Audio Similarity

Mark Zadel

MUMT 611

March 8, 2004

Overview

- MFCCs
- Foote – Content-Based Retrieval of Music and Audio (1997)
- Logan, Salomon – A Music Similarity Function Based On Signal Analysis (2001)

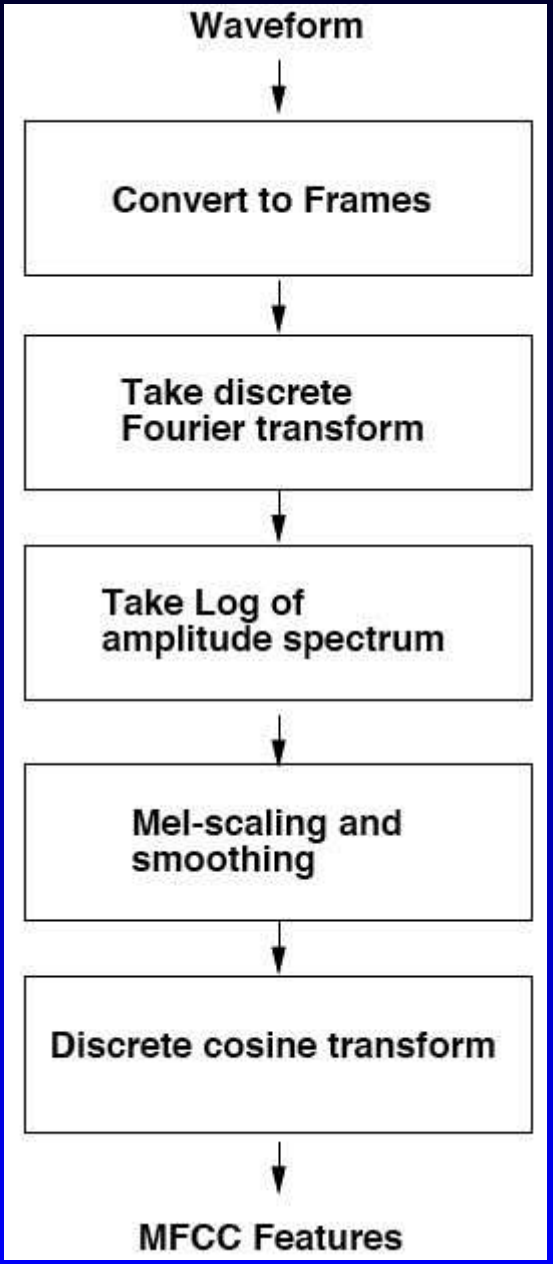
Motivation

- MP3s, music on the Internet
  - Large collections of songs
  - How to search?
- Digital music libraries
- Commercial applications

