ESANN'2007 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2007, d-side publi., ISBN 2-930307-07-2.

Bat echolocation modelling using spike kernels with Support Vector Regression.

Fontaine Bertrand[1], Peremans Herbert[1] and Schrauwen Benjamin[2] *†

1-Antwerp University, APL, Prinsstraat 13, 2000 Antwerpen, Belgium

2-Ghent University, ELIS, St-Pietersnieuwstraat 41, 6000 Gent, Belgium

Abstract. From the echoes of their vocalisations bats extract information about the positions of reflectors. To gain an understanding of how target position is translated into neural features, we model the bat's peripheral auditory system up until the auditory nerve. This model assumes multiple threshold detecting neurons for each frequency channel where the inter-spike times are linked to the location of the reflector. To show that this coding process can be reversed we compute the kernel product of the spike trains using a non-binned spike kernel function. This approach allows doing regression on azimuth and elevation using Support Vector Machines.

1 Introduction

Bats use their biosonar to navigate through foliage and to localize and recognize prey. They emit ultrasound calls and from the modified received echoes extract information regarding the geometry of their environment. Experiments have shown that bats use spectral clues to determine the spatial position of the reflecting target [4]. 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 [5], corresponding with the Head Related Transfer Functions (HRTF) studied in human hearing research. Below, we will show that these spectral cues are preserved/enhanced by cochlear processing [2] and can be used to extract reflector position information from the neural code present at the level of the auditory nerve.

Although it is customary in audition research to assume that the information at the level of the auditory nerve is represented using rate coding, we propose that in the case of bat echolocation a time coding perspective is more appropriate. Indeed, due to the very short durations of both the calls and the echoes, 2-10 ms, only 1 or 2 spikes are generated in the Inferior Colliculus i.e., auditory midbrain, of the bat [7] when presented with a single call-echo pair. Furthermore, experiments have shown that single call-echo pairs are sufficient for the bat to base behavioral decisions upon. Hence, as firing rates loose their meaning when only single spikes are present we conjecture that the inter-spike times of the generated spike burst code for the relevant information. See [3] for a thorough discussion of time vs. rate coding issues.

*This research is funded by the EC, IST-FET Initiative as part of the CILIA project.

†We would like to thank C. Moss and A. Murat for making available the Eptesicus Fuscus HRTF measurements.

Because of the time coding, existing spike kernels [6] cannot be used to embed the processing of these spike trains in standard Support Vector Machine theory as they require the computation of firing rates. Hence, a new spike kernel, introduced recently in [9], that keeps the precise timing information intact was used to investigate to what extent classical regression on azimuth and elevation using SVM can be efficiently done on spike burst codes generated in the context of echolocation.

2 Peripheral auditory system processing of echoes

Call The emitted ultrasound is based on the call of a particular FM bat i.e., Eptesicus Fuscus. It consists of a downward 2ms FM sweep whose fundamental ranges from 50kHz to 20 kHz.

Reflecting objects The reflectors are simulated as a set of point reflectors. For each observation a new, random, reflecting object is generated. Hence, when transmitting the signal $s(t)$, the filtered echo signal $z(t)$ received from such a composite reflector (point reflectors at $(r_i, \theta_i, \varphi_i)$ with $i = 1 \dots N_{refl}$) can be written as

$$z(t) = \sum_{i=1}^{N_{\mathrm{refl}}} h_{\mathrm{HRTF}}\left(t; \theta_i, \varphi_i\right) * a_i s\left(t - \frac{2r_i}{v_{\mathrm{sound}}}\right) \qquad (1)$$

with the speed of sound given by v_{sound}, the distance by r, the azimuth by θ, the elevation by φ and the angular dependent filtering (HRTF) denoted by $h_{\mathrm{HRTF}}(.)$. The number of point reflectors (N_{refl}) is drawn uniformly between 1 and 5. Once this number is known, one of the point reflectors is set to have a reflection coefficient $a_k = 1$ and the ratio of the others reflection coefficients a_i with respect to the latter are uniformly distributed between 0.01 and 0.4, reflecting coefficient range we extracted from our measurements. The delay times ($\tau_i = 2r_i/v_{\mathrm{sound}}$) between point reflectors are generated from a Poisson process (simple model for a cloud reflector) whose mean depends on the number of reflectors. From Eq. 1 we conclude that even when disregarding the angular dependent filtering $h_{\mathrm{HRTF}}(.)$ whenever the target is no longer a single point reflector the echo signal $r(t)$ will be a filtered version of the transmit pulse $s(t)$.

