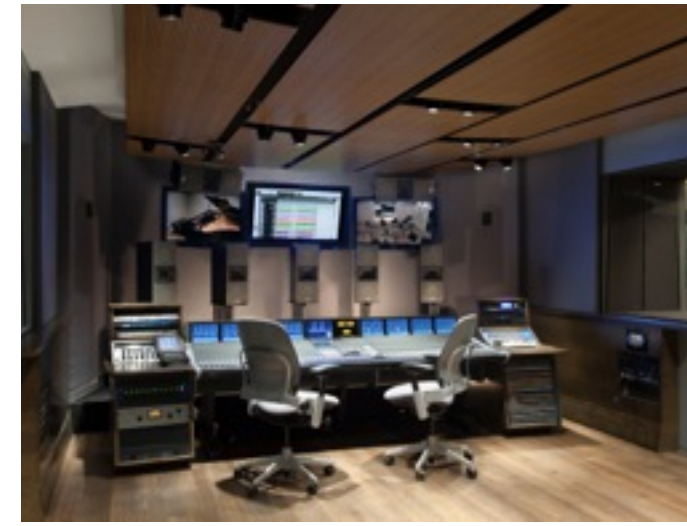
Music Informatics @ NYU

Music and Audio Research Laboratory (MARL)
New York University

Eric J. Humphrey
25 January, 2014

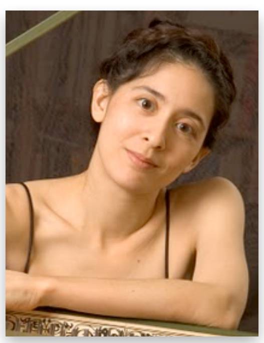


MARL: founded in late 2008, moved to new facilities in 2009
14+ researchers, Funded by NSF, IMLS, NYU

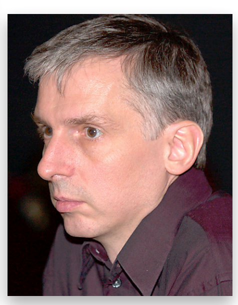
http://marl.smusic.nyu.edu

MARL: Areas of Interest

Immersive Audio (A. Roginska)



Music Cognition (M. Farbood and P. Mavromatis)



