The spectrogram correlation and transformation receiver, revisited

H. Peremans and J. Hallam

September 3, 1997

Mailing address:
Department of Artificial Intelligence
University of Edinburgh
Forrest Hill, 5
EH1 2Ql Edinburgh, UK

Tel: 0131-6511212
Fax: 0131-6506899
email: herbertp@dai.ed.ac.uk

Abstract

The spectrogram correlation and transformation receiver (SCAT) has been proposed as a model for the receiver structure used by fm-bats. This paper analyses which features of the proposed model are responsible for its high accuracy in estimating arrival times of overlapping echoes. Apart from providing an answer to this question the analysis will also indicate the limitations of the SCAT receiver. In particular, it is shown that the temporal block of the SCAT receiver returns erroneous results for inter-echo delays < 20 μs and for inter-echo amplitude ratios > 6dB. It is also shown that the spectral block of the SCAT receiver generates spurious arrival times if more than two overlapping echoes are present. Finally, it is discussed how the characteristic errors predicted by this analysis could be used to further verify the validity of the SCAT receiver as a model of the receiver structure used by fm-bats.

1 Introduction

In many naturally occurring situations bats seem to show a remarkable capacity to track a single reflecting object among a densely packed group of reflectors [1]. Furthermore, in addition to tracking the object of interest, the bats must also have some notion of the positions of the other reflectors as shown by their successful chase of insects in vegetation [2] without colliding with the branches and leaves around them. This seems to indicate that bats are somehow capable of decomposing the summation of highly overlapping echoes picked up by their ears into its constituent parts.

Simmons and his coworkers propose that this decomposition occurs in the temporal domain, i.e. the bat's brain builds a monaural time-domain representation of the signal picked up by each ear [1, 3, 4]. In this representation each echo is mapped onto its arrival time. Information about the 3D position [5, 6, 7] of a reflector is derived from the interaural and intra-aural time differences between the arrival times corresponding with that reflector. Because of the smallness of these time differences it has been argued by other authors [8, 9, 10] that instead of using time differences bats are more likely to use inter-aural and intra-aural intensity information to derive reliable angular position estimates. However, it should be noted that the mechanisms based on intensities proposed so far [10] assume isolated echoes and cannot explain the bat's performance in highly cluttered environments. Simmons et al. [1] point out that to derive the intensities of the individual echoes from the composite received signal the arrival times of the echoes should be extracted first. Hence, even if the time differences prove too small to be of use to the bat for position estimation, the time domain decomposition of the signal would still be required to extract reliable intensity information in highly cluttered environments. Experimental evidence for this temporal domain decomposition is derived from two-glint experiments as described in [1, 11, 4]. These experiments also indicate that the bat is capable of measuring time-delays between two overlapping echoes that are as small as 2 μs.

In [4] Saillant et al. propose a receiver structure (SCAT) that combines a time domain, coarse time information, and frequency domain, fine time information, analysis of the received signal to arrive at the temporal decomposition referred to above. The advantages of combining time and frequency domain analyses of the received signals had already been described by Beuter in [12] but the SCAT model introduces the additional assumption that the final representation is solely a time domain one. The model's performance is very similar to the one shown by bats in the two-glint experiments. However, although they show through numerous examples how well the results predicted by the proposed receiver structure match the actual results from experiments performed on bats, the authors do not explain where its resolving power comes from. Hence, to get a better understanding of where its power comes from and, equally important, what the

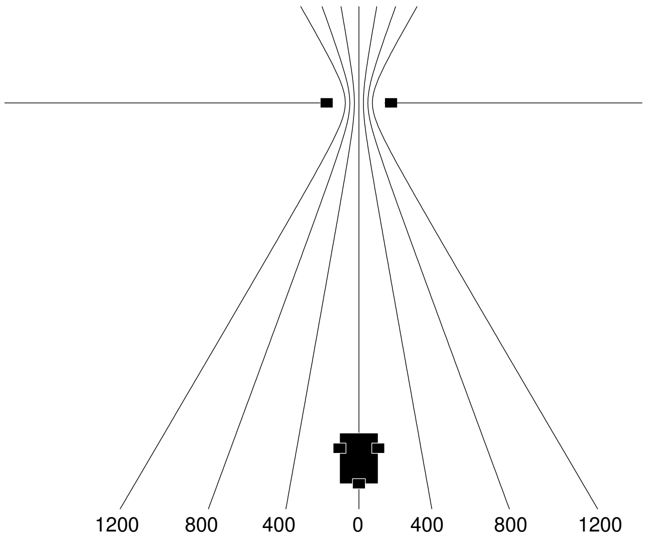

Figure 1: Different paths the robot can take to drive through an opening. Each path is an hyperbola along which the delay, numbers are in μs, between the echoes from the edges is constant.

limits to its correct operation are we present a more formal analysis of this receiver structure. An important result from this analysis is the suggestion of new experiments that can help to decide on the validity of the SCAT as a model for the receiver structure used by fm-bats.

So far, we have looked upon the SCAT solely as a model of a biological sonar system. However, our research [13, 14] has shown that the sensor interpretation problems encountered by bats are very similar to the ones encountered by mobile robots fitted with ultrasonic sensors. In particular, we have shown that the single most important source of errors is the occurrence of highly overlapping echoes. We have also found, in agreement with Simmons et al. [1], that many frequently occurring situations give rise to such overlapping echoes. Furthermore, particular echoes overlap quite often for considerable periods of time making it impossible to remove the possibly erroneous measurements by out-lier rejection techniques. This can be appreciated by considering a mobile robot driving through an opening while looking at the two edges of the objects defining the opening. As can be seen from Fig. 1 the echoes from the two edges will overlap considerably at both ears along any reasonable path that leads through the opening. Hence, a better understanding of the principles from which the SCAT model derives its resolving power can lead to a significant improvement in robotic sonar systems.

In section 2 we briefly describe the spectrogram correlation and transformation receiver. The analysis of the temporal and spectral components of the SCAT is given in section 3. This analysis will expose particular assumptions made by the SCAT. Section 4 contains a discussion of a set of experiments designed to bring out the specific errors resulting from violating those assumptions. These experiments can help decide on the biological plausibility of the SCAT.