Head Related Transfer Function The HRTF used was taken from measurements on an FM-bat, Eptesicus Fuscus [5]. The reflector positions considered span from -30 degrees to 30 degrees azimuth and the same for elevation. From this spatial domain we choose 150 non-evenly distributed locations.

The HRTF results in spatially dependent filtering of the received echo. The spectral features of the received echo signal e.g., peaks and notches, introduced by the HRTF can thus code for the position of the reflecting object. The frequency dependent nature of the HRTF (see Fig. 1) allows for the ambiguity i.e., multiple locations mapped onto the same echo strength, present in a single frequency channel to disappear if multiple frequency channels are combined.

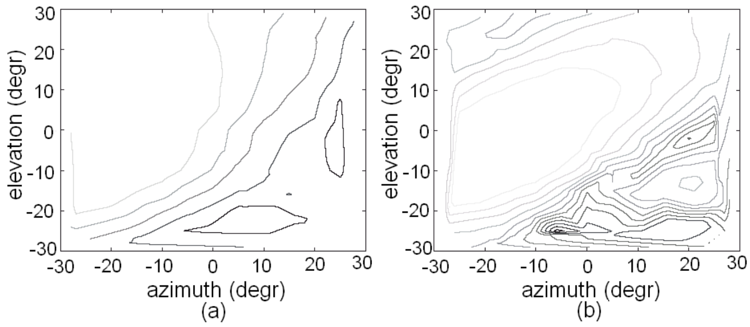

Fig. 1: 3dB contours of the monaural HRTF at (a) 40kHz and (b) 68kHz.

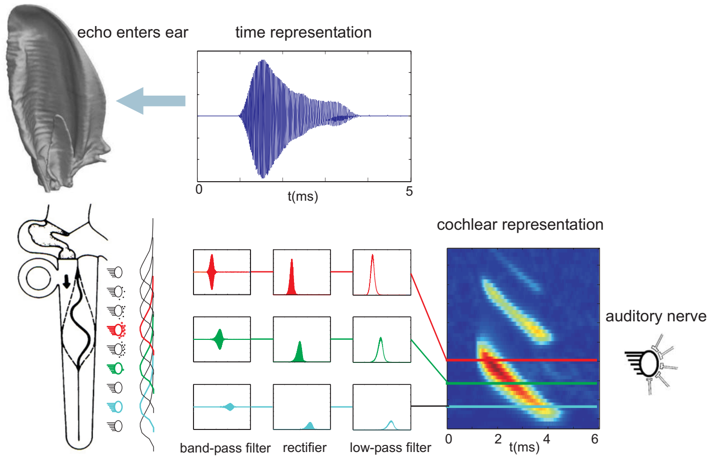

Fig. 2: The model of the peripheral auditory processing

Cochlea The joint time-frequency analysis performed on the incoming signal is modelled on the transduction stage located in the inner ear (cochlea) of the bat. A simple, yet functionally adequate, model of this analysis [2] is a filter bank consisting of parallel band-pass filters with subsequent envelope extraction (half way rectification and low pass filter) in each channel. This model is illustrated in Fig. 2. The band-pass filters cover a range from 20kHz to 80kHz and have a quality factor of 25. The overlapping factor is set such that 100 filters cover the frequency range of interest.

Spike Generation As illustrated in Fig. 2 Inner Hair Cells (IHC) are situated along the basilar membrane. The IHC convert local motion of the basilar membrane into amount of neurotransmitter released. Several spiral ganglion cells (SGC) synapse with the same IHC. The SGC fire an action potential if their membrane voltage due to the neurotransmitter released by the IHC exceeds a threshold. Different IHC-SGC synapses have different thresholds.