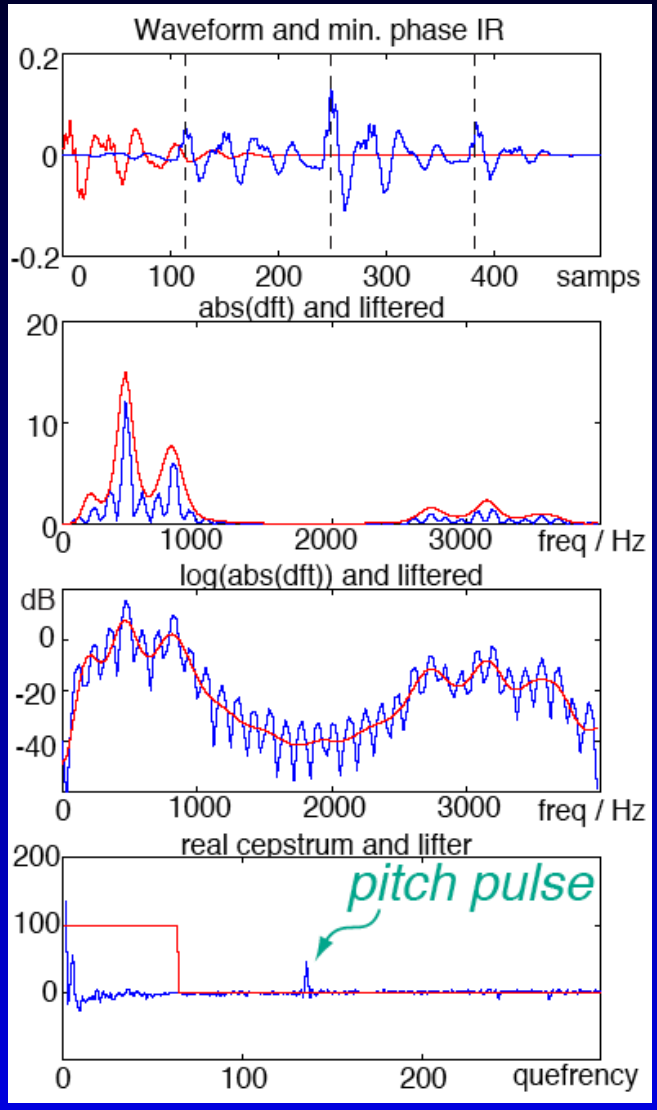
MFCCs

- “Mel-frequency cepstral coefficients”
- Popular in speech analysis community
- Feature vector characterizing one frame of audio
- Gives the spectral envelope for the frame
- Emphasizes perceptual aspects: mel frequency scale, logarithmic amplitude

