THE ANNALS OF “DUNAREA DE JOS” UNIVERSITY OF GALATI
FASCICLE III, 2006 ISSN 1221-454X
ELECTROTECHNICS, ELECTRONICS, AUTOMATIC CONTROL, INFORMATICS

SONAR TARGET LOCALIZATION BASED ON SPIKE CODED SPECTROGRAMS

Herbert Peremans, Bertrand Fontaine

Universiteit Antwerpen, 13 Prinsstraat 2018, Antwerpen, België

Abstract: Target location is coded into the pattern of spikes that run up the auditory nerve to the bat's brain. Realistic scenes containing multiple, closely spaced, reflectors give rise to complex echo signals consisting of multiple filtered copies of the bat's own vocalisation. Some of this filtering is due to the directivity of the bat’s reception system i.e., the outer ears, and some of it is due to sound absorption and the reflection process. The analysis below concentrates on the conspicuous ridges (notches) these filter operations give rise to in the time-frequency representation of the echo as produced by the bat's inner ear. Assuming multiple threshold detecting neurons for each frequency channel it is shown how the distribution of spike times within the generated spike bursts is linked to the presence and characteristics of these notches. A neural network decoding the spike bursts in terms of target location is described.

Keywords: Auditory system, Neural networks, Robot sensing systems, Sonar signal processing,

1. INTRODUCTION

Bats when flying in a natural environment, e.g. a forest, are routinely confronted with large numbers of (randomly) distributed scatterers of various sizes, shapes and orientations. Hence the bat receives at both ears a complex echo composed of many closely spaced and hence interfering echoes. Nevertheless, they manage to extract the spatial information they need in order to maneuver with great skill through their environment while at the same time using their sonar to hunt for prey (Popper and Fay, 1995). Hence, mobile robot researchers desiring to implement navigation in realistic outdoor environments i.e., containing trees, and other natural vegetation, while solely relying on in-air sonar systems have recognized these bio-sonar systems as interesting sources of inspiration (Reijniers and Peremans, 2004). We believe that understanding better how bats achieve this impressive performance is a worthwhile exercise for engineers also.

The importance of spectral information for bats is well established: their inner ear performs a joint time-frequency analysis of the received echoes, as is the case for all mammalian cochlea’s (Pickles, 1982) Moreover, experiments have shown that bats do indeed complement the binaural Inter-aural Intensity Differences (IID) and Inter-aural Time Differences (ITD) clues with information extracted from the systematic variations in the spectrum of the received echo depending on the spatial position of the reflecting target (Wotton *et al.*, 1995). This direction dependent filtering is to a large extent caused by diffraction of the sound waves around the particular shape of the bat outer ear and head (Müller and Hallam, 2005), corresponding with the Head Related Tranfer Functions (HRTF) studied in human hearing research.

In this paper we describe a bio-inspired time-frequency representation of received, monaural, echo signals and show how this representation allows the extraction of target bearing information both when confronted with single, isolated reflectors as well as with complex, multifaceted, reflectors. Furthermore, it will be shown that monaural HRTF information is sufficient for these purposes.

In Sec.2 we describe the bio-inspired time-

This paper was recommended for publication by Dorel Aiordachioaie

frequency representation. Next, we develop our approach in Sec.3 Finally, in Sec.4 we describe experimental results obtained with this sonar system. In Sec.5 we summarize our findings.

2. SPIKE CODED SPECTROGRAM

2.1 Realistic reflector model

Realistic reflectors consist of multiple closely spaced reflecting facets e.g., the leafs of a tree. Hence we introduce the 6-tuple model $(N_{refl}, a_i, r_i, \alpha_{az}, \alpha_{el}, \theta_i)$ for a realistic reflector (Reijniers and Peremans, 2004), with N_{refl} the number of reflecting facets, $(r_i, \alpha_{az}, \alpha_{el}, \theta_i)$ the polar coordinates with respect to the sonar system as illustrated in Fig. 1 and a_i the strength of the echo reflecting of the i^th^ facet.

