

Human Echolocation Waveform Analysis

Graeme E. Smith and Christopher J. Baker

The Ohio State University, 2015 Neil Ave, 205 DL, Columbus, OH 43201 USA
e-mail: {baker.1891, smith.8347}@osu.edu

Keywords: Radar, Cognition, Target Recognition, Waveforms, Ambiguity Functions

Abstract— All humans have the ability to exploit acoustic echoes for cognitive sensing of the environment. There are blind people who use active versions of the technique as an augmentation to the long stick and can even perform remarkable tasks such as riding a bicycle. Those who are expert in this practise can evaluate range, location, size, shape and texture of objects. This provides a very powerful basis for perception and cognition and is somewhat beyond that which is routinely achieved by radar systems. This example of “sight by sound” makes an ideal candidate to study in order to understand and articulate the cognitive methods used. Subsequently artificial forms of cognition can be synthesised and applied to radar and sonar sensing. In this paper we report on an initial examination of human echolocation by presenting and analysing transmitted and received waveforms generated via “tongue-clicking”.

I. INTRODUCTION

Supa presented the first studies of human echolocation in 1944 [1]. Blindfolded participants were asked to walk towards an obstacle and stop when they were able to sense it. They were only allowed to make noises via scuffing their heels on the ground (i.e. the experiments were bistatic). The obstacle could be detected between 3m and 5m with best performance being by blind participants. Rice conducted the most extensive program of research on human echolocation in the 1960s, e.g. [2], [3] that studied both blind and sighted people. They were allowed to “vocalize” sounds making any noise they wished. Most participants chose either a long hissing sound or a punctuated tongue click.

In radar parlance we might associate these noises with waveform modulations designed to provide the desired information in the returned echo. Participants were able to detect an object 20cm in diameter at a range of 2.75m more than 60% of the time. Participants also demonstrated accurate spatial localization and discrimination of objects having the same area but different shapes. Subsequent studies, e.g. [4], [5], have confirmed these findings and speculate that the spectral composition of echoes provides a vital source of environmental information for tasks such as traversing an aperture such as a doorway. This demonstrates the concept of a “perception-action” cycle that contains many of the key cognitive processes that are highly desirable in future radar sensor systems.

More recently Thaler et al. [6] have examined the neuronal excitation that arises when humans, expert in the use of echolocation, detect the presence or absence of an obstacle using “tongue-clicking” to generate the transmitted waveform. They found specific activity in the middle temporal and nearby cortical regions of the brain when participants listened to echoes from objects. They conclude that humans recruit regions of the brain that would otherwise be devoted to visual interpretation rather than only the auditory part of the brain.

A study distilling the cognitive processes of humans using echolocation is considerably beyond the scope of this paper, and here we restrict ourselves to an examination of the form of the signals transmitted by expert human echolocators. These are subsequently examined in terms that are more meaningful within the radar community, via the wideband ambiguity function (WAF). We use the same waveforms generated by the experts that took part in the experiments reported in [6].

II. WAVEFORM ANALYSIS

Fig. 1 shows the time domain and frequency domain representations of a human echolocation tongue-click from the early blind (EB) participant in [6]. EB was born with only partial sight due to retinal cancer, and after about one year had his eyes removed to prevent the cancer spreading. The waveform was digitized with a sampling frequency of 44.1 kHz using an in ear microphone on the left side of the head. The signal was Hilbert transformed and downsampled by a factor of 2 in order to obtain I and Q samples, the real part is shown. The waveform was also high-pass filtered to remove a 60Hz component originating from the electrical power supply.

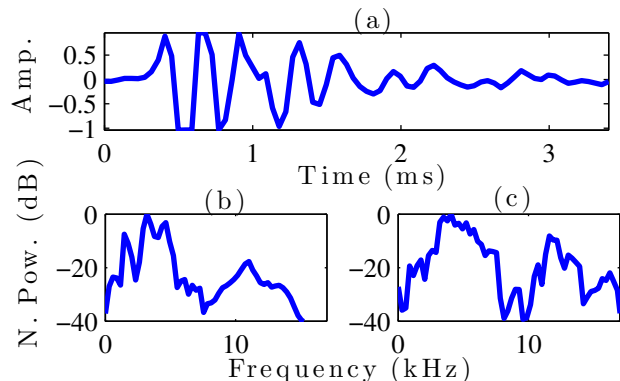


Fig. 1: The human echolocation tongue click (a) time domain Tx signal, (b) Tx power spectrum & (c) Rx power spectrum