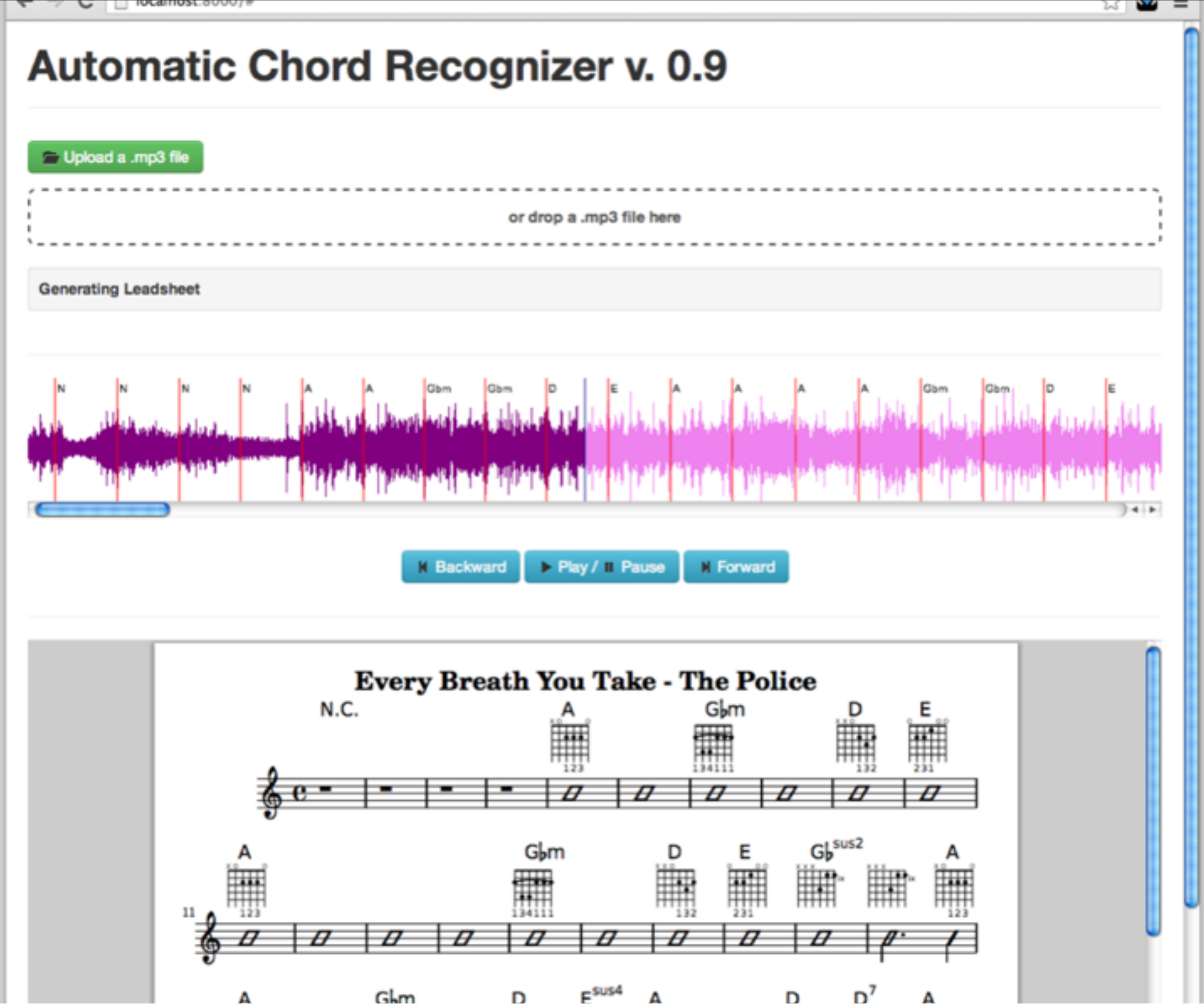
Computer Music (T.H. Park and R. Rowe)

Music Informatics (J.P. Bello)

http://marl.smusic.nyu.edu

MARL – PhDs and Post-Doc*

http://marl.smusic.nyu.edu



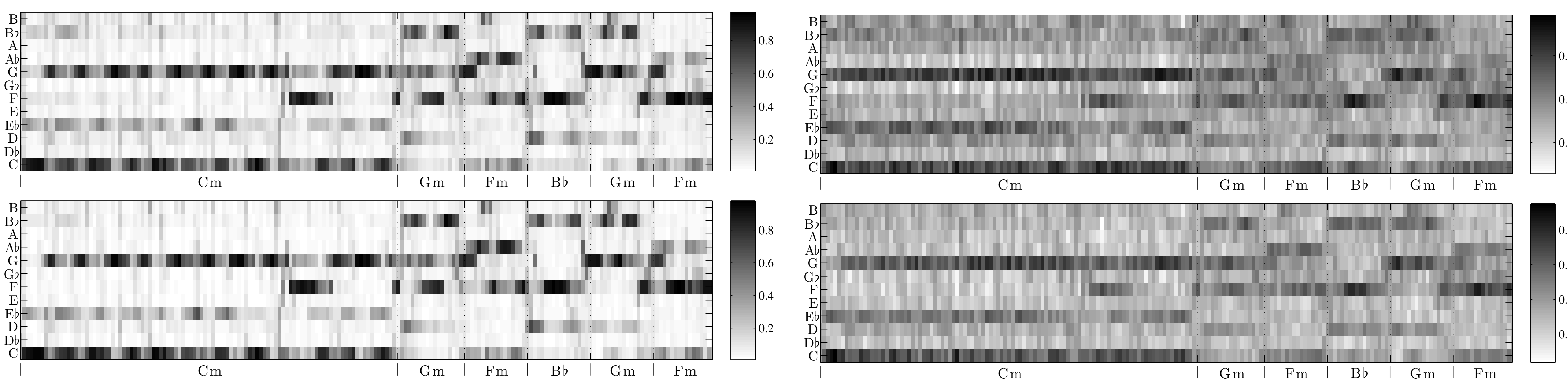
Content-Based MIR

Chord Recognition
Deep Feature Learning
Rhythmic Similarity
Melody Extraction
Pattern Discovery & Segmentation

