Applied Acoustics 71 (2010) 1107–1112

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Subunit definition and analysis for humpback whale call classification

Federica Pace [a,*], Frederic Benard [b], Herve Glotin [b], Olivier Adam [c], Paul White [a]

[a] Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton, Hants SO17 1BJ, UK
[b] CNRS UMR6168 Lab Sciences de l'Information et des Systèmes (LSIS), Univ. Sud Toulon Var R229-BP20132-83957, La Garde CEDEX, France
[c] University of Paris, Bioacoustics Team, NAMC-CNRS, UMR8620, University of Paris Sud, Bat. 446, 91405 Orsay, France

ARTICLE INFO

Article history:
Available online 12 June 2010

Keywords:
Humpback
Subunit
Song
Megaptera novaeangliae
Classification

ABSTRACT

Songs of humpback whales (*Megaptera novaeangliae*) have been studied for several years to gain a deeper insight on the intraspecific social interactions. Such a complex acoustic display is indeed thought to play an important role in both the mating ritual and male to male interaction. Hence, the need to classify the unit constituents of a song objectively and systematically has become crucial to allow processing large data sets. We propose a new approach for song segmentation based on the definition of subunits. Songs of humpback whales collected in Madagascar in August 2008 and 2009 were segmented using an energy detector with a double threshold and classified automatically with a clustering algorithm using MFCCs: the results, which were checked against a manual classification, showed that the use of subunit as the basic constituent of a song rather than the unit produces a more accurate classification of the calls. Such results were expected given that subunits are generally shorter in duration and less variable in terms of their frequency content and so their characteristics are more easily captured by an automatic classifier. Analysis of songs from other years and various areas of the World is necessary to corroborate the repeatability of the method proposed.

1. Introduction

Automatic classification methods for marine mammal vocalisations have been intensively studied in the last decade. The interest in this field of research has arisen from the need to objectively describe vocalisations and to be able to do it in a fast and effective way. However, the vocal repertoire of marine mammals is extremely variable from species to species and in some cetaceans, such as bottlenose dolphins and humpback whales, intraspecific calls are very complex as they might convey information about the signaller [1–8]. Moreover, the signals recorded are often buried in noise. For these reasons, the task of automatic classification can be very challenging.

In this paper, we propose a new approach for the analysis of humpback whale calls from the North East coast of Madagascar; to date little research has been carried out studying the animals in this region.

It is well known that during their winter migration to low latitudes for breeding purposes, male humpback whales engage in the production of complex songs. These were defined by Payne and McVay in 1971 [9] as an association of themes which are repeated in a specific sequence. The basic building blocks of a song were named units and were defined as the continuous sounds between two silences. Based upon this definition, the calls have been characterised using a variety of techniques including: linear prediction coding (LPC) coefficients [10], energy content in specific time windows [11], spectrographic analysis [8,12], Mel Frequency Cepstrum Coefficients (MFCCs) [13] and Cepstrum coefficients [10,14]. Classification of units has been attempted using *k*-means clusters [15], self-organising maps (SOM) [10,12], Hidden Markov Models (HMM), entropy estimation [12] and other neural network approaches.

The application of methods developed for human speech analysis to humpback whale vocalisations is widespread. The suitability of these tools stems from the acoustic similarities between human speech and humpback whale songs: they occupy a similar frequency band, both exhibit tonal (voiced) and broadband (unvoiced) elements. Like speech, humpback songs are composed of vocalisations of various durations which are punctuated by silences, i.e. units, as defined by Payne. In this sense the structure of a unit is comparable to that of a word in human speech.

One should take considerable care not to infer too much from these acoustic similarities: they do not imply that the songs of the humpback whale form a language; they are merely structural parallels which mean that the speech analysis tools are natural candidates for the analysis of humpback whale song. In addition the area of speech processing has been one of the most actively studied and consequently the methods applied there are amongst the most advanced.

* Corresponding author.
E-mail addresses: fp@isvr.soton.ac.uk (F. Pace), prw@isvr.soton.ac.uk (P. White).

0003-682X/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apacoust.2010.05.016

The great variety of methods used by researchers to analyse humpback whale vocalisations reflects the great diversity of these sounds. The goal of a preliminary analysis was to determine which of the most commonly used methods could characterise the majority of calls more accurately for classification purposes.

Analysis is usually conducted on sound units according to Payne's definition; however, throughout the duration of a unit it is possible to observe significant variations in the signal's characteristics. Therefore, we propose a new building block, referred to as a subunit, which forms the constituent part of a unit. Extending the structural analogy with human speech, the subunit has a role which is the counterpart of a phoneme, in the sense that in speech phonemes are combined to create words. Like phonemes, subunits can occur with different durations. Automated speech recognition systems are hierarchical, in that they identify phonemes, not words, since there are fewer phonemes which represent uniquely the movements and positions of the vocal apparatus during sound articulation. The structure proposed for humpback whale recognition follows this same pattern; classifying subunits should simplify the classification aspect of the task, albeit that potentially complicates the segmentation process.