2 The spectrogram correlation and transformation receiver

In this section we will summarise the receiver structure proposed by P. Saillant et Al. [4] as it was implemented by us. The SCAT receiver consists of three blocks. The cochlear block

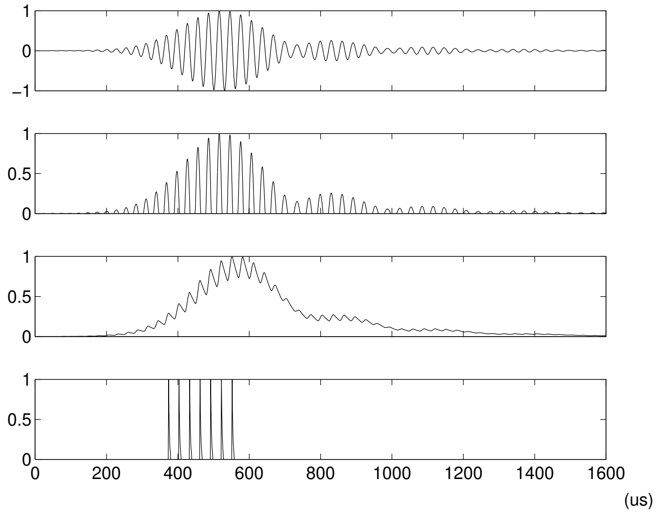

Figure 2: The outputs at different points in the cochlear block. From top to bottom: the bandpass filter, the half-wave rectifier, the low-pass filter and the neural coding.

takes in the signal as received by the ear and analyses it in a number of frequency channels. The temporal block will use the timing information about corresponding events in the different frequency channels to derive an estimate for the arrival time of a group of overlapping echoes. This arrival time of the group is then further refined by the spectral block which uses the relative amplitudes of the outputs of the different frequency channels to derive estimates for the delays between the individual echoes that make up the group.

In all subsequent examples the transmit cry is a frequency-swept sine wave, its period linearly modulated, starting at 110 kHz and going down to 15 kHz in 2.0 ms.

2.1 The cochlear block

This block derives a spectrogram representation from the original signal. The implementation consists of a filterbank, 81 bandpass filters, a set of half wave rectifiers and a low-pass filtering stage. The bandpass filters are 10th order Butterworth IIR filters with a fixed 3dB bandwidth of 4 kHz. The central frequencies of the bandpass filters range from 20 kHz to 100 kHz. The frequency scale is hyperbolic, i.e. the centre periods of the bandpass filters are linearly spaced. The output of each of the bandpass filters is half-wave rectified and low-pass filtered. The low-pass filter is a 1st order Butterworth IIR filter with cut-off frequency equal to 3 kHz.

The outputs of the low-pass filters are then converted into discharge patterns, each discharge is a 1 μs pulse. We will consider 'peak detection' neurons only. They mark the occurrence of an event by discharging whenever a local peak in the low-pass filter output is above the threshold for that particular frequency channel. During its refractory period, lasting 1 ms, the neuron can only be retriggered by a new local peak if this new peak is larger than the peak that last triggered the neuron.

The outputs at different stages in a particular frequency channel from the cochlear block are shown in Fig. 2.

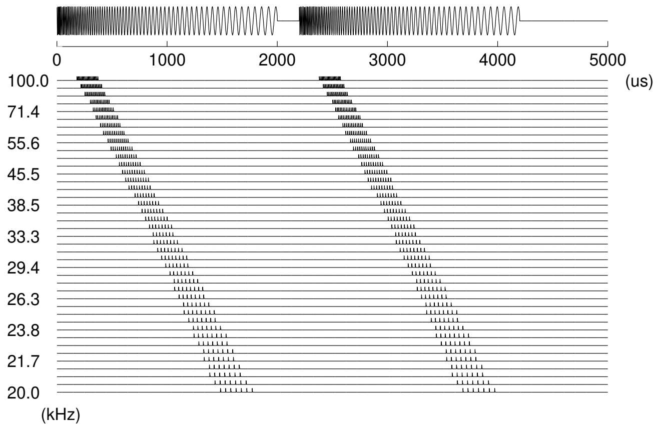

Figure 3: Activity of the peak detecting neurons due to the emission and a single echo.

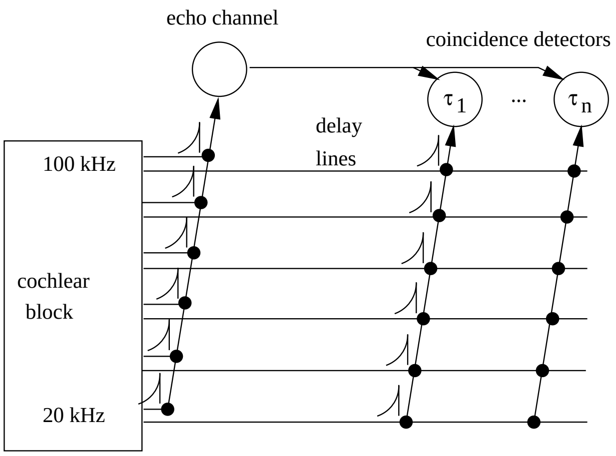

Figure 4: The architecture of the spectrogram correlation block.

2.2 The spectrogram correlation block

This block, also called the temporal block, determines the delay between the emission and the echo by making use of tapped delay lines. If multiple highly overlapping echoes are received, i.e. delays between echoes smaller than the integration time 350 μs, this block returns the delay between the emission and the first echo.

First, the sonar emission itself is processed as the model assumes that the ears pick up the emitted cry and use this signal to make the receiver self-calibrating. So, the pulse volleys caused by the emission are fed into a set of tapped delay lines, taps at 1 μs separation, one delay line for each frequency channel. Next, the delay between the volley produced by the emission and the volley produced by the echo is measured, Fig. 3. This is done by a series of coincidence detectors each tuned to a different delay as shown in Fig. 4. One set of inputs of each coincidence detector is connected to the tapped delay lines. They add together the neural activation, i.e. the pulses caused by the emission, travelling along the different delay lines. The other set of inputs is connected directly to the outputs of the cochlear block and sum the neural activation in the different frequency channels due to the echo. However, because of the fm-swept nature of the cry the high frequency channels are activated first as the emission frequency sweeps down, see Fig. 3. Hence, to sum the activity in the different frequency channels due to the same echo these delays have to be compensated for. Therefore the inputs to the coincidence detecting neuron are taken from points that are shifted by these characteristic delays with respect to each other, Fig. 4.

To implement a variable coincidence-detection resolution the neural pulses, duration 1 μs, are converted into decaying exponentials with a variable time constant before the summation is performed. The default value for this time constant is set to 2μs. Finally, the summed activity on each set of inputs is compared with a threshold, when the threshold is exceeded on both the delayed emission inputs and the echo inputs the coincidence detector triggers and signals an echo at the particular delay value represented by it.

2.3 The spectrogram transformation block

This block, also called the spectral block, determines the delay between the first and the second echo using the outputs of the low-pass filters from the cochlear block. Adding this delay to the arrival time of the first echo extracted by the temporal block results in the arrival time of the second echo. Note that the results from the spectral block are only used when the delay between the two echoes is smaller than the integration time of the cochlear block, i.e. 350 μs. For delays larger than the integration time the temporal block on its own returns accurate estimates for all the arrival times.

