	2006 International Conference on Multimedia & Expo
	Tutorial:
Music Information Retrieval	
	ICME 2006
George Tzanetakis Assistant Professor Computer Science Department (also in Music, ECE) University of Victoria, Canada gtzan@cs.uvic.ca www.cs.uvic.ca/~gtzan	
	1 Copyright 2006 G.Tzanetakis











## Observations

- Analysis/Understanding require multiple representations – no "best" one
- Coding/Compression/Processing typically search for the "best"/"optimal" way to do things
- Paradigm shift is necessary to make music and multimedia more than just a lot of numbers – content "understanding"

7

> MACHINE LEARNING is crucial











































































## Rhythm



- Rhythm = movement in time
- > Origins in poetry (iamb, trochaic...)
- Foot tapping definition
- > Hierarchical semi-periodic structure at multiple levels of detail
- Links to motion, other sounds
- Running vs global

### 43









































62



O) 

Ó

Gaussian Mixture Models













- > Timbral texture, rhythmic structure, harmonic content
- > Automatic Musical Genre Classification
  - Evaluate musical content features
  - Structure audio collections 71











# Comparison of human and automatic genre classification

Best feature set: MFCC + Rhythm Less influence of classifier e.g. MFCC + Rhythm / Gaussian



Automatic classification

Lippens et al. ICASPP 04

performs as well as the "worst" human

"best" human is 20% more accurate

Difference is probably smaller today (better features, SVMs)



















































## Audio Fingerprinting



- Each song is represent as a fingerprint (small robust representation)
- > Search database based on fingerprint
- Main challenges
  - > highly robust fingerprint extraction
  - > efficient fingerprint search strategy
- Information is summarized from the whole song – attacks degrade unlike watermarking

101









## Philips Research

Haitsa & Kalker Ismir 2002

- > 32-bit subfingerprints for every 11.6 msec
- overlapping frames of 0.37 seconds (31/32 overlap)
- > PSD -> logarithmic band spacing (bark)
- bits 0-1 sign of energy
- > looks like a fingerprint
- > assume one fingerprint perfect hierarchical database layout (works ok)

105













## Content & Context Aware User Interfaces

- Automatic results not perfect
- Music listening subjective
- Browsing vs retrieval
- "Overview, Zoom and Filter, Details"
- > Adapt UI to audio "Content & Context"
  - Computer audition
  - Visualization

111



























