

Automatic speech recognition and speaker identification of animal vocalizations

P.J. Clemins and M.T. Johnson

Electrical and Computer Engineering, Marquette University, Milwaukee, U.S.A.

There have been many studies on the relationship between animal behavior and auditory communication. Many of these show that there is a strong correlation between the animal's vocalizations and the action(s) it is performing at the time. Others have shown that this correlation varies, depending on the individual. For instance, tamarins have been known to vary their calls depending on what foods they prefer, and elephants have different vocalizations for greeting other elephants and for showing aggression. While research has shown correlations between vocalizations and behavior, there has been little effort to build automatic classifiers to categorise vocalizations or determine which animal made them.

There has also been much research on human speech processing. Two tasks receiving great attention are speech recognition and speaker recognition. Well-founded techniques, such as Hidden Markov Models (HMMs), can translate spoken utterances to written language (speech recognition), and others, such as Gaussian Mixture Models (GMMs), can identify the speaker of a given utterance (speaker identification). These two tasks correspond directly to what many researchers in the animal behavior field are attempting to do with animal vocalizations. The present research aims to adapt these 'human' techniques for use with animals, to perform speaker identification and speech recognition.

This project aims to create a framework in which animal vocalization classifiers can be built to perform the tasks of speaker identification and speech recognition. Elephants, tamarins and aquatic mammals will be the first species explored. The initial step is to identify the features of vocal utterances carrying the most information. In human speech, spectral characteristics have proven to be the most effective, but animal behavior research shows that temporal characteristics are also important. Numerous statistical measures, including autocorrelation and mutual information, can indicate the importance of specific features.

The second part of the framework will involve identifying the model for the classifier. Currently, two of the most popular models in human speech research are HMMs for speech recognition and GMMs for speaker identification. However, it is expected that simpler models, such as Dynamic Time Warping (DTW), will also prove effective for animal vocalizations.

The development of a framework from which to build these animal vocalization classifiers will have a profound impact on animal behavior research. By using an automated algorithm to understand what animals are trying to say, software or devices could be made to provide a human language transcription of animal vocalizations. The ability to identify individual animals could even lead to the development of an animal tracking system, whereby animals could be tracked in the wild using microphones, instead of implanted devices.