Each frequency channel in the spectral block casts a series of votes in favour of certain delays and against other delays between the two echoes. This is implemented by representing the delay-dependency of the votes cast by a frequency channel by a cosine wave. The positive regions of the cosine wave correspond with delay values favoured by the frequency channel and the negative regions with delay values it disagrees with. The frequency of the cosine wave is equal to the central frequency of the channel. All the cosine waves are started at the time indicated by the temporal block as being the arrival time of the group of echoes. Furthermore, each voting pattern, i.e. cosine wave, is scaled by the strength of the echo at the central frequency of the channel. This strength is determined by integrating the part of the output of the low-pass filter falling within a window with length 350 μs and centred around the arrival time of the first echo as detected by the temporal block. This value is normalised with respect to the emission strength. After adding the voting patterns from the different frequency channels together, the spectral block returns the delays that exceed a threshold.

Finally, the cosine vote patterns are modified somewhat to guarantee good performance of the receiver for very closely spaced echoes as well. Because of the finite frequency range spanned by the frequency channels the spectral block does not return a single impulse at zero delay when a flat spectral response, meaning only one object is present, is measured. To avoid this phenomenon from interfering with the detection of very small delays (< 16 μs), a new set of basis vectors is derived from the cosine waves by adjusting them such that for every delay value their sum equals zero. Furthermore, to compensate for the nonuniform contribution of the energy per frequency caused by the hyperbolic spacing of the constant bandwidth frequency channels, the basis vectors corresponding with the different frequency channels are weighted appropriately.

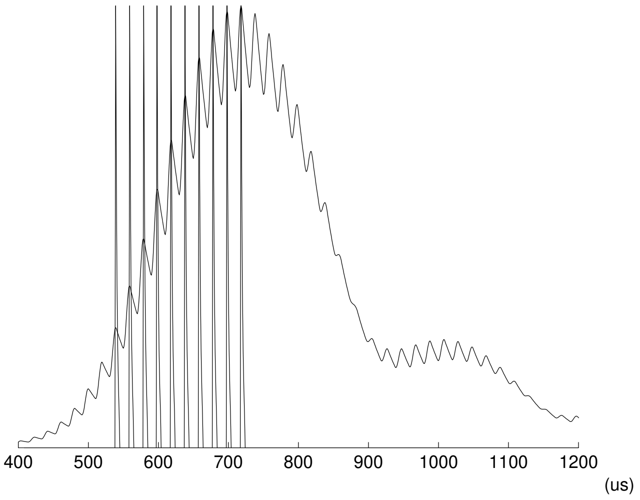

Figure 5: Activity of the peak detecting neurons is synchronised with the ripple on the outputs of the low-pass filters

3 Analysis of the SCAT receiver

In this section we want to look at the mechanisms by which the receiver structure described above works as claimed. This analysis will also shed some light on the limits beyond which the receiver will no longer operate correctly. First, we will look at how the temporal block manages to return a remarkably accurate estimate of the arrival time for a group of closely spaced echoes and next we will see how the spectral block refines this result by providing estimates for the delays between the arrival time of the first echo, i.e. the arrival time of the group, and the arrival times of the subsequent echoes.

3.1 The temporal block

To help us gain a better understanding of its operation in the presence of multiple, overlapping, echoes we look at the temporal block's handling of single echoes first.

3.1.1 Non overlapping echoes

From Fig. 3 we note that the separation between subsequent pulses from a burst increases along with the centre period for the different frequency channels. Looking at Fig. 5, showing the output of the low-pass filter together with the neural pulses triggered by it, we can be even more specific. The pulses generated by the cochlear block are synchronised with the ripple still present on the outputs of the low-pass filtered signals. This behaviour is a consequence of the way the 'peak detection' neurons are defined.

To make the previous statements more precise we now derive the exact positions of the neural pulses when only a single echo is present. The output of each of the bandpass filters in response to an echo [15] can be written as

$$o_i(t) = a(t)\cos(2\pi f_i t),$$

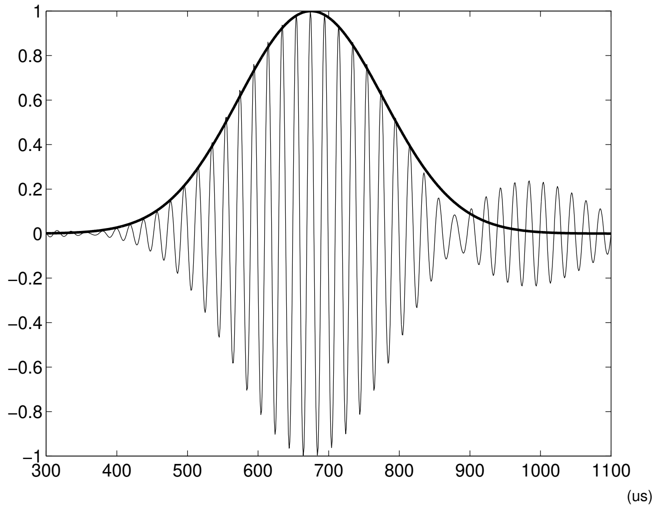

Figure 6: Approximating the envelope of the output of a cochlear bandpass filter by a gaussian.

for $i = 1 \cdots N$, where f_i denotes the central frequency of the ith frequency channel, and $a(t)$ the envelope of this output. To simplify the calculations we approximate the envelope around its maximum $t = t^+$ by

$$|o_i(t)| = a(t) = \exp(\frac{(-(t-t^+)^2}{\sigma^2}), \tag{1}$$

where σ is determined by the bandwidth of the bandpass filter, as shown in Fig. 6. The extrema of the output of the bandpass filter occur at times $t = t^*$,

$$0 = a'(t^*)\cos(2\pi f_i t^*) - a(t^*) 2\pi f_i \sin(2\pi f_i t^*), \tag{2}$$

where the prime indicates time derivatives. From Eqs. (1) and (2) we get

$$\tan(2\pi f_i t^*) = \frac{-2(t^* - t^+)}{2\pi f_i \sigma^2}. \tag{3}$$

Putting $t^* = t^+ + \Delta t$ and $t^+ = (m+\alpha)T_i$ with $\alpha \in [-1/2, 1/2]$ and m an integer, we want to find the smallest Δt that locally maximises $o_i(t)$ as this will correspond with the global maximum of $o_i(t)$. From Eq. (3) and approximating $\tan(\theta) \simeq \theta$ for small θ we get

$$\Delta t = \frac{-\alpha T_i}{1 + \frac{2}{(2\pi f_i \sigma)^2}} \simeq -\alpha T_i.$$

Hence $t^* \simeq mT_i$, i. e. the global maximum occurs at the maximum of the modulating cosine-wave closest to the maximum of the envelope and all other local maxima coincide with the maxima of the modulating cosine-wave. In terms of the phasor representation of the signal $o_i(t)$, Fig. 7, this means that the maxima occur for the phase of the phasor $\gamma = k2\pi$ for $k = 0, 1, 2, \cdots$ despite the time dependency of the envelope $a(t)$.

