Instrument Recognition

Spectral Analysis of Musical Instruments and Pitch

Patrick Kruse, Kyle Ringgenberg, Yi-Chieh Wu

Purpose

To detect the pitch and instrument of a monophonic signal. To decompose polyphonic signals into their component pitches and instruments by analyzing the waveforms and spectra of each instrument.

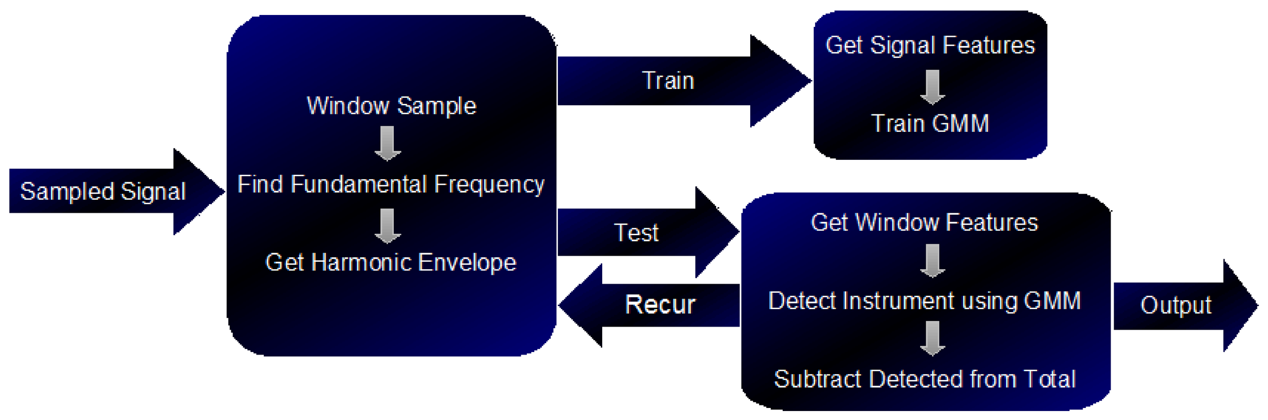
Applications
- Understanding Musical Timbre
- Automatic Music Transcription
- Music Information Retrieval

Background
- •Techniques from speech processing
- •Focus on monophonic recognition
- •Limited successes
  - •Limited number of instruments
  - •Known pitch, detect instrument
  - •Specially-arranged ensemble recordings

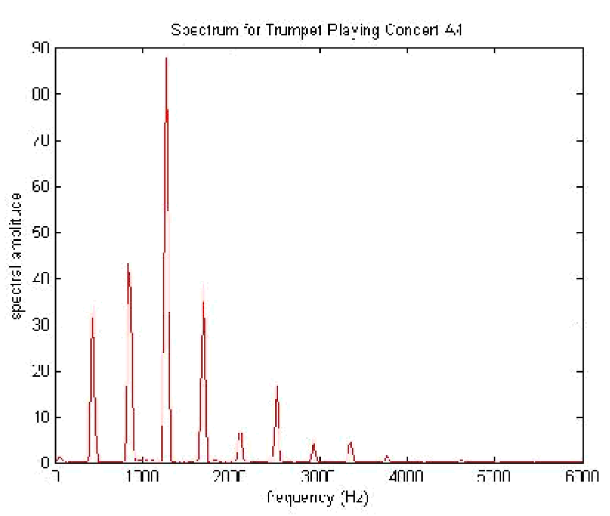
Approach

- Capture characteristics (features) of the signal correlated to the instrument
- Classify using a Gaussian Mixture Model
- Determine unknown instrument using signal features

Why not use Matched Filters?
- Good in Concept, Bad in Practice
- Player Differences
- Very Frequency-Sensitive
- Instrument Signals Too Similar
- Hard to reproduce sounds in lower and upper ranges



Pitch Detection



Technique
- Detect frequency of lowest peak in spectrum

Advantages
- Simplistic and easy to code
- Works for harmonic signals

Disadvantages
- Useless in presence of noise or other inharmonic frequencies

Other Algorithms
- Autocorrelation
- Harmonic Product Spectrum
- Maximum Likelihood Estimation

Advantages
- Works with signals that are not purely harmonic

Disadvantages
- Frequency-halving or frequency-doubling

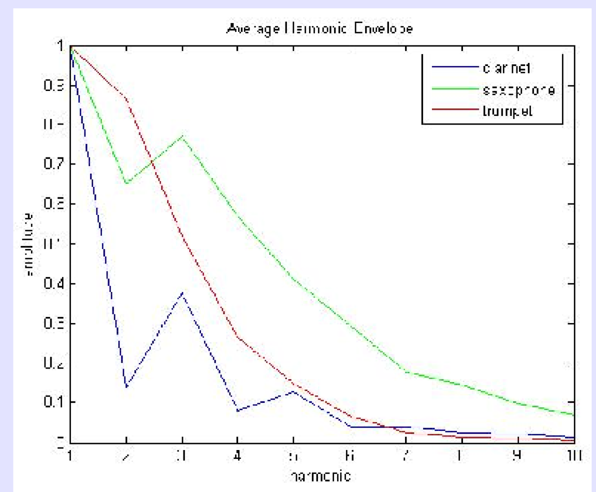
Features

Cepstral Features
- Mel-frequency Cepstrum Coefficients (MFCC), k = 2:13

Spectral Features
- Slope
- Roll-Off
- Centroid
- Spread
- Skew
- Kurtosis
- Odd-to-Even Harmonic Energy Ratio (OER)
- Tristimulus