The time domain plot, Fig. 1(a) shows the overall envelop to include two components: i) an initial short, large amplitude component; and ii) an extended lower amplitude component. The power spectrum, Fig. 1(b), has a maximum peak at 3.33 kHz with strong secondary components at 1.54 kHz and 4.62 kHz. There is a further weaker component at 11.03 kHz. Fig. 1(c) shows the power spectrum of the echo from a target straight ahead of EB. The same signal components are present, but offset by ≈ 500 Hz. The origin of the frequency offset is uncertain. In [6] the target is described as stationary. In [4] the head movements made by human echolocators while sensing is discussed. Speculatively, these motions could induce a Doppler shift. However, a 500 Hz shift requires a head movement speed exceeding 20 ms^{-1} . Note that there are significant differences between the transmitted signal and the returned echo. It is these differences that provide information about the target size and shape.

The echo was extracted by matched filtering the digitized signal with the tongue-click to create a range profile, Fig. 2. A maximum-of CFAR was used for detection, and Fig. 1(c) presents the power spectrum of the echo from the object at 1.51 m (note the detection at 0 m is the “cross-talk” response). In the experimental description of [6], the target is placed 1.5 m in front of EB. Furthermore, the detections at 2.88 m and 3.71 m were observed in several test cases suggesting they could be undocumented objects in the test area.

Fig. 3 presents the spectrogram of the tongue click, generated using a 2.3 ms Kaiser window, with shape parameter 3. It has a very unusual form of modulation compared to those typically used in radar and sonar systems. The pulse has duration of the order of 2 ms and has a primary bandwidth of approximately 3.8 kHz, equivalent to a range resolution of 5.5 cm. However, this bandwidth is made up of

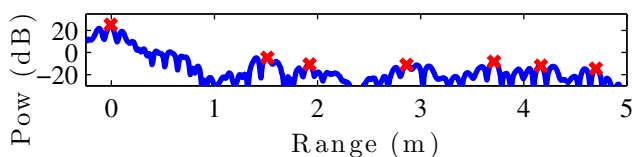


Fig. 2: Range profile (blue line) and target detections (red x).

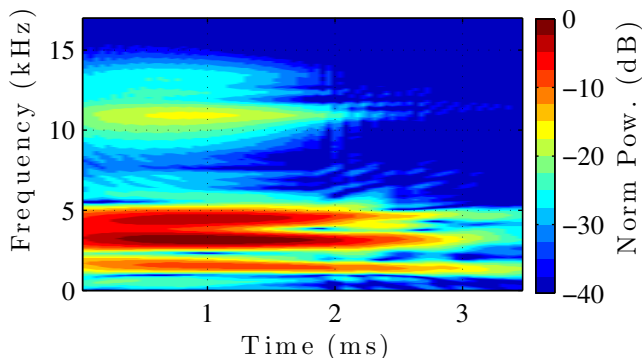


Fig. 3: Spectrogram of Tx tongue click.

three separate constant frequency components and is not truly narrowband. We use primary bandwidth to denote high power components of the signal between DC and 5 kHz. The full bandwidth would include the 11.03 kHz line giving a potential range resolution of 3 cm. The full primary bandwidth is exhibited over the entire time span of the pulse, quite different the common linear FM chirp. Fig. 3 also shows that the precise duration of the pulse is somewhat unclear with energy appearing to be transmitted and steadily decreasing levels for at least a further 1 ms. The three components identified in Fig. 1(b) are visible and it is noted that they have a roughly harmonic relationship. Harmonics in transmitted waveforms, albeit with different modulations, have been observed in the signals routinely transmitted by bats [7].

Fig. 4 presents the wideband ambiguity function (WAF) of EB’s tongue-click. It is calculated using

$$|\theta(\alpha, \tau)|^2 = \left| \alpha^{0.5} \int u(t)u^*(t/\alpha - \tau) dt \right|^2 \quad (1)$$

where $u(t)$ is complex sampled tongue-click. α is the Doppler compression factor calculated as

$$\alpha = (c - v)/(c + v) \quad (2)$$

where, c is the speed of sound in air (343 ms^{-1}) and v is the target speed, which is positive during approach. In the electromagnetic case, $\alpha = 1$ to four significant figures even when $v = 1 \times 10^3 \text{ ms}^{-1}$, but for in-air sonar $\alpha = 0.84$ for $v = 30 \text{ ms}^{-1}$ and hence cannot be ignored. In Fig. 4 the y-axis is velocity and has been calculated using (2).