Note that we have determined the positions of the maxima of the outputs of the bandpass filters. However, the positions of the neural pulses are determined by the outputs of the low-pass filters, Fig. 5. The latter can be easily derived from the results we have got since the half-wave

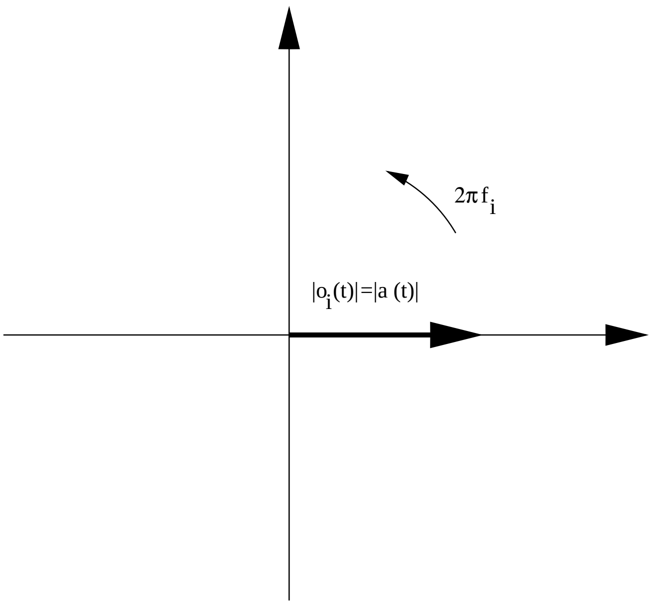

Figure 7: The phasor representation of the output signal $o_i(t)$ if a single echo is present.

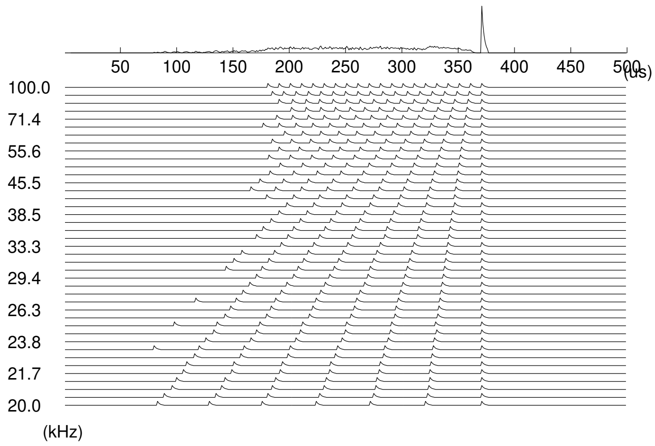

Figure 8: Activity of the peak detecting neurons due to a single echo (realigned time axes). The top line shows the summed activity.

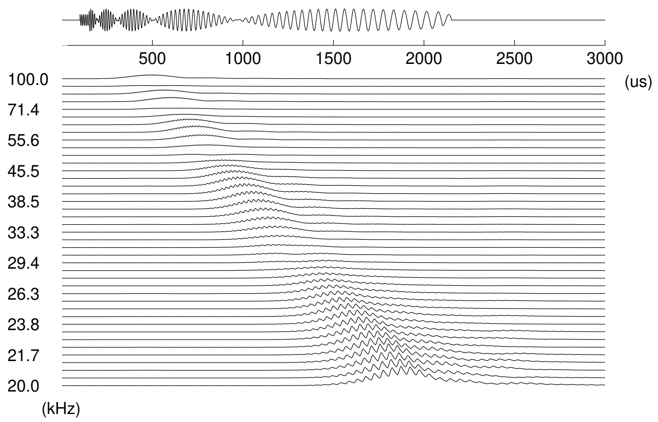

Figure 9: The outputs of the low-pass filters for two overlapping echoes, $\tau = 50\mu s$.

rectifier changes nothing to the positions of the maxima, as can be seen from Fig. 2, and the low-pass filter introduces a constant shift only.

In Fig. 8, we zoom in on the activity burst due to the echo. The time axes are realigned taking into account the different delays corresponding with the different frequency channels. Note that these delays were arrived at by a self-calibration procedure that measures the delays between the position of the peak (global maximum) of the emission response in the first channel and the position of that same peak in the other channels. Hence, by construction, the pulses produced by the peak (global maximum) of the echo response will be the only ones that line up after realigning the frequency channels. The other pulses, being separated by a multiple of the centre period of their frequency channel from the true peak, will be smeared out after realignment, as can be clearly seen from Fig. 8. Hence, the well defined peak in the summed activity of the different frequency channels will reliably trigger the coincidence detecting neuron in the case of an isolated echo.

3.1.2 Two overlapping echoes

We have shown in Fig. 9 the signals at the outputs of the low-pass filterbank when two highly overlapping, i.e. delay < 350 μs, echoes occur. Contrary to an isolated echo, there is now no longer a single peak in all frequency channels. Furthermore, in those channels that do contain a single peak the position of that peak has shifted with respect to where it would be if either the first or the second echo were the only echo present. As we saw in the previous section the neural pulses from a burst are separated by the centre period of the frequency channel and the last spike from the burst corresponds with the peak itself, Fig. 5. Therefore, shifting the peak will shift all the neural pulses in the burst as well. Since the amount of shift of the peaks will vary from one frequency channel to another, the correct operation of the spectrogram correlation block is no longer guaranteed. However, despite the shifted and possibly doubled peaks the model still returns a surprisingly accurate estimate of the first echo's delay. How is that possible? It turns out that the positions of the shifted peaks, both double and single peaks, are still related to the positions of the peaks that would occur if only one of the two echoes were present. Below we analyse this phenomenon more quantitatively.

