THE VIRTUAL BAT: ECHOLOCATION IN VIRTUAL REALITY

Dean Waters and Husam Abulula

School of Biology
University of Leeds
Leeds, LS2 9JT, Great Britain
D.A.Waters@leeds.ac.uk, H.H.Abulula@leeds.ac.uk

ABSTRACT

Work in progress is being presented on the effectiveness of using sonar to orientate and navigate in a virtual reality system. The sonar system is based on those of bats, using ultrasonic frequency modulated signals reflected from simple targets. The model uses the reflectivity characteristics of ultrasound, but the frequency and temporal structure used are scaled down by a factor of ten to bring the frequency range and temporal resolution within the capabilities of the human auditory system. Orientation with respect to the ensonified target is achieved by time of flight time delays to give target range, and binaural location information derived from interaural timing differences, interaural intensity differences, and head-related transfer functions. Data on the ability to locate targets as a function of signal frequency, bandwidth, duration and sweep pattern is presented.

1. INTRODUCTION

Thomas Nagel's philosophical essay on what it is like to be a bat [1] asks whether it is possible for a human to perceive the world as a bat would. While it is impossible to know what it is like for a bat to be a bat, it is now possible to simulate the sonar environment that bats inhabit to model what it is like for a human to be a bat.

Bats use biosonar to navigate in darkness and to find their insect food using ultrasonic calls. These calls have a range of temporal and frequency structures, primarily in the ultrasonic range from 15 kHz up to 200 kHz and lasting a few milliseconds. In general, the call structures are frequency modulated sweeps, starting at a high frequency, and terminating at a lower frequency. The calls are emitted at very high amplitude, often 110 dB SPL at 10 cm from the bat's mouth [2] and by timing the delay until the return of the echo, the bat is able to determine the range of the target. It has been determined in some experiments that bats are able to resolve timing differences of the order of 500 ns [3], equivalent to an absolute range resolution of 85 μm, though these results are still controversial [4]. It has also been determined that bats can differentiate echoes with multiple wavefronts from those without, where the wavefronts differ in timing of 10 ns [5]. These astonishing feats of resolution are carried out by a brain that can weigh less than half a gram. To account for this, a method of extracting spectral information and converting to temporal information was proposed, the "Spectrogram Correlation and Transformation Receiver (SCAT)" model [6]. While this model has some virtues, there are limitations [7]. Surprisingly, most of the adaptations of the bat's brain and auditory system are concerned with the high frequencies and fast processing time needed, rather than sophisticated spatial tasks. At a fundamental level, the structure of the bats auditory system is not unlike that of other mammals, albeit occupying a much greater proportion of the brain [8]. Only in the lower brainstem are there any clear specializations for echolocation [9]. If it were possible to slow down the speed of sound by a factor greater than ten, the frequency and temporal range would be brought within the capabilities of human listeners and it may be possible to emulate the sonar world of bats. Such work has three benefits: the first is to allow biosonar researchers insight into how bats perceive targets from their perspective, secondarily, to allow those with visual impairment access to virtual reality (VR) environments using sonar to navigate and interact, and finally, to use sonar rendering to provide a novel interface for complex spatial tasks.

1.1. Human Location of Azimuth and Elevation

There are numerous studies of the properties of sound in pinpointing the azimuthal angle of a sound source. Both the timing of the arrival of the signal at the ear in the form of interaural timing differences (ITD) and the intensity of the sound at both ears in the form of interaural intensity differences (IID) can provide information on azimuth. Since the hearing system can detect differences in phase at frequencies below 1 kHz, timing differences caused by the different path lengths to each ear can be detected by phase differences at each ear. At low frequencies however, the wavelength of sound is large compared to the diameter of the head, and diffraction of the sound around the head means that IIDs are of little use in detecting azimuth. Above 1.5 kHz, diffraction effects produce a significant sound shadow, meaning at IIDs can be used as an azimuth cue. Above the limit of phase resolution, timing differences are of use in identifying the time of arrival of the envelope of signals, but this appears limited in relation to IIDs and ITDs with their associated frequency constraints [10][11]. Since a sound source at left 20° azimuth will provide the same IID and ITD cues as one placed at left 160° azimuth, head movements can also be used to remove this ambiguity. As the head moves left, a source at 20° will become more centered, while one at 160° will become more obviously placed at the left.

