

Audio Retrieval


David Kauchak
cs458
Fall 2012

Administrative

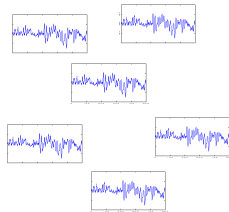
- Assignment 4
 - Two parts
- Midterm
 - Average: 52.8
 - Median: 52
 - High: 57
- In-class "quiz": 11/13

Audio retrieval

text retrieval corpus



audio retrieval corpus



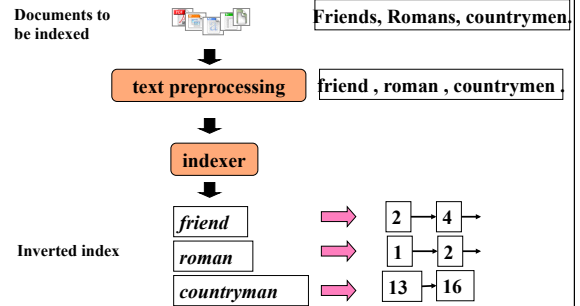
What do you want from an audio search engine?

- **Name:** You might know the name of the **song** or the **artist**
- **Genre:** You might try "Bebop," "Latin Jazz," or "Rock"
- **Instrumentation:** The tenor sax, guitar, and double bass are all featured in the song
- **Emotion:** The song has a "cool vibe" that is "upbeat" with an "electric texture"
- Some other approaches to search:
 - musicoverly.com
 - pandora.com (song similarity)
 - Genius (collaborative filtering)

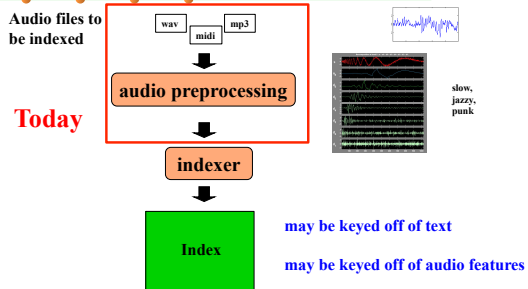
Current audio search engines

- What are they?
- What can you search by?
- How well do they work?
- How could they be improved?
- Challenges?

Text Index construction



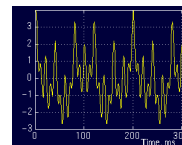
Audio Index construction



Sound

What is sound?

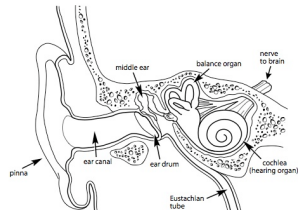
- A longitudinal compression wave traveling through some medium (often, air)
- Rate of the wave is the frequency
- You can think of sounds as a sum of sign waves



Sound

How do people hear sound?

The cochlea in the inner ear has hair cells that "wiggle" when certain frequency are encountered



<http://www.hcchildres.ca/NR/rdonlyres/9A4B4AD64-A01F-4469-8C CF-EA2B88617C39/16128/theear.jpg>

Digital Encoding

Like everything else for computers, we must represent audio signals digitally

Encoding formats:

- WAV
- MIDI
- MP3
- Others...

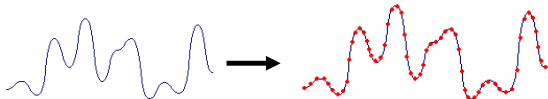
WAV

Simple encoding

Sample sound at some interval (e.g. 44 KHz).

High sound quality

Large file sizes



MIDI

Musical Instrument Digital Interface

MIDI is a language

Sentences describe the channel, note, loudness, etc.

16 channels (each can be thought of and recorded as a separate instrument)

Common for audio retrieval and classification applications

MP3

Common compression format

3-4 MB vs. 30-40 MB for uncompressed

Perceptual noise shaping

- The human ear cannot hear certain sounds
- Some sounds are heard better than others
- The louder of two sounds will be heard

Lossy or lossless?

- Lossy compression
- quality depends on the amount of compression
- like many compression algorithms, can have issues with randomness (e.g. clapping)

MP3 Example

How MP3 Files Work

If there is a loud sound in one band, the compression algorithm can ignore all of the other bands.

© 2003 HowStuffWorks

Features

→

Weight vectors

- word frequency
- count normalization
- idf weighting
- length normalization

→

?

