

Discussion: Mel Frequency Cepstral Coefficients for Music Modeling

**Abram Hindle**

**Software Engineering Group**

**Department of Computer Science**

**University of Victoria**

**abez@uvic.ca**

**February 6, 2005**

# MFCC for Music Modeling

- Mel Frequency Cepstral Coefficients (MFCCs)
  - 20ms frames
  - Hamming window
  - DFT
  - Take logs of amplitude
  - Mel Scaling and smoothing
  - DCT (KL Transform)

## MFCC for Music Modeling

- What did the paper attempt to test?
  - The paper attempted how appropriate MFCC was for discriminating music and speech.
  - How appropriate is the MFCC for music, it works on speech but does it work better than alternatives such as linear scaling?
  - \* Faired better than the Linear Scaling at Music
  - Is the DCT transform still appropriate for music?
    - \* They claimed it was still appropriate

# MFCC for Music Modeling

- Possible Issues with the experiment
  - Mel Vs Linear
    - \* The did not show us what linear buckets they used
    - \* Sampling Rate very low
    - \* Why use 25.6ms frames when 20ms frames were typically used in speech?
    - \* Sample data - a broadcast news show - is this appropriate?
    - \* Why not use other classifiers for music/speech segmentation as well
  - Is the DCT transform still appropriate for music?
    - \* The comparison of the two transform is poor
    - \* They claim the eigenvectors look “coslike” but do they test the difference between using a KL transform and DCT music?

## Summary

- MFCC seemed to fair better than linear but it was not compared against other technique currently in use.
- Authors claim DCT is appropriate but it seems to lack actual testing.
- The MFCC seems appropriate for the the discrimination of speech and audio.
- Mel-Frequency Cepstral Coefficients for Music Modeling” - Beth Logan, Proc. Int. Conf. on Music Information Retrieval (ISMIR), Plymouth, Massachusetts, 2000