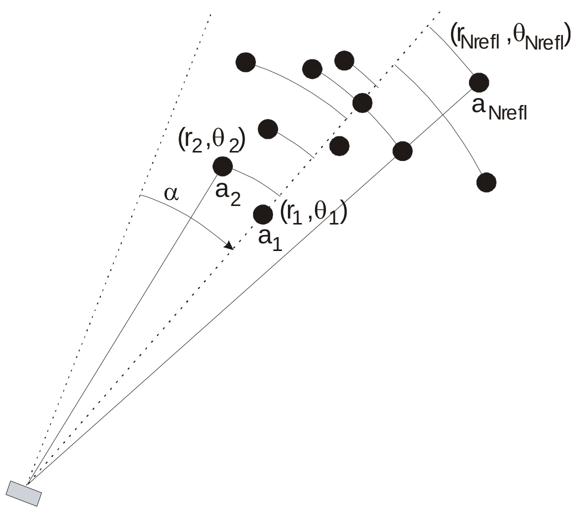


Fig. 1. 2D view of a realistic reflector modeled as a stochastic cloud of point reflectors

For the simulation results presented below we choose the distributions for these stochastic parameters:

$(N_{refl} \sim U[1,7],\ a_i \sim U[0.1,0.5],\ \alpha_{az}, \alpha_{el} \sim U[0^\circ, 60^\circ],\ \theta_i = 0)$

Hence, when transmitting the signal *s(t)*, the echo signal *r(t)* received from such a reflector can be written as

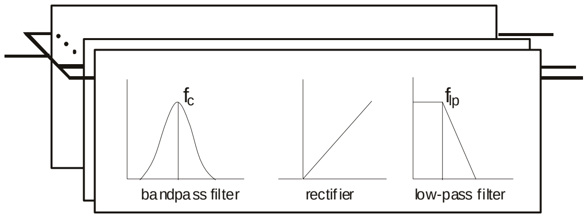
$$r(t) = \sum_{i=1}^{N_{refl}} h_{tr-rec}(t; \alpha_{az}, \alpha_{el}, \theta_i) \otimes a_i \cdot s(t - \frac{2r_i}{v_{sound}}) \quad , \qquad (1)$$

with speed of sound given by v_{sound} and angular dependent filtering (HRTF) denoted by $h_{tr-rec}(t; \alpha_{az}, \alpha_{el}, \theta_i)$. We assume here that the HRTF is mostly due to the outer ear.

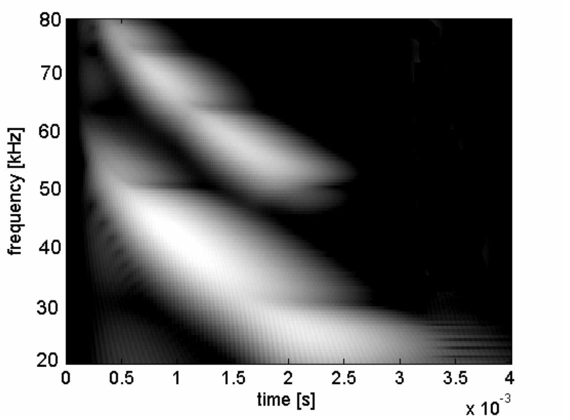
2.2 Spike coded spectrogram representation

The joint time-frequency analysis performed on the incoming signal is modeled on the transduction stage located in the inner ear (cochlea) of the bat. A simple, yet functionally adequate, model of this analysis (Siallant *et al.*, 2005) is a filter bank consisting of parallel band-pass filters with subsequent amplitude demodulation in each channel. This model is illustrated in Fig. 2 showing the output when the received echo signal consists of two identical, shifted copies of the transmit pulse:

$$r(t) = s(t - \frac{2r_1}{v_{sound}}) + s(t - \frac{2r_2}{v_{sound}}) \qquad (2)$$



(a)



(b)

Fig. 2. (a) A simple inner ear model leads to (b) a spectrogram representation of the received echo signal (blue: low amplitude, red: high amplitude), filtering due to complexity reflector.

