A Biologically Plausible Acoustic Azimuth Estimation System

Sofia Cavaco
Departamento de Informática
Universidade Nova de Lisboa
2825-114 Monte da Caparica, Portugal
e-mails: sc@di.fct.unl.pt,
scavaco@mail.telepac.pt

John Hallam
Division of Informatics
University of Edinburgh
5 Forrest Hill
EH1 2QL Edinburgh, U.K.
e-mail: john@dai.ed.ac.uk

Abstract

In this paper we present an acoustic azimuth estimation system to be used in a small robotic 'cat'. In addition to being reliable and having an acceptable degree of accuracy the system must therefore be computationally cheap. Also an effort was made to build a biologically plausible system. In the system developed the sound source's azimuth is estimated using interaural time differences, which are found using zero crossings, a property of waves proposed by Marr in the context of computer vision [Marr, 1982] as a suitable abstraction of the detailed wave shape. The system, which can be used with more than one source of sound, can estimate the azimuth better than humans.

1 Introduction

Sounds arriving at one ear are slightly different from the same sounds received by the other ear. It is these interaural differences, along with other spectral cues, that are used by the auditory system to find out the sound sources' locations.

The difference in time between the arrival of the sound at each of the two ears is called the Interaural Time Difference (ITD) and is used in several parts of the auditory system. The azimuth of the source can be extracted from this difference[1]. If the sound comes from 0° in azimuth or 180° it will reach both ears synchronously (ITD=0); the maximum absolute value of ITDs (max_{ITD}) is obtained whenever the sound comes from 90° or -90°. Other positions will have ITDs between 0 and max_{ITD}.

If the wavelength is less than or equal to the distance between the ears there is ambiguity when extracting the azimuth of the source from ITDs. For this reason and also from results of experiments it is thought that ITDs are only used for frequencies below a certain value which depends on the size of the head.

In this project it was decided to use ITDs to localise the sound source's azimuth. Not only is it well understood how ITDs vary with azimuth but also a system based on these cues to localise sound can be easily implemented in hardware and computationally cheap software. Also electronic ITD measurement can in principle be higher resolution than neural.

2 The SSAE System

After recording the sounds, with two microphones displaced by $d_{ears} = 9.5cm$, into files[2] which can be directly used by MATLAB, the signals were analysed by the Sound Source's Azimuth Estimation (SSAE) system, which was built in MATLAB.

Given a signal, the SSAE system computes the ITDs of that signal in order to produce a topographic map of the azimuth. This map, which is represented using a vector in which each position corresponds to a small part of the azimuth, resembles the organization of the medial superior olive.

The superior olivary nuclei are responsible for the localisation of sound (on the contralateral side of the head). While the lateral superior olive uses Interaural Intensity Differences (IIDs), the medial superior olive uses ITDs to localise sound. Neurons close to one extreme of the medial superior olive are maximally activated by short delays between the arrival of the signal at each ear, while neurons close to the opposite end of the nucleus are maximally activated by long delays. The cells which lie between the two ends of the nucleus respond maximally to intermediary delays [Guyton, 1984].

The SSAE system's vector is organised in the same way as the neurons of the medial superior olive nuclei. However, instead of having two maps, one to localise sounds coming from each side of the brain, the system joins both maps into one larger vector. The middle position of the vector corresponds to the middle (frontal) position in azimuth (0°). The rightmost position is associated with 90° and the leftmost position with $-90°$. Therefore, just as in the medial superior olive, there is a spatial pattern of stimulation; sound that comes directly

[1] Strictly speaking, it cannot. The sound can be located on a hyperboloid of revolution, the surface for which the difference in path lengths to the ears is fixed. However, assuming that the sound comes from a certain elevation we can extract the azimuth from the ITD.

[2] The format of the files is PCW and the sampling frequency is $f_s = 44100Hz$.

from ahead stimulates the middle position of the vector while sound coming from the sides stimulates the lateral regions of the vector.

Using just ITDs to estimate the position of sound sources does not allow us to distinguish whether a sound comes from ahead or behind. For instance, in fig. 1 (a) the ITDs of sounds coming from A and B are equal. However, if the head rotates the ambiguity disappears (fig. 1 (b)). In fact, in biological systems, head movements play an important role to help localising sound sources.

The vector works as a voting system. Each position of the map contains the number of votes for a given region of the azimuth and therefore the values of each position of the map are a measure of how certain the system is that sound comes from that region [Tan, 1996]. The map is initialised to zeros, which means that at the beginning the system thinks there is no sound, and each time an ITD is produced the map is updated. Every ITD contributes a vote to the construction of the map. For instance, if the arriving ITD says that sound comes from 45°, a vote will be added to the position of the map that corresponds to 45° in azimuth.

