Sound Identification of Loons

Sajeev C. Krishnan (Advisor: Prof. Bruce Land)

Electrical and Computer Engineering ,Cornell University, Ithaca, NY 14850

Introduction

Loons are a group of aquatic birds found in many parts of North America and northern Eurasia (Europe, Asia and debatably Africa). The loons are the size of a large duck or small goose, which they somewhat resemble in shape when swimming. The focus of this project is sound identification of loons through their calls or songs. Most sound identification schemes that exist today are constrained by the fact that they either require individuals to have similar call-types or do not take into account the change in vocal characteristics over time. The objective is to determine a scheme of sound identification with feature extraction that is stable over time and location, which would greatly increase the ease with which these birds are studied.

Figure 1. The Pacific Loon and the Yellow-billed Loon

Issues

- **Call independent identification**: The loon yodel changes with a change in location, hence a scheme of feature extraction has to be devised that is not dependent on the yodel itself.
- **Temporal stability:** Although the variability in loon yodels due to changes over time are not as great as the variability caused due to a change in location, this remains a factor in identification and must be appropriately handled.
- **Background Noise:** The scheme to be used for feature extraction, MFCC - is not very robust in the presence of additive noise. Hence, the proposed solution should contain some scheme that would sufficiently mitigate the effect of background noise without a significant loss in the quality of the recording.

Approach

- **MFCCs**: Physiological studies have shown that human auditory system does not follow a linear scale. Thus for each tone with an actual frequency, fHz, a subjective pitch is mapped on a scale called the mel scale. The mel-frequency scale is a linear frequency spacing below 1000 Hz and a log spacing above 1000 Hz. The main advantage of using mel frequency scaling is that it approximates the frequency response of human auditory systems and can be used to capture the phonetically important characteristics of speech.

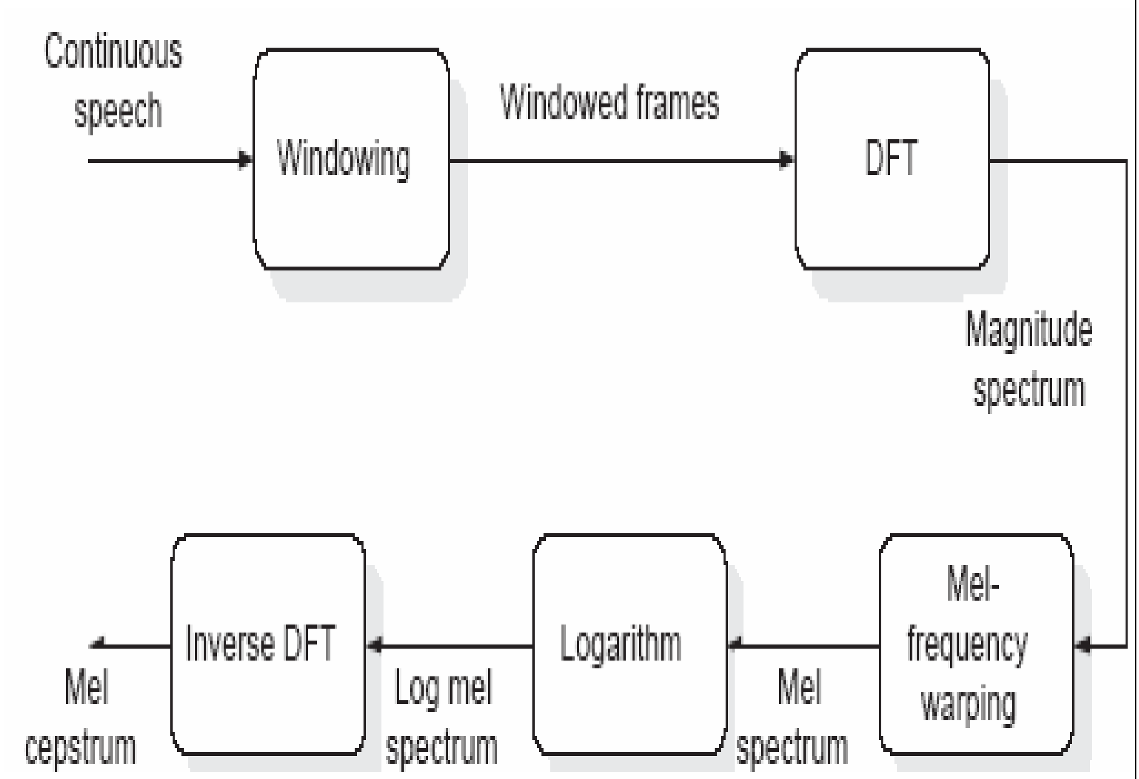

Figure 2. Illustration of the extraction of MFCCs.

