is then mixed with a carrier in the upconverter, which shifts the centre frequency of the waveform to the carrier frequency. The transmitter then amplifies the upconverted waveform and energizes the transmit antenna.

The signal from the transmitter propagates into the surrounding environment, illuminating objects in the surroundings according to their individual characteristics. In this context, an object is arbitrarily defined as anything large enough to be resolved by the system, and it is therefore acceptable to define any large object as a collection of smaller objects. The reflected signals from the surroundings return to the receiver array. The combination of all reflections from the environment is a complicated waveform, containing information regarding the distance to each object, the rate of change of that distance, and the characteristics affecting how objects are illuminated. Because the receiver Array contains more than one receiver element arranged in an intelligent geometry, the direction-of-arrival of each reflected signal can also be determined.

The combined waveform from the receiver array is then mixed with the carrier in the downconverter, and band-limited by digital downconversion to yield multiple superimposed baseband copies of the original waveform created in the waveform generator, adjusted in time according to the range to each object and in centre frequency according to the rate of change of that range.

In the correlator, the combined waveform is compared with the original waveform to yield correlation peaks corresponding to each object, which are in turn classified based on centre frequency and range. This process reduces the effect of other sources of illumination of unknown origin. In the beamformer, the information in the geometry of the receiver array is used to further classify the signal based on direction-of-arrival.

In the detector, a power threshold is applied to the classified signal to remove weak background noise. The result is a group of objects classified based on range, bearing, and range rate. In the plot extractor, co-ordinates for the centre of each object are determined, which are used by the tracker to associate objects as time passes. The information is then supplied to the user graphically.
Figure 1  Comparison of Approaches
HYBRID SYSTEM
In human echolocation, the signal processing which takes place between the correlator and the tracker is effectively done in the human brain, and is a very efficient albeit not completely known process. A system designed to capitalize on the inherently efficient signal processing in the human brain, then, must be designed without all of the information necessary to determine, analytically, a suitable waveform for transmission.

In the Hybrid System shown in the centre column of Figure 1, the verbal click of human echolocation is replaced with the waveform generator of the radar system, though the waveform generated could be identical to conventional echolocation. Transmission of this waveform at an ultrasonic centre frequency provides several advantages, apart from the elimination of the social disturbance. For instance, the Doppler Effect is more pronounced, allowing resolution of the rate of closure with an object. Furthermore, the degree to which the surroundings can be illuminated is limited only by the power and dynamic range of the hardware, providing the opportunity for greater previun than is available from human echolocation.

CONSIDERATIONS
To effectively reproduce echolocation, there are several aspects that we must consider in the design of the system to ensure that masking and annoyance are minimized. Since we don’t often model interfaces in the auditory domain without some aspect of vision, we must consider the differences. Also, echolocation is usually only needed when an individual must travel in their environment. While stationary, it is not necessary to create a spatial map. How does movement affect the ability to localize and how can we design to incorporate this issue?

There are two main differences between vision and sound. First, sound is omnidirectional. Where the sound is coming from cannot be limited as vision can by turning away. The sound reflections can provide information and we must capitalize on this information, but it is not possible to avoid certain sounds. All information to the individual should provide additional information about the environment. Secondly, sound is transient. A specific sound cannot be reproduced in exactly the same way. Even a recorded sound will have a different environment in which it is played every time. This transience requires that sounds are developed to provide quality information to an observer continuously.

Now let’s look at movement within the environment and how understanding it can aid in the design process. Ashmead et al. [2] found that movement toward sounds increased the ability to localize distances. Turning the head toward the source has shown to be most effective at localising sound sources [7]. Naturally, individuals move their head to hear the source of the auditory stimulus [8]. Dynamics of the head are intuitively and elegantly incorporated in the Hybrid System in exactly the same way so long as the necessary instrumentation moves with the observer’s head. Rosenblum et al [10] also found that moving toward the obstacle while localising enabled better distance perception than localising in a stationary position.

Understanding the differences between vision and sound and being able to apply dynamics of movement allows for a good basis for a prototype echolocation device.

CURRENT RESEARCH
This research aims to provide visually impaired people with a means to visualize surroundings using echolocation. The use of ultrasound, which is then converted to audible sound for the user, affords the opportunity to provide this capability without distraction to others. In this system, much of the “signal processing” takes place in the human brain, so our greatest challenge is to determine just how the brain processes sound information so that we may direct the design of well-suited assistive technology.

This system has been developed such that direct digital downconversion of ultrasound signals allows any movement in the immediate environment to be illuminated. If the traveller is moving within the environment, obstacles within the effective range will be heard. This includes any movement relative to the head in a reference frame which rotates with the head, and as such, the use of this device closely simulates echolocation. If an obstacle is moving in the environment, a pedestrian or a car for instance, only that obstacle will provide any feedback to the individual.

PHYSICAL EMBODIMENT
The instrumentation developed for this purpose is comprised of a lightweight, portable set of ultrasonic receivers, a transmitter, a purpose-built frequency synthesiser, and a processing unit that allows for the greatest degree of flexibility in the transmitted waveform and the processing of the received information.