Computing MFCCs

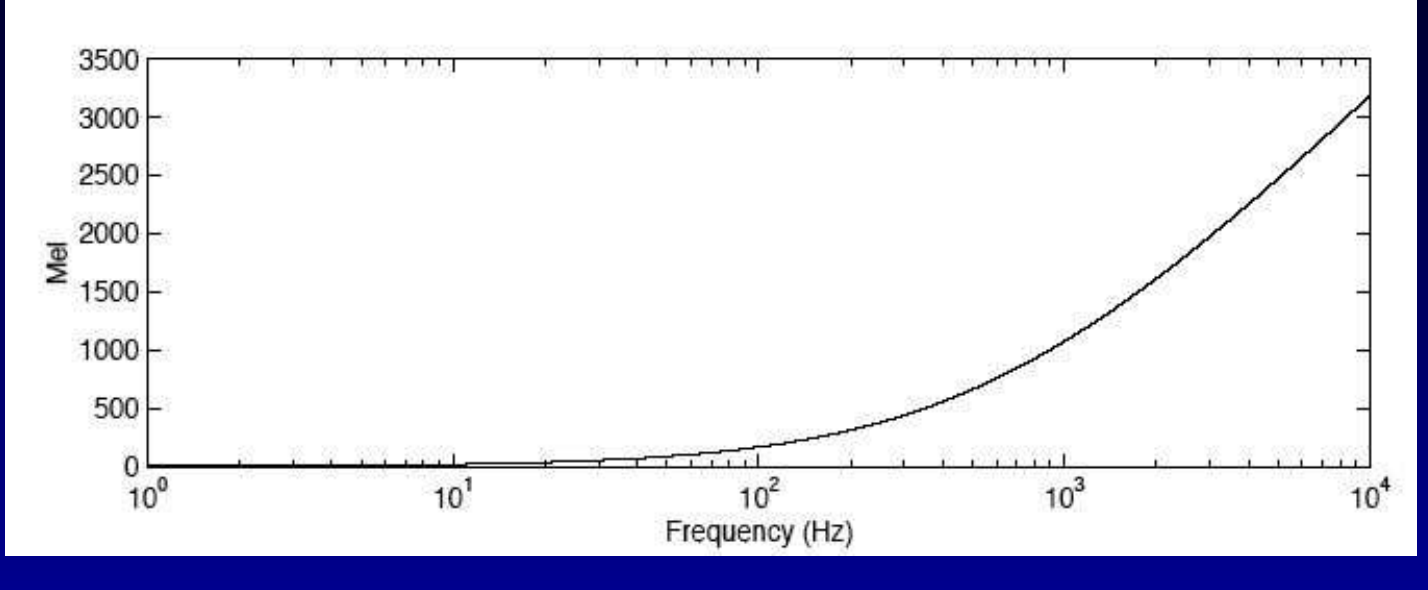


Cepstrum

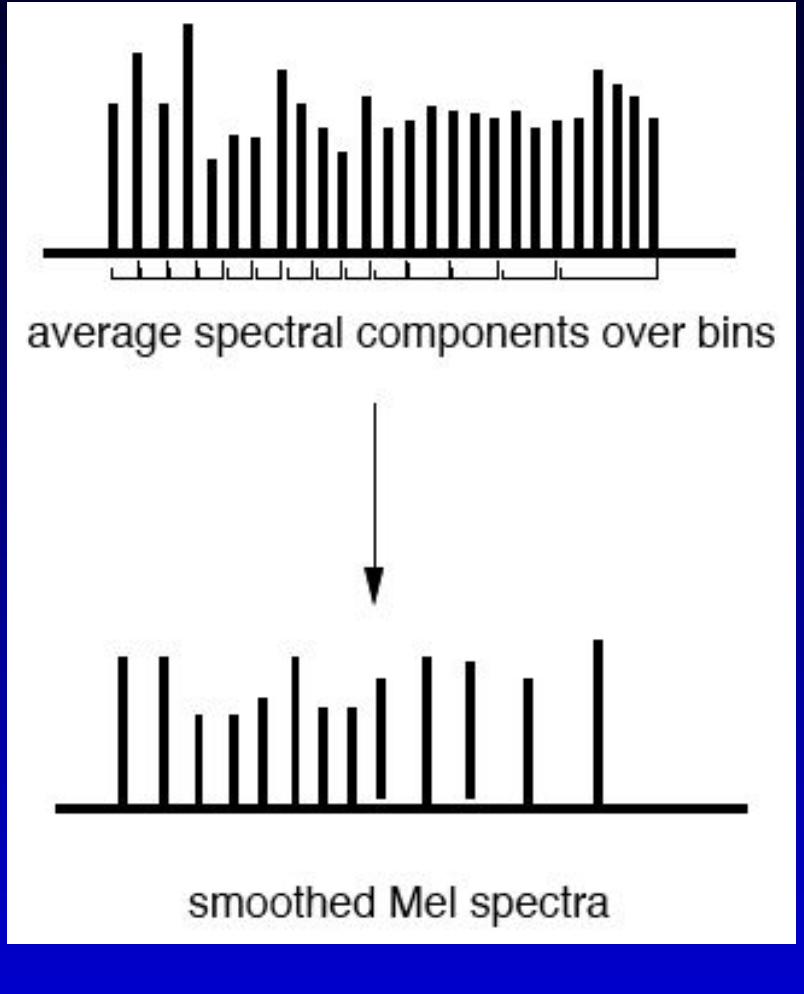


From http://www.cs.biu.ac.il/ aronowc/speech/features.pdf

Mel Frequency



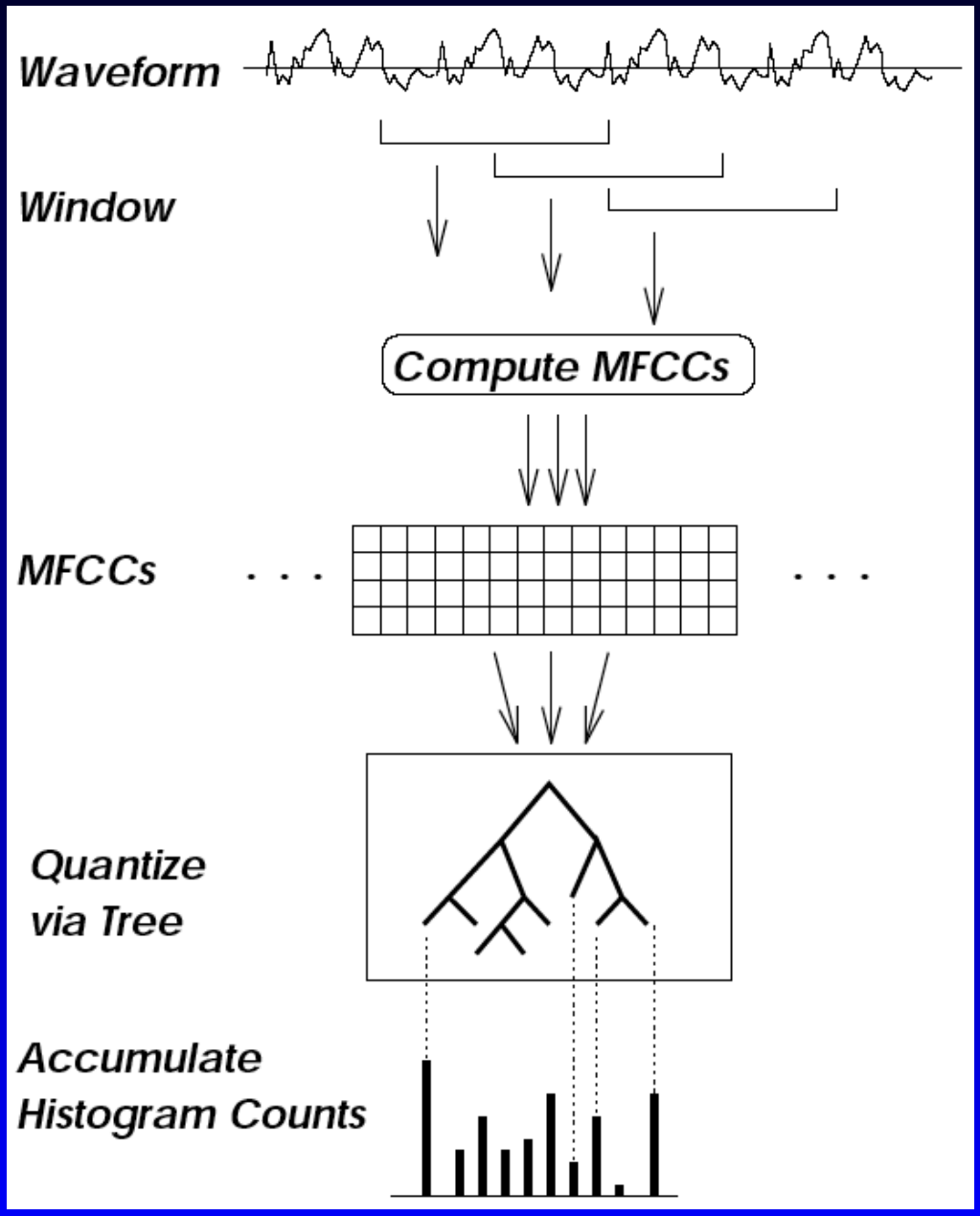
Mel Spectra


average spectral components over bins
smoothed Mel spectra

Foote – Content-Based Retrieval of Music and Audio

- Assess acoustic similarity of audio segments, use this to search database
- System trained by human input
- Uses vector quantization
- Extract feature vectors, quantize, generate template
- Compute distance between templates

