A Classification Approach to Melody Transcription

Graham Poliner and Dan Ellis

{graham,dpwe}@ee.columbia.edu

OUTLINE

1. Motivation and task definition
2. Classification approach
3. Experimental results

MOTIVATION

Content-based retrieval:

- Query by humming, etc
- Recurring melodic themes

Music Analysis:

- Transcription
- Reduction

TASK DEFINITION

f_0 transcription of predominant melody

(10 ms grid)

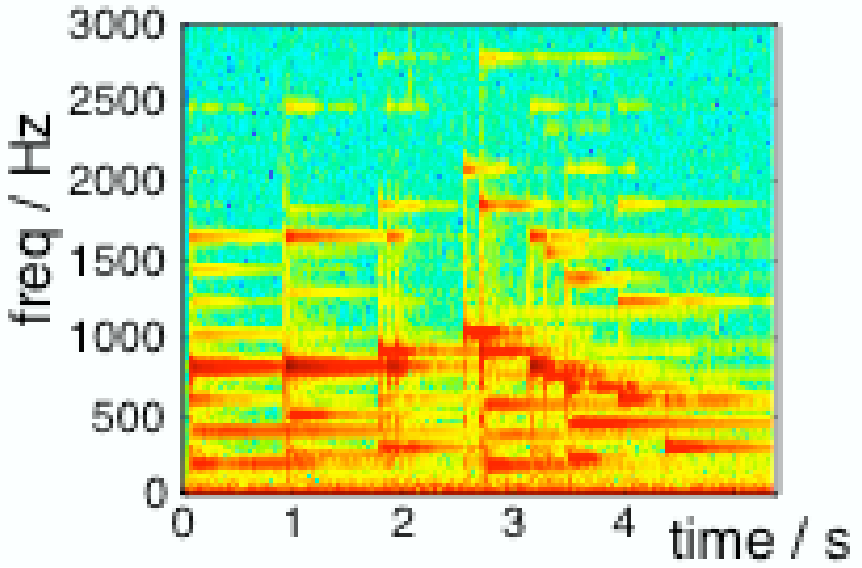
Evaluation metrics:

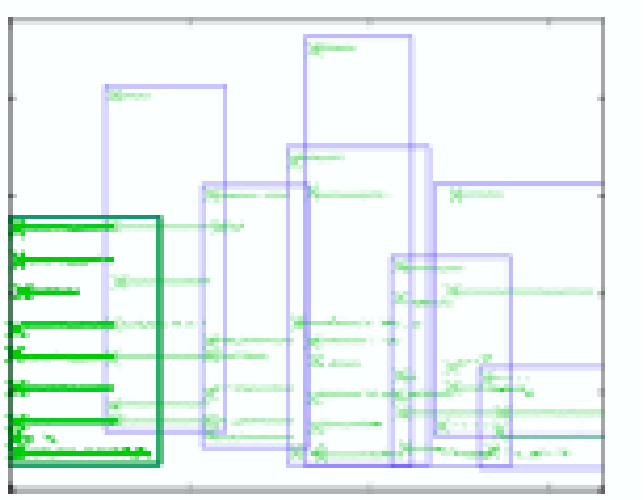
- Success = f_0 prediction within ±1/4 tone
- Chroma = integer (sub)multiple of f_0 (octave errors)

At this time, consider only voiced segments

Model-Based Approach

- Previous systems use e.g. sine models:

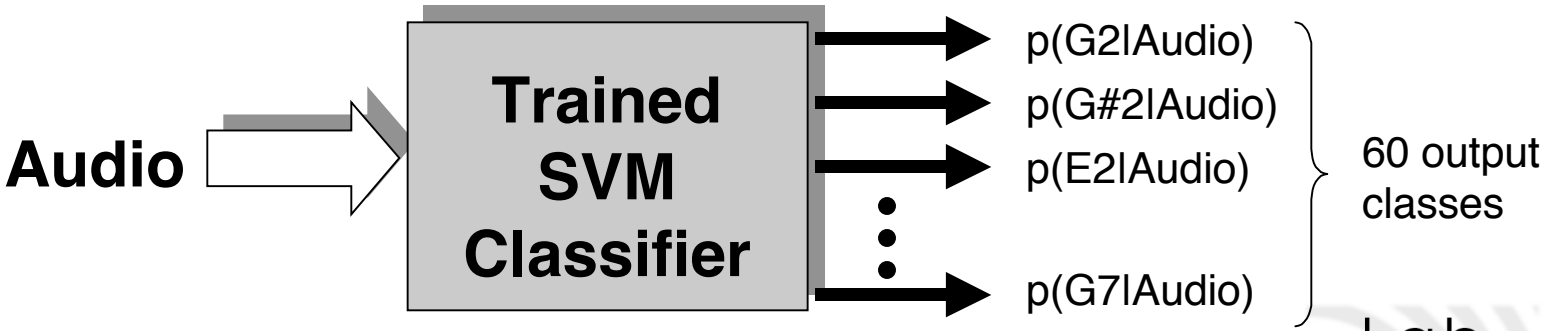




Our Approach: Classification

- Signal models may not capture everything
- Instead, trade domain knowledge for data