However, an accumulating votes system has a drawback which is that, when working on-line, eventually an overflow will take place. Fortunately, that drawback can easily be overcome with the introduction of a decay expression. If at regular time intervals the voting system suffers some decay, the overflow can be avoided.

Moreover, the introduction of a decay expression has another advantage, which is forgetting the past. On the one hand, if in some interval of time there is enough noise to lead the voting system to produce a faulty solution, the decay expression allows the system to forget the noise and to be free to produce a more accurate response. On the other hand, once a sound is extinguished there is no point in keeping information about it in the map for much longer (otherwise, at the end of the day the map would contain information about every single sound the system had heard during the day).

To compute the ITDs the system uses the zero crossing algorithm along with a matching method (both explained further below). The zero crossing algorithm, which looks for some properties (the zero crossings (ZCs)) in the sound waves, is applied to the waves of both channels. When the ZCs have been found, a matching method tries to match pairs of ZCs (one ZC from each channel) in order to produce the ITDs that will be used to generate the topographic map of the azimuth.

Since the estimation of the azimuth is only based in ITDs, the system must take as much information as possible from the signal in order to produce those ITDs. Dividing the signal into several frequency bands and applying the zero crossing and matching methods to each of the bands allows the system to base itself on more information in order to produce the final result [Babeanu, 1994; Tan, 1996]. A set of ITDs is produced for each band and thus the overall response of the system can have a higher degree of certainty.

Moreover, if more than one sound source is present and they have components in different bands the system is prepared to distinguish them. Also having bands $[f, 2f]$ allows us to proceed by looking for and matching ZCs, which uniquely determine a signal after such filtering (theorem of Logan (1977) cited in [Marr, 1982]).

The complete SSAE system is illustrated in fig. 2. After the signal has been recorded, both waves (the wave from the left channel and the wave from the right channel) are filtered by a set of band pass filters. The zero crossing method is then used to find the ZCs of each wave thus generated. Afterwards, the matching algorithm uses the ZCs of each pair of left and right waves filtered by the same band pass filter, to generate a set of ITDs, which work as votes to be added to the map. As the map is being updated, a decay expression is used to allow it to have some degree of forgetfulness.

Since a set of ITDs is computed for each band of the set of bandpass filters, a map is built for each band. Those maps can be analysed separately or their results can be added (with a simple sum or a normalised sum, since the higher bands will produce a stronger result[3]).

Each of the steps the system goes through in order to produce the topographic map of the azimuth is described in more detail in the following sections.

2.1 Filters - choices and design

The SSAE system has a set of bandpass filters, inspired by the cochlear mechanical filtering. The signal is filtered into each of those bands and for each pair of waves (the left and right waves in each band) the zero crossing and matching algorithms are used to find the ITDs.

One of the characteristics of filters is their groupdelay. When a signal is filtered the resulting wave is shifted in time. However most filters do not apply the same delay to every frequency. As a result, when a signal composed of several frequencies is filtered, its different components will be shifted by different amounts. Consequently, the ZCs found when such a filter is used can be desynchronised, originating erroneous ITDs.

The SSAE system uses a bank of 8th order bandpass digital Bessel filters (which is an adaptation of the MATLAB lowpass analog Bessel filter). Apart from having a constant groupdelay (for frequencies lower than the cut-off frequency) this filter has three other main advantages: its flat band pass characteristics (it produces a predictable response), its commercial availability in hardware and its sweepable clock (the cut-off frequencies can be easily changed by means of reseting the frequency of the timer clock) [Tan, 1996]. Since this filter shifts all the frequencies by the same amount, the ZCs found are synchronised and can be subtracted to find the ITDs, which therefore do not suffer deviations from their true values due to interferences caused by the filter.

[3] Higher frequencies will produce more ZCs and thus more ITDs, so their maps will have larger values than low frequency maps.