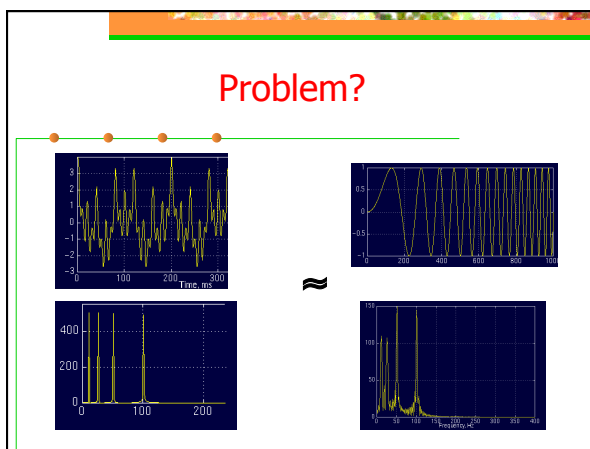
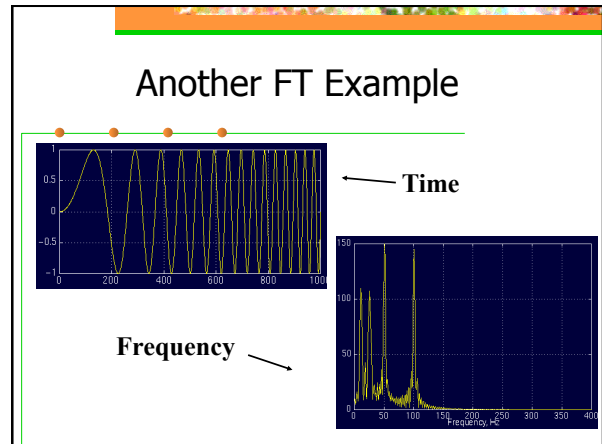
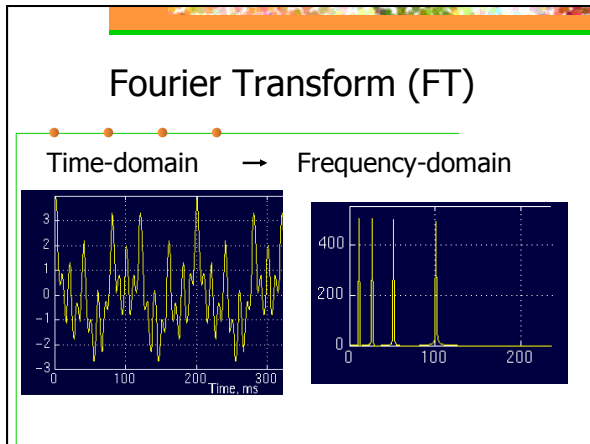
Tools for Feature Extraction

Fourier Transform (FT)

Short Term Fourier Transform (STFT)

Wavelets

Paul Antonson



Problem with FT

- FT contains only frequency information
- No **time** information is retained
- Works fine for stationary signals
- Non-stationary or changing signals cause problems
 - FT shows frequencies occurring at all times instead of specific times

Ideas?

Short-Time Fourier Transform (STFT)

Idea: Break up the signal into discrete windows
 Treat each signal within a window as a stationary signal
 Take FT over each part

STFT Example

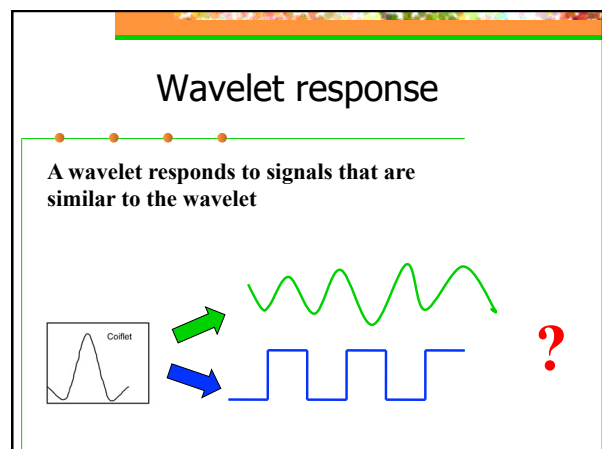
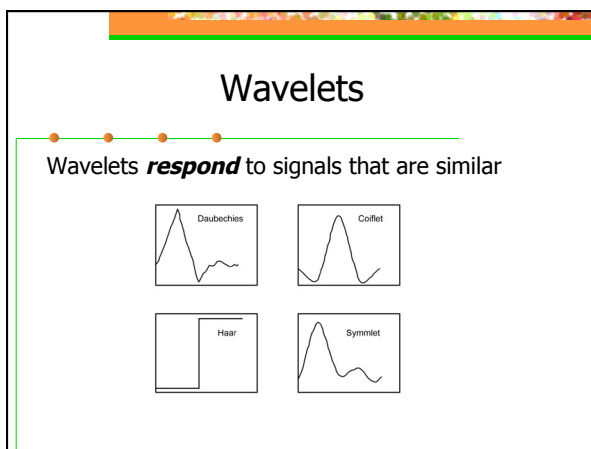
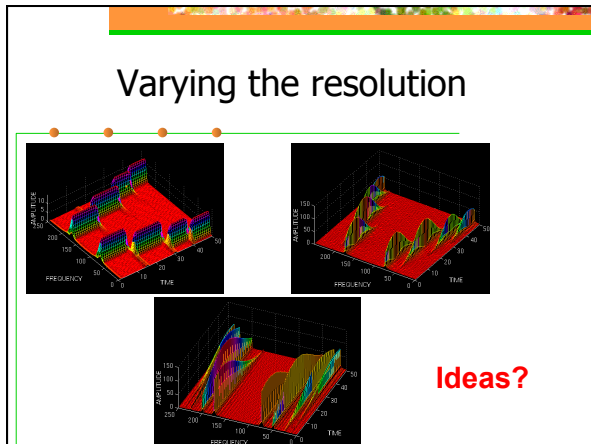
STFT Example