Foote: Procedure



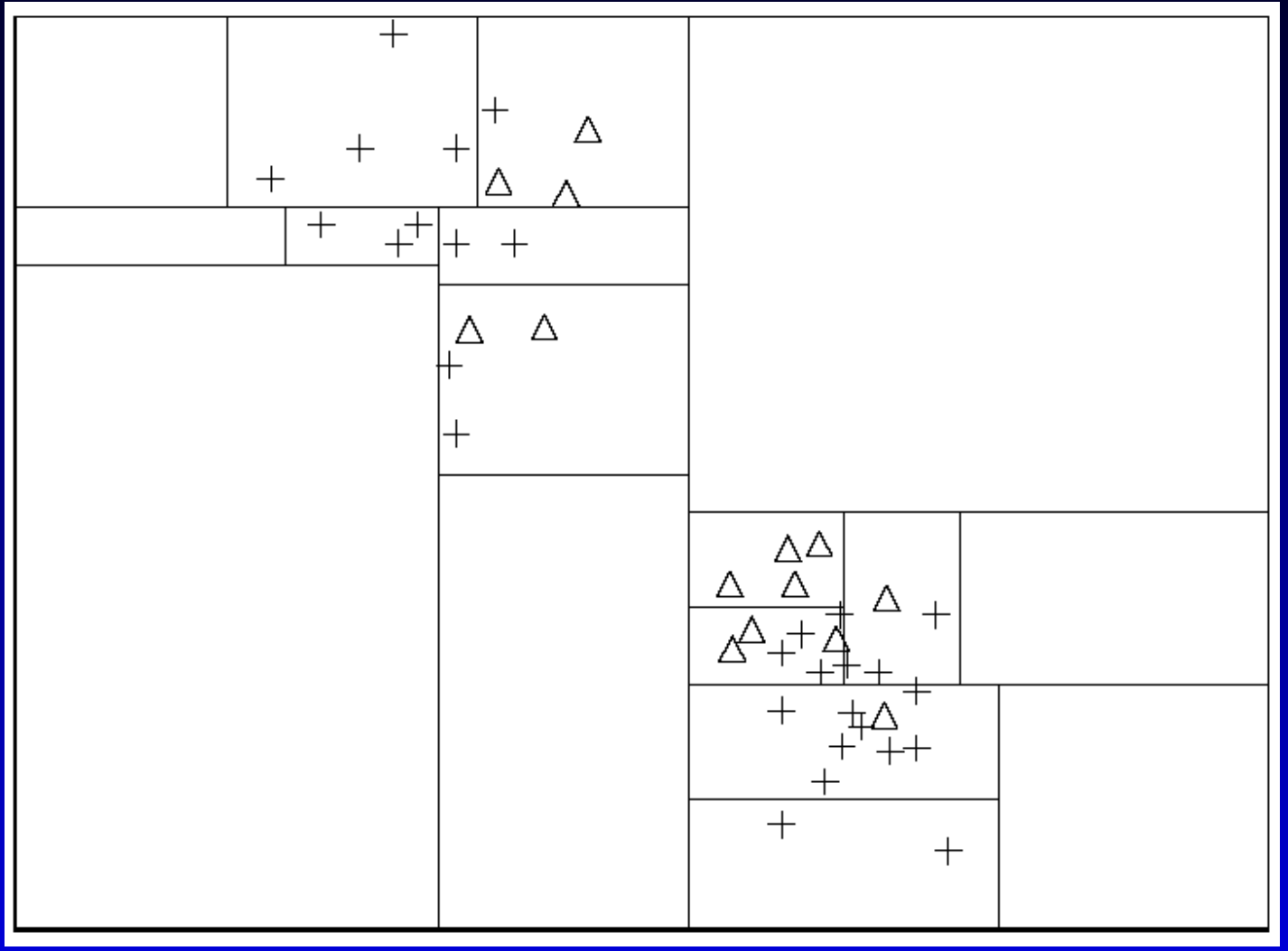
Foote: Training

- Give a set of labeled examples to the system
- These labels drive tree-based quantization
- Training deemphasizes irrelevant information

Foote: Tree-Based Quantizer

- Feature space partitioned into cells
- Cells have maximally different class populations
- Recursively split space along each dimension
- Maximize mutual information $\rightarrow$ probability that the different cells contain different classes