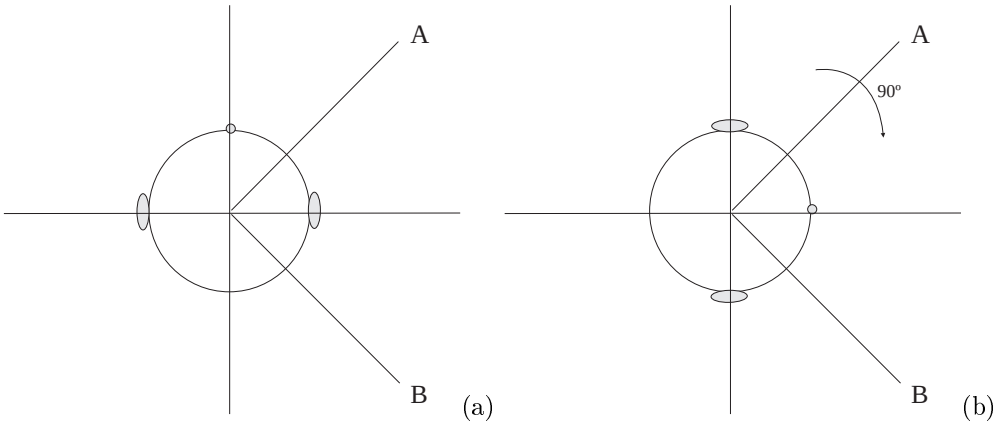

Figure 1: (a) Sources A and B are not distinguishable. (b) With a head rotation it is possible to decide whether the sound comes from A or B.

Since the filters are not ideal, a certain amount of information that is outside the cut-off is passed through to the output signal (the cut-off is not a vertical line in the graph of amplitude against frequency and the cut-off slope depends on the filter's order). In order to have as much unique information in each band as possible, the cut-off frequencies of neighbouring bands were split apart. The SSAE system uses two frequencies (f_1 and f_2, with $f_1 > f_2 > f_1/2$) to make the bank of band-pass filters [Babeanu, 1994; Tan, 1996], which has the following structure[4]:

band 1: $[f_1, f_2]$
band 2: $[f_1/2, f_2/2]$
.....
band i: $[f_1/2^{i-1}, f_2/2^{i-1}]$.

2.2 Detecting Zero Crossings

To compute the time difference between the arrival of the signal at each ear some common properties of the waves are required for matching between the left and right waves. When a pair of such properties is found the delay between their occurrences is computed, by the matching algorithm, giving rise to an ITD. The properties looked for are ZCs, which consist of the positions (in time or sample number) at which the wave crosses zero. In other words, given x, such that $w(x) < 0$, where w is the wave function, z is a ZC iff $w(z) \geq 0$ and there is no y such that $x < y < z$ with $w(y) \geq 0$.

Marr [1982] uses ZCs to detect intensity changes in image data. He argues on various theoretical and practical grounds that as a way of detecting intensity changes in image data, biological systems might use a kind of ZC detection.

The zero crossing is therefore a biologically acceptable method, which can be used to serve the purposes of the SSAE system. When we have two channels, the ITDs can be calculated by matching the ZCs between these channels:

$$ITD = ZC_{left\_channel} - ZC_{right\_channel}.$$

Apart from being biologically acceptable, the zero crossing algorithm has several other advantages. To begin with it is an easy, simple, cheap and competent algorithm, which does not require too much computation, to find the ITDs between two channels. It is guaranteed to find pairs of samples that can be matched to find the ITDs. Given a wave whose frequency is f_1 we know for sure that on average it crosses the time axis f_1 times every second and given a wave whose frequency belongs to $[f_1, f_2]$ we know for sure that this wave crosses the time axis not less than f_1 and not more than f_2 times in one second, on average.

Had the envelope of the waveform been used, just one ITD would have been produced. On the contrary, using ZCs allows the system to rely on more data (more ITDs) to produce the final result. Also, this method is much cheaper computationally than a Fourier Transform method. Despite relying on much less information than the natural auditory systems, the zero crossing method was used since the first purpose of the SSAE system was to be engineering efficient (rather than a faithful copy of the mammalian auditory system).

The zero crossing algorithm used by the SSAE system is not exactly the one explained in the first paragraph of this section. In the real world sounds do not reach the ears with their original waveform. One of the interferences a wave can experience is background noise. The wave that reaches the ears is thus the sum of the original wave with the background noise, which can lead the system to produce faulty ITDs.

Even though it is the ZCs that characterise the signal, the interference caused by background noise (with a lower amplitude than the original wave) can be overcome if a suitable gate is chosen when looking for *gate*

[4]The number of channels is many fewer than in the cochlear system to keep the computation low.