- **Linde-Buzo-Gray (Vector Quantization):** VQ is a process of mapping vectors of a large vector space to a finite number of regions in that space. Each region is called a cluster and is represented by its center (called a centroid). A collection of all the centroids make up a codebook. The amount of data is significantly less, since the number of centroids is at least ten times smaller than the number of vectors in the original sample. This will reduce the amount of computations needed when comparing in later stages. In this project the MFCCs derived from the songs of reference subjects and test subjects are vector quantized to obtain satisfactory codebooks. These codebooks are then compared to determine identity of the test subject.

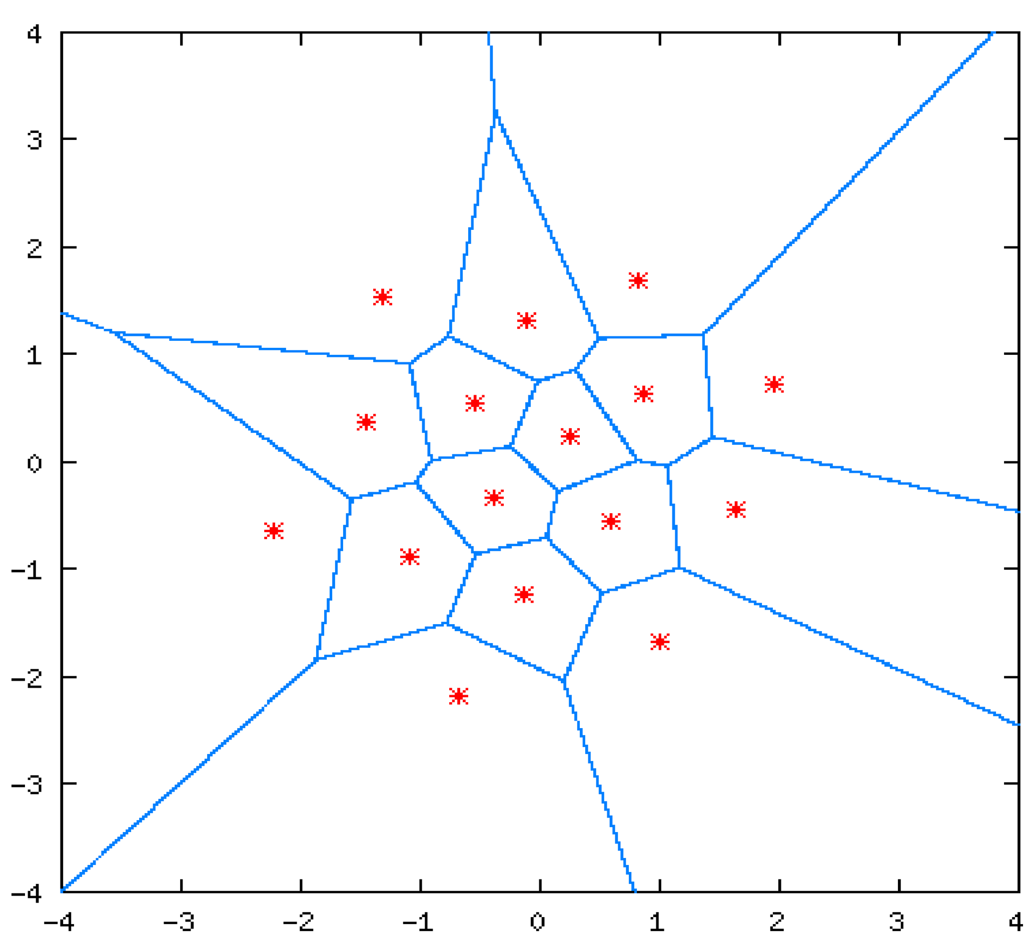

Figure 3. Obtaining a Codebook using the LBG Algorithm

- **Noise/Silence Removal:** High frequency (>2Khz) and Low frequency (<200Hz) is removed using a high pass and a low pass filter respectively. The song itself, which lies between 300Hz – 1.5Khz, was isolated using a band pass filter. The term "silence" here refers to sections of the recording in which the loon remains silent, but there is still some noise, possibly in the same frequency range as the loon call. We notice that the histogram is considerably wider in the presence of noise, i.e. has a greater standard deviation than sections of the recording that have just the song. The song is divided into 50ms windows and based on a predefined threshold for standard deviation, a window is either maintained or deleted.

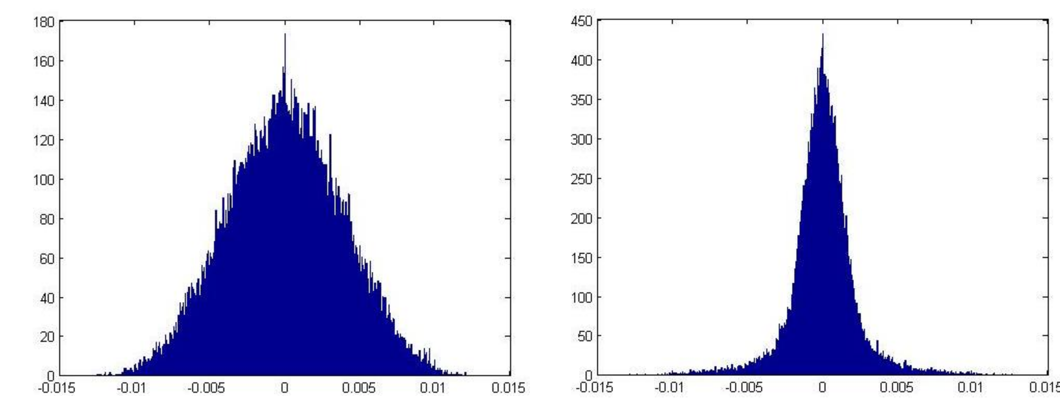

Figure 4. Histograms of windows with and without noise.