We model a simplified version of this spike generation process using 15 threshold neurons per frequency channel that fire deterministically when the signal exceeds their threshold. The thresholds are chosen logarithmically following psychoacoustic custom. Each threshold neuron can trigger only once, but there is not that much information lost as the time scale of the signal is the same as the refractory period of a neuron. Finally, the spikes coming from the 15 threshold neurons per frequency channel are embedded in one spike train.

3 Spike Kernel

Techniques using kernel functions for classification and regression problems have been proven to be very efficient. Here we want to use standard Support Vector Machines (SVM) to do regression on azimuth and elevation [1]. However, kernels are traditionally applied to vectors whereas for our purposes we need a kernel function that can be applied to a set of spikes. Kernels specifically developed for spike data have been described [6], but they require binning thereby losing the precise timing information of the spike trains. Moreover, as stated before, the occurrence of only 1 or 2 spikes in each neuron's output does not allow computing a firing rate either.

The new spike kernel introduced in [9] does allow us to preserve the temporal resolution needed for our application. In this paper, defining a spike train x as the set of threshold crossing times $\mathbf{x} = \{t_1, t_2, ...t_N\}$, the linear spike kernel operating on two spike trains $\mathbf{x}$ and $\mathbf{y}$ containing N and M spikes respectively is defined as:

$$K_{linear}(\mathbf{x}, \mathbf{y}) = \sum_{i}^{N} \sum_{j}^{M} \left[\max \left(1 - \frac{\lambda}{2} |t_i - t_j^{'}|, 0 \right) \right] \tag{2}$$

where t_i is the i-th spike of $\mathbf{x}$ and $t_j^{'}$ is the j-th spike of $\mathbf{y}$. The λ parameter determines the temporal width of the kernel. A large λ leads to a kernel comparing only close-by spikes whereas a small one results in a kernel comparing each spike with every other spike. This kernel comparison could possibly be executed in the Superior Colliculus (SC) which links sensory based spatial information with orienting responses [8]. SC receives inputs from multiple auditory centers among which the cortex.

In our echolocation application we use the different spike trains from the different frequency channels as input for our regression system. The kernel method has to be slightly extended to take such structured collections of spike train data into account. If we have P input spike trains $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_3 ... \mathbf{x}_P)$ from P frequency channels, a new multi spike train kernel $K^{'}$ can be built by taking the sum of the P individual kernels:

$$K_{linear}^{'}(\mathbf{X}, \mathbf{Y}) = \sum_{i}^{P} K(\mathbf{x}_i, \mathbf{y}_i) \tag{3}$$

Using multiplication instead of addition, a single kernel product which yields zero would cancel out the contributions from all other frequency channels.

4 Simulations and results

The data set is built as follows: for each position from the 150 considered, 10 observations are simulated with the reflecting object complexity being drawn randomly (see Sec. 2). It gives thus 1500 observations in total. Each observation consists of maximum 100 spike trains generated for the 100 frequency channels. A single spike train consists of maximum 15 spikes, less if not all the thresholds are crossed. Of the 1500 observations, 1000 are used for learning and the rest for testing. Within those 1000 observations, 200 are used to do a grid search on the parameters of the model using 5 cross-validations. The parameters to optimize are the cost C, the epsilon of the regression SVM [1] and the scale parameter λ of the kernel. Once the parameters are found 5 others cross-validations are done with the testing data. Azimuth and elevation are estimated independently.

First the system takes its input from one frequency channel situated in the middle of the frequency range (50kHz). Next, the number of frequency channels is increased by distributing channels evenly over the full frequency range while keeping channels used in a previous simulation in the new set (monotonous increase of information). The results of those simulation are plotted in Fig. 3.