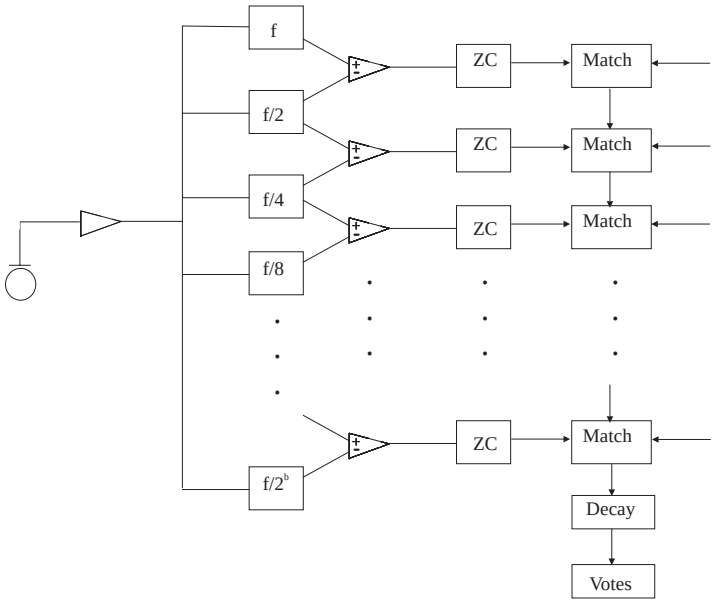

Figure 2: The SSAE system. Just the processing of the left channel is represented in the figure. The matching boxes receive the ZCs of both left and right waves.

crossings[5], i.e., instead of looking for the wave's abscissae that first cross the time axis one can look for the wave's abscissae that first cross some other straight line parallel to the time axis [Babeanu, 1994]. If the gate is high enough so that all the peaks that are a result of noise interference are lower than the gate and if the gate is low enough so that all the interesting peaks are higher than the gate, only relevant ZCs will be chosen.

This algorithm can be used with two gates as well. The higher gate is used as explained in the last paragraph while the lower gate is used to find the first samples that are lower than the gate's value (fig. 3). The lower gate can be, for instance, the negative of the higher gate.

The SSAE system uses two gates in order to be able to rely on more information than that obtained with just one gate.

2.3 The Decay Expression and Normalisation of Data

The decay expression depends on the last vote's age. The system suffers a small decay when a vote has been generated recently (and therefore the system is not deceived by *noisy* votes) whilst a larger decay is used when no vote has been added to the vector for a long time (and in this manner an overflow can be avoided). The longer the time since the last vote was generated is, the bigger the decay is. The original decay expression is:

$$votes(t) = votes(t-1) * e^{\Delta t/\tau}, \tau < 0, \qquad (1)$$

where Δt is the time between the sample being processed and the last vote[6], which therefore is always increasing until a new vote appears.

With expression 1 a problem still remains. Since different bands correspond to different frequencies, their ZCs occur at different rates. Therefore, the system gives more importance to higher bands. A normalisation of the values of each vector (one vector per band) can be obtained either by normalising the votes or by normalising the decay expression. Normalised votes have the form:

$$vote_b = f_1/f_b, 1 \leq b \leq bands, \qquad (2)$$

where f_1 stands for the highest frequency considered (i.e. the highest cutoff frequency of the highest band) and f_b stands for the vote generating band's highest cutoff frequency. A normalisation of the decay expression can be $votes(t) = votes(t-1) * e^{\Delta t * f_b/\tau}$ with $\tau < 0$, which is equivalent to:

$$votes(t) = votes(t-1) * e^{\Delta t/(\tau * 2^{b-1})}, \tau < 0. \qquad (3)$$

Both possibilities were successfully tried. It could be observed that the lower bands had their heights increased. To conclude, the SSAE system uses expression 3 both as a decay and normalisation expression.

2.4 Matching Zero Crossings

Once the zero crossings of each channel are found, they have to be paired in order to be used to generate ITDs.

[5] From now on whenever we refer to ZC we mean *gate* crossing.

[6] Since the sample frequency (f_s) is known and the number of samples since last vote (Δn) can be easily found, it is straightforward to compute Δt: $\Delta t = \Delta n/f_s$.

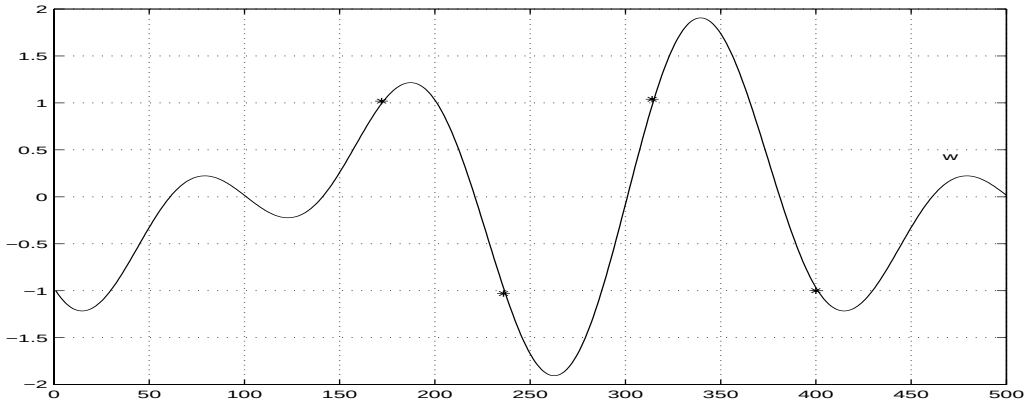

Figure 3: The stars (*) are the gate crossings of wave w. The gates are $+1$ and -1.