In the presence of two echoes the output of the i-th bandpass filter from the cochlear block

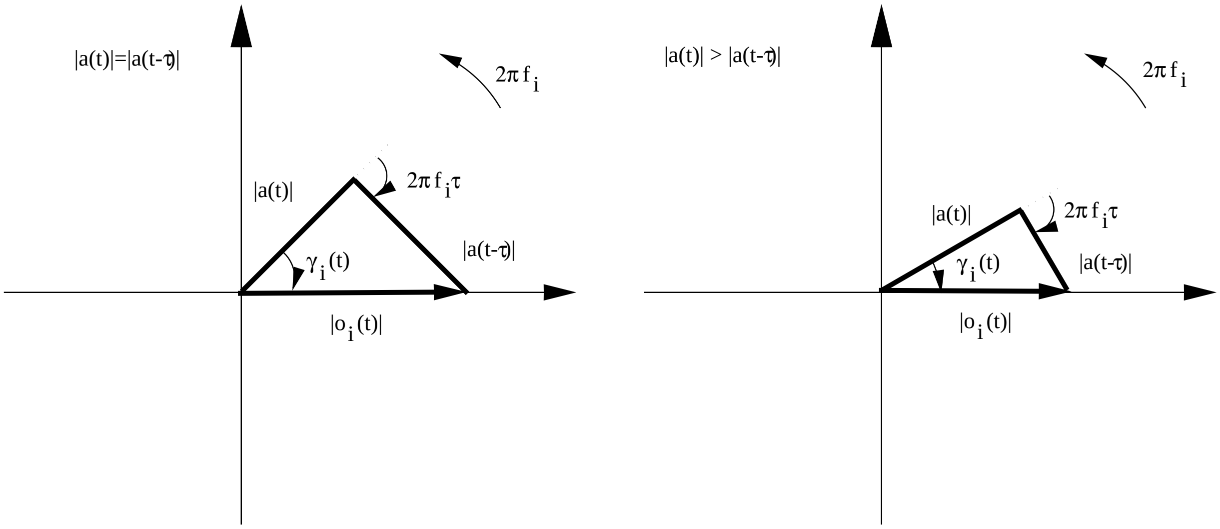

Figure 10: The phasor representation of the output signal $o_i(t)$ if two echoes are present, for $a(t) = a(t-\tau)$ and for $a(t) > a(t-\tau)$.

can be written as

$$o_i(t) = a(t)\cos(2\pi f_i t) + a(t-\tau)\cos(2\pi f_i (t-\tau)),$$

for $i = 1 \cdots N$, where τ denotes the delay of the second echo with respect to the first echo. Fig. 10 shows the phasor representation of $o_i(t)$, i.e. its squared envelope

$$|o_i(t)|^2 = a^2(t) + a^2(t-\tau) - 2a(t)a(t-\tau)\cos(\pi - 2\pi f_i \tau), \tag{4}$$

and its phase

$$\gamma_i(t) = \arctan(\frac{\sin(2\pi f_i \tau)}{\cos(2\pi f_i \tau) + a(t)/a(t-\tau)}). \tag{5}$$

From the analysis in the previous section we conclude that the phase difference $\gamma_i(t^*)$ corresponds with the time difference between the maxima occurring when a single echo is present and the maxima for a two echo signal. We have again denoted the times at which $o_i(t)$ is maximal by $t = t^*$. We can derive a number of properties of $\gamma(t^*)$ from Eq. (5), all of which can be verified from the graphical representation shown in Fig. 10. As $2\pi f_i \tau$ varies for different frequency channels, $\gamma_i(t^*)$ will vary as well but we will always have

$$\gamma_i(t^*) \in [-\frac{\pi}{2}, \frac{\pi}{2}]. \tag{6}$$

Assuming an envelope which is symmetric with respect to its maximum, we will have $a(t^* - \tau) = a(t^*)$ for $t^* = \tau/2$ and hence

$$\gamma_i(t^*) = \frac{2\pi f_i \tau}{2}.$$

In all other cases for $t^* < \tau/2$ we have $a(t^* - \tau) < a(t^*)$ and hence

$$\gamma_i(t^*) < \frac{2\pi f_i \tau}{2}, \tag{7}$$

(μs)	t_1	τ	t_1	τ	t_1	τ	t_1	τ	t_1	τ	t_1	τ	t_1	τ
true value	1000	200	1000	100	1000	50	1000	25	1000	20	1000	10	1000	5
SCAT	1001	200	1002	100	1002	50	1000	25	1011	20	1005	11	1003	6

Table 1: Comparison of the estimated arrival times of the two echoes and the true ones (arrival time first echo: t_1, delay: τ).

and $\gamma_i(t^*) \to 0$ for $a(t^* - \tau)/a(t^*) \to 0$. For $t^* > \tau/2$ we have $a(t^* - \tau) > a(t^*)$ and if we redefine $\gamma(t^*)$ to represent the phase difference between the maxima of the second echo and those of the two echo signal Eq. (7) remains valid.

Eq. (6) indicates that the positions of the maxima when two echoes are present are maximally shifted with respect to the maxima for a single echo over $T_i/4$ and most of the time by much less, as shown by Eq. (7). Again, by the same reasoning as described in the previous section, this result about the positions of the maxima of the outputs of the cochlear bandpass filters is equally valid for the maxima of the outputs of the low-pass filters. Hence, we conclude that the positions of the neural pulses when two echoes are present will also be maximally shifted by $T_i/4$ with respect to the neural pulses caused by a single echo. Note that this is not strictly true for the global maximum as in this case there can be an additional shift mT_i, i.e. an unknown but integer multiple of the centre period of the frequency channel.

Knowing these properties of the positions of the neural pulses we can explain the robustness of the temporal block in the presence of overlapping echoes. The coincidence detection scheme in the temporal block is a voting scheme with two important features. First, each frequency channel votes for a number, i.e. more than one, of possible peak positions, see Fig. 8. Second, the neural spikes are exponentially decaying pulses, time constant 2μs, which means that the vote for a particular peak position is actually spread out over its close neighbours in an exponentially decaying fashion. Despite being fed shifted peaks only, the neural pulses are shifted by $(m + \alpha)T_i$, $\alpha \in [-1/4, 1/4]$ with respect to the single echo case, these two features allow the temporal block to produce an accurate estimate of the true position of the first echo nevertheless. First, the temporal block compensates for the unknown integer m by casting multiple votes, at multiples of the centre period, for each frequency channel. Next, the equally unknown but much smaller, maximally $T_i/4$, shift is compensated for by convolving each vote with an exponentially decaying function. Note that this latter compensation is only a partial one as the decaying exponential has a fixed length. This makes it impossible to spread out each vote over the known uncertainty region $[-T_i/4, T_i/4]$ which increases along with the centre period of the frequency channels. Furthermore, whereas the true uncertainty region is symmetrically distributed around the measured pulse position, the decaying exponential is non-zero in the positive half of this region only.

