OBJECTIVITY IN THE STUDY OF MARINE MAMMAL VOCALISATIONS: A WAVELET APPROACH

E. Delory[(1)] and J. R. Potter[(2)]
Acoustic Research Laboratory, EE Dept & TMSI, 10 Kent Ridge Crescent
National University of Singapore, 119260 Singapore
[(1)]eric@arl.nus.edu.sg
[(2)]johnp@arl.nus.edu.sg

INTRODUCTION

Sounds produced by marine mammals generally have a non-stationary behaviour, ranging from short transients to long tonals, with a huge variety of chirp-like intermediates. Analysis of these sounds needs adapted tools in order to characterise the signals efficiently in both time and frequency dimensions In numerous studies and for well-known reasons, the spectrogram has been used as the interface between sound and feature extraction.

The spectrogram popularity in our type of study stems from at least three reasons: firstly, a spectrogram is a two-dimensional picture displaying sounds along two physical dimensions (time and frequency) which are familiar to our auditory system. Little training is needed to recreate an approximate sound image out of a spectrogram and vice-versa. Secondly, real-time implementation can now be performed on off-the-shelf platforms thanks to software availability and the Fast-Fourier-Transform (FFT), which greatly reduces the computational load. Thirdly, as a spectrogram is the squared modulus of a Short-Time-Fourier-Transform (STFT), the underlying parameters, say the window size, the amount of overlapping and window shape are the only parameters affecting the result. The user interface is therefore very simple and little mathematical background is necessary to handle it. For example, Figs.1 & 2 show significant differences when parameters are changed. If more frequency resolution is needed, the window must be longer. As time resolution is lower when long windows are used, a higher overlap is needed in order to attenuate the blocking effect in time-domain. Choosing the window type is more subtle: a higher contrast is obtained with a simple rectangular window but spectral high energy peaks tend to spread over the whole bandwidth (side lobes). The use of smoother windows gives lower contrast but attenuates these side-lobe artefacts.

Nevertheless, in many applications, such as in behavioural and communication studies, the spectrogram does not provide enough freedom to the user. When the interest is to obtain a “brain picture” of the sound a marine mammal actually hears, it is indispensable to be able to implement and account for a maximum number of parameters, which describe its auditory perception. Hearing sensitivity and frequency discrimination capabilities are two additional parameters we decided to consider in order to create a new representation. Though hearing sensitivity is straightforwardly computed, frequency discrimination is not. When a system has more resolution in frequency, it has less in time, and vice-versa, this principle is known as the Heisenberg uncertainty principle. Here we introduce a method based on a wavelet approach, which respects this principle and still provides a spectrogram-like representation of the signal. Moreover, unlike STFT, the transform is orthogonal, so that extraction of coefficients (e.g. whistle time-frequency pattern) allows perfect reconstruction of the corresponding information in the temporal domain.

METHOD

To develop the idea introduced above, sounds from a *Tursiops truncatus* were used, as some information is available for both hearing sensitivity and frequency discrimination (Au, 1993). Fig. 3 & 4 show that this information is frequency dependant, implying that the processing method must be performed in the frequency domain. Meyer wavelet subband filtering is an efficient tool for frequency domain processing because frequency windows of arbitrary size can be chosen (Mallat, 1998). The processing is performed in the frequency domain: hence, given a signal s(t) (length: power of 2), and a Meyer wavelet transform operator W adapted to the data in Fig. 4, the procedure is quite simple:

1) The spectrum S(f) is calculated by Fast-Fourier-Transform (FFT)

2) W(S(f)) is computed to separate frequency windows. The result is a matrix of N vectors, each representing the spectral content of a given frequency bin. The use of Meyer wavelets insures that no redundancy occurs, a consequence of orthogonality.

3) spectral information from these frequency bins is inverse-FFT to reconstruct the temporal information.

4) a 2D representation is built out of each (temporal information-frequency range) pair.