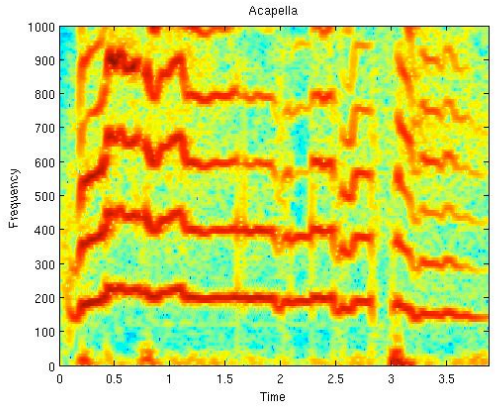
N-Way discrimination for melody



TRAINING DATA

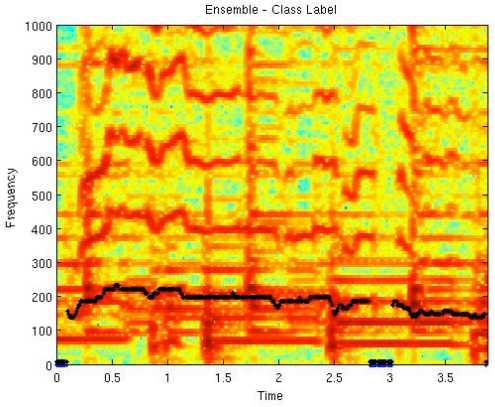
Audio ⇨ Labeled Data

- Multi-track Recordings
- MIDI Audio



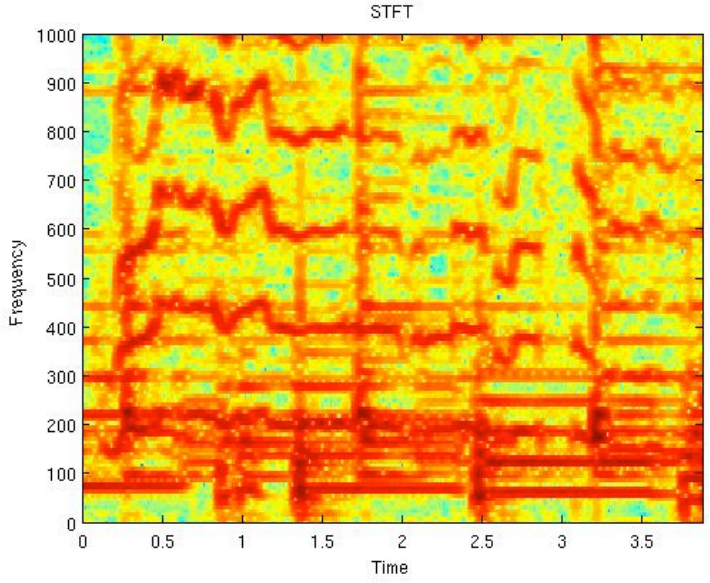
Features: column of STFT

Labels: MIDI Note Number



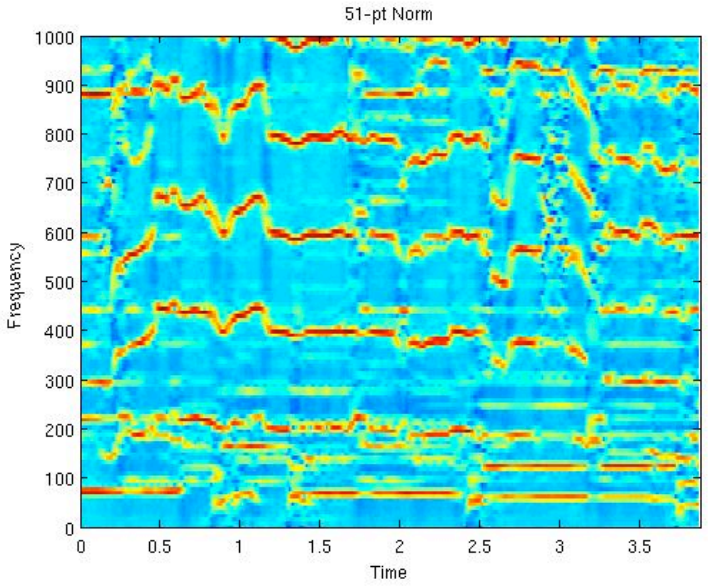
Feature Normalization

Normalization	Rate
STFT	59.0
51-pt Norm	**62.7**
Cube root	62.4
Autocorr	59.0
Cepstrum	52.1
Liftering Ceps	60.3



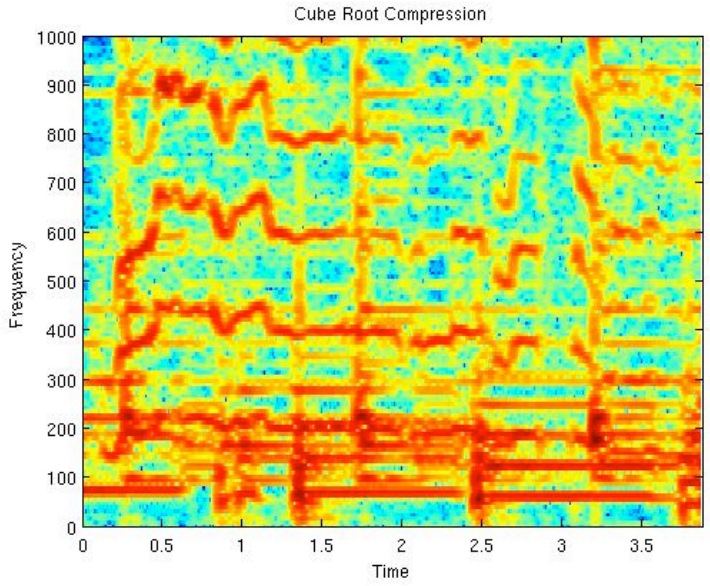
Feature Normalization

Normalization	Rate
STFT	59.0
51-pt Norm	**62.7**
Cube root	62.4
Autocorr	59.0
Cepstrum	52.1
Liftering Ceps	60.3



Feature Normalization

Normalization	Rate
STFT	59.0
51-pt Norm	**62.7**
Cube root	62.4
Autocorr	59.0
Cepstrum	52.1
Liftering Ceps	60.3