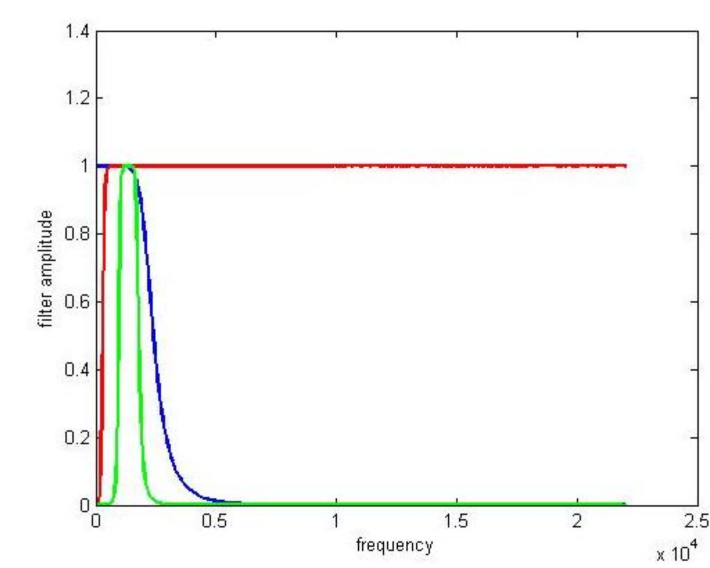

Figure 5. Filters

Results and Conclusions

•Noise Removal

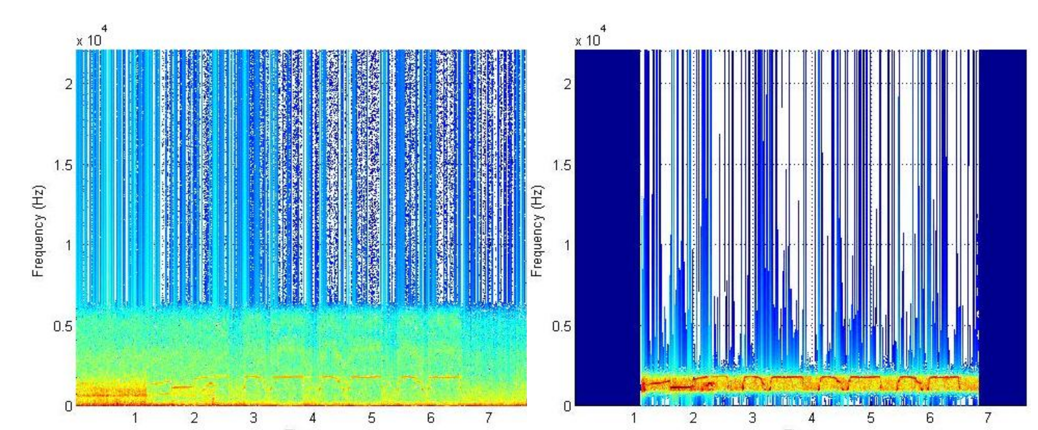

Figure 5. Power Spectra before and after noise removal

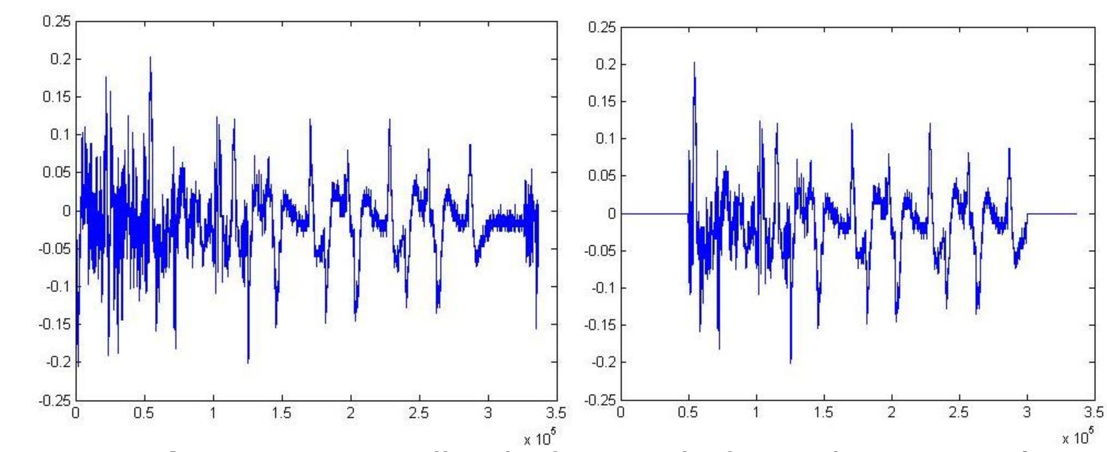

Figure 5. Recording before and after noise removal

•Sound identification

While comparing one loon to a training set of loons from 2 lakes the program was 100% accurate, i.e. matched every loon correctly. When trained against a set of 5 lakes with loon recordings obtained over several years, the program accuracy drops to about 88%. This can be attributed to several factors such as

- Poor recording quality: MFCCs are not too robust in the presence of additive noise, so it can throw off the centroids in the codebook.
- Over-training: Using too many samples for training that results in a highly clustered codebook which may lead to inaccuracies in prediction.

Overall the program succeeds at matching the loons correctly, but depends on the choosing the right training set. Some improvements could include using independent component analysis to considerably reduce the influence of noise, One could also use a different scheme for identification, ex: Neural Net/HMM

References

Walcott C, Mager J, Piper W. (2006) Changing territories, changing tunes: Male loons, *Gavia immer*, change their vocalizations when they change territories. *Anim Behav* **71**:673–83

T. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantization. *IEEE Trans. Communications,* 28(1):84–95, 1980

Fox, Elizabeth J. *Call-independent Identification in Birds*. Thesis. University of Western Australia, 2008

Acknowledgments

I would like to thank Prof . Bruce Land and Prof. Charles Walcott for giving me the opportunity to work on this project and providing me help and support whenever I needed it. I would also like to thank Ms. Laura Hou for her patience and support, without whom the project would not have been successful.