RESULTS

Unfortunately, the signals used for the experiment were sampled at 48 kHz, limiting the spectral representation at 24 kHz, a bandwidth where no really interesting phenomenon occurs in the dolphin auditory system. However, a substantial change in the representation can be noticed in Fig 5., where two types of vocalisations are expanded. The signal includes whistles and echolocation clicks. The most striking change is a tremendous loss in the low frequency part, where hardly any information remains compared to the upper part. Another interesting phenomenon is the fact that clicks seem to take a "drop" shape, with very low time resolution in the low frequencies. Globally, the representation is far less resolved in time than the upper spectrogram. Although the new representation is far less attractive in terms of perceptual pattern extraction, it gives a more precise picture of what the animal experiences. Moreover, extraction of shapes to reconstruct the corresponding signals in the temporal domain is not only possible but also more adapted, which is of major interest in communication and behaviour studies.

CONCLUSION

A new method to represent marine mammal vocalisations with an insight in their auditory system parameters has been introduced. It is believed that this adapted decomposition will be useful for scientists who want to study intra-specific behaviour and communication. Certainly an efficient tool for an intuitive insight in sonar capabilities too, this wavelet approach will hopefully open new perspectives in the study of marine mammal vocalisations.

REFERENCES

Au, W. 1993. Characteristics of the receiving system for simple signals, Pp 32-58. In *The Sonar of Dolphins*, Springer-Verlag, New York.

Delgutte, B.1996, Physiological models for basic auditory percepts, Pp 157-220, In *Auditory Computation* (Eds. H. L. Hawkins, T. A. McMullen, A. N. Popper & R. R. Fay). Springer-Verlag, New York.

Mallat, S. 1998, *A Wavelet Tour of Signal Processing*, Academic Press, New York.

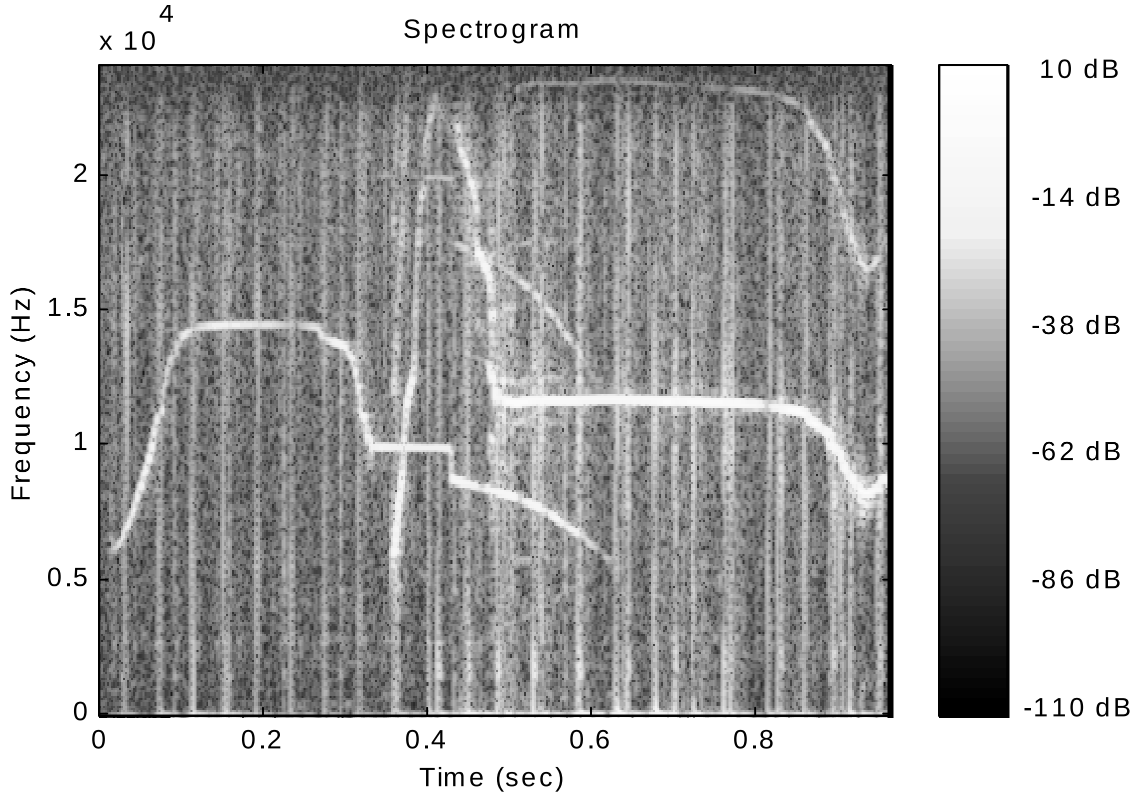

Fig.1 Spectrogram of *Tursiops truncatus* sounds (whistles and echolocation clicks), using a Kaiser window size of 512 samples, with 80% overlap.