Chord Recognition

- Feature variations have a considerable impact (~10%)

	C			C^W			C^W_{Log}		
filter	BT	$G1$	$G25$	BT	$G1$	$G25$	BT	$G1$	$G25$
N/A	47.0	46.5	48.8	52.1	49.4	51.7	55.5	58.3	57.7
avg / med-filter + Viterbi	66.7	66.1	72.0	71.1	68.7	74.6	73.1	75.6	77.6
Beat-sync + Viterbi	64.4	61.5	67.5	67.9	67.9	73.4	72.7	76.7	77.5



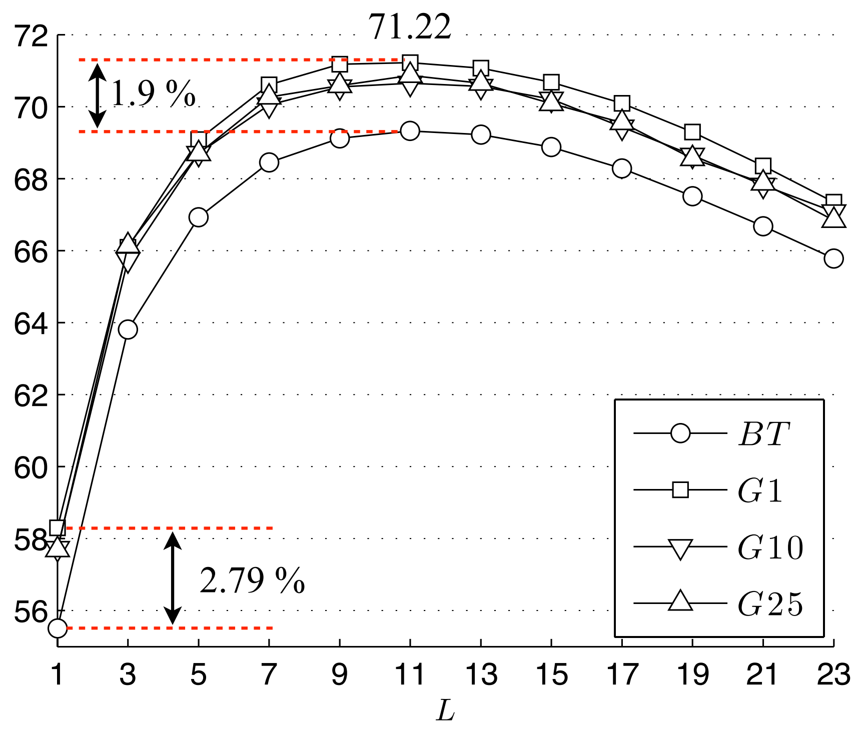
[Cho, T. // Bello, J.P.]