In an idealized spherical head, a cone of confusion exists where for any two opposite positions on the surface of the cone extending from the ear the IIDs and ITDs are the same [12]. However, such a simple model takes no account of the asymmetric structure of the head and the pinna shape, all of which generate spectral cues for localizing a sound position [13]. These head related transfer functions (HRTFs) are used to remove the ambiguity from positions on the cone of confusion,

since each pair of ambiguous positions show different filtering characteristics.

The ability to determine azimuth is also a function of the duration of a sound. Short duration sounds being more difficult to locate than those of a longer duration [14]. Click transients of 1 ms could only be localized to approximately 3°.

Determining elevation may be a simple task of using HRTF's to decode the position of a sound source on the cone of confusion in association with IIDs and ITDs. However, even in the medial plane when IIDs and ITDs are identical, and HRTFs should be identical, pinna cues improved vertical localization [15][16]. Humans do not show a uniform ability to locate a sound source irrespective of position. With broadband sources, azimuth and elevation error is reduced in front of the head, with an accuracy of localization of between 4° and 10° [17], while behind, the error can be as high as 40°. Minimum Audible Angle (MAA) studies show accuracy of 1° for sounds at 0° azimuth, decreasing as the source moved to the periphery. For concurrent sound sources, where there is temporal overlap, the concurrent MAA shows similar results to the MAA, but with much decreased resolution at the periphery [18].

1.2. Human Perception of Range

While the perception of azimuthal and elevational cues in the human auditory system is relatively well understood, the perception of range to a sound source is less easy to define. The obvious cue for distance is the intensity of the sound source, a louder sound is more likely to be nearer that a quiet one. However, for unfamiliar sounds it may be impossible to differentiate a loud, distant signal from a nearby quiet one. The ratio of direct to reverberant sound can be used, and removal of this cue in an anechoic chamber can remove the ability to discern sound distance [19]. The change in intensity as a listener moves towards a sound source may also be important in determining range [20], as are changes in spectral content due to excess atmospheric attenuation [21].

1.3. Bat Location of Azimuth and Elevation

Aside from humans, the auditory system of the bat is possibly the best known model for auditory processing. Since it is not possible to ask bats where they think a sound source is located, experimenters must find different mechanisms for investigating the basic parameters of hearing. In general, it seems that the ability for bats to localize the azimuth and/or elevation of a signal is lower than that for humans. Some of the best estimates give MAA values of 3°-4° for sounds at 0° azimuth [22]. Other studies suggest this figure is closer to 10-14° [23][24]. The complex pinna shape of bats appears to restrict sound detection to a 60° cone in front of the bat's head [25], which given the directional nature of ultrasonic emissions from the bat's mouth is perhaps not surprising. The bat may sacrifice omnidirectional hearing for enhanced gain in the forward-facing direction. Perhaps due to the complex pinna shape of bat's ears, and the ultrasonic nature of returning echoes, bats have pronounced HRTF responses [26][27].

1.4. Bat Perception of Range

Unlike humans, bats have an inherent mechanism for locating sound distance, that of time of flight from an emitted sound to the time of reception. Estimates for the accuracy in determining range of targets, and hence of accuracy in determining time of flight remain controversial. Timing accuracy of 500 ns have been reported, suggesting an ideal coherent receiver with phase detection well into the ultrasonic region. Other studies have suggested that ITDs of the order of 70 μs can be distinguished, although this is at frequencies above the ability of the bat to phase lock auditory neurons [28].