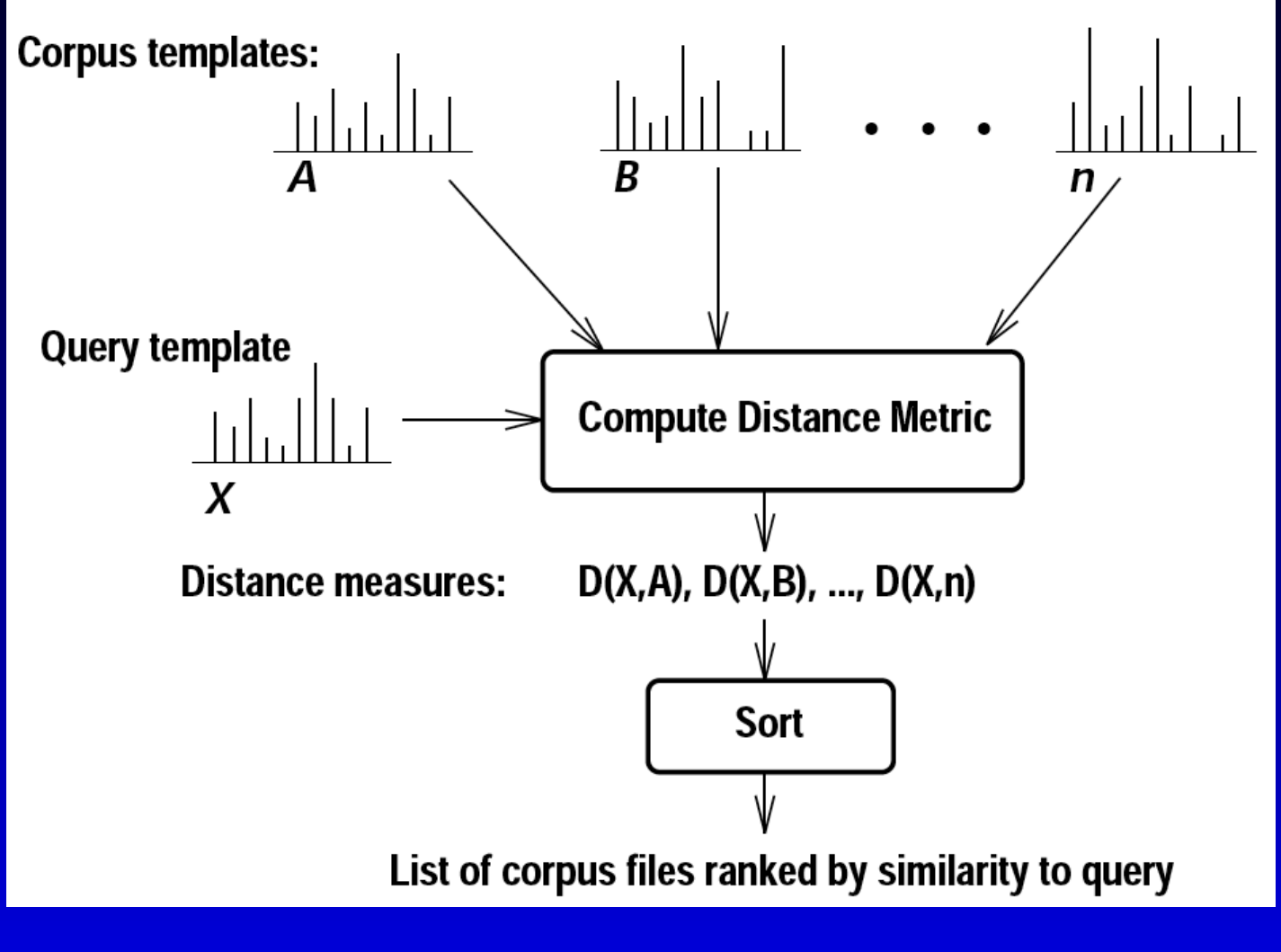
Foote: Tree-Based Quantizer



Foote: Comparing Templates

- Make a template (histogram) based on the frequency of each cell
- Similar templates will be close to each other
- Define distance: Euclidean distance, cosine distance
- Search: compute distance to audio samples in database, sort

