A New Method for DCLDE

A High-precision, Image-matching System

Neil J Boucher[1], Michihiro Jinnai[2], Hollis Taylor[3]

[1]SoundID, Maleny, Australia, [2]Nagoya Women's University, Japan,
[3]University of Technology, Sydney, Australia

Introduction

A system of sound recognition using 1-d (one-dimensional) spectrums has been demonstated by us to perform at expert level, in both speed of execution and accuracy for most most sounds. More difficult sounds (whales are prime examples) proved problematic for 1-d analysis mainly because of the high-noise level and the fact that much of the noise is in the same band as the call. Here we describe our new 2-d (2-dimensional) technique which will be ideal for whale studies.

In essence, our technique is not new; we simply match spectrogram images of sounds. What is new is that we use high-precision spectrograms and a new, fast method to automatically compare those images. We show that this technique can produce sound recognition accuracy that can match or exceed that of a human expert and do so at the rate of 100,000 matches per second.

Objectives

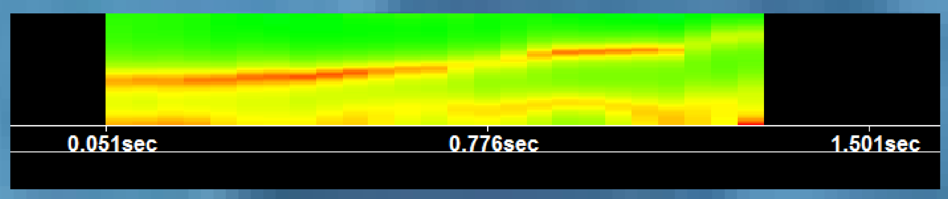
Our objective is to produce a sound recognition system that can match a human expert for any sound recognition excercise, including whales. Attempts to do so until now have been hampered by poor performance in recognition accuracy, false positives and slow execution speed.

Our software will run on any 64-bit Windows machine and is coded in .NET.

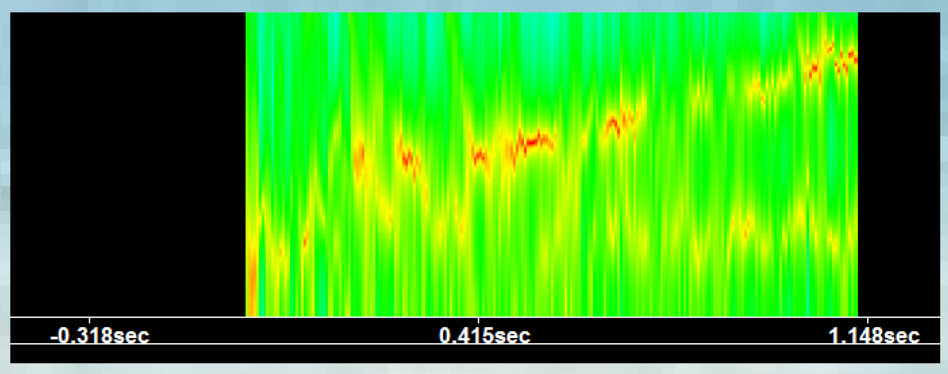
Materials & Methods

Our first step was to abandon the FFT for the more precise LPC spectrogram. When run at low resolution the LPC gives images that are similar to those of the FFT.

Below we show the LPC image of the upsweep from the conference data set at low resolution (upsweep file 328 at 890 seconds, 501 points frame width and order 15).



Now we step the resolution of the same call up by using a frame width of 51 points. So that the detail can be seen better we have also zoomed in on it.

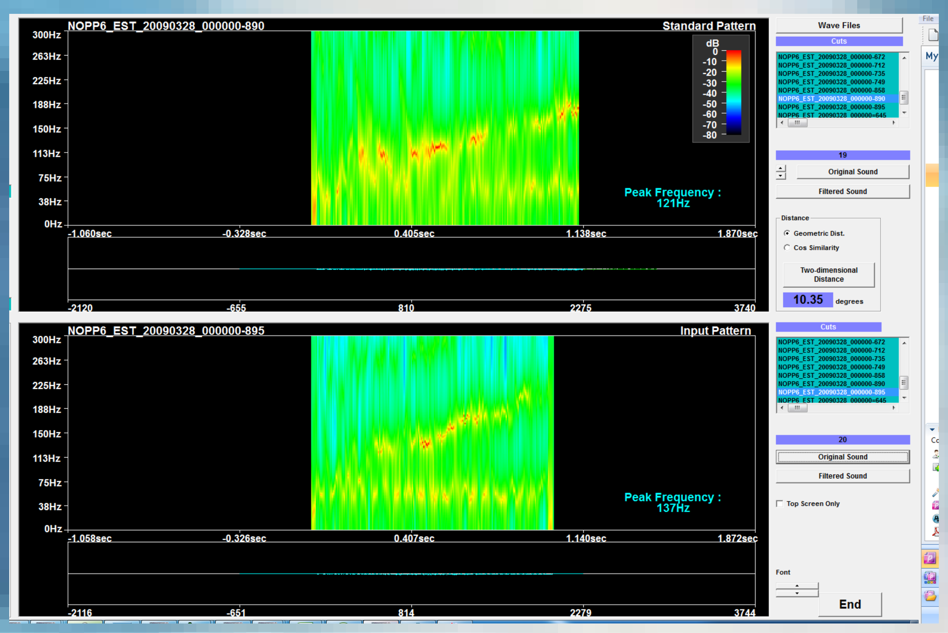


One important point to notice is that in the low-resolution image there is a lot of low-frequency noise. A portion of this is due to the DC offset of the recording but the bulk of it is the various components of the 1/f background noise.

Now look at the high-resolution image and most of the noise has gone. This has been done without filtering and is a natural consequence of the small frame sizes.

Notice also that the call is no longer a smooth upsweep, but consists of a series of bursts that have an overall trend of rising frequency.

The next objective is to compare such images. At the top left of each spectrogram we see the file name from the conference data set and notice that these calls were consecutive and about 5 seconds apart. In the middle right of the image we see the match is 10.35 degrees. Any two images that are more than 6 degrees apart are generally different sounds. The main difference here is that the lower spectrogram shows a faster upsweep and it has a significant set of relatively constant frequency pulses at around 65 Hz.



Results

Running a cluster analysis on 20 upsweep calls set for 6 degrees of difference, we find that in 20 calls there are 9 distinct calls of which there is one group each of 3, 4 and 5 members. For 92 gunshot calls (data set 20080125) 22 clusters were found with the largest set having 20 members. This would tell us that for good recognition of that set we would need to have an example of at least one member of each set.

Conclusions

This new 2-d method of sound recognition will accurately identify difficult sounds with the precison of an expert. The recognition module is fully operational and the detection module is under development.

A full, high-speed (real-time or faster) accurate sound recogniser will be available later this year. Like the 1-d software the 2-d will be designed for both real-time processing and for terabytes of pre-recorded data. The first test for it will be the DCLDE data set.