The WAF, Fig. 4, has a clear central peak as required for

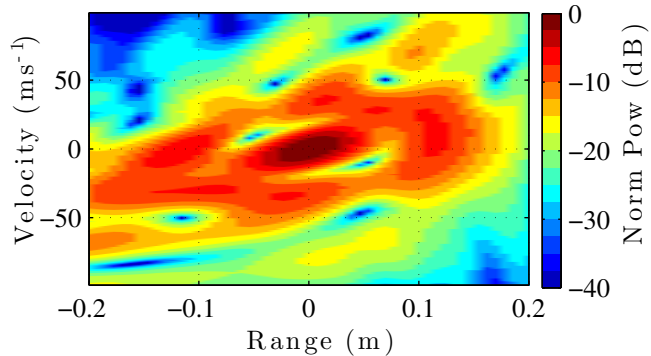


Fig. 4: WAF of the tongue click.

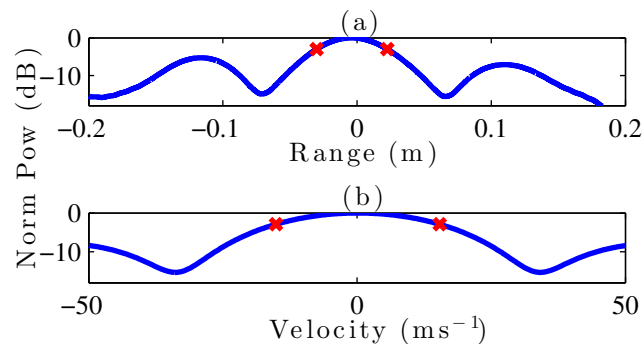


Fig. 5: The (a) zero-velocity and (b) zero-range cuts with highlighted 3 dB points.

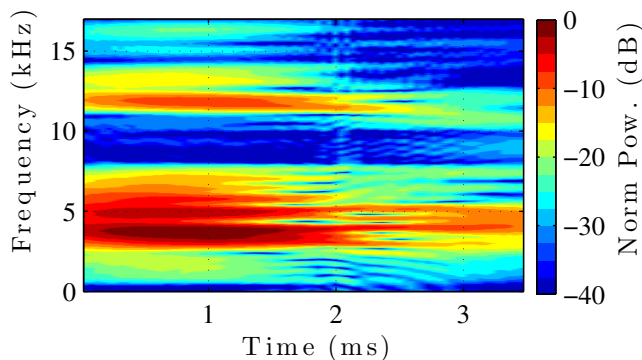


Fig. 6: Spectrogram of the received click.

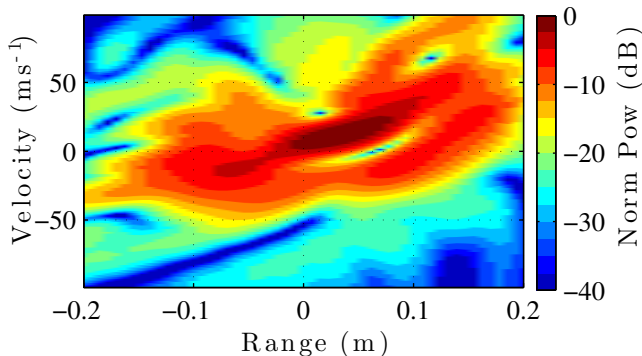


Fig. 7: The narrowband cross-ambiguity function.

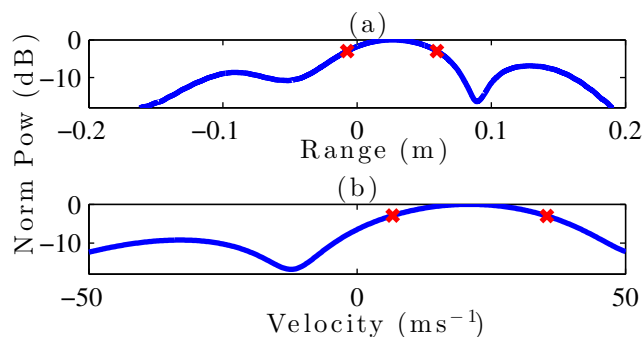


Fig. 8: The (a) range and (b) velocity cuts through the cross-ambiguity function peak with highlighted 3 dB points.

making positional and velocity estimates. However, the close in range side-lobes are only ≈ 6 dB down on the main-lobe, although they rapidly fall away as range increases.