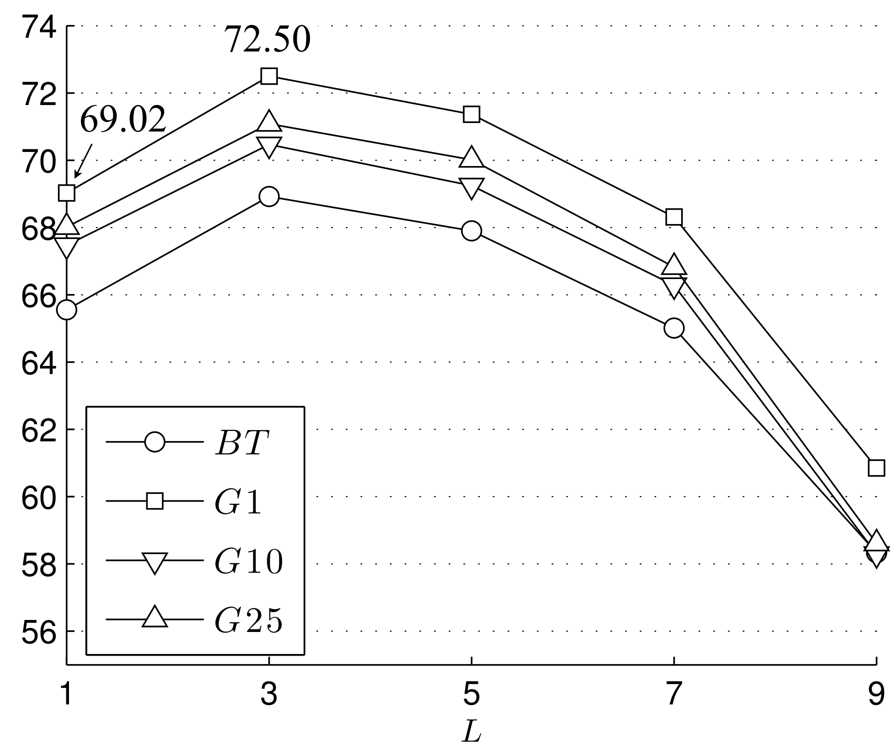
http://marl.smusic.nyu.edu

Chord Recognition

- Feature filtering has a huge impact (~20%)

	C			C^W			C^W_{Log}		
filter	BT	$G1$	$G25$	BT	$G1$	$G25$	BT	$G1$	$G25$
N/A	47.0	46.5	48.8	52.1	49.4	51.7	55.5	58.3	57.7
avg / med-filter + Viterbi	66.7	66.1	72.0	71.1	68.7	74.6	73.1	75.6	77.6
Beat-sync + Viterbi	64.4	61.5	67.5	67.9	67.9	73.4	72.7	76.7	77.5





[Cho, T. // Bello, J.P.]

http://marl.smusic.nyu.edu

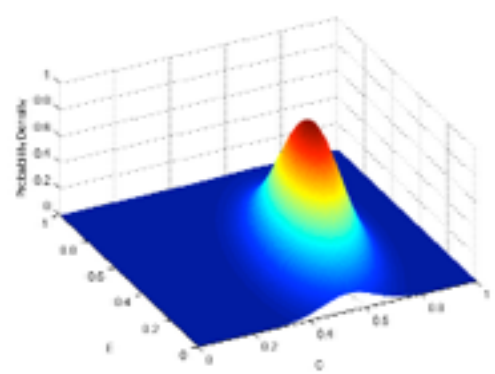
Chord Recognition

- Complexity of models has a modest impact (~5%)

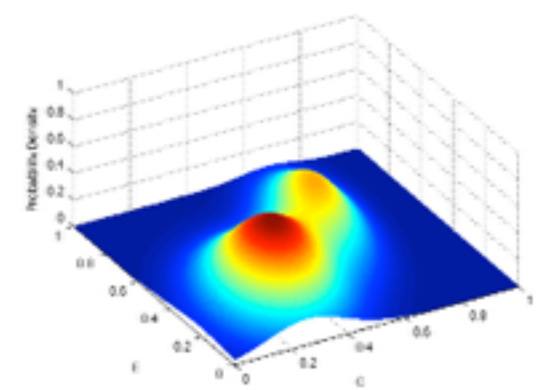
	C			C^W			C^W_{Log}		
filter	BT	$G1$	$G25$	BT	$G1$	$G25$	BT	$G1$	$G25$
N/A	47.0	46.5	48.8	52.1	49.4	51.7	55.5	58.3	57.7
avg / med-filter + Viterbi	66.7	66.1	72.0	71.1	68.7	74.6	73.1	75.6	77.6
Beat-sync + Viterbi	64.4	61.5	67.5	67.9	67.9	73.4	72.7	76.7	77.5

Binary Template (BT)

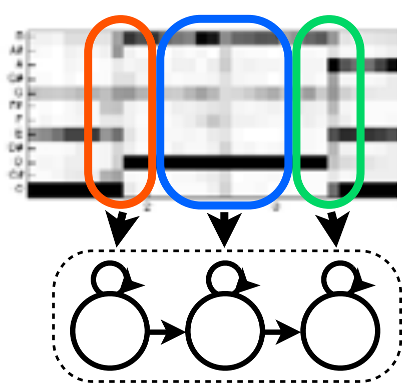
Single Gaussian (G1)



