

Unsupervised Segmentation and Classification of Orca Vocalizations using the Fundamental Frequency Variation Spectrum

Florian Hönig¹, Reyhan Sonmez², Elmar Nöth^{1,3}, Martha Manser²

¹Pattern Recognition Lab, Friedrich-Alexander University Erlangen-Nuremberg, Germany

²Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Switzerland

³Electrical & Computer Engineering Department, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia.

hoenig@informatik.uni-erlangen.de

Orca acoustic dialects have been studied and recorded intensively for the past 20 years. The enormous amount of audio data available for this species calls for efficient and intelligent methods for automatic segmentation and classification into meaningful categories. We suggest applying automatic speech recognition (ASR) algorithms to this problem. Key components of ASR systems are Hidden Markov Models (HMMs) and an acoustic frontend. The HMMs model the temporal evolution of probability densities for observations created by the acoustic frontend. The latter captures the coarse shape of a short-time spectrum and its temporal dynamics. In order to successfully apply this technology to orca vocalizations, a specialized acoustic frontend needs to be developed. Fundamental frequency plays a crucial role in mammal vocalizations. In orca calls, pitch and its temporal trajectory seem to be one of the main information carriers, resulting in complex and diverse pitch curves. However, automatic pitch tracking is principally error-prone, especially in low signal-to-noise conditions. We therefore propose the application of the Fundamental Frequency Variation (FFV) Spectrum, which is an implicit representation of the derivative of pitch along time for a given short-time patch of audio. It has the advantage of being robust to noise, and does not suffer from the (octave) ambiguity in harmonic pitch detection. Thus, it seems an ideal candidate for the acoustic frontend needed for ASR techniques. The present study investigates the feasibility of this approach by using the FFV, a clustering technique and a simple single-state HMM to carry out an unsupervised segmentation task of continuous orca recordings. Resulting classes correspond well to human segmentations of the same data, with a precision approaching inter-annotator agreement. This automatic detection and classification system can easily be applied to other species and may enhance objectivity and comparability in the study of animal communication.