As illustrated by Fig. 2(b), when the target is no longer a single point reflector the echo signal will be a filtered version of the transmit pulse. However, from (1) we conclude that in addition to the filtering due to the complexity of the reflector the real spectrogram will exhibit extra features due to the angular dependent filtering of the outer ear. Fig. 3 shows how even a single point reflector can return a filtered copy of the transmit pulse where the filtering is now location dependent.

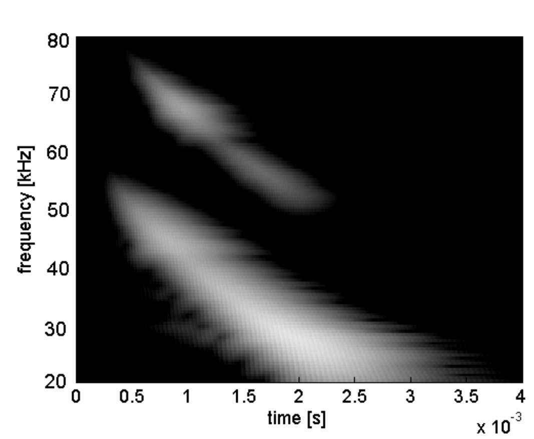


Fig. 3. Spectrogram representation of the received echo signal (blue: low amplitude, red: high amplitude), filtering due to the outer ear.

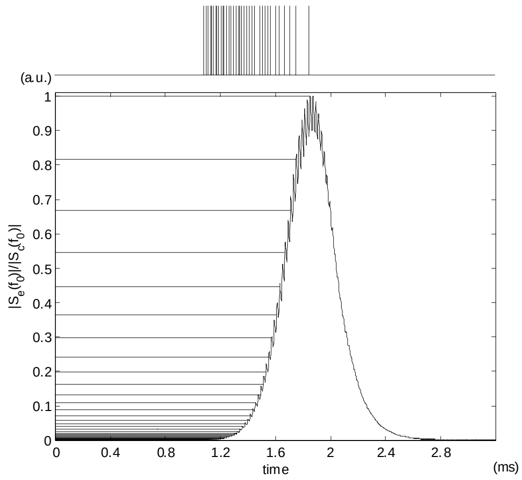
This location dependency of the outer ear filtering is the phenomenon we want to make use of to localize targets. Looking at the spectrograms shown in Fig. 2(b) and Fig. 3 the most prominent features are the notches i.e., the weak response regions. Hence, tracking the movements of these notches (Fig. 3) as the target moves around in the sonar system’s field of view will allow the classifier described below to correctly locate the target if it can ignore the notches (Fig. 2(b)) caused by the complexity of the reflector. Also, it has been suggested that these notches are indeed important clues for elevation estimation by bats (Wotton *et al.*, 1996).

However, it should be noted that the information in the spectrograms shown here is not available to the bat. Indeed, the inner ear produces patterns of neural spikes and sends them up the auditory nerve into the bat’s brain. Hence, to keep the model biologically plausible we perform a final spike conversion. A set of thresholds is applied to the output of each frequency channel producing a spike whenever the up-going edge of the signal crosses one of the threshold levels. Combining these spikes into one spike train for each frequency channel (see Fig. 4) we take this set of spike trains as input for the target localization system. As can be seen by comparing Fig. 4 (a) and (b), both the number of spikes in a spike train as well as their relative positions code for the shape of the output of the respective frequency channels.