Mixtures of Gaussians (GN)



Networks of HMMs

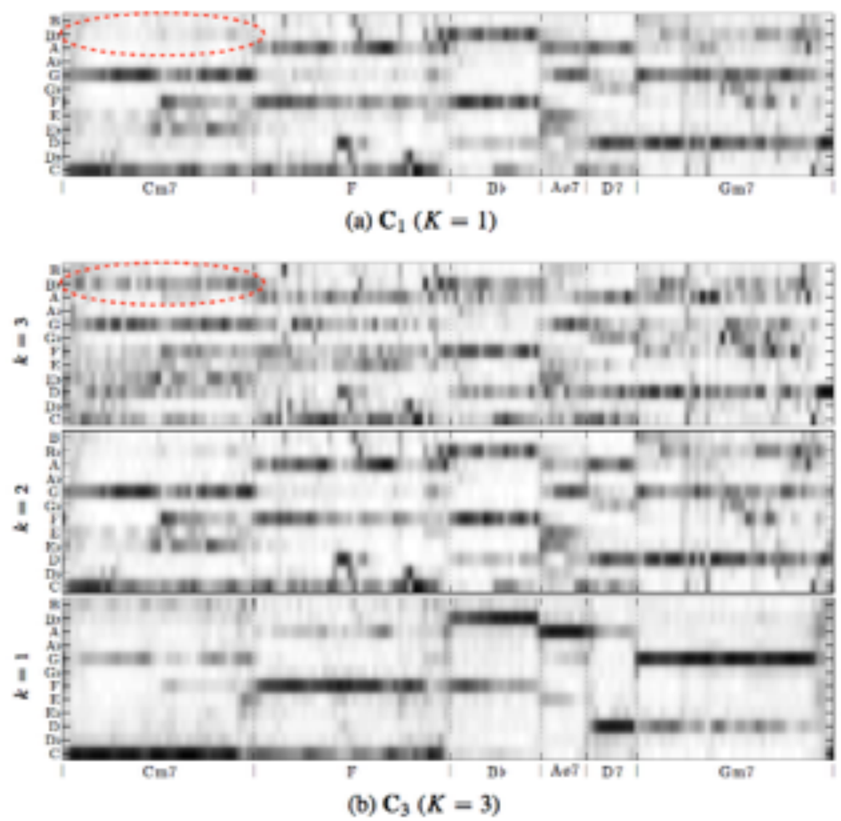


[Cho, T. // Bello, J.P.]

http://marl.smusic.nyu.edu

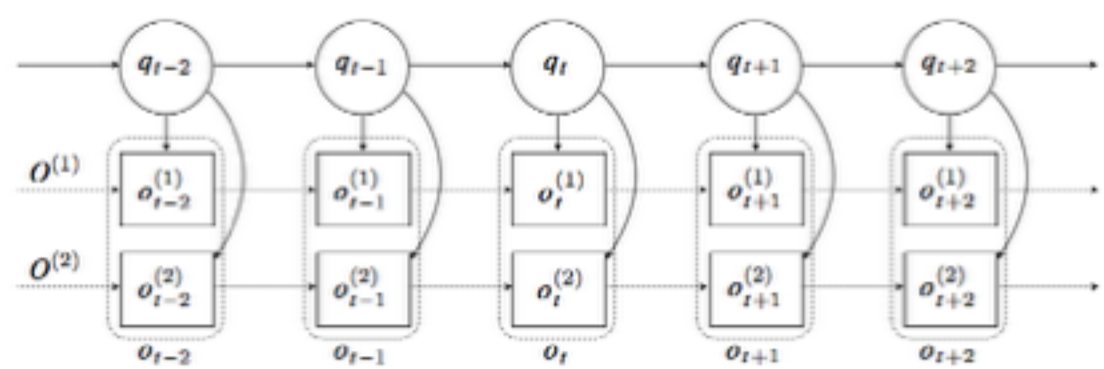
Large Vocabulary Chord Recognition

- Chroma may be insufficient to discriminate complex chord types:
  - Subband (K) chroma features
  - K-stream Hidden Markov Model



Subband Chroma

	# of chords	description
Lexicon 1	61 chords	(maj, min, maj7, min7 and 7) × 12 keys plus no-chord
Lexicon 2	157 chords	(maj, min, maj7, min7, 7, maj6, min6, dim, aug, sus4, sus2, hdim7 and dim7) × 12 keys plus no-chord



K-Stream HMM

[Cho, T. // Bello, J.P.]