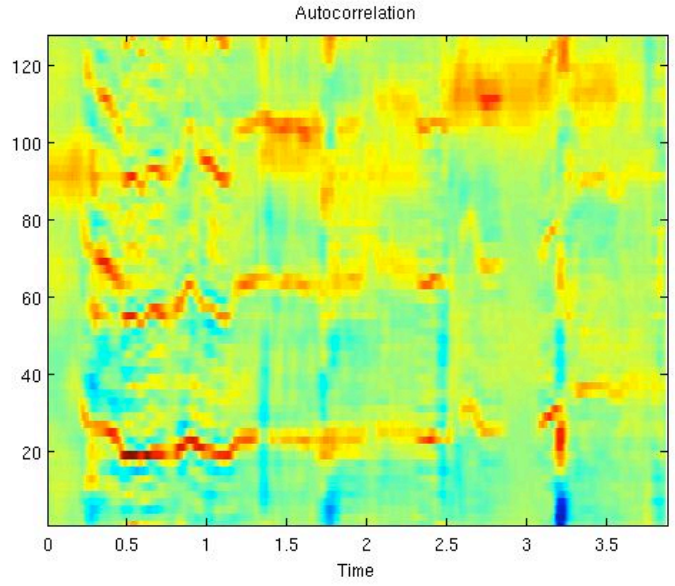


Columbia University

Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

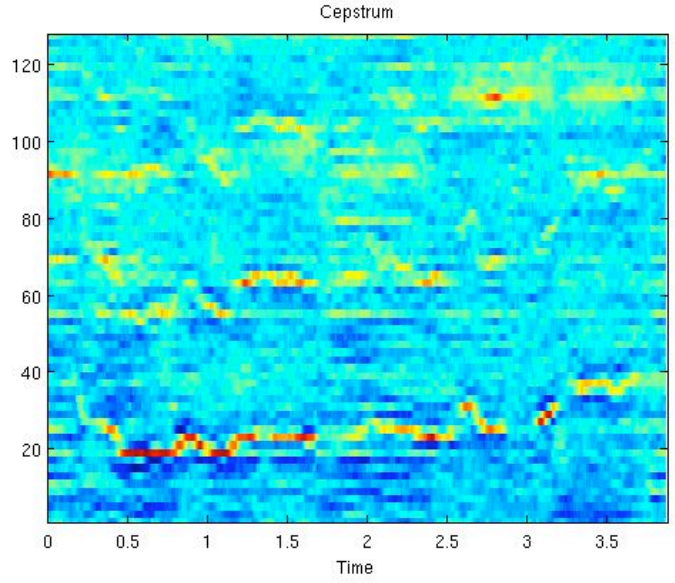
Feature Normalization

Normalization	Rate
STFT	59.0
51-pt Norm	**62.7**
Cube root	62.4
Autocorr	59.0
Cepstrum	52.1
Liftering Ceps	60.3



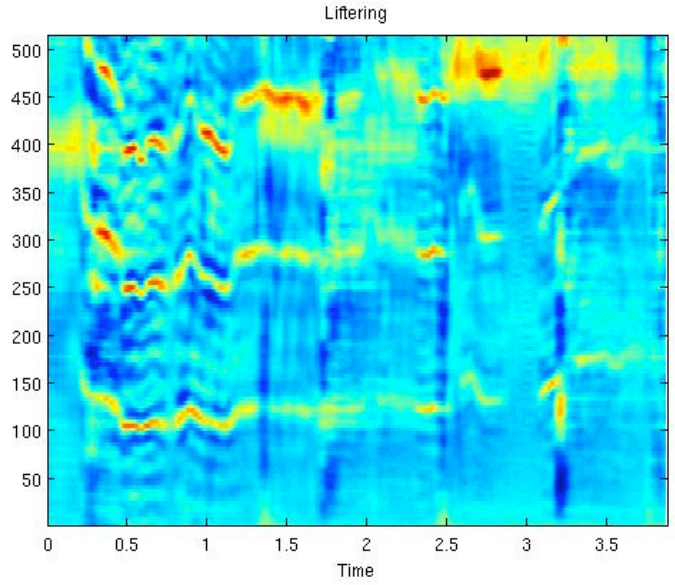
Feature Normalization

Normalization	Rate
STFT	59.0
51-pt Norm	**62.7**
Cube root	62.4
Autocorr	59.0
Cepstrum	52.1
Liftering Ceps	60.3