The high side-lobes are surprising, but the detailed way in which neurons are excited combined with processing in the human brain appears to result in a system able to cope adequately. Indeed the way in which neurons are excited is known to be non-linear and may have the effect of imparting a weighting function that effectively reduces side-lobes to a manageable level. Fig. 5(a) presents the zero-velocity cut of the WAF and shows the 3 dB range resolution to be 5.3 cm. As noted above, EB’s primary bandwidth is ≈ 3.8 kHz, which corresponds to a range resolution of 5.5 cm, suggesting the higher frequency components play little part in the range resolution observed in Fig. 5(a). Conversely, the velocity

resolution is quite coarse and the 3 dB resolution measured from Fig. 5(b) is 30.5 ms^{-1} .

Whilst the self-ambiguity function of the tongue-click tells us about the potential accuracy for measuring spatial and velocity properties, it is the received echo that contains the sensed information presented to the brain. As stated earlier it is the difference between the transmitted signal and the received signal that contains the very information that we seek to understand and exploit.

Fig. 6 shows an example echo in the form of a spectrogram. This is for an echo received at the ear prior to processing by the brain. Again it is clear when comparing with Fig. 3 that the echo has been quite substantially altered through interaction with the target beyond the approximately 500 Hz increase in frequency observed earlier. The form of the received echo spectrogram has the appearance of being more complex than the transmitted signal. As stated before these differences must contain information about the target and it is this interpretation by the human brain that ultimately is a driver for further research of this phenomenon.

We postulate that the velocity offset may form a method to separate the “cross-talk” tongue-click, which propagates through the body, from target echoes. It was theorised above that the origin of the offset was due to EB subconsciously making the head motions described in [4] even when he had been asked to remain motionless. The in-air length of the click is ≈ 85 cm and in the literature [1-6] targets closer than 40 cm are successfully detected indicating the human echolocators are not subject to the pulse blanking common in radars. Comparable techniques are used by bats to overcome pulse ambiguity [8], although the bat can shift the transmit frequency directly rather than relying on head motion.

The more complex form of the echo has the three relatively distinct spectral lines at 1.5, 3 and 4.5 kHz being much more merged, although they can still be discerned. The strongest echo is at the lowest of the frequencies in contrast to the transmission case. This may be a result of higher attenuations at higher acoustic frequencies. Conversely, the higher frequency component, at 11 kHz, has greater power relative to the three low frequency components compared to the transmitted click. Such variation may provide insight into the nature of the reflecting target.

The wideband cross-ambiguity function (WCAF) of the echo and click is presented in Fig. 7. The form of this figure is not dissimilar to that of the WAF of Fig. 4, but with a peak offset from the (0, 0) position. The velocity offset is consistent with the above discussions on head movement. The range offset is a result of mismatch in the extraction of the echo signal, relative to transmitted click, from the digitized signal and can be ignored. The corresponding range and velocity cuts through the peak are presented in Fig. 8. The asymmetry visible in the cuts, and the WCAF relative to the WAF, are due to (1) not being symmetrical about the zero Doppler compression point.

It is not at all clear how the nature of Figs. 7 and 8 assist detection, shape and texture recognition as reported in [1-6].

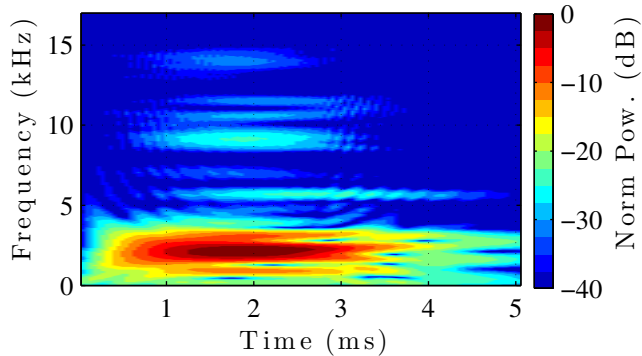


Fig. 9: Spectrogram of LB's Tx tongue click.

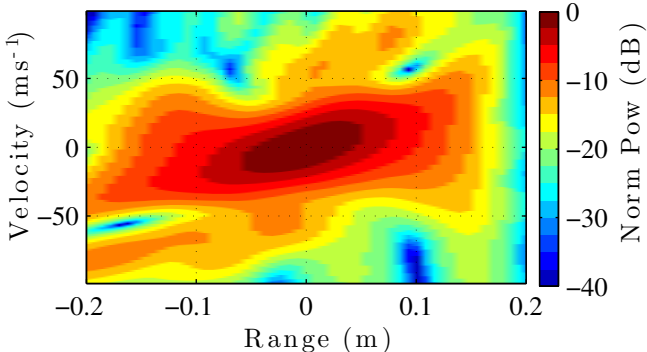


Fig. 10: WAF of LB's tongue click.