Large Vocabulary Chord Recognition

- Performance is much lower than in the classic MIREX formulation (24M/m+N)

(a) Frame-based features with $G5$ and A_F

	Lexicon 1				Lexicon 2			
	Recognition rate		Avr. ind. chord		Recognition rate		Avr. ind. chord	
K	$\mathbf{C}_K$	K-stream	$\mathbf{C}_K$	K-stream	$\mathbf{C}_K$	K-stream	$\mathbf{C}_K$	K-stream
1	60.78		63.31		57.50		41.83	
4	62.11	63.72	63.75	65.59	60.97	62.65	39.46	43.75

(b) Beat-synchronous features with $G5$ and A_B

	Lexicon 1				Lexicon 2			
	Recognition rate		Avr. ind. chord		Recognition rate		Avr. ind. chord	
K	$\mathbf{C}_K$	K-stream	$\mathbf{C}_K$	K-stream	$\mathbf{C}_K$	K-stream	$\mathbf{C}_K$	K-stream
1	62.14		62.29		60.21		39.95	
4	63.69	65.30	61.74	63.87	62.85	65.24	35.58	39.19

[Cho, T. // Bello, J.P.]

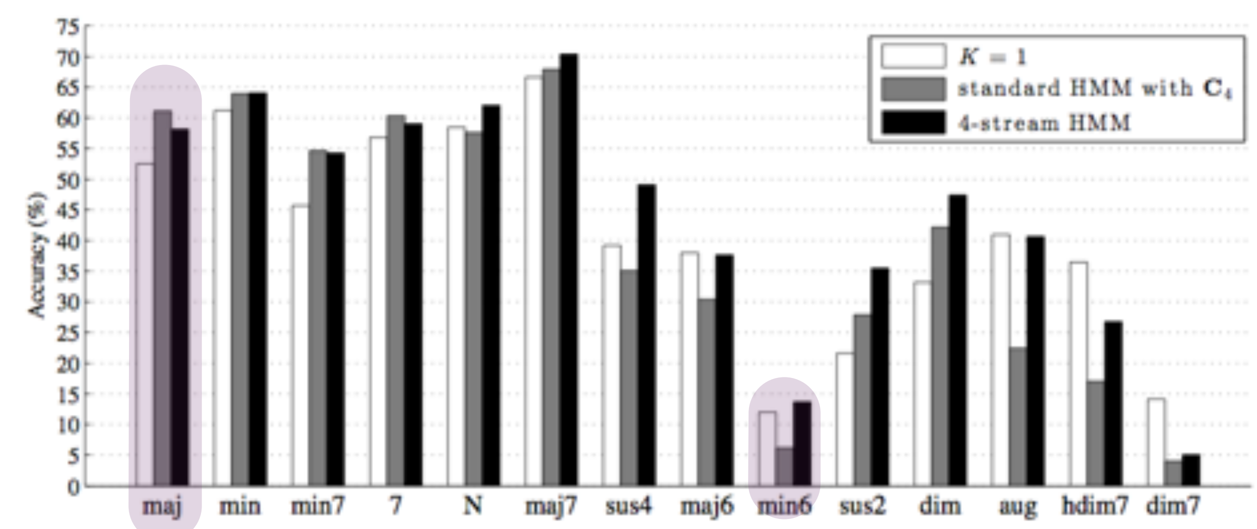
Large Vocabulary Chord Recognition

- Different metrics tell different stories
  - Framewise Recognition Rate (FWRR)
  - Average Chord Quality Accuracy (ACQA)

(a) Frame-based features

	maj	min	min7	7	N	maj7	
	511848	138845	75927	70894	41223	32458	
sus4	**maj6**	**min6**	**sus2**	**dim**	**aug**	**hdim7**	**dim7**
13909	10832	4826	4653	3214	2863	1715	1432

>100X more Major than minor6!