The matching cannot be done arbitrarily. Some restrictions must be obeyed to produce physically correct matchings (and consequently, reliable data). Marr [1982] presents some constraints on the matching of image points, which are here adapted for the sound case. Differences arise as the data is now analysed not in a time instant but over time and the interaction of more than one object in an image and more than one sound source in a sound wave give rise to different results.

One sound source

If the sound comes just from one source those constraints can be stated as follows:

a) To begin with there is the *compatibility constraint*, which says that two points can be matched if they correspond to the same position along the original wave.

b) The *uniqueness constraint* states that each point in one channel can match only one point in the other channel. This constraint can be inferred from the first: two points can only be paired if they are the mapping of the same point in the original wave. Since each point in the original wave is mapped to just one point in the wave that reaches each ear (or channel), it follows that for each point in one channel there is only one point in the other channel it can be matched with.

As a consequence of the compatibility and uniqueness constraints and because the precedence relation between two points in the original wave is inherited by their mappings it follows that no two different pairs of points correctly matched can cross over each other.

c) Finally, comes the *continuity constraint*. The different pairs of points found in a wave produced by a static sound source must be equally displaced. In the case that the source is not static[7], the source is expected to have some kind of continuous motion. Therefore the displacement between matched points must vary smoothly. In other words, the matched points' displacement graph must show a continuously changing pattern. For instance, when the source is at 0° and it moves to 90°, the displacement of pairs of points will monotonically increase from 0 to a maximum value (max_{ITD}). The points' displacement changing rate depends on the sound source's velocity.

The matching of ZCs not only has to respect the constraints discussed above but it also must take into consideration that there is a maximum displacement between two matched ZCs. Two ZCs can only be paired if they fall within a time window (or, in implementation terms, a samples window), whose size was defined to be the time that sound takes to travel directly from one ear to the other, along the axis that crosses the ears (*timeout*) [Babeanu, 1994].

When a ZC is found in one channel, the matching algorithm memorises the time at which it took place and it continues looking for more ZCs. When a ZC is found in the other channel, the ITD is calculated and the vector of votes is updated accordingly. However, if no ZC is found in the other channel within the time window, the ZC is forgotten.

This algorithm does not allow a single ZC to be used twice. Each ZC can only be used to form one pair of ZCs and consequently one ITD.

Several sound sources

The three constraints stated above are not so clear when more than one sound is present. The problem with more than one sound source arises from the fact that more than one wave is being mapped into just one wave by each channel and the way the waves are combined is different in each channel (they are added with different phase differences at each channel).

[7] Just cases with moving sources that are continuously producing sound are considered. If a source produces sound at a given location, then changes its location and produces sound again, but no sound is produced while the source is moving from one position to the other, the first and second sounds are thought to have been made by two different sources.

Nevertheless, the three constraints are here restated for a multi-source system:

a) The *compatibility constraint* says that two points in the image waves[8] can be matched if they correspond to the same position along one of the original waves.

b) The *continuity constraint* says that if we consider static sounds and we divide the pairs of matched points into as many sets as the sound sources, the continuity constraint of a one-source system applies to each set separately.

In a multi-source system, the uniqueness constraint does not apply. Consider the next example from fig. 4. The compatibility constraint says that l_1 and r_1 can be matched because they correspond to the same position (a_1) along wave a. That constraint also allows l_2 and r_1 to be matched since they map the same point (b_1) along wave b. Therefore, point r_1 can be matched both with l_1 and l_2!

Provided it were possible to find matching points in the waves of both channels that respect these restated constraints, the position of different sources would be accurately found. However, ZCs do not respect those constraints in a multi-source system because a point in the original wave that is mapped as a ZC in one of the channels can be mapped in the other channel as a point which is not a ZC. Therefore, the distance between two ZCs may be different from the distance between the two true mappings of the point.