Note: Most of these features have perceptual interpretations.
centroid → sound "brightness"
tristimulus → equivalent to visual color attributes

Sinusoidal Harmonic Modeling
- Estimate the harmonic peaks
- Produce the "typical" spectrum of the instrument independent of fundamental frequency



Instrument Characteristics

Clarinet
- Fast Decline, Low Roll-Off Frequency
- High OER due to Closed Cylinder at One End
- First Tristimulus Proportionately Higher than Second and Third

Saxophone
- Slow Decline, High Roll-Off Frequency
- More Evenly Distributed Tristimulus

Trumpet
- Medium Decline, Mid Roll-Off Frequency

Gaussian Mixture Model

- Models the probability density function of observed variables by a multivariate Gaussian mixture density
- Independent variables are measured as fractions of a total
- K-means clustering
- Refine using Expectation-Maximization
- Missing Features Approach

$$p(x)=\sum_{i=1}^{N} p_i\phi_i(x,\mu_i,\Sigma_i) \longrightarrow p(x)=\sum_{i=1}^{N}\prod_{j\in M'}\phi_i\left(x_j,m_{ij},\sigma_{ij}^2\right)$$

(assuming independence of features and missing properties *M*)

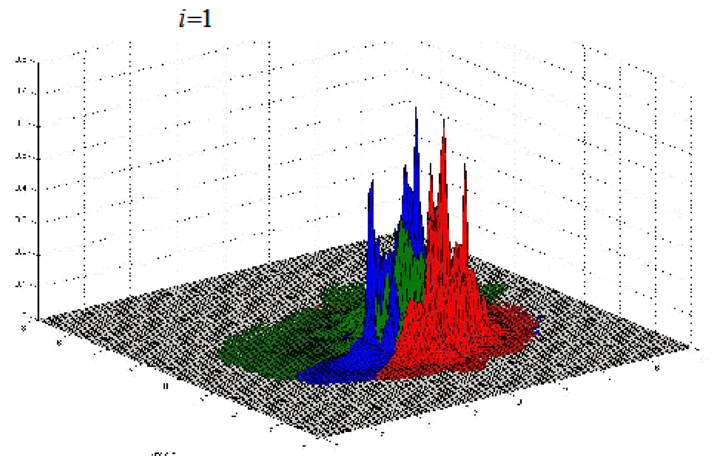


Figure 1: Gaussian Mixture Model for Clarinet (blue), Saxophone (green), and Trumpet (red). Signals with features falling in a colored area are classified as a particular instrument. (Gray represents indeterminate instrument.)

Sound Data

Training
- Monophonic signals
- One full chromatic scale per instrument

Testing
- One short monophonic tune per instrument
- Two short polyphonic tunes with each instrument combination