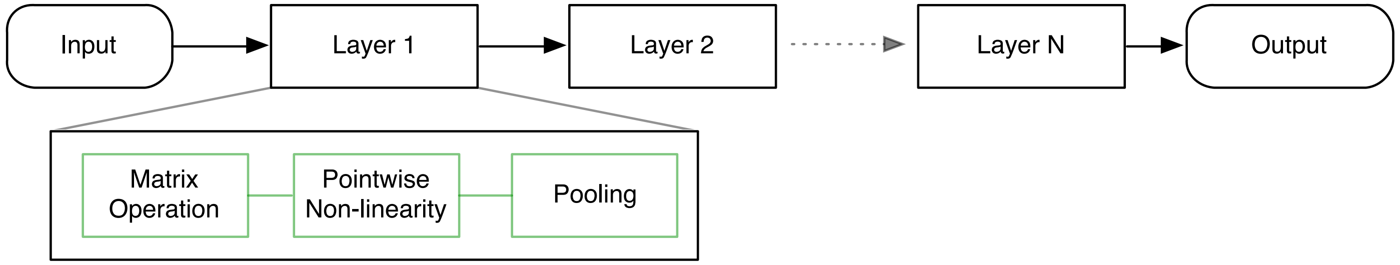
(a) Frame-based

[Cho, T. // Bello, J.P.]

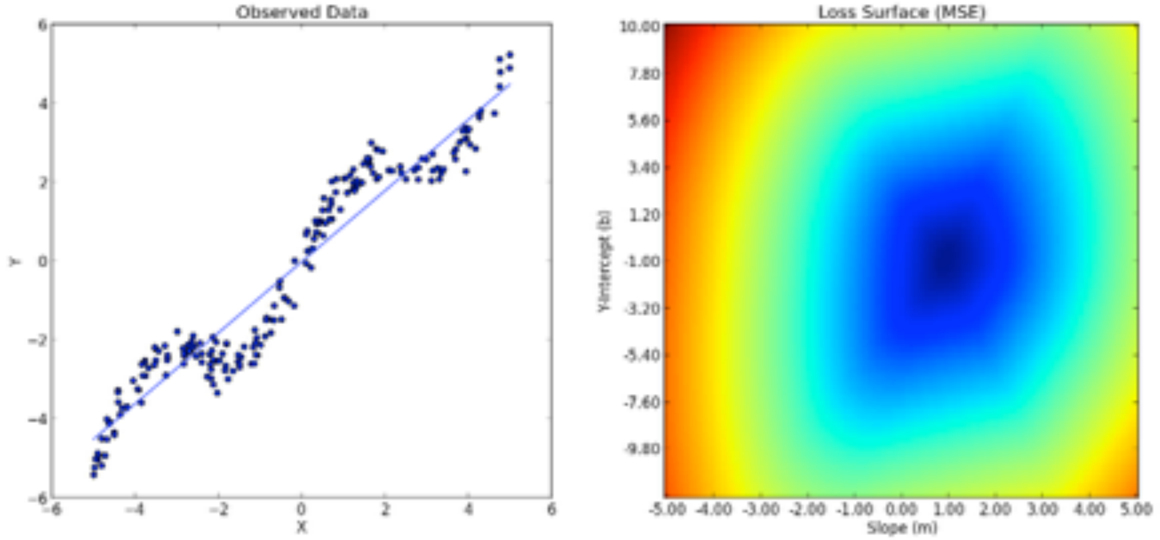
http://marl.smusic.nyu.edu

Deep Learning - A Slightly Different Approach to Design

- Cascade of multiple layers, composed of a few simple operations
  - Linear algebra
  - Point-wise nonlinearities
  - Pooling



- Learning leverages numerical methods to *find* good parameters.