Problem: Resolution

How do we pick the window size? Trade-offs?

We can vary time and frequency accuracy

- Narrow window: good time resolution, poor frequency resolution
- Wide window: good frequency resolution, poor time resolution



Wavelet response

Scale matters!

The diagram illustrates the concept of scale in wavelet analysis. On the left, a small wavelet labeled 'Cofflet' is shown. A green arrow points to a larger wavelet that matches the high-frequency oscillations of a signal. A blue arrow points to a smaller wavelet that matches the low-frequency oscillations of the same signal. A red question mark is placed to the right, suggesting the need to choose the appropriate scale for analysis.

Wavelet Transform

Idea: Take a wavelet and vary scale

Check response of varying scales on signal

This slide introduces the wavelet transform by showing four common wavelet functions: Daubechies, Cofflet, Haar, and Symmetlet. Each function is shown in a separate plot, illustrating their unique shapes and how they vary with scale.

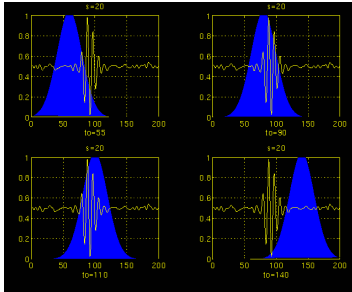
Wavelet Example: Scale 1

This slide shows four subplots of wavelet responses at scale 1. Each subplot displays a signal (yellow) and a wavelet response (blue) over a specific time interval. The intervals are labeled as $t \in [0, 90]$, $t \in [50, 90]$, $t \in [100, 140]$, and $t \in [100, 140]$. The responses show how the wavelet captures local features of the signal at this scale.

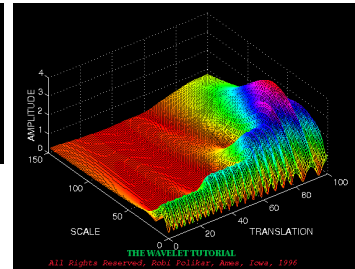
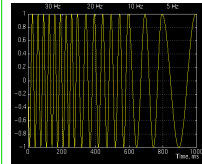
Wavelet Example: Scale 2

This slide shows four subplots of wavelet responses at scale 2. Each subplot displays a signal (yellow) and a wavelet response (blue) over a specific time interval. The intervals are labeled as $t \in [0, 20]$, $t \in [50, 80]$, $t \in [100, 110]$, and $t \in [100, 140]$. The responses show how the wavelet captures local features of the signal at this finer scale.

Wavelet Example: Scale 3



Wavelet Example

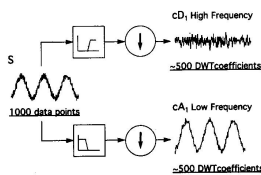


Scale = 1/frequency
Translation \approx Time

THE WAVELET TUTORIAL
All Rights Reserved. Robb Patton, Ames, Iowa, 1996

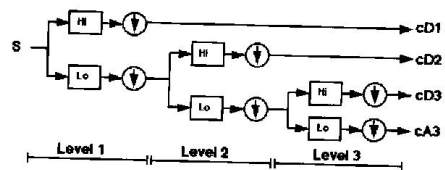
Discrete Wavelet Transform (DWT)