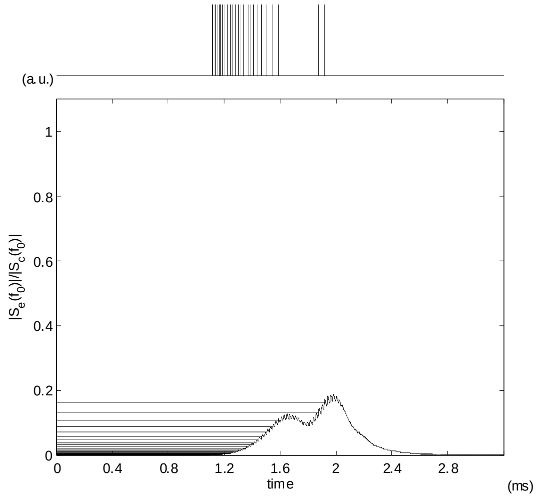
3. TARGER LOCALISATION

3.1. Distinguishing spike trains

To distinguish between spike trains i.e., quantify the difference between the spike trains shown in Fig. 4 (a) and (b), we require a distance measure. Hence,



(a)



(b)

Fig. 4. Running the spectrogram representation of Fig. 2 (b) through a set of threshold detectors (red lines indicate threshold levels) results in a set of spike trains, one for each frequency channel; (a) 44kHz , (b) 93kHz.

let S be a spike train i.e., an ordered sequence of times $t_1, t_2, \dots, t_k$, denoted by $S = \{t_1, t_2, \cdots, t_k\}$. We can then, as detailed in (Victor and Purpura, 1997), define the metric $D(S_a, S_b)$ as the minimum ‘cost’ required to transform the spike train S_a into the spike train S_b via a path of elementary steps where $K(S_j, S_{j-1}) \geq 0$ is equal to the cost of an elementary step from S_j to S_{j-1}

$$D(S_a, S_b) = \min_{S_0, S_1, \cdots, S_r} \left(\sum_{j=1}^{r} K(S_j, S_{j-1}) \right), S_0 = S_a, S_r = S_b \quad (3)$$

The elementary steps that are allowed to transform one spike train into another are: inserting a single spike (cost = 1), deleting a single spike (cost = 1) and shifting a spike by Δt (cost = $q.|\Delta t|$) with q a parameter that controls the trade-off between inserting/deleting spikes and shifting existing spikes.

One such path of elementary steps transforming spike train S_a into the spike train S_b is shown in Fig. 5.

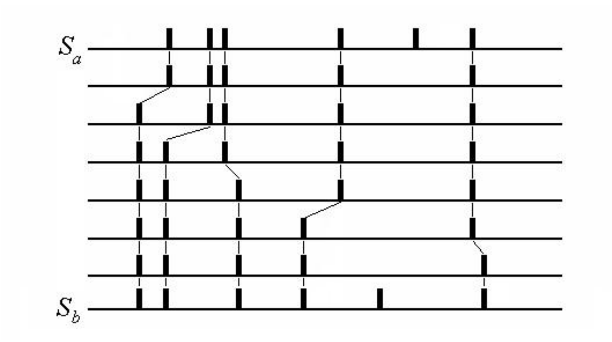


Fig. 5. Path of elementary steps to transform spike train S_a into spike train S_b.

It is shown in [8] that this procedure does indeed result in a mathematically well defined distance function.

3.2. *MDS pre-processing of spike trains*

Using standard Multi Layer Perceptron (MLP) networks to localize a target, the spike trains cannot be used for input directly. Doing so would make the dimension of the input space too large and thus prohibit convergence of the training of the MLP network. Hence, we introduce a pre-processing step to reduce the dimensionality of the input to the classifier. We chose to use a Multi Dimensional Scaling (MDS) approach to map the distances, calculated with the procedure described above, between a new spike train and a set of prototype (fixed) spike trains into a 3D metric space. The coordinates in this 3D space are then used as input to the MLP classifier.

The prototype spike trains are chosen in the following manner. Starting from the spatial sensitivity pattern of the sonar system i.e., its directivity, at the central frequency of each channel in the inner ear model, we determine the most sensitive and the least sensitive location (see Fig. 6). Next we choose 5 intermediate locations on the line connecting these two points. Together, the spike trains corresponding with these 7 points make up the set of prototype spike trains for this frequency channel. By selecting the prototypes in this manner we ensure that the different filter operations performed by the outer ear in the frequency band covered by this frequency channel are reasonably well represented by the prototypical spike trains. The same procedure is repeated for all frequency channels.
Once the prototypes for all frequency channels have been selected the MDS pre-processing step can be executed.