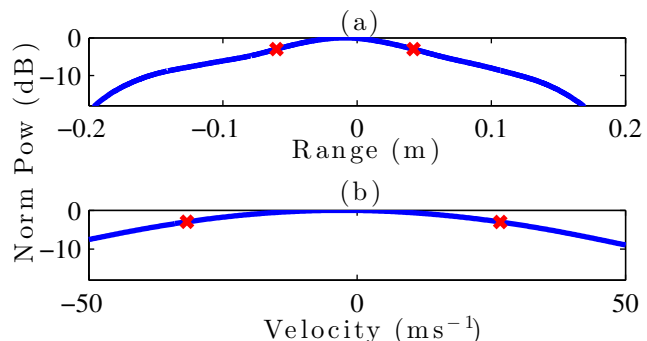


Fig. 11: The (a) zero-velocity and (b) zero-range cuts of LB's WAF with highlighted 3 dB points.

It might be safely concluded that it is likely the brain is processing the data in a very different way from a conventional radar system. Given the structure of the auditory tract, the “borrowing” of parts of the brain normally used for visual processing and the multiple parallel processing paths often employed in the mammalian brain and the ability of humans to discriminate it comes as no surprise that additional processing is at play.

Figs. 9, 10 and 11 show the spectrogram, WAF and zero velocity and range cuts for the second volunteer from [6]. The measurement conditions are the same as before. This volunteer lost his sight later in life, and is referred to as late blind (LB), but is still an expert at echolocation who is also capable of target detection as well as shape and texture

recognition. Fig. 9 shows that the overall click duration was comparable to EB's and again energy was present at distinct frequencies for the duration of the click. For the frequency components below 5 kHz, where the majority of energy is located, LB's individual frequency components were not as distinct as EB's. Above 5 kHz LB had more frequency components than EB. Although no supporting figure is presented here, the echoes from LB's click were also offset in frequency despite the targets being stationary. For LB, the offset was ≈ 300 Hz.

The differences in the LB's click resulted in the WAF, Fig. 10, having a broader central peak, but lower sidelobes. Measuring from the zero velocity and range cuts presented in Fig. 11, LB was found to have a range resolution of 10 cm and a velocity resolution of 55 ms^{-1} . These resolutions are of the same order of magnitude as EB's, but slightly coarser.

In [6] EB is reported as having greater sensitivity to target azimuth location than LB, and we speculate this is attributable to his greater control of the spectral content of his click. EB's click has two clear frequency regions, one around 3.33 kHz and the other 11.03 kHz. It is reported in [9] that the cry of the big brown bat contains frequency components at distinct separations also. The transmit and receive apertures (the mouth and ears) have a fixed area during a cry so the 3 dB angular beamwidths differ for the different frequencies. Humane tests on big brown bats, reported on in [9], indicate that the subsequent neurological processing of the frequency-beamwidth diversity in the echo allows the bat to detect straight ahead targets that would normally be masked by clutter at the same range but different azimuth. It is interesting, therefore, to speculate that the human echolocators have a similar capability and that EB's better frequency separation results in him being more sensitive to aspect angle than LB.

III. DIRECTIONS FOR RADAR

Modern radar systems are capable of sophisticated waveform design. The use of digital synthesis permits direct implementation of polyphase coding schemes that including large symbol dictionaries and long sequences. Despite this sophistication, the principal objective of such waveforms is the same as the classic linear frequency modulation or chirp: compressive the received pulse to give fine range resolution and maximize the signal to noise ratio. The forms of radar waveforms tend to be very different from those employed by echolocating mammals such as humans and bats. We postulate that these natural waveforms were selected to maximize information content in the received echo and so facilitate cognitive sensing.

The analysis of section II is of interest to radar designers because of the capabilities of human echolocators. The literature on human echolocation [1-6], and references therein, demonstrates detection and recognition capability in controlled conditions. However, less formal sources, e.g. [9] and [10], report blind people to be capable of riding bicycles and cross-country hiking all through the use of echolocation

(referred to as “FlashSonar” in [9] and [10]). In this context echolocation is being used as part of a cognitive sensing activity and can serve as a guide for cognitive radar. We therefore review how the human echolocation tongue click differs from a typical radar waveform.