2. Materials and methods

2.1. Data collection

The data were collected in the Sainte Marie Island Channel which is located between the Island of Sainte Marie and the North East Coast of Madagascar. The Ste Marie Channel was surveyed during August 2008 and 2009 between the coral reef in the South of the Island and the Northern part up to the submarine canyon in front of Coco Bay, i.e. the closest point between Madagascar and Ste Marie Island. The water depth throughout the channel varies between 30 and 40 m, with exception of a canyon, where water reaches a depth of 60 m.

A total of 18 days were spent at sea and 21 h of recordings were collected and stored. The recordings were taken from a 4 m long boat using a COLMAR Italia GP0280 hydrophone (omni-directional, [5 Hz, 90 kKz], sensitivity −170 dB re 1 V/μPa) connected to its amplifier and a HD-P2 TASCAM recorder. The sampling frequency chosen was 44.1 kHz as the harmonics of the vocalisations of humpback whales have been observed up to 20 kHz.

Songs were recorded in variable sea sates and weather conditions; however, only one high quality song with a good signal to noise ratio recorded on the 12th of August 2009 at 8:50 am for 1 h was selected for the analysis described in this paper. This period was selected as the recording vessel was close to the singer. The boat was estimated to be ca 100 m from the singer although the depth of the singer and its relative position to the hydrophone were not measured. Also during the recording the geometry between vessel and the whale varied as a result of wind and tidal currents. Other singers were audible in the recording; nevertheless, the level of their calls was insignificant compared to the level of the calls emitted by the focal animal.

2.2. Data analysis

The song was initially segmented using an energy detector with a double threshold, i.e. threshold of start (TS) of the vocalisation and threshold of end (TE) of the vocalisation to detect the sound units present within the song (Fig. 1).

The manually selected value for TS was quite high, to reflect the high Signal to Noise Ratio (SNR) of the recording and to ensure that the calls detected by the algorithm were those emitted by the focal singer in the proximity of the boat; whereas, the value of TE was lower to allow the algorithm to capture the majority of the energy in the vocalisations.

The units obtained were then checked manually by the main author; this resulted in a total of 424 vocalisations. Where appropriate the units were then subdivided into more basic components, i.e. subunits. Subunits were identified based on visual inspection of the spectrograms coupled with listening to three recordings with high SNRs from different years. Only subunits observed in the 2009 recording are presented in this paper.

Subunits were classified as such when they occurred on their own in between two silences (in which case the concepts of a subunit and a unit coincide) and/or in association with other sound subunits with no silence in between them; in the latter case, the differentiation between subunits was marked by a change in the acoustic properties of the call, such as the fundamental frequency, the envelope or the minimum and maximum frequencies. For in-

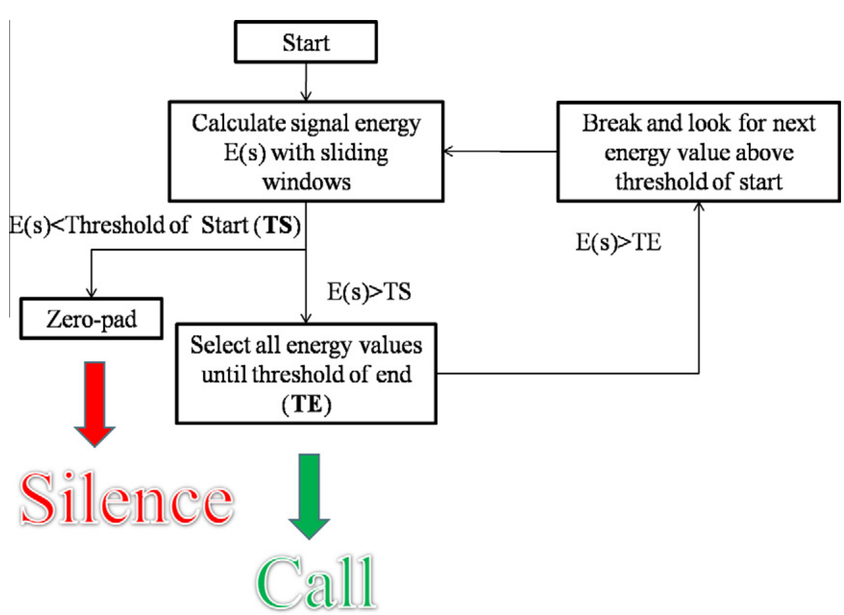

Fig. 1. Diagram showing the operation of the energy detector.