Comparing the summed activity pattern of a single echo shown in Fig. 8 with that of two overlapping echoes ($\tau = 25\mu$s) shown in Fig. 11, it is clear that it gets increasingly difficult to recognise the peak from the background activity as the delays get smaller. This can be understood by noting that as the delays τ get smaller, the ratio $a(t^*)/a(t^* - \tau) \to 1$ and $\gamma(t^*)$ increases as can be seen from Eq. (5). When $\gamma(t^*)$ increases, the amount of overlap of the exponential vote patterns decreases thereby lowering the peak to background ratio. Nevertheless, in this example the temporal block correctly estimates the arrival time of the first echo to be at 1000 μs. The results of a more comprehensive test, Tab. 1, indicate that the temporal block can reliably estimate delays $\tau \leq 25\mu$s. If the delays get smaller errors occur as will be explained below.

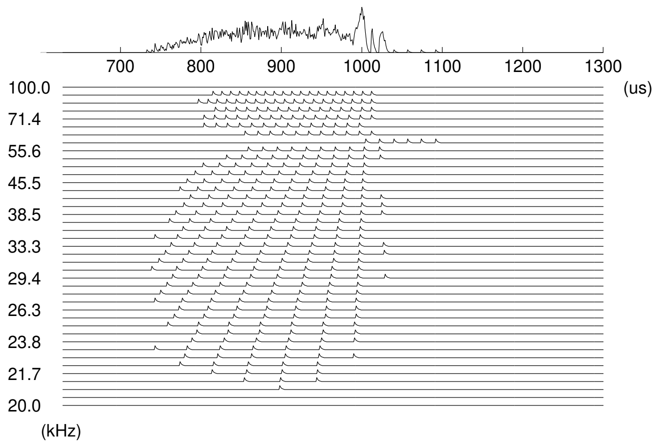

Figure 11: Activity of the peak detecting neurons due to two echoes at 1000 μs and 1025 μs (realigned time axes). The top line shows the summed activity.

3.1.3 Limits to the correct operation of the temporal block

From the results in Tab. 1 it is clear that as the delays get smaller than 20 μs, the spectrogram correlation block returns the time halfway in between the two echoes instead of the arrival time of the first echo. This behaviour can be understood using the results derived above. First we note that, assuming a symmetrical envelope, the two echo signal will have an extremum at the point halfway in between the two echoes. This extremum will be a maximum for all the frequency channels for which the delay $\tau \in [-T_i/2 + 2mT_i, T_i/2 + 2mT_i]$ with T_i the centre period of the channel and $m = 0, 1, 2 \cdots$. For all the channels for which $\tau < T_i/2$ the global maximum in the two echo signal is precisely this maximum halfway between the two echoes. After realigning the frequency channels, the peaks generated by these global maxima will all be lined up since they are shifted over a fixed delay $\tau/2$ with respect to the global maxima of the first echo signal. This phenomenon can be clearly observed in Fig. 12. All the frequency channels with centre period $T_i/2 > 15\mu$s generate a peaks halfway between the two echoes that line up perfectly. The frequency channels with centre period $T_i/2 < 15\mu$s generate peaks that lie around the true arrival time of the first echo. For delays $\tau \leq 20\mu$s the number of channels that contribute votes to the true arrival time falls below the background vote level. At the same time, the number of channels contributing votes to the time halfway between the two echoes rises. These two effects explain why for $\tau < 20\mu$s the peak corresponding with the true arrival time of the first echo is no longer detected and why instead the peak corresponding with the time halfway in between the two echoes is.

Another problem arises if the amplitudes of the overlapping echoes are no longer equal, as they were in all our examples so far. The derivation of Eq. (7) assumes that the magnitude of the envelope of the first echo is larger than that of the second echo when the two echo signal reaches its maxima. This would no longer be guaranteed if the second echo gets stronger than the first one. When the amplitude of the second echo is larger than that of the first echo Eq. (7), as noted before, remains valid but it now describes the separation of the peaks of the two echo signal from those of the second echo. Hence, the voting operation performed in the temporal block will favour the arrival time of the second echo in this case. Such a situation is shown in Fig. 13, the amplitude of the second echo is twice that of the first echo, and as expected the

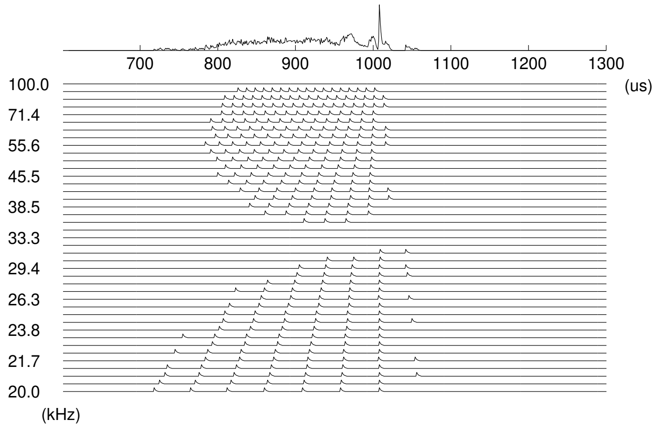

Figure 12: Activity of the peak detecting neurons due to two echoes at 1000 μs and 1015 μs (realigned time axes). The top line shows the summed activity.

arrival time returned is that of the second echo 1100 μs.

3.2 The spectral block

This block converts the spectral representation, i.e. magnitude of the outputs of the filterbank, of the composite echo signal into a series of delay values. As mentioned above, the normalised vote patterns, i.e. the warped cosines, are defined such as to transform a flat spectral representation into a zero vote for all possible delays. Hence, in that case a single echo will be detected and its arrival time will be given by the estimate returned by the temporal block. However, all non-flat spectra will result in the detection of multiple echoes and their arrival times will be given by the delay estimates from the spectral block added to the estimate from the temporal block.

3.2.1 Two overlapping echoes

We start by examining the spectra of signals consisting of two overlapping echoes. When two echoes $e(t)$, equal magnitude a and delays τ_1, τ_2, are present the received signal $s(t)$ is given by

$$s(t) = ae(t-\tau_1) + ae(t-\tau_2),$$

and its spectrum is

$$S(f) = aE(f)\exp^{-j2\pi f\tau_1}(1+\exp^{-j2\pi f(\tau_2-\tau_1)}), \tag{8}$$

with $E(f)$ the spectrum of the echo. From Eq. (8) it is clear that this non-flat spectrum contains notches at frequencies f^- for which

$$f^- = \frac{(2m+1)}{2}\frac{1}{(\tau_2-\tau_1)}, \tag{9}$$

with m an integer. Hence, from the description of the operation of the spectral block, section 2.3, we know that the frequency channels corresponding with these particular frequencies will generate no, or only a very weak, cosine wave. Consequently, the delays d corresponding with odd multiples of half the period p^- of those frequency channels