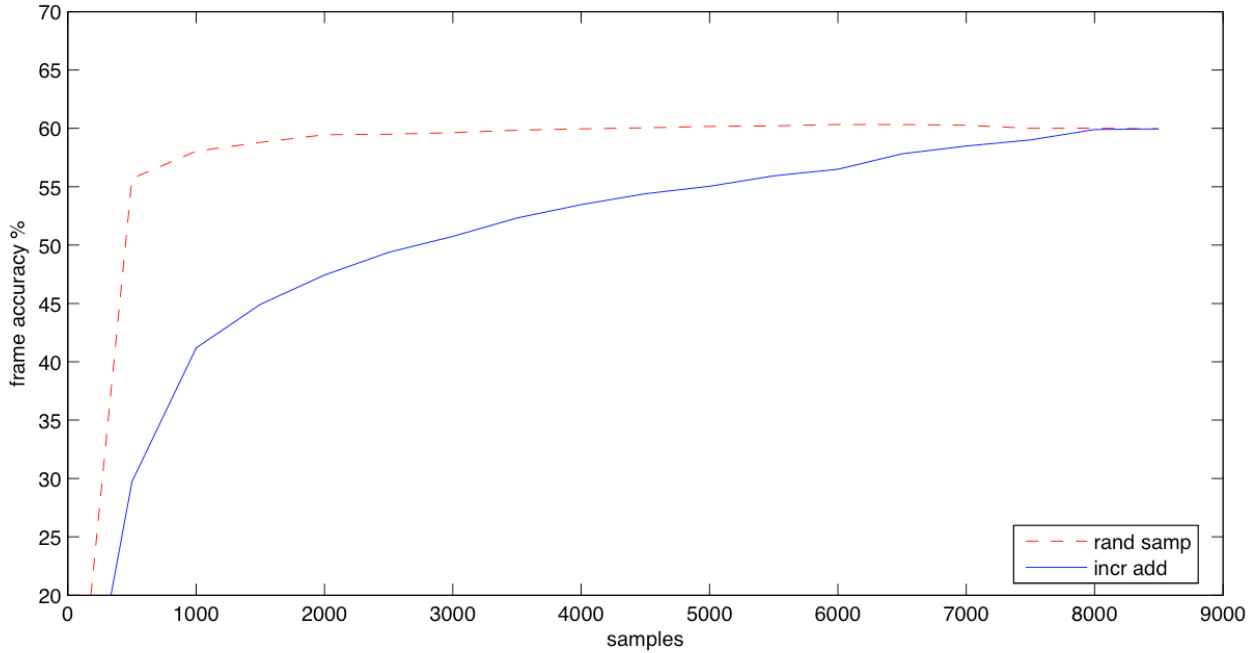
Feature Normalization

Normalization	Rate
STFT	59.0
51-pt Norm	**62.7**
Cube root	62.4
Autocorr	59.0
Cepstrum	52.1
Liftering Ceps	60.3