Foote: Comparing Templates



Foote: Performance

Distance	Q Tree (D_C) "unsupervised"	Q Tree (D_C) supervised	Muscle Fish (no DPL)	Muscle Fish (+ DPL)
Laughter (M)	0.68	0.82	1.00	1.00
Oboe	0.11	0.43	0.69	0.94
Agogo	1.00	1.00	0.53	0.58
Speech (F)	0.77	0.87	0.69	0.94
Touchtone	0.61	1.00	0.44	0.73
Rain/thunder	0.22	0.35	0.30	0.42
Mean AP	0.580	0.772	0.608	0.768

Distance:	Euclidean (D_E) supervised	Euclidean (D_E) "unsupervised"	Cosine (D_C) supervised	Vector distance
AP	0.35	0.32	0.40	0.31

Logan – Music Similarity

- Automatically determine music similarity
- Builds on work of Foote. Differences:
  - Histogram bins local to each song
  - Uses “Earth Mover’s Distance”

Logan: Procedure

- Compute signature based on spectral features
  - Generate MFCCs
  - Cluster using K-means technique
  - Set of clusters (mean, covariance, weight) is song's signature
  - NB: clustering is local to each song
- Compare signatures using EMD

Logan: K-means Clustering

- Randomly assign MFCCs to K clusers
- For each point
  - Calculate distance to the centroid of each cluster
  - Move it to the closest cluster
- Sum of distances smaller at each step
- Stop when no other moves required
- Clusters non-hierarchical, non-overlapping
- Every member closest to its own cluster

Logan: Earth Mover's Distance

- Calculates the minimum amout of work required to transform one signature into the other
- Cluster p_i expressed as $(\mu_{p_i}, \Sigma_{p_i}, w_{p_i})$
- Uses distance $d_{p_i q_j}$ (Kullback Leibler), "flow" $f_{p_i q_j}$ between clusters
- Solve for flow subject to constraints
  - Minimize $W = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{p_i q_j} f_{p_i q_j}$
- $EMD(P, Q) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d_{p_i q_j} f_{p_i q_j}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{p_i q_j}}$

Logan: Performance

Nr. MFCC Features	Average number of songs in the same genre		
	Closest 5	Closest 10	Closest 20
12	3.43	6.53	12.4
19	3.44	6.57	12.5
29	3.36	6.44	12.3

Table 2: Average number of closest songs with the same genre as the seed song

Nr. MFCC Features	Average number of songs by the same artist		
	Closest 5	Closest 10	Closest 20
12	1.13	1.72	2.46
19	1.17	1.80	2.59
29	1.16	1.80	2.64

Table 3: Average number of closest songs by the same artist as the seed song

Nr. MFCC Features	Average number of songs on the same album		
	Closest 5	Closest 10	Closest 20
12	0.84	1.21	1.61
19	0.86	1.26	1.68
29	0.81	1.21	1.69

Table 4: Average number of closest songs on the same album as the seed song

Logan: Performance

Algorithm	Average Number of Similar Songs		
	Closest 5	Closest 10	Closest 20
Random	0.2	0.6	0.9
Proposed	2.5	4.7	8.2

Table 5: Average number of similar songs in playlists generated at random and by our similarity measure as judged by 2 users on 20 queries

Nr. MFCC Features	% of times original song returned within:		
	Closest 1	Closest 5	Closest 10
12	98.8	99.2	99.3
19	99.8	100.0	100.0
29	97.2	97.6	97.8

Table 6: Percentage of times the original song is returned as one of the closest 1, 5 and 10 songs when the query is a clipped version of the original

Further Reading

- Logan – Mel Frequency Cepstral Coefficients for Music Modeling (2000)
- Logan – Toward Evaluation Techniques for Music Similarity (2003)
- Liu, Huang – Content-Based Indexing and Retrieval-By-Example in Audio (2000)