Radar waveforms are currently constrained to a single frequency at each time instant, while the tongue-click has multiple frequencies. In both EB and LB’s clicks there were multiple pure tones, potentially with amplitude weighting, present for the duration of the click. There appeared to be three higher power tones between 1 kHz and 5 kHz with additional lower powered tones above 10 kHz. The tones were more distinct in EB’s click and it was speculated that this allowed him greater angular localization in line with the findings of [9].

The human echolocators appeared to induce a Doppler shift in the received echo signal by a mechanism that is unknown. We speculated that the head motions described in [4] could account for Doppler offset and that it could help discriminate the target echo from cross-talk for close range targets. Implementing such capability in radar could allow physically extended pulses to be used—putting more power on the target—without creating a long pulse blanking range.

The tongue-click is wideband, while the majority of radar waveforms are narrowband. Considering the primary bandwidth, EB’s click spanned the interval 1.5 kHz to 4.6 kHz, a bandwidth (BW) of 3 kHz at a centre frequency of 3 kHz and comparable results were obtained for LB. This gives a fractional bandwidth of 100%, which would rise to nearly 200% if full bandwidth were considered. Such high bandwidths would give excellent range resolution for radar imaging and many researchers are already striving to develop hardware capable of achieving these requirements. However, for human echolocators, the bandwidth is occupied with discrete tones and not continuously as is more common in manmade systems.

IV. CONCLUSIONS

In this paper we have examined previous research on human echolocation that highlight the remarkable abilities of human to detect objects and provide information regarding their shape and texture.

Specifically we have presented and examined the waveforms generated by tongue-clicking humans expert in echolocation. We have shown that the waveform is wideband, and complex with a range resolution between 5 cm and 10 cm and a velocity resolution between 30.5 ms^{-1} and 55 ms^{-1} depending on the individual. Furthermore, the properties of the range-Doppler ambiguity surface are not obviously ideal for the sophisticated processing and perception that appears to follow. Naturally this is only a first foray into what is a very complex area but one that promises much insight and advantage to future radar systems and hence a fertile area of research.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the state of Ohio and are grateful to Lore Thaler for the waveforms generated and received by their highly expert human echolocators.

REFERENCES

- [1] M. Supa, M. Cotzin and K.M. Dallenbach, “Facial vision”, the perception of obstacles by the blind”, *American journal of psychology*, **57**, pp. 133-183, (1944)
- [2] C.E. Rice, “Human echo perception”, *Science*, **155**, pp. 656-664, (1967)
- [3] C.E. Rice, “Perceptual enhancement in the early blind”, *Psychological record*, **19**, pp. 1-14, (1969)
- [4] T.A. Stoffregen and J.B. Puttenger, “Human echolocation as a basic form of perception and action”, *Ecological psychology*, **7(3)**, pp. 181- 216, (1995)
- [5] B. Hughes, “Active artificial echolocation and the nonvisual perception of aperture passability”, *Human movement science*, **20**, pp. 371-400, (2001)
- [6] L. Thaler, S.R. Arnott and M.A. Goodale, “Neural correlates of human echolocation in early and late blind echolocation experts”, *Plos One*, **6(5)**, pp. 1-16, (2011)
- [7] M Vespe, G Jones and C J Baker, “Diversity Strategies: Lessons from natural systems”, Chapter in *Principles of waveform diversity and design*, Ed M Wicks, SciTech Publishing, pp25-50, (2010)
- [8] S. Hiryu, M. E. Bates, J. a Simmons, and H. Riquimaroux, “FM echolocating bats shift frequencies to avoid broadcast-echo ambiguity in clutter.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 15, pp. 7048-53, Apr. 2010.
- [9] M. E. Bates, J. A. Simmons, and T. V. Zorikov, “Bats use echo harmonic structure to distinguish their targets from background clutter.”, *Science (New York, N.Y.)*, vol. 333, no. 6042, pp. 627-30, Jul. 2011.
- [10] “World Access for the Blind.” [Online]. Available: <http://www.worldaccessfortheblind.org/>. [Accessed: 06-Jul-2012].
- [11] D. Kish, “Daniel Kish: Blind Vision,” *PopTech PopCasts*, 2011. [Online]. Available: http://poptech.org/popcasts/daniel_kish_blind_vision. [Accessed: 06-Jul-2012].