CLASSIFICATION EXPERIMENTS

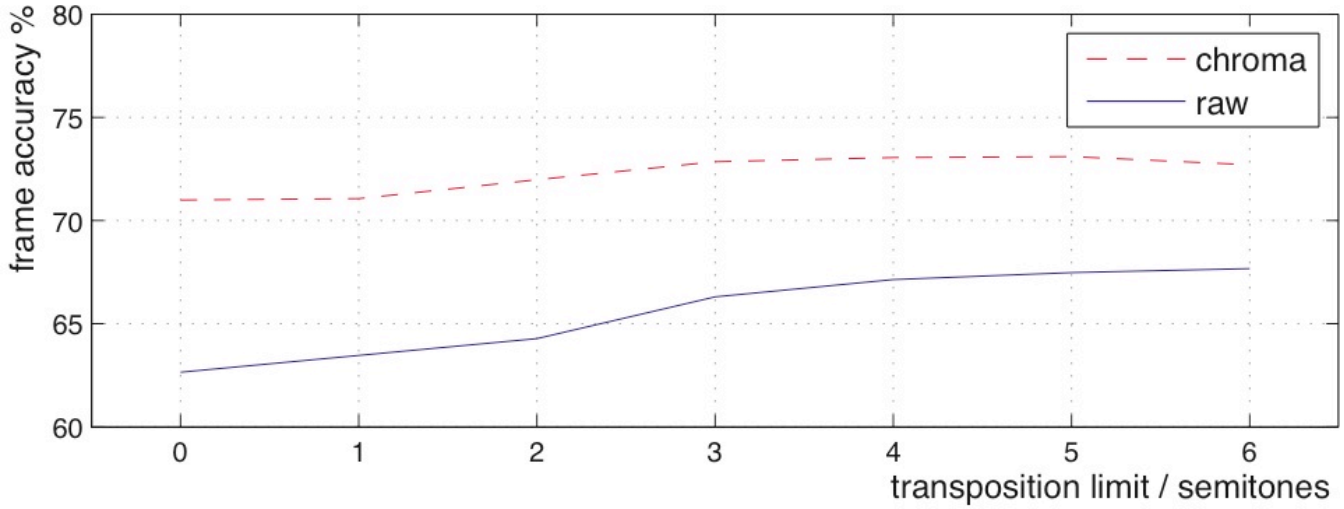
- Greatest influence - **more data**



Resampled Audio

Reduce cost of data

- Resample labeled audio ±1…6 semitones
- Classification accuracy increases over 5 %



Columbia University

Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

MIREX RESULTS

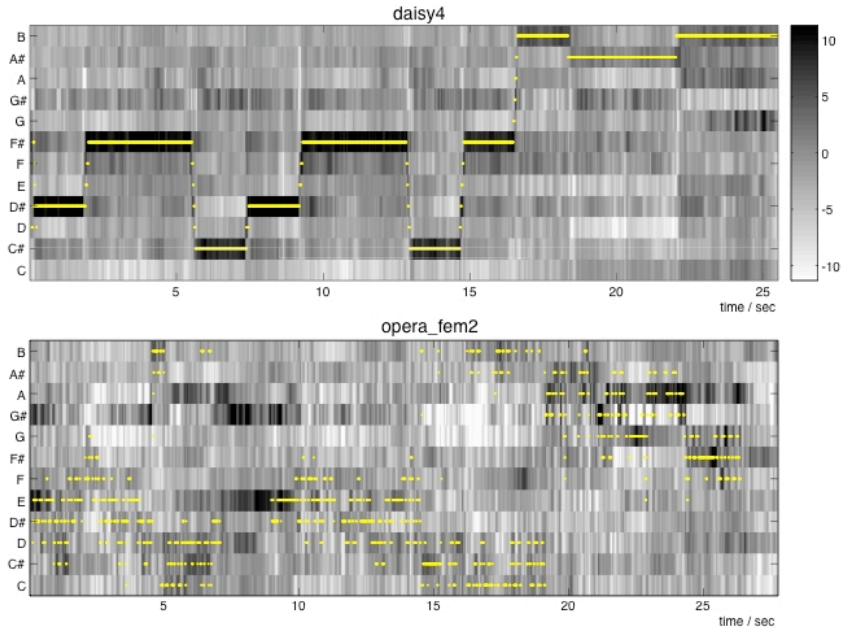
Excellent candidate identification

Rank	Participant	Overall Accuracy	Voicing d'	Raw Pitch	Raw Chroma	Runtime / s
1	Dressler	**71.4%**	**1.85**	68.1%	71.4%	32
2	Ryynänen	64.3%	1.56	**68.6%**	**74.1%**	10970
3	Paiva 2	61.1%	1.22	58.5%	62.0%	45618
3	Poliner	61.1%	1.56	67.3%	73.4%	5471
5	Marolt	59.5%	1.06	60.1%	67.1%	12461
6	Paiva 1	57.8%	0.83	62.7%	66.7%	44312
7	Goto	49.9%*	0.59*	65.8%	71.8%	211
8	Vincent 1	47.9%*	0.23*	59.8%	67.6%	?
9	Vincent 2	46.4%*	0.86*	59.6%	71.1%	251
10	Brossier	3.2%* †	0.14 * †	3.9% †	8.1% †	41

Classification Posteriors

Exploit temporal structure of music

- N Binary classifiers
  - (one vs. all classification)
- Fit logistic models to distance from class boundary
- Pseudo-posteriors



Conclusion

- Simple implementation
- Low training data requirements
- Competitive transcription method
- Excellent potential for future work

http://labrosa.ee.columbia.edu/projects/melody/