It follows that for multiple sound sources, the zero crossing is not a suitable method to find the ITDs. Nevertheless, if the sounds do not have a similar spectra, that is, if they have frequencies within different bands of the bank of bandpass filters, the SSAE system is expected to localise the different sources.

3 Results

Several tests were performed to check how accurate the system is. The sounds, which were tried in several positions with frontal azimuth, consisted of voices, coughs, whistles, clapping hands, crumpling and tearing paper, feet beating on the floor, rolling chairs, among others. The tests were analysed individually and as a whole: the results of several sounds in the same position were compared as well as the results of the same sound in different positions. In addition, tests were done using more than one source at the same time. Some of the recordings were done with more than one source and also some files were created from the combination of the single sound files. In that way it was possible to compare the results of the sound by itself and when mixed with another sound.

As stated in section 1, ITDs can only be unambiguously used with wavelengths bigger than the distance between the ears. In this case, with the prototype cat head used, the minimum wavelength that can be used is: $\lambda_{min} = d_{ears} = 9.5cm$. The frequency associated with λ_{min} is f_{max}, which is defined as follows:

$$f_{max} = v_{sound}/\lambda_{min} = 3578.9Hz.$$

Nevertheless, the system has been tested with frequencies higher than f_{max}, to see how it would react to ambiguous inputs. The bands used were:

band 1: [7300, 14316] Hz,
band 2: [3650, 7158] Hz,
band 3: [1825, 3579] Hz,
band 4: [912.5, 1789.5] Hz,
band 5: [456.25, 894.75] Hz,
band 6: [228.125, 447.375] Hz.

When the wavelength is bigger than twice the distance between the ears, for each ZC there is only one possible pair in the other channel. For instance, imagine that there is a ZC in the left channel (ZC_l) and two ZCs in the right channel: one that preceded ZC_l (ZC_{r1}) and one that comes after ZC_l (ZC_{r2}). If $|ZC_l - ZC_{r1}| \leq timeout$ then $|ZC_l - ZC_{r2}| > timeout$ (because $|ZC_l - ZC_{r1}| + |ZC_l - ZC_{r2}| > 2 * timeout$). The frequency which corresponds to the wavelength $\lambda = 2 * d_{ears}$ is $f_{max}/2$.

For the same reason there are no ghosts in the vector of votes of bands with highest frequency lower than $f_{max}/2$. When the wrong pairs of ZCs are chosen, a peak in the wrong side of the graph can appear. However, if the wavelength is bigger than twice the distance between the ears, it will not be possible to choose wrong pairs of ZCs and therefore there will be no ghosts in the graph.

As expected it turned out that for bands 5 and 6 the algorithm has good performance.

Despite the fact that between the highest frequency of band 4 and $f_{max}/2$ there are some frequencies that can produce ambiguous matchings of ZCs, this band performs quite well.

Even though band 3 is higher than $f_{max}/2$, it turned out that this band produces good results in general. Some ghosts can appear on the wrong side of the azimuth. However, their height is insignificant compared to the height of the correct peaks.

Although the algorithm was not expected to perform well in band 2, some interesting results were obtained in this band. From the experiments done it was possible to see that the algorithm produced some peaks in the wrong places. Nevertheless the difference in height between the correct peak and the wrong peaks is still significant.

Finally, band 1 has, in general, poor results.

It turned out that when $\lambda < \lambda_{min}$ the algorithm chooses the right pairs of ZCs provided that the displacement of the waves is smaller than the wavelength, i.e. the interaural phase difference is less than 2π ($IPD < 2\pi$).

Comparing the azimuth estimation for different sounds at the same position it was easy to see that the accuracy depends on the kind of sound. For instance, table 1 shows the results for four sounds at around -48°. The best performance is obtained with a voice (the error is zero) and then with a cough (the error is around

[8] Image waves are the waves received by the channels.