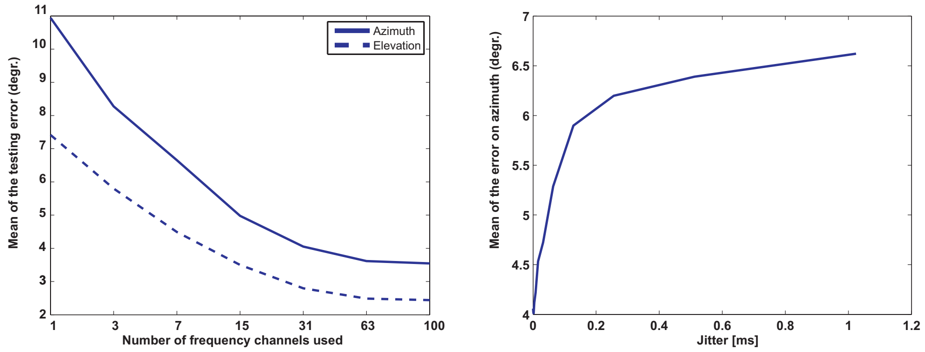

Fig. 3: (a) Resulting mean testing error (deg.) using a growing number of frequency channels over 5 validations; (b) Mean error (15 frequency channels) in elevation as a function of spike time jitter.

5 Discussion

As expected the mean error goes down as the number of frequency channels is increased (see Fig. 3(a)), due to a reduction of ambiguity when combining HRTF's of different frequencies (see Fig. 1). For the biological system this would mean more and more accurate position information is available moving up the monaural auditory pathway starting from the cochlear nucleus through the inferior colliculus ending in the cortex. This is a consequence of the systematic broadening of the frequency integration properties of the tonotopically organized neurons along this pathway.

We also notice that the elevation estimate is more accurate than the azimuth estimate for the same number of frequency channels. This is also expected from behavioral experiments on bats as azimuth estimation relies on binaural cues whereas elevation estimation can be performed using monaural spectral information [4, 5]. Furthermore, the actual accuracies achieved are also in accordance with measured accuracies for elevation in bat echolocation [4]. A previous model based on a backpropagation network [10] making use of the same monaural cues achieved performances of 7.5 for azimuth and 8.9 for elevation. Our model clearly achieves better performances and is less complex.

From Fig. 3(b) we conclude that the kernel indeed makes use of temporal information to estimate the position. On the other hand, the asymptotic behavior for larger jitter (>1ms) values shows that even in the absence of precise timing information the performance of the elevation estimation is better than chance behavior.

Future work should investigate and use more biologically plausible kernels. For instance, spike kernels should consider the absence of spike as informative whereas the present used kernel does not.

References

[1] B. Schokopf and A.J. Smola. *Learning with kernels: Support Vector Machines, Regularization, Optimization and Beyond*, MIT press, 2002.

[2] J. O. Pickles, *An Introduction to the physiology of hearing*, Academic Press, 1982.

[3] F. Rieke and D. Warland and R. de Ruyter van Steveninck and W. Bialek, *Spikes*, MIT Press, 1999.

[4] J. Wotton, T. Haresign, and J. Simmons, Spectral cues and perception of the vertical position of targets by the big brown bat, eptesicus fuscus, *J. Acoust. Soc. Am.*, 107(2), 1034-1041, 2000.

[5] A. Murat, E. Grassi, M. Sahota and C. Moss, The bat head-realted transfer function reaveals binaural cues for sound localization in azimuth and elevation, *J. Acoust. Soc. of Am.*, 116, 3594-3605, 2004.

[6] L. Shpigelman, Y. Singer, R. Paz and E. Vaadia, Spikernels; Predicting arm movements by embedding population spike rate patterns in inner-product spaces, *Neural Computation*, 17(3), 671-690, 2005.

[7] M. Sanderson and J. Simmons, Neural responses to overlapping FM sounds in the inferior colliculus of echolocating bats,*J. Neurophysiology*, 83, 1840-1855, 2000.

[8] D.E. Valentine and C.F. Moss, Spatially Selective Auditory Responses in the Superior Colliculus of the Echolocating Bat, *The Journal of Neuroscience*, 17(5), 1720-1733, 1997.

[9] B. Schrauwen and J. Van Campenhout, Linking Non-binned Spike Train Kernels to Several Existing Spike Train Metrics. *Neurocomputing*, doi:10.1016/j.neucom.2006.11.017, 2007.

[10] J.M. Wotton and R.L. Jenison, A backpropagation network model of the monaural localization information available in the bat echolocation system, *J. Acoust. Soc. Am.*, 101(5), 2964-2971, 1996.