$$d = \frac{2n+1}{2}p^-,$$

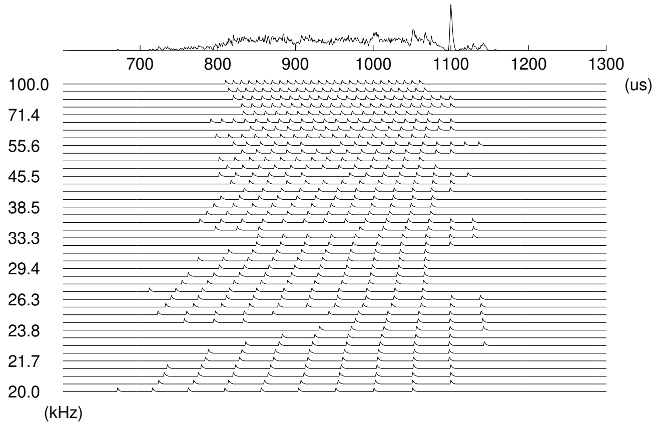

Figure 13: Activity of the peak detecting neurons due to two echoes at 1000 μs and 1100 μs (realigned time axes), the amplitude of the second echo is twice that of the first one. The top line shows the summed activity.

with n an integer, will be inhibited less, i.e. supported more. Similarly, the spectrum will contain peaks at frequencies f^+

$$f^+ = m\frac{1}{(\tau_2 - \tau_1)}. \tag{10}$$

Hence the frequency channels corresponding with those frequencies will generate strong support for the delays

$$d = np^+,$$

with n an integer. Note that Eqs. (9), (10) guarantee that the true delay $\tau_2 - \tau_1$ will always be among the delays d that are supported more. Hence, adding the voting patterns for all the frequency channels together, the true delay $\tau_2 - \tau_1$ will get the highest support. Tab. 1 shows the results of the spectral block on the more comprehensive test. As can be seen from these results, the spectral block returns very good estimates of the delay between the first and second echo down to 5 μs.

However, the conditions as described in this section are quite idealised: echoes have equal amplitudes, echoes are perfect time-delayed copies of emission, only two overlapping echoes. We will see below that the algorithm to convert spectra into delays is quite sensitive to those assumptions.

3.2.2 Limits to the correct operation of the spectral block

The result of analysing the spectrum generated by three overlapping echoes with the spectral block is shown in Fig. 14. The spurious peak in the output of the spectral block corresponds with the delay between the second and the third echo. This behaviour of the spectral block can be understood by examining a spectrum of a signal consisting of three overlapping echoes. If three echoes are present

$$s(t) = ae(t - \tau_1) + ae(t - \tau_2) + ae(t - \tau_3),$$

the spectrum is

$$S(f) = aE(f)\exp^{-j2\pi f\tau_1}(1 + \exp^{-j2\pi f(\tau_2-\tau_1)} + \exp^{-j2\pi f(\tau_3-\tau_1)}).$$

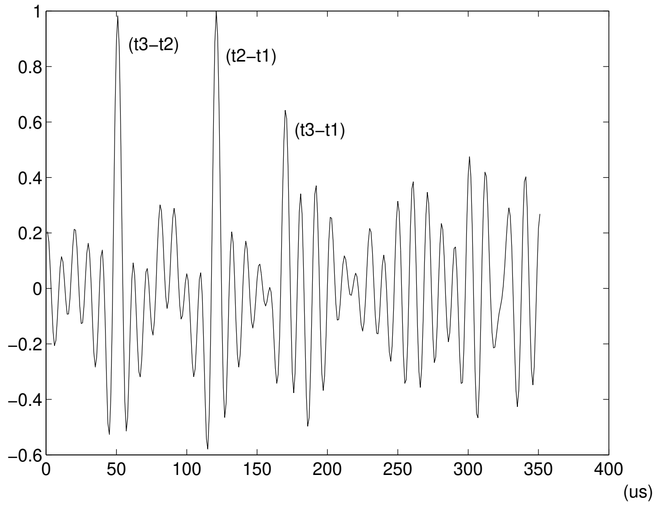

Figure 14: The peaks in the output of the spectral block, indicating the delays between the three overlapping echoes, contain a spurious delay estimate (true delays: 120 μs and 170 μs; estimated delays: 50 μs, 120 μs and 170 μs).

The frequencies f^* that maximise this expression have to comply with the combined constraint

$$\begin{aligned} 2\pi f^*(\tau_2 - \tau_1) &= m2\pi, \\ 2\pi f^*(\tau_3 - \tau_1) &= n2\pi, \end{aligned}$$

with m, n integers. However, all frequencies f^* that do so will also comply with

$$2\pi f^*(\tau_3 - \tau_2) = k2\pi = (n - m)2\pi.$$

Hence, the corresponding frequency channels will not only provide strong support for the delays $\tau_2 - \tau_1$ and $\tau_3 - \tau_1$ but also for the delay $\tau_3 - \tau_2$. Therefore, we conclude that the spectral block will generate spurious delay estimates whenever more than two echoes are present.

As shown by Schmidt [16] varying the ratio of the amplitudes of the second and the first echo results in a reduction of the depth of the notches and the height of the peaks in the spectrum of the two echo signal. We now analyse how this could affect the delay estimate returned by the spectral block. The strength of the two echo signal, delay τ, at the central frequency f_i of the different frequency channels is given by

$$|o_i| = (1 + a_r^2 + 2a_r \cos(2\pi f_i \tau))^{\frac{1}{2}}. \tag{11}$$

with the amplitude of the first echo $a_1 = 1$ and $a_r = a_2/a_1$. The output of the spectral block is proportional to

$$\sum_{i=1}^{N} |o_i|^2 g_i(t), \tag{12}$$

where the number of filters is denoted by N and the warped basis vectors by $g_i(t)$. For a two echo signal, the delay estimate is given by the point where this sum reaches its maximum. From Eqs. (11), (12) and noting that

$$\sum_{i=1}^{N} g_i(t) = 0,$$

by construction, the sum calculated by the spectral block $y(t)$ can be written

$$y(t) = 2\sum_{i=1}^{N} \cos(2\pi f_i \tau) g_i(t),$$

for $a_r = 1$. If the two echoes no longer have the same amplitude the sum calculated by the spectral block can be similarly shown to be proportional to

$$a_r y(t).$$

This scaling however will not influence the position of the maximum. Hence, apart from making the delay estimates more susceptible to noise the scaling will have no direct impact on the results returned by the spectral block.

4 Discussion

The analysis presented above clearly shows that the SCAT receiver as described by Saillant et al. [4] performs very well given the right conditions are fulfilled: not more than two overlapping echoes are present, the echoes have equal magnitudes, the delay between the echoes is more than 20μs. Different parts of the SCAT receiver rely to different degrees on those conditions for their correct operation. Violating these conditions results in very specific errors as shown by the analyses from sections 3.1.3 and 3.2.2. Hence, we propose using the predicted errors to verify the validity of the SCAT as a model for the receiver structure of fm-bats.