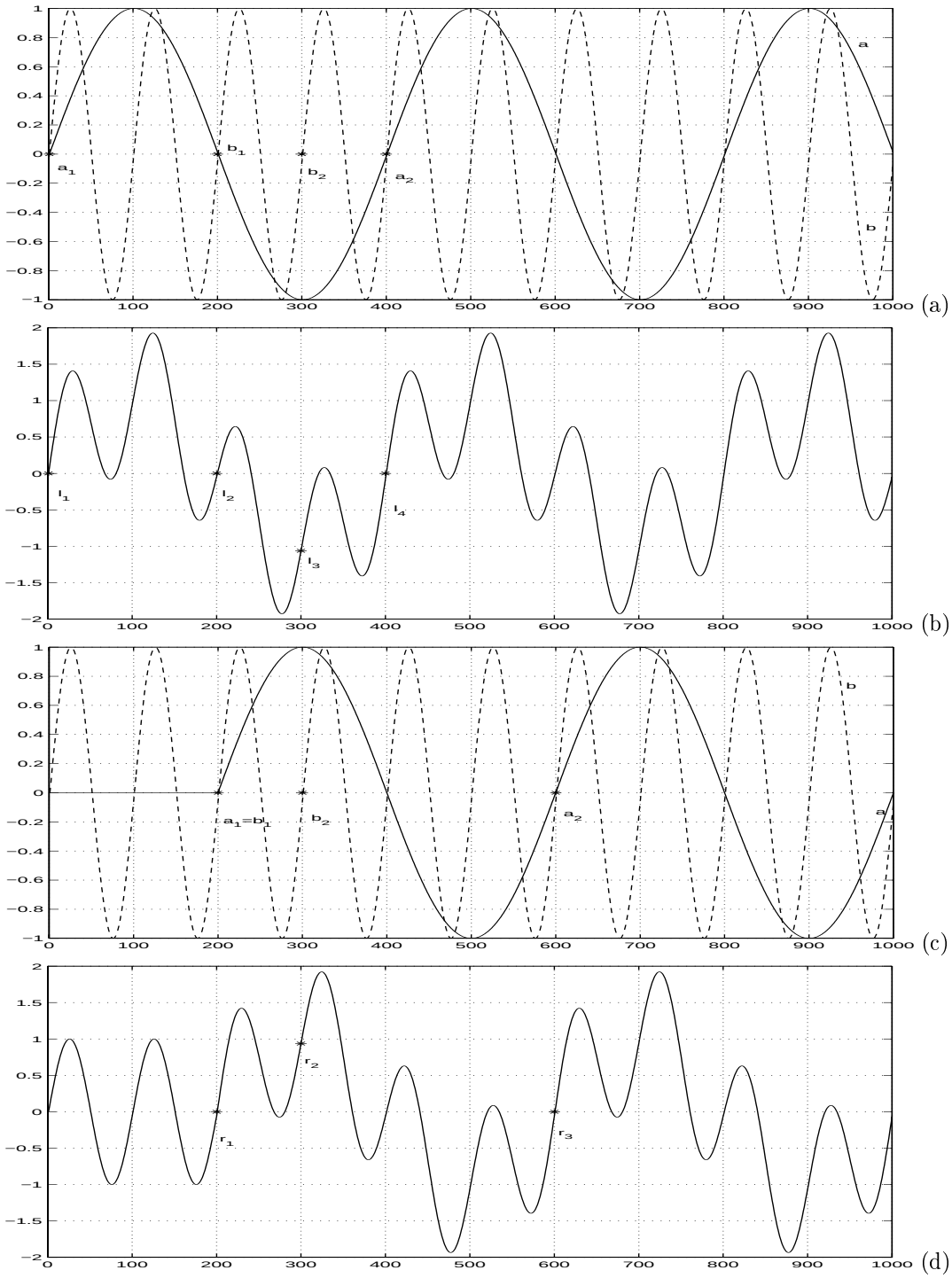

Figure 4: A system with two sources at different locations. One of the sources generates wave a and the other wave b. (a) Wave a and b arrive at the same time at the left ear. (b) The wave that results from the combination of wave a and b at the left ear. (c) Wave a arrives later than b at the right ear. (d) The wave that results from the combination of wave a and b at the right ear. Point a_1 is mapped into l_1 and r_1, point a_2 is mapped into l_4 and r_3, point b_1 is mapped into l_2 and r_1 and point b_2 is mapped into l_3 and r_2.

6.67°). The worst results are obtained with crumpling and tearing paper samples (the error is around 13.34°). The error for this position is therefore between 0° and 13.34°. Though that may seem a lot, in fact it is a very good result. Jeffress [1975] reports that for humans the average error of the azimuth localisation is between 4.6° (at 0° in azimuth) and 16.3° (at 45° in azimuth), then it decreases to 15.6° at 75° in azimuth and it increases again to 16° at 90° in azimuth[9]. Additionally, Jeffress reports that the average error of the sound source's azimuth localisation also depends on the frequency of the sound.

sound	region with more votes
voice	−7
cough	−8
crumpling paper	−9
tearing paper	−9

Table 1: Several sounds at the same location. The correct region for these sounds is -7, which corresponds to -53.36° to -46.69° in azimuth. The values in the table show the region with most votes in the vector of votes.