The frequency synthesizer and signal processing unit both currently reside in a processor unit, shown in Figure 2 along with the two receivers and the transmitter. Each of the three enclosures measures approximately 14 centimetres on
the longest dimension, however no attempt has been made to miniaturize the components, most of the space currently dedicated to the containment of easily reconfigurable experimental circuits and standard rechargeable batteries. The self-powered receiver units include audio amplifiers, while the transmitter is simply an ultrasonic transmitting transducer with a cable connector. The processing unit accepts a waveform generator signal, generates the necessary transmit signal, accepts the dual receiver signals, and provides the downconverted output. The waveform generator is a device that can play a recorded a click from the user or any otherwise generated audible sound for conversion to ultrasound. Echoes are downconverted in the processor by simple intentional aliasing and relayed to the user via headphones. Those signals that are within the auditory range are not disturbed by aliasing, but those from the ultrasonic range are audible after aliasing.

DIGITAL DOWNCONVERSION

The Nyquist Theorem states that the highest frequency which can be accurately represented is less than one half of the sampling rate. This allows for sufficient information to be present to draw out all components of a frequency spectrum. Aliasing is a form of sampling that does not allow all the components to be present and as a result, signals that are very different have similar frequency responses. To gain a better understanding of aliasing, we can look at a simple sinusoid. In this case, at the Nyquist frequency we would get two sample points for each period. If we sample at a frequency lower than Nyquist, there will be fewer points per period and it will appear as though the points came from a sinusoid that was of a lower frequency.

The Hybrid System uses a form of intentional aliasing common in radar and sonar systems, referred to as digital downconversion. However, as limited digital signal processing takes place in the Hybrid System, a simplified form of digital downconversion is used, where the reflected signal is sampled at the carrier frequency to bring the signal to the auditory range. The transmitted signal is at a frequency of 40 kHz. The signal that returns as an echo is then sampled at a rate of 40 kHz, phase locked with the transmitted signal. Based on the Nyquist Theorem, only tones at 20 kHz would be able to be accurately represented. The aliased signal is present at a point predictable in the frequency spectrum. For the echo that is unmodified as a result of non-movement, the difference between the two signals is zero, and therefore the direct signal and reflections from relatively stationary objects is obliterated.

On the other hand, if there is movement, the reflected tone is aliased down to the difference between the two signals. So if movement creates a tone of 40 400 Hz, an auditory signal is heard at 400 Hz.

DOPPLER RESOLUTION IN ULTRASONIC ECHOLOCATION

Doppler shift in the auditory range is a change in the wavelength or frequency of sound as a result of motion. In the case of an approaching ambulance, the pitch of the siren appears to increase as it nears an observer and decrease as it fades away. This allows the observer to determine the relative location of the ambulance and respond accordingly.

The general wave equation for the relationship between the perceived frequency (f′) and the emitted frequency (f) is as follows:

\[ f' = \frac{v}{v \pm v_o} f \]

Where v is the speed of the waves in the source and v_o is the motion of the observer.

In our case, the observer creates a Doppler effect by moving toward or away from the object. The obstacle is stationary with the observer walking toward it. Thus, the following equation applies:

\[ f = f_0 \left[ 1 - \frac{v_o}{v} \right] \]

where f is the perceived frequency, f_0 is the frequency of the transmitted sound, v is the speed of wave in the medium, v_o is the velocity of the observer.

Assume that sound travels at 344 m/s in air. First let’s look at the situation in which the person is echolocating in the auditory range. If the person sends out a signal at 100 Hz while travelling toward a stationary obstacle at an average walking speed of 1.3 m/s the frequency of the return signal will be 99.62 Hz. This difference in frequency from the sent signal is 0.38 Hz. This difference will be virtually undetectable to the untrained observer. It is almost impossible to detect Doppler shift by normal human movement.

If the person is “echolocating” in the ultrasound range we can apply the same formula as we did when discussing Doppler shift in the auditory range. The signal is transmitted at 40 kHz and the frequency of the return signal will be 39849 Hz. Although this is a difference of 151 Hz, it is not within the audible range. Now, this signal is sampled at a frequency of 40 000 Hz and the resulting alias is within the auditory range at 151 Hz. An untrained observer can easily detect a frequency difference of 151 Hz, so this provides a means to observe the closure rate of objects within range.

The auditory signal reflected from different obstacles in the environment will be dependent on the speed of motion, the distance to the object and the size of the object. A person can thus manoeuvre around obstacles in their environment without having a complete picture of the environment. This is a method of direct perception of locomotion that resembles that of Gibson [5]. Gibson’s theory of vision while locomoting is based on the concept of a visual array which changes as one approaches or recedes away from an environmental obstacle.
REFERENCES

1. Anderson, David B. and Casey, Michael A. Sound work through a Doctoral Research Award and the CHPSTP Canadian Institutes for Health Research who supports this Strategic Training Initiative.

2. Ashmead D.H., Wall, R. S., Susan B.Eaton, Kiara A.


ACKNOWLEDGMENTS

We wish to acknowledge the generous contribution of the Canadian Institutes for Health Research who supports this work through a Doctoral Research Award and the CHPSTP Strategic Training Initiative.

FUTURE WORK

Our next task involves the identification of a set of waveforms to which human beings are particularly sensitive through adaptive feedback. The authors will draw on their strengths in human motion analysis and applied optimal estimation to develop a closed-loop algorithm for the identification of candidate ultrasonic waveforms. Specifically, the physically measurable kinematic response of human subjects equipped with the above noted ultrasonic instrumentation will be used to adjust the waveform transmitted by the instrumentation in real time. We anticipate the opportunity for discovery of new methods to synthesise sound stimuli based on the observation of human motion. If successful, the results of this investigation will be widely applicable to human-machine interface design and will influence our future work in the development of multi-modal aircraft instrumentation and human assistive technologies especially for the visually impaired.