1.5. Adaptations of call structure to task and environment

Two constraints on the design of echolocation calls in bats are excess atmospheric attenuation and pulse-echo overlap. For sonic sound up to 18 kHz, the effect of absorption of sound energy by the atmosphere is negligible. At 20 kHz it is 0.5 dBm^{-1}. At 40 kHz it is 1.2 dBm^{-1} and at 100 kHz it is 3 dBm^{-1} over the loss due to spherical spreading. This effectively limits bat echolocation to frequencies below 150 kHz. Such a limit has implications for target detection, since the target strength of objects such as insect prey is frequency dependent – a low frequency call may miss small insects [29]. Bats have a mechanism by which the auditory system is protected from the high intensity call when emitted by the attenuation of the signal within the inner ear, thus making the reception of an echo during the outgoing pulse unlikely. To compensate for the inability to determine location or range during signal emission, bats reduce their call duration when approaching a target to minimize the effects of pulse-echo overlap [30]. These issues mean that bat echolocation calls have a variable structure, both between different species living in different environments, and within species depending on different tasks e.g. a long duration low frequency signal may be used by a high, fast flying bat for long range detection of large prey, while a short duration broadband signal may be used when foraging in clutter to detect small prey.

2. MATERIALS AND METHODS

2.1. Hardware and Software Environment

The experimental apparatus consists of two elements. An SGI O2 machine running IRIX 6.5 which hosts the reality engine, and a Pentium II machine running NT4 which hosts the acoustic rendering hardware (Lake CP4).

The SGI machine renders a VR environment programmed in WoldToolKit from Sense8. The environment consists of an unbounded space with a simple object located at an arbitrary position. In the visual representation, this object is a moth. Where required, the visual representation can be relayed to the user via a head mounted display (Virtual Research V6). The absolute position and orientation of the user is encoded by a Polhemus Fastrack 6D sensor mounted on a pair of Sennheiser HD565 Headphones attached to the head mounted display. The sensor allows > 1mm range accuracy over the X, Y and Z axes, and 0.15° angular resolution of Yaw, Pitch and Roll. The Polhemus Long Ranger Transmitter allows tracking of head position and orientation over a 10m range from the transmitter, but at reduced positional and orientation accuracy. The geometry of the virtual world and the real world are mapped 1:1 such that a movement of the user forward by 1 m results in a movement in the virtual world of 1 m.

Attached to the moth position is a sound file consisting of a FM sweep with a sinusoidal amplitude envelope, generated from custom written software.

The acoustic rendering hardware is a Lake CP4 card hosted in a PC running NT-4. This card communicates with the SGI using TCP/IP. Given a listener position and orientation, and a sound position, the CP4 convolves the sound with a set of HRTF filters, and also reduces the sound level in accordance with the inverse square law. The rendered output is then amplified (Samson Q5 headphone amplifier), before being passed to the headphones.

2.2. Echolocation in the VR environment

In order for the individual to echolocate within the VR environment, a software trigger initiates the playback of the FM sweep sound file at a rate of 0.5 Hz. The sound file is acoustically rendered at the position of the user's head at the moment of triggering to provide an reference time for the output signal, and a reference spectrum. The distance between the user and the target object is calculated from the Fastrack data, and the playback delayed to simulate two-way time of flight from the user to the object and back again assuming a speed of sound of 8.5 ms^{-1} to allow sufficient temporal scaling for the human auditory system. After the delay, the user position and orientation, and the target object position is relayed to the CP4 via TCP/IP as the sound signal is output from the line output of the SGI. This signal is passed through the audio I/O of the CP4, convolved with the HRTF sets, and played back to the user via the Samson audio amplifier and Sennheiser headphones.

2.3. Experimental Procedure

The experimental subject is shown a marked trial arena which is a semi-circle of 10m radius with the Polhemus transmitter at its origin. The arena represents the area over which it is possible to encode the subjects position and orientation. Since the Polhemus transmitter exhibits an inverted magnetic field when traversing one hemisphere of the transmitter, it is not possible to use a full 10m radius circle. Due to interference from metal supports in the building, it is also necessary to use large open spaces with no metal support in the floor in order to encode position and orientation with any accuracy. In our trials we used the Leeds Grand Mosque.