The temporal block shifts the arrival time of the first echo to the time halfway in between the two echoes if the delay τ gets smaller than 20μs. According to the SCAT model this phenomenon shifts both arrival times backwards by $\tau/2$. Experiments similar to the ones described in [1, 17] could be used to verify that both arrival times of the echoes in the two echo signal are shifted backwards by the predicted amount. In this type of experiment the arrival time of a single echo probe is varied around the values of the arrival times from a two echo signal. The bats show a marked increase in errors whenever the single echo's arrival time coincides with one of the perceived arrival times from the two echo signal.

The temporal block erroneously takes the arrival time of the second echo to be that of the two echo signal if the amplitude of the second echo is sufficiently larger than that of the first one. Furthermore, as shown in section 3.2.2, the delay between the two echoes should not be affected much by varying the amplitude ratio. Hence, as in the previous experiment, the bats should respond as if only the apparent range to the reflecting object was shifted. Previous studies have shown [18] that many bat species are capable of reliably discriminating between two targets at ranges that differ by less than 100 μs. Hence, one possible way of setting up such an experiment could be to have two target echoes each one consisting of two copies of the recorded cry separated by 100 μs. For the first target the first echo's amplitude would be twice as large as that of the second one whereas for the second target it would be the other way around. The target with the loud first echo would be delayed (< 100 μs) with respect to the other target to compensate for any amplitude induced latencies. The SCAT model predicts that despite this delay the target with the loud first echo would still be judged to be the closest target.

The spectral block introduces spurious echoes in situations where more than two overlapping echoes occur. In the simplest case of three overlapping echoes the model predicts that a spurious echo will be detected at $t_1 + t_3 - t_2$ where t_1, t_2 and t_3 denote the arrival times of the three echoes. As before, an experiment that varies the arrival time of a single echo probe around the values of the arrival times from the three echo signal should make it possible to verify this prediction.

As mentioned in the introduction, apart from SCAT's relevance for biology, its capability of coping with multiple overlapping echoes would constitute a significant improvement over current ultrasonic sensor systems used for robotic applications. However, to determine whether SCAT could play such a role we need to replace the simulated echoes used in this study by actually measured ones. Only real world data can tell us whether the limitations of SCAT, as derived above, are minor inconveniences or whether they present important obstacles to its use by a real robot. Furthermore, all echoes have been taken to be unfiltered copies of the emitted cry so far. But in the real world, absorption, the directivity of emission and reception, reflection upon a real target will all filter the echo to varying degrees. Since the spectral block translates all deviations from a flat spectrum into hypotheses about particular delays between the echoes it is not clear what will happen when real echoes are processed. Hence we are currently implementing SCAT on a real sonar system allowing us to verify its usefulness for robotic applications.

5 Acknowledgement

This work was supported by a Marie Curie fellowship in the framework of the TMR programme initiated by the European Commission.

References

[1] J. Simmons, M. Ferragamo, P. Saillant, T. Haresign, J. Wotton, S. Dear, and D. Lee, *Hearing by Bats*, ch. Auditory Dimensions of Acoustic Images in Echolocation, pp. 146–190. Springer Verlag, 1995.

[2] F. Webster, "Performance of echolocating bats in the presence of interference," in *Animal Sonar Systems: Biology and Bionics.* (R. Busnel, ed.), (France), pp. 673–713, 1967.

[3] J. Simmons and L. Chen, "The acoustic basis for target discrimination by fm echolocating bats," *Journal of the Acoustical Society of America*, vol. 86, pp. 1333–1350, October 1989.

[4] P. Saillant, J. Simmons, and S. Dear, "A computational model of echo processing and acoustic imaging in frequency-modulated echolocating bats: The spectrogram correlation and transformation receiver," *Journal of the Acoustical Society of America*, vol. 94, pp. 2691–2712, november 1993.

[5] N. Suga, "Cortical computational maps for auditory imaging," *Neural networks*, vol. 3, pp. 3–21, 1990.

[6] Z. Fuzessery, "A mechanism for horizontal and vertical target localisation," in *Animal SONAR Processes and Performance (NATO ASI Series)* (P. Nachtigall and P. Moore, eds.), pp. 471–475, Plenum Press, 1988.

[7] B. Lawrence and J. Simmons, "Echolocation in bats: The external ear and perception of the vertical positions of targets," *Science*, vol. 218, pp. 481–483, October 1982.

[8] Z. Fuzessery, D. Hartley, and J. Wenstrup, "Spatial processing within the mustached bat echolocation system: possible mechanisms for optimization," *Journal of Comparative Physiology A*, no. 170, pp. 57–71, 1992.

[9] G. Pollak, "Time is traded for intensity in the bat's auditory system," *Hearing Research*, vol. 36, pp. 107–124, 1988.

[10] R. Kuc, "Sensorimotor model of bat echolocation and prey capture," *Journal of the Acoustical Society of America*, vol. 96, pp. 1965–1978, 1994.

[11] J. Simmons, "Evidence for perception of fine echo delay and phase by the fm bat, eptesicus fuscus," *Journal of Comparative Physiology A*, vol. 172, pp. 533–547, 1993.

[12] K. Beuter, "A new concept of echo evaluation in the auditory system of bats," in *Animal Sonar Systems.* (R. Busnel and J. Fish, eds.), pp. 747–761, 1980.

[13] H. Peremans, K. Audenaert, and J. V. Campenhout, "A high-resolution sensor based on tri-aural perception," *IEEE Transactions on Robotics and Automation*, vol. 9, pp. 36–48, February 1993.

[14] H. Peremans, "A maximum likelihood algorithm for solving the correspondence problem in tri-aural perception.," in *IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, (Las Vegas), pp. 485–492, October 1994.

[15] Berkowitz, ed., *Modern radar: analysis, evaluation and system design.* New York: John Wiley & Sons, Inc., 1965.

[16] S. Schmidt, "Perception of structured phantom targets in the echolocating bat, megaderma lyra," *Journal of the Acoustical Society of America*, vol. 91, pp. 2203–2223, April 1992.

[17] J. Simmons, C. Moss, and M. Ferragamo, "Convergence of temporal and spectral information into acoustic images of complex sonar targets perceived by the echolocating bat, eptesicus fuscus," *Journal of Comparative Physiology A*, vol. 166, pp. 449–470, 1990.

[18] C. Moss and H. Schnitzler, *Hearing by Bats*, ch. behavioral Studies of Auditory information processing, pp. 87–145. Springer Verlag, 1995.