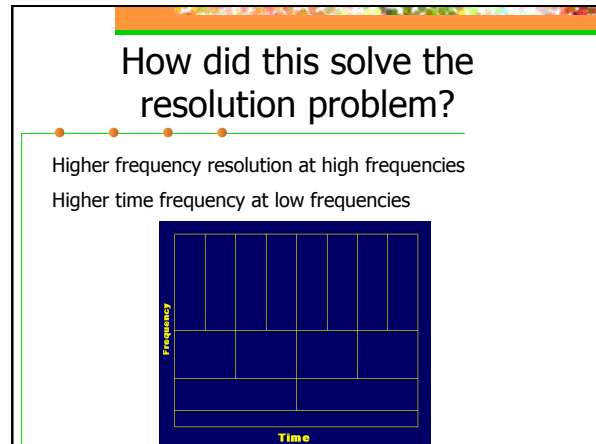
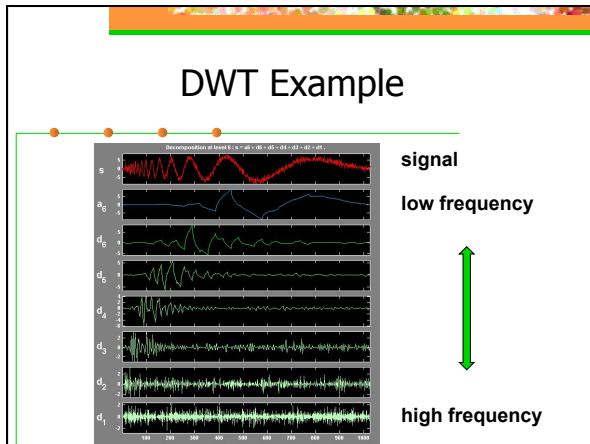
Wavelets come in pairs (high pass and low pass filter)
Split signal with filter and downsample



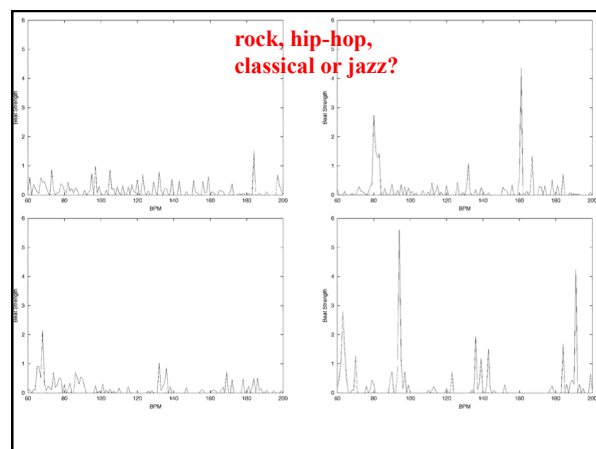
DWT cont.

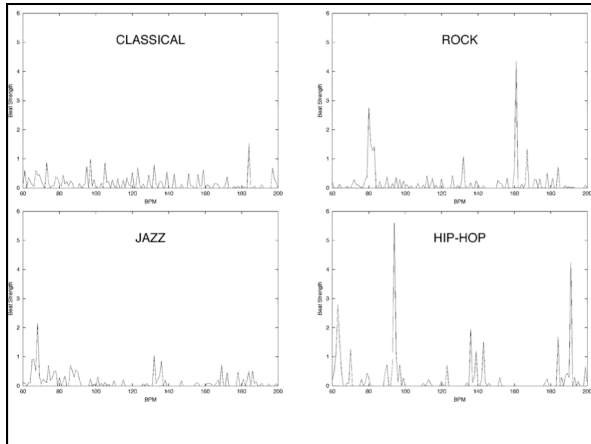
Continue this process on the low frequency portion of the signal





- ### Feature Extraction
- All these transforms help us understand how the frequencies changes over time
- Features extraction:
- Mel-frequency cepstral coefficients (MFCCs)
 - Attempt to mimic human ear
 - Surface features (texture, timbre, instrumentation)
 - Capture frequency statistics of STFT
 - Rhythm features (i.e the "beat")
 - Characteristics of low-frequency wavelets





Music Classification

Data

- Audio collected from radio, CDs and Web
 - Speech vs. music
 - Genres: classic, country, hiphop, jazz, rock
 - 4-types of classical music
- 50 samples for each class, 30 sec. long
- Task is to predict the genre of the clip

Approach

- Extract features
- Learn genre classifier

Music Classification

Data

- Audio collected from radio, CDs and Web
 - Speech vs. music
 - Genres: classic, country, hiphop, jazz, rock
 - 4-types of classical music
- 50 samples for each class, 30 sec. long
- Task is to predict the genre of the clip

How well do you think we can do?

General Results

	Music vs. Speech	Genres	Classical
Random	50%	16%	25%
Classifier	86%	62%	76%

Results: Musical Genres

	Classic	Country	Disco	Hiphop	Jazz	Rock
Classic	86	2	0	4	18	1
Country	1	57	5	1	12	13
Disco	0	6	55	4	0	5
Hiphop	0	15	28	90	4	18
Jazz	7	1	0	0	37	12
Rock	6	19	11	0	27	48

Pseudo-confusion matrix

Results: Classical

	Choral	Orchestral	Piano	String
Choral	99	10	16	12
Orchestral	0	53	2	5
Piano	1	20	75	3
String	0	17	7	80

Confusion matrix

Thanks

- Robi Polikar for his old tutorial
(<http://www.public.iastate.edu/~rpolikar/WAVELETS/WTtutorial.html>)