The subject is allowed to walk around the arena wearing the headphones, and connected to the rendering hardware by a 15m set of leads for the Polhemus receiver mounted on the headphones, and the headphone lead. At a suitable position, the subjects head position is encoded, which will serve as the position of the virtual sound source. The subject is unaware of where that position will be. The subject is then asked to return to the origin point of the arena, and playback of the virtual sound source is begun. The sound for each run is randomly selected from a set of 22 different sounds. The task for the subject is to walk towards the sound source, and when they believe their head is centered on the source, to stop walking. Trials last thirty seconds for each sound type. The position and orientation of the user's head is encoded every 50 ms and recorded to a text file for later analysis. The next sound is then selected, a new virtual sound position defined, and the trial run again. Each test subject receives each test sound three times, making a total of 68 sound presentations per subject.

2.4. Sound Stimuli

Twenty two sound files are used in the presentations. These can be defined as nine in a broad-band set and nine in a narrow-

band set, plus three white noise controls and a constant sound training set. The broadband set consisted of hyperbolic frequency sweeps with sinusoidal envelopes, starting at 20 kHz and sweeping down to 2 kHz, starting at 11 kHz and sweeping down to 2 kHz and starting at 20 kHz and sweeping down to 11 kHz Each frequency set is presented at three durations, 125 ms, 250 ms and 500 ms.

The narrowband set followed the same pattern except that the frequency sweeps were from 4.4-3 kHz, 1.5-0.1 kHz and 6-0.7 kHz. Sonograms of the two sets are shown in Figure 1 and 2.

Prior to the presentation of the bat-like sound signals, the user is trained to localize a piece of music in virtual space. We chose Michael Nymans 'The Draughtsman's Contract' as it is both broadband and uses repeated units.
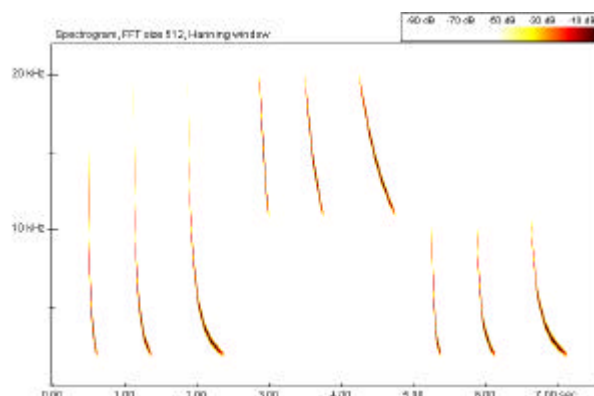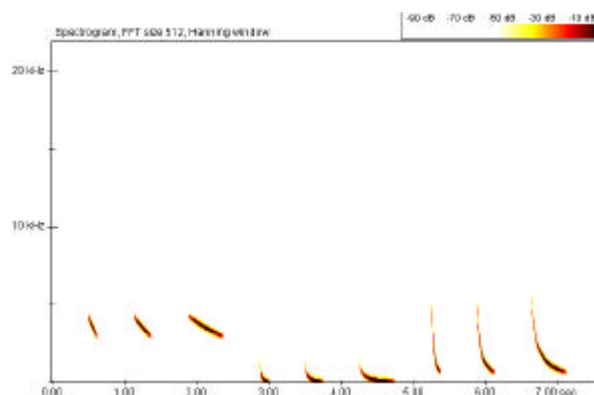

Figure 1. *Sonograms of the broadband set of stimuli.*

Figure 2. *Sonograms of the narrowband set of stimuli.*

2.5. Data Analysis

The text files of positional and orientation information are extracted to Excel. In order to assess the effectiveness of a particular sound stimulus, we calculated the index of tortuosity, given by:

$$T = \frac{A}{D} \tag{1}$$

where:
T is the index of tortuosity

A is the actual length of the path taken by the subject from their start position to their finishing position

D is the direct path length from their start position to the object position.

Since movement in the vertical plane is limited, this can effectively be ignored as a contribution to the overall path length.

For the presentation within this paper, only the initial results are described from four individuals.