From all the other tests done the same results were obtained (the error was somewhere between 0° and 13.34°). Exceptions were found in the case of clapping hands and footsteps (with rubber sole shoes), for which the system was unable to localise the source. That may be due to the fact that too many echoes were generated with this kind of sound.

From the tests done with the same sound at different positions it was possible to observe that the error can change with azimuth. See for instance table 2. The maximum error in this example is at -48° and 48° in azimuth, while no error is obtained at 0° in azimuth. That agrees with Jeffress' statement that the minimum average error is at 0° in azimuth and the maximum at about 45° [Jeffress, 1975].

region	region with more votes
-7	-9
0	0
7	9

Table 2: The same sound at different locations. The sound, crumpling paper, was recorded at -48° (region -7), 0° (region 0) and 48° (region 7). The values in the table show the region with more votes in the vector of votes.

Also different sounds have more votes in different bands. For instance, sounds like crumpling papers have more votes in higher bands (bands 1, 2 and 3) whereas sounds like voices have more votes in lower bands (bands 4, 5 and 6). As confirmed by experiments, that leads to good results when more than one sound is heard and the filters are used to analyse them in different bands. Moreover it was observed that the system can also have good results with sounds that do not have such a different spectra.

[9]Humanity was thought to be the species that could localise sound sources more accurately until recently, when Payne [1962] discovered that barn owls are even more accurate than humans.

4 Conclusion

In this paper we presented a biologically plausible system that uses ITDs to estimate the azimuth of sound sources. The system developed can estimate azimuth better than humans. When dealing with some sounds (like voices) it can achieve a close to ideal response. The SSAE system has shown to have an estimation error between 0° and 13.34°. It was also observed that, just as with humans, the estimation of the azimuth accuracy depends on the region (the estimations are more accurate at 0° in azimuth than at around 45°).

The system is also able to identify and localise different sources emitting sound at the same time (with the same range of errors as the ones described in the last paragraph). Though it was not expected, different sounds can even be identified and localised within the same band of the bank of bandpass filters.

Despite its ability to localise a great range of sounds, the system is not able to localise sounds that generate a great quantity of echoes, like clapping hands.

Even though the duplex theory states that ITDs are used just for low frequencies (such that the wavelength is bigger than the distance between the ears) the system developed allows that in certain cases (if $IPD < 2\pi$) ITDs can be used with higher frequencies to identify the source's azimuth.

To conclude, in our opinion, the next step should be the introduction of the analysis of intensity and IID cues into the system. Firstly, the system would not only improve its azimuth estimation of sounds with some high frequency components but would be able to localise sounds composed only of high frequencies as well. Moreover, the mammalian auditory system uses both ITDs and IIDs to localise sound. Also the system could use the changes in intensity at each ear to distinguish sources moving towards or away from it.

Acknowledgments

This research was performed at the University of Edinburgh and was supported by a grant from the Junta Nacional de Investigação Científica e Tecnológica.

References

[Babeanu, 1994] A. Babeanu. Steerable ears for robotic cat. Master's thesis, Dept. of Artificial Intelligence, University of Edinburgh, 1994.

[Guyton, 1984] A.C. Guyton. *Fisiologia Humana*. Guanabara, 1984.

[Handel, 1989] S. Handel. *Listening, An Introduction to the Perception of Auditory Events.* A Bradford Book, the MIT Press, 1989.

[Jeffress, 1975] L.A. Jeffress. Localization of sound. In W.D. Keidel and W.D. Neff, editors, *Auditory System Physiology (CNS). Behavioural Studies Psychoacoustics*, chapter 10, pages 449–459. Springer-Verlag, 1975.

[MacFarland, 1993] D. MacFarland. *Animal Behaviour.* Addison Wesley, 1993.

[Marr, 1982] D. Marr. *Vision.* Freeman, 1982.

[Mills, 1972] W. Mills. Auditory localization. In J.N. Tobias, editor, *Foundations of Modern Auditory Theory*, volume 3. Academic Press, 1972.

[Musicant *et al.*, 1990] A.D. Musicant, J.C.K. Chan, and J.E. Hind. Direction-dependent spectral properties of cat external ear: New data and cross-species comparisons. *Journal of the Acoustical Society of America*, 87(2):757–781, February 1990.

[Payne, 1962] R.S. Payne. How the barn owl locates prey by hearing. In *Living Bird*, volume 1, pages 151–159. 1962.

[Tan, 1996] W.J. Tan. The cat ears project 1996. Master's thesis, Dept. of Artificial Intelligence, University of Edinburgh, 1996.

[Zemlin, 1988] W.R. Zemlin. *Speech and Hearing Science, Anatomy and Physiology.* Prentice Hall, 1988.