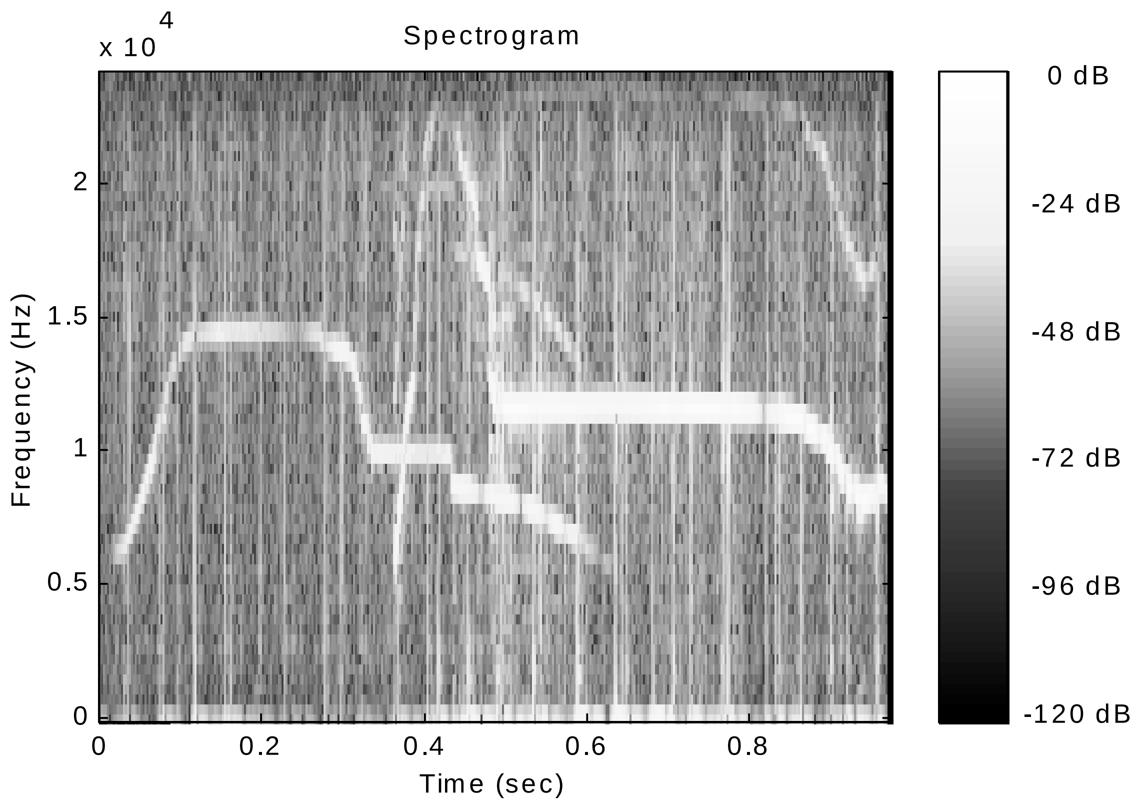

Fig.2 Same sound as above, using a Hamming window size of 128 samples, with 50 % overlap. The representation is less redundant, but the result is far less attractive.

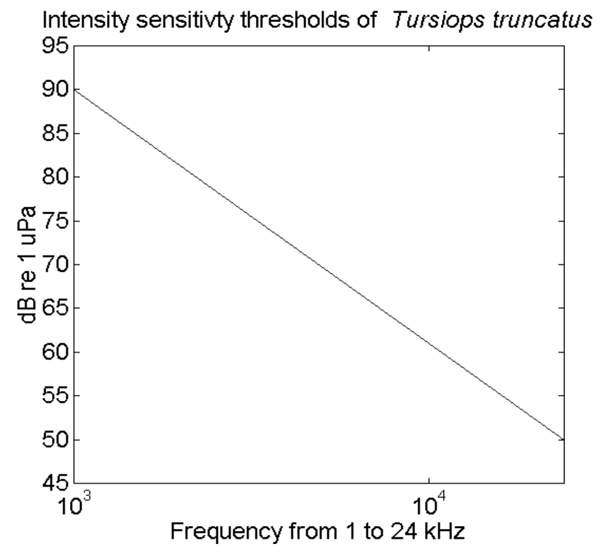

Fig. 3 Linear interpolation of hearing sensitivity for *Tursiops truncatus* between 1 and 24 kHz.