3. RESULTS

The mean index of tortuosity for all treatment presentations was 2.99 ±0.51 meaning that on average, the path taken to the virtual sound position was three times the length of the direct path. For the data so far collected, there are no significant differences in the ability to orientate and walk towards the different sound types (Repeated measures ANOVA $p > 0.05$). Mean indices of tortuosity for four subjects for the broadband and narrowband sets are shown in figures 3 and 4.
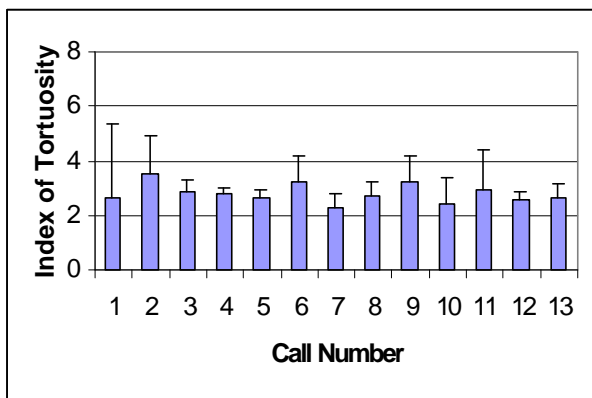

Figure 3. *Indices of tortuosity for the broadband set of stimulus calls. Call 1 is the musical control, 2 ,3 and 4 are the 20-2 kHz set, 5, 6 and 7 the 20-11 kHz set, 8, 9 and 10 the 11-2 kHz set, and 11, 12 and 13 the white noise controls. All sets are shown with the 125 ms duration first, followed by the 250 ms and 500 ms presentations.*
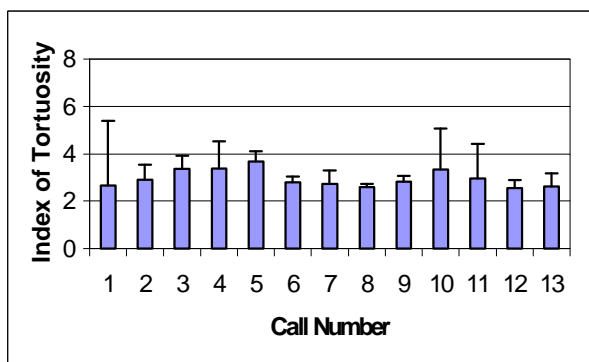

Figure 4. *Indices of tortuosity for the narrowband set of stimulus calls. Call 1 is the musical control, 2 ,3 and 4 are the 4.4 - 3 kHz set, 5, 6 and 7 the 1.5 – 0.1 kHz set,*

8, 9 and 10 the 6 – 0.7 kHz set, and 11, 12 and 13 the white noise controls. All sets are shown with the 125 ms duration first, followed by the 250 ms and 500 ms presentations.

Sample tracks from a single situation when the localization was accurate is shown in figure 5, and where the localization is poor is shown in figure 6. As shown in figure 5, the listener originally sets off on the wrong track, but quickly corrects and walks directly to the sound position and stops. Figure 6 shows a listener who originally sets off on the correct vector, passes through the virtual sound position, and then is unable to relocate it.
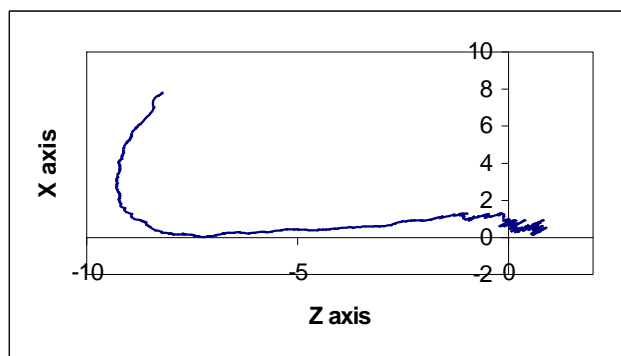

Figure 5. *The track of a listener viewed from above who is able to locate the virtual sound position within the allowed 30 seconds. The axes are marked in meters, with the sound source located at 0,0. Axis conventions are those of WorldToolKit.*
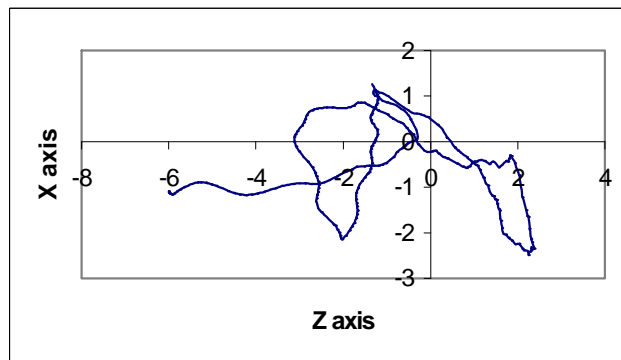