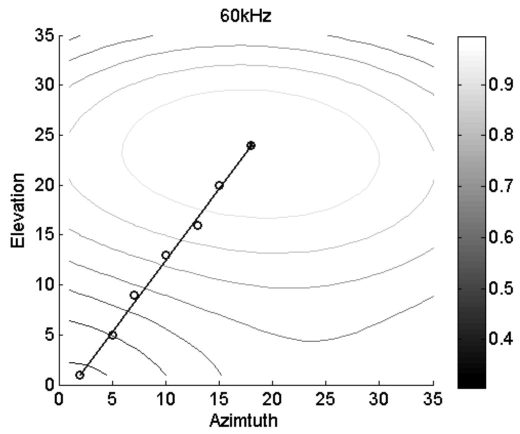


Fig. 6. Selection of prototype spike trains (blue circles) spanning the full range of the 60 kHz frequency channel's possible responses.

This results in a 3D representation for each spike train produced by every frequency channel in the inner ear model. Hence, every new received echo will be mapped onto a vector of dimension=N_{freq}x3, with N_{freq} the number of frequency channels in the inner ear model. This vector will be the input to the MLP classifier.

As stated before, real targets will not be single point reflectors, hence it is instructive to study to what extent the complexity of the reflector will affect the quality of the target localization. Indeed, as indicated by (1), in addition to the filter associated with the directivity, complex reflectors introduce additional filtering which will also affect the spike train.

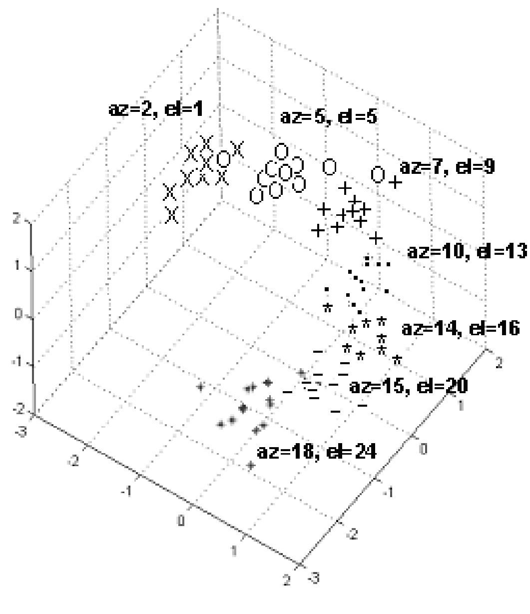


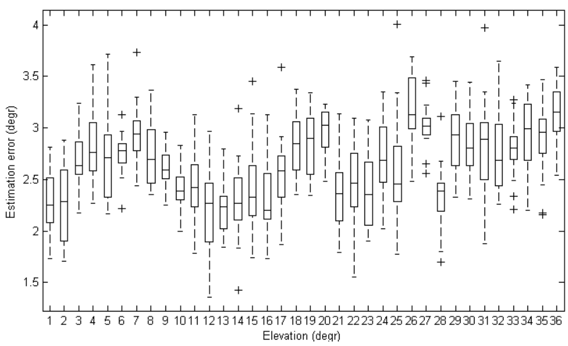
Fig. 7. Clusters formed in the metric 3D space of a typical frequency channel after MDS pre-processing of the spike trains generated by targets of varying complexity at fixed positions. The (az,el)-pairs indicate the position of the complex reflector.

Fig. 7 shows a result of a simulation study that varied the complexity of reflectors at various places in the field of view of the sonar system. Similar results are obtained when processing the spike trains generated by the other frequency channels. Hence, we conclude that although the complexity of the reflector will place a limit on the accuracy of target localization i.e., the size of the clusters, it will not make useful localization impossible.