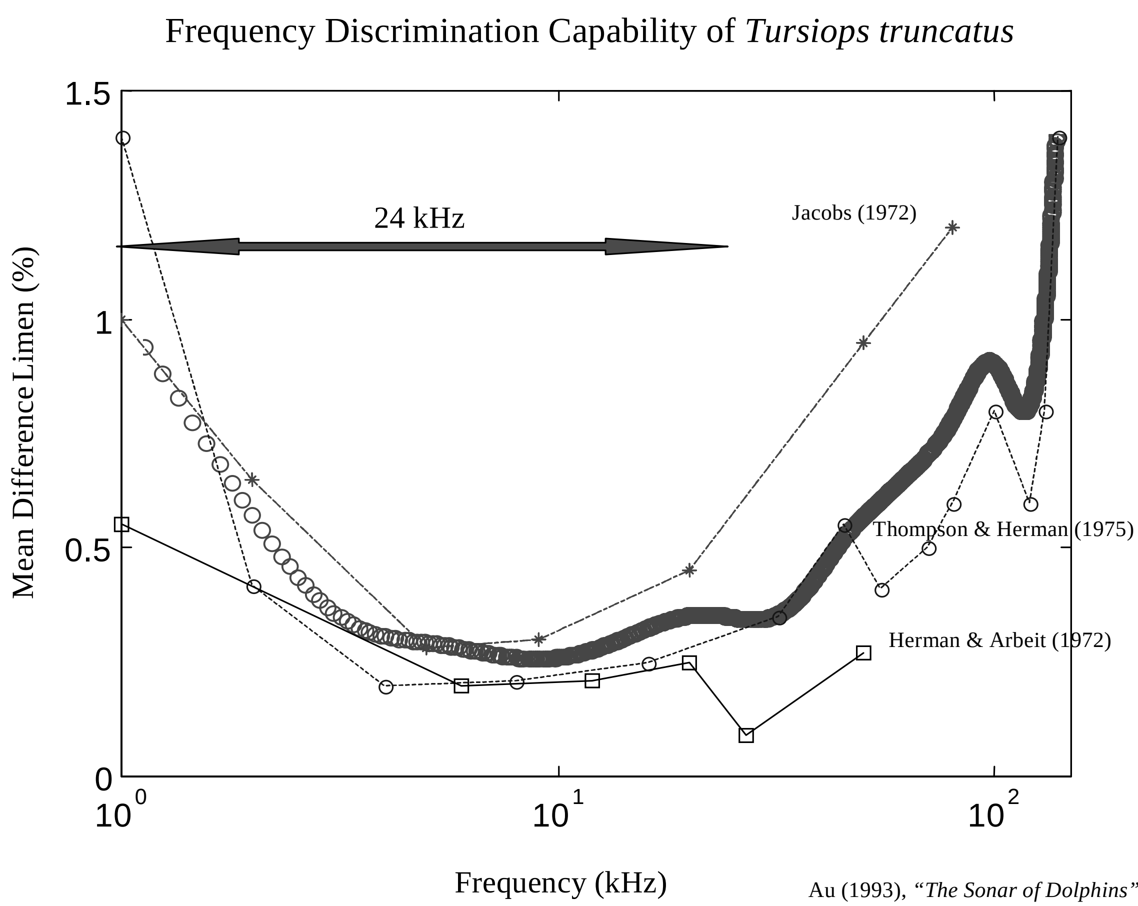

Fig. 4 Interpolation (circles) of frequency discrimination capability data for *Tursiops truncatus* . Difference limen is the ratio between frequency window size and central frequency. In the current study, only the 24 kHz bandwidth information was used.

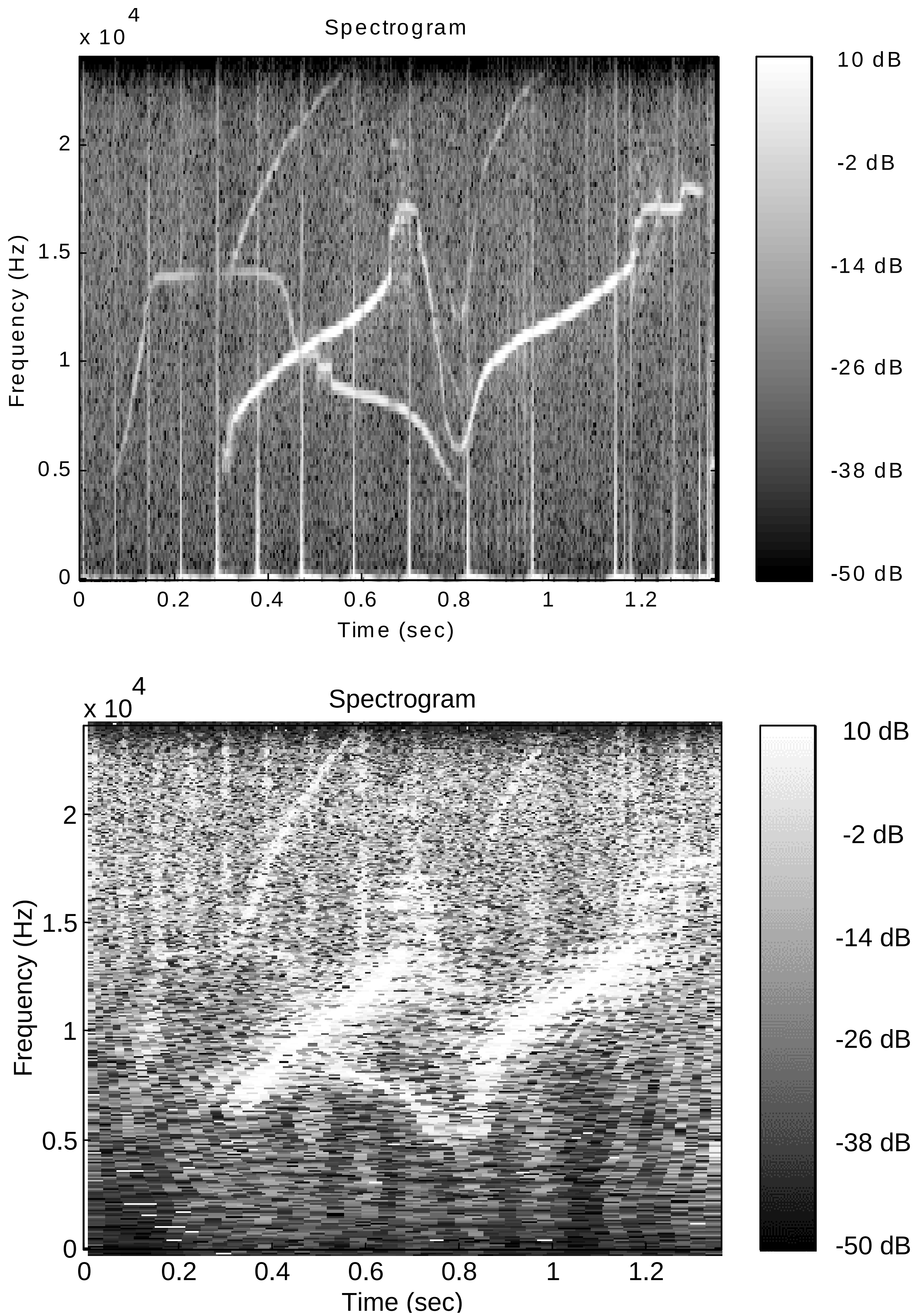

Fig 5. Top: Spectrogram of a set of whistles and echolocation clicks from *Tursiops truncatus*, using a 256 point hamming window and 50% overlap. Bottom, representation obtained considering the auditory system sensitivity and frequency discrimination