Figure 6. *The track of a listener viewed from above who is unable able to locate the virtual sound position within the allowed 30 seconds.*

4. DISCUSSION

It was anticipated that signals which are shorter in duration and higher in frequency would be more difficult to locate. As was found from these preliminary results, there are no large differences readily observable. Clearly, much more data needs to be collected before making any generalizations on the types of signal which are easiest to locate. Subjective impressions are that the 20-11 kHz signals are very difficult to localize. These

signals will only contain IID cues, and should be harder to locate than the 6-0.7 kHz signals which should contain both ITD and IID cues. However, with the current dataset, this does not appear to be the case. The timing of pulse-echo delay certainly provides a useful cue in identifying the range of the virtual sound source. In fact, rather than being a problem, entry into the zone of pulse-echo overlap helps to identify that the listener is close to the sound source position. However, within the zone of pulse echo overlap, identifying the exact sound source position (i.e. to localize it within the head) is very difficult. Future experiments will test the value of pulse-echo detection for identifying range by suppressing the outgoing reference pulse. Also, since orientation data is logged by the system as well as positional data, it is possible to investigate how the head is turned to aid location. Input from the user as to when a pulse needs to be emitted, and also user control of the structure of outgoing pulses will be implemented so that pulse-echo overlap can be minimized by using shorter duration pulses when the listener is close to the target.

5. CONCLUSIONS

A system has been developed which allows the simulation of bat echolocation in virtual space. Initial results suggest that there are no large scale effects which make one signal easier to localize than another, even though subjective comments suggest that some sounds are easier to localize than others. Development of more user-control, the analysis of orientation data and larger data sets will allow the investigation of how the human auditory system behaves in an echolocation environment.

6. REFERENCES

[1] T. Nagel, "What is it like to be a bat?" *Philosophical Review*, vol. 83, pp. 435-450, 1974.

[2] D.A.Waters and G. Jones "Echolocation call structure and intensity in five species of insectivorous bats". *J. Exp. Biol.,* vol. 198, pp. 475-489, 1995.

[3] J.A. Simmons, "The perception of phase information in bat sonar". *Science*, vol. 171, pp. 1335-1338, 1979.

[4] G.D. Pollack, "Some comments on the proposed perception of phase and nanosecond time disparities by echolocating bats". *J. Comp. Physiol. A.*, vol 172, pp 523-531, 1993.

[5] J.A. Simmons, "Discrimination of jittered sonar echoes by the echolocating bat, *Eptesicus fuscus*: the shape of target images in echolocation". *J. Comp. Physiol. A.*, vol 167, pp 589-616, 1990.

[6] P.A. Saillant, J.A. Simmons and S. Dear, "A computational model of echo processing and acoustic imaging in frequency-modulated echolocating bats: The spectrogram correlation and transformation receiver". *J. Acoust. Soc. Am.*, vol 94, pp 2691-2712, 1993.

[7] H. Peremans and J. Hallam, "The spectrogram correlation and transformation receiver, revisited". *J. Acoust. Soc. Am.*, vol 104, pp 1101-1110, 1998.

[8] G. Baron, H. Stephan and D. Heiko, *Comparative neurobiology in Chiroptera*. Birkhäuser Verlag, Germany, 1996.

[9] G. Baron, "Differential phylogenetic development of the acoustic nuclei among chiroptera", *Brain, Behav. Evol.*, vol. 9, pp. 7-40, 1974.

[10] J.C. Middlebrooks and D.M. Green, "Sound localization by human listeners". *Ann. Rev. Psychol,.* vol. 42, pp. 135-159. 1991.