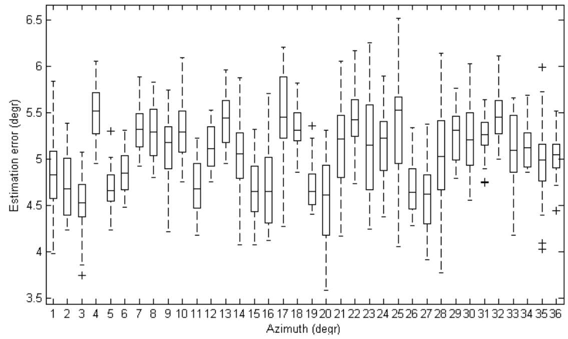
4. EXPERIMENTAL RESULTS

Simulation experiments were performed using monaural spectral information only. The HRTF was taken from a simulated Nyctalus Plancyi outer ear (Müller and Hallam, 2005). As described above, an MDS pre-processing step to lower the dimensionality of the classifier's input space was applied to the spike trains produced by the inner ear model. From the set of frequency channels calculated by the full inner ear model we mapped 20 frequency channels evenly distributed over the full frequency range (30-130kHz). A set of 20x3 training vectors was created by varying the complexity of the target as well as its position.

Two standard MLP networks with 1 hidden layer were trained following a batch method using a Levenberg-Marquardt minimization of mean-square error criteria. The output of the two networks was respectively the azimuth and the elevation estimate of the target. Separate networks were chosen to allow studying azimuth and elevation estimation separately. Optimization of the network architecture was performed for one channel and the result was used in the other channels. The fully connected network consists of 20 neurons with sigmoid activation in the hidden layer. The output neuron has linear activation.



(a)



(b)

Fig. 8. Target location estimation errors as a function of location: (a) box and whisker plot of elevation error, (b) box and whisker plot of azimuth error.

Behavioural experiments on bats indicate that monaural information results in fairly good elevation estimation whereas binaural information seems to be required to get good azimuth estimation. The results shown in Fig. 8 (a) and (b) are in good agreement with this empirical observation by indicating an average elevation error of ~2.5° which is considerably smaller than the average azimuth error of ~5°.

5. CONCLUSION

In this paper it is shown that target localization based on spike coded spectrograms is feasible. The information used by the sonar system is the self-induced i.e., by the outer ear, filtering that varies systematically with azimuth and elevation. It is also shown that the method described here can cope with moderately complex reflectors i.e., up to 7 closely spaced glints, despite the additional filtering such reflectors give rise to. To be able to work with spike trains as produced by a first order model of the bat's inner ear a dimension reducing MDS pre-processing step was proposed. This allows standard MLP networks to localize the target, both azimuth and elevation, with realistic and useful accuracy.

6. REFERENCES

Pickles J. O. (1982) *An Introduction to the physiology of hearing*. London: Academic Press.

Popper, A. and Fay, R., Eds. (1995) *Hearing by Bats*. Springer Verlag.

Müller R. and Hallam J. (2005) Knowledge mining for biomimetic smart antenna shapes, *Robotics and Autonomous Systems,* **50**, 131–145.

Reijniers, J. and Peremans H. (2004) *Towards a theory of how bats navigate through foliage,* in Proc. of the eighth Int. Conf. on Simulation of Adaptive Behavior, Los Angeles, pp. 77–86.

Saillant P. , Simmons, J. and Dear S. (1993) A computational model of echo processing and acoustic imaging in frequency-modulated echolocating bats: The spectrogram correlation and transformation receiver, *Journal of Acoustical Society of America*, ***94(*5)**, 2691–2712.

Victor J. D. and Purpura K. P. (1997) Metric-space analysis of spike trains: theory, algorithms and application, *Network: Comput. Neural Syst*. **8(2)**, 127-164.

Wotton J. , Haresign, T. and Simmons J. (1995) Spatially dependent acoustical cues generated by the external ear of the big brown bat, eptesicus fuscus, *Journal of Acoustical Society of America*, **98**, 1423–1445.

Wotton J., Haresign T., Ferragamo M. and Simmons J., (1996) Sound source elevation and external ear cues influence the discrimination of spectral notches by the big brown bat Eptesicus fuscus, J. Acoust. *Journal of Acoustical Society of America*, **100**, 1764–1776.