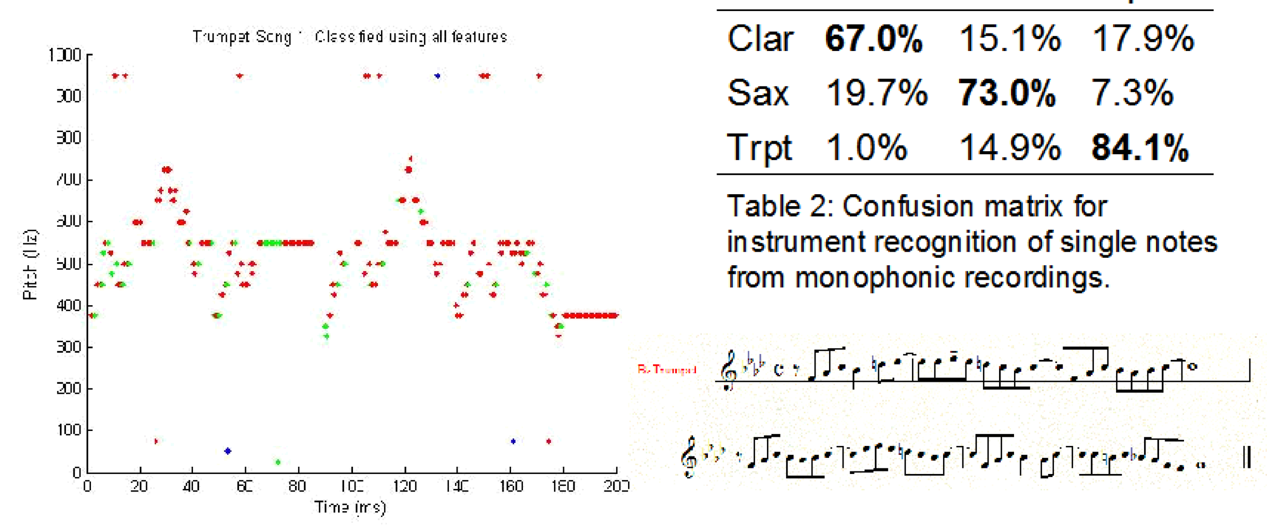
Results

Self-Validation

		Predicted		
		Clarinet	Saxophone	Trumpet
Actual	Clarinet	**90.0%**	7.5%	2.5%
	Saxophone	2.9%	**92.3%**	4.9%
	Trumpet	0.9%	11.5%	**87.5%**

Table 1: Confusion matrix for instrument recognition with training data.

Monophonic Recordings

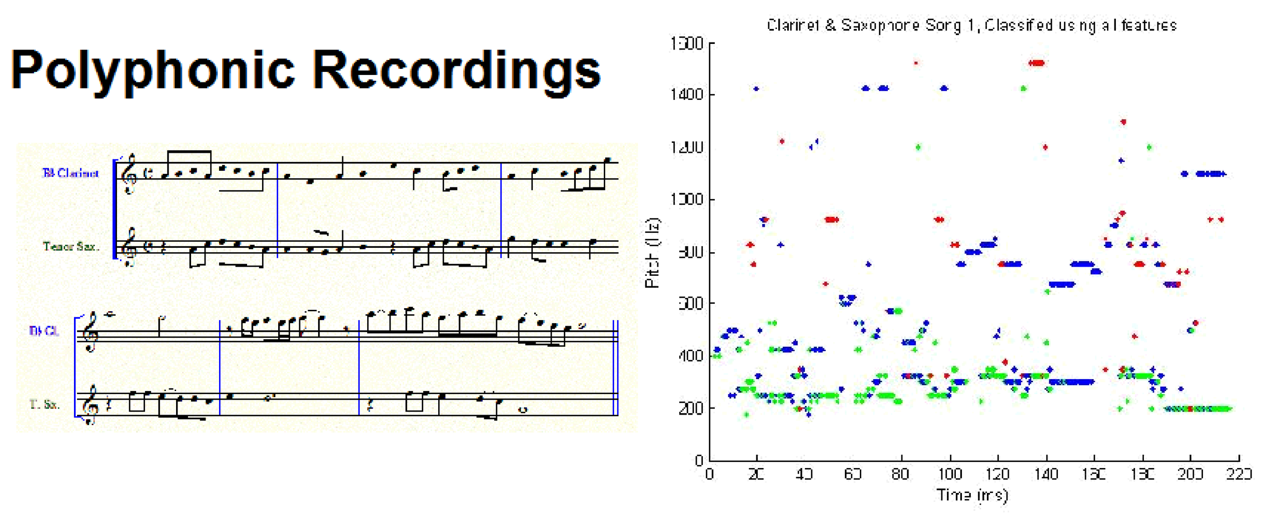


	Clar	Sax	Trpt
Clar	**67.0%**	15.1%	17.9%
Sax	19.7%	**73.0%**	7.3%
Trpt	1.0%	14.9%	**84.1%**

Table 2: Confusion matrix for instrument recognition of single notes from monophonic recordings.

- Average instrument identification: 75%
- Much better than guessing!
- In test data, clarinet and saxophone are most similar (same instrument family), and clarinet and trumpet are most dissimilar (very different spectrum).

Polyphonic Recordings



Discussion

- Detected pitch through fundamental frequency analysis.
- Classified characteristics of each instrument based on feature analysis and training.
- Future work
  - Model additional temporal, spectral, harmonic, and perceptual features
  - Stronger training data – different players, environments, musical genre; polyphonic music
  - Capture other instruments and instrument families (strings, woodwinds, percussion, etc)

References

- J. Eggink and G.J. Brown. "A Missing Feature Approach to Instrument Identification in Polyphonic Music," in IEEE International Conference on Acoustics, Speech, and Signal Processing, Hong Kong, April 2003, 553-556.
- A.A. Livshin and X. Rodet. "Musical Instrument Identification in Continuous Recordings," in Proc. of the 7th Int. Conference on Digital Audio Effects, Naples, Italy, October 5-8, 2004.
- G. Peeters. "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," 2003. URL: http://www.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf.

For questions, comments, and preprint requests:
{pakruse, kringg, yjw}@rice.edu

Acknowledgements
- •Dept. of Electrical and Computer Engineering, Rice University
- •Richard Baraniuk
- •William Chan
- •Music Classification by Genre. Elec 301 Project, Fall 2003. Mitali Banerjee, Melodie Chu, Chris Hunter, Jordan Mayo
- •Auditory Toolbox. Malcolm Slaney
- •Netlab. Neural Computing Research Group. Aston University