[11] C. Trahiotis and L.R. Bernstein, "Lateralisation of bands of noise and sinusoidally amplitude-modulated tones: effects of spectral locus and bandwidth". *J. Acoust. Soc. Am.*, vol. 79, pp. 1950-1957, 1986.

[12] W. Mills, "Auditory Localisation" in *Foundations of Modern Auditory Theory* (J.V. Tobias ed.). Academic Press New York, 1972.

[13] J. Blauert, *Spatial hearing. The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA. 1983.

[14] W.A. Yost and E.R. Hafter, "Lateralisation" in *Directional Hearing*" (W.A. Yost and G. Gourevitch eds). Springer Verlag, New York, 1987.

[15] A.D. Musicant and R.A. Butler, "The influence of pinnae-based spectral cues on sound localisation". *J. Acoust. Soc. Am.*, vol. 75, pp. 1195-1200, 1984.

[16] S.R. Oldfield and S.P.A. Parker, "Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions". *Perception,* vol. 13 pp. 581-600. 1984.

[17] S.R. Oldfield and S.P.A. Parker, "Acuity of sound localisation: a topography of auditory space. II. Pinnae cues absent". *Perception* vol. 13 pp. 601-617. 1984.

[18] D.R. Perrott, "Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity". *J. Acoust. Soc. Am.*, vol. 75, pp. 1201-1206, 1984.

[19] D.H. Mershon and J.N. Bowers, "Absolute and relative cues for the auditory perception of egocentric distance", *Perception*, vol. 8, pp 311-322, 1979.

[20] D.H. Ashmead, D. Davis and A. Northington, "The contribution of listeners' approaching motion to auditory distance perception". *J. Exp. Psychol.: Human Percept. Perf.*, vol. 21, pp 239-256. 1990

[21] A.D. Little, D.H. Mershon and P.H. Cox, "Spectral content as a cue to perceived auditory distance", *Perception*, vol 21, pp 405-416. 1992

[22] J.M. Wooton and J.A. Simmons, "Spectral cues and perception of the vertical position of targets by the big brown bat, *Eptesicus fuscus*", *J. Acoust. Soc. Am.*, vol. 107, pp. 1034-1041, 2000.

[23] G. Koay, D. Kearns, H.E. Heffner, R.S. Heffner, "Pasive sound localisation ability of the big brown *bat Eptesicus fuscus*". *Hearing Res.*, vol. 119, pp 37-48, 1998.

[24] R.S. Heffner, G. Koay, H.E. Heffner, "Sound localization in an old-world fruit-bat (*Rousettus aegyptiacus*): Acuity, use of binaural cues and relationship to vision". *J. Comp. Psychol.*, vol. 113, pp 297-306, 1999.

[25] J.M. Wooton, R.L. Jension and D.J. Hartley, "The combination of echolocation emission and ear reception enhances directional spectral cues of the big brown bat, *Eptesicus fuscus"*. *J. Acoust. Soc. Am.*, vol. 101, pp. 1723-1733, 1997.

[26] J.M. Wooton, T. Haresign and J.A. Simmons, "Spatially dependent acoustic cues generated by the external ear". *J. Acoust. Soc. Am.*, vol. 98, pp. 1423-1445, 1995.

[27] M.K. Obrist, M.B. Fenton, J.L. Eger and P.A. Schlegel, "What ears do for bats- a comparative study of pinna sound pressure transformation in chiroptera". *J. Exp. Biol.* vol. 180, pp 119-152, 1993.

[28] Z.M. Fuzessery, "Acute sensitivity to interaural time differences in the inferior colliculus of a bat that relies on passive sound localization". *Hearing Res.* vol. 109, pp 46-62, 1997.

[29] D.A. Waters, J. Rydell and G. Jones "Echolocation call design and limits on prey size: a case study using the aerial-hawking bat *Nyctalus leisleri". Behav Ecol. Sociobiol*, vol. 37, pp 321-328, 1995.

[30] E. Kalko and H.U. Schnitzler, "The echolocation and hunting behaviour of Daubenton's bat, *Myotis daubentoni. Behav. Ecol. Sociobiol.* vol 24, pp 225-238, 1989.