Humpback whale

Song analysis

UNIVERSITY OF Southampton
Institute of Sound and Vibration Research

Author: F Pace

▪ Introduction :

Humpback whales are a very widespread species; their characteristic songs have been intensively studied over the last few decades as a way to gain a further insight on the population dynamics.

In 1971, Payne defined the structure of humpback whale songs: they are formed by themes that are repeated in specific patterns and the basic building blocks are termed units – i.e. the shortest continuous sound between two silences (Fig. 1) [1].

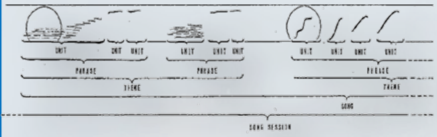

Figure 1: song structure drawn by Payne in his paper published in 1971.

Typically songs last for several hours and most of their energy is contained below 4 kHz, although sound harmonics extend up to 25 kHz [2].

▪ Aims :

- ➢ Develop a segmentation algorithm based on the energy content of the signal
- ➢ Characterise sound units using LPC, MFCC and cepstrum coefficients and evaluate their performance against a manual classification of the vocalisations
- ➢ Introduce novel approach based on the definition of subunits.

▪ Segmentation algorithm

The song analysis was carried out on a section of a recording carried out in August 2008 in the channel between the East coast of Madagascar and Ste Marie Island. The energy of the signal was calculated and a threshold of start and one of end were applied to identify the vocalisations.

The algorithm was efficient for segmenting the song, in particular when the signal was pre-filtered to improve the signal to noise ratio (Fig. 2).

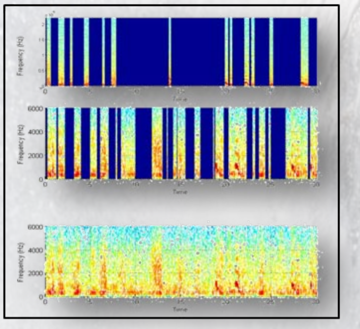

Figure 2: spectrogram of a 30 seconds segment of the original signal (bottom), and of the units identified using the algorithm on unfiltered (top) and filtered (middle) data.

▪ Processing methods

The sound units identified through the segmentation algorithm were characterised using three different models, which are commonly employed in speech processing. This approach is justified by the fact that humpback vocalisations present similar characteristics to voiced and unvoiced signals.

The models were:

- Linear prediction coefficients (LPCs)
- Cepstrum coefficients
- Mel-frequency cepstrum coefficients (MFCCs)

All of them have previously been used to describe humpbacks' calls in the published literature.

Linear prediction coefficients are used to represent speech signals based on the assumption that a speech sample can be approximated as a linear combination of past speech samples [3]. Hence:

Equation 1

$$S(n) = e(n) * \sigma(n)$$

On the other hand, cepstrum coefficients are based on the Fourier transform of the signal:

$$C(n) = FT^{-1}\{\log | FT\{s(n)\} |\}$$

Equation 2

The MFCCs are based on a homomorphic filter where the frequency bands approximate the logarithmic hearing of human listeners (Fig. 3) [3].

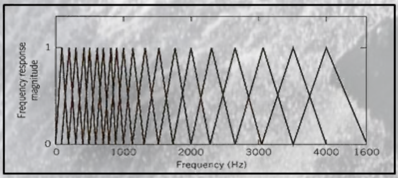

Figure 3: Mel-spectrum filter bank.

The coefficients obtained were then fed into the *k-means* clustering algorithm to classify the units. for instance for further analysis of the song structure.

▪ Coefficients performance

The results obtained were presented splitting the vocalisations into 5 main groups proposed by Dunlop and colleagues in their analysis of social vocalisations of humpback whales of the West coast of Australia [4].

The results showed that MFCCs were the best at characterising sound units in all cases except low frequency and amplitude modulated calls (Fig.4)

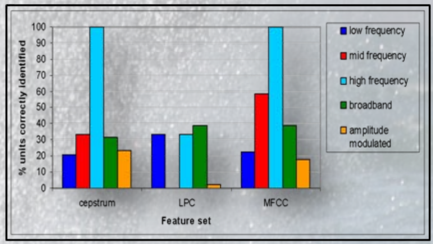

Figure 4: percentage of units correctly identified using the three models. The total number of units in this analysis was 149 and the model order 36.

▪ Discussion and perspectives

The fact that such an anthropogenic approach outperformed the other two models was not expected although research on the songs of other humpback populations showed that MFCCs are successful in this task. This can be due to the complex harmonic structure of many vocalisations and the energy of the signals being concentrated in a frequency range similar to the range of human hearing.

However, in general the performance of the classification algorithm was below 50% with independently of the feature set employed.

There are several reasons to explain this outcome:

- ❑ The sample size used was small due to the fact that the classification was compared against a manual classification which was very time consuming to be carried out. A larger data set should improve the results, especially if a train and test algorithm is employed.
- ❑ The recordings were quite noisy; a higher signal to noise ratio is necessary to improve the results and to be able to include a larger number of vocalisation in the analysis. In the study presented here, units with a low energy content were discarded by selecting a high threshold of start in the segmentation process in order to avoid noise to be detected.

Furthermore, the analysis showed that the characteristics of some sound units, particularly the long ones, may vary significantly over their duration, making it harder to capture their essential features for classification purposes.

For this reason, we propose a definition of subunits based on the changes in their frequency content through time (Figure 5). Typically units are formed are a combination of subunits that have a distinct structure; hence, subunits can be found on their own throughout a recording.

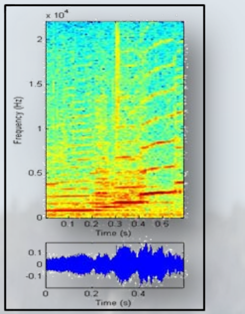

Figure 5: example of a unit composed of several subunits.

▪ References

1. Payne, R.S. and S. McVay, *Songs of Humpback Whales.* Science, 1971. **173**(3997): p. 585-597.

2. Au, W.W.L., et al., *Acoustic properties of humpback whale songs.* The Journal of the Acoustical Society of America, 2006. **120**(2): p. 1103-1110.

3. Deller, J.R., J.G. Proakis, and J.H.L. Hansen, *Discrete-time processing of speech signals*. 1993, New York: Macmillian Publishing Company.

4. Dunlop, R.A., et al., *The social vocalization repertoire of east Australian migrating humpback whales (Megaptera novaeangliae).* The Journal of the Acoustical Society of America, 2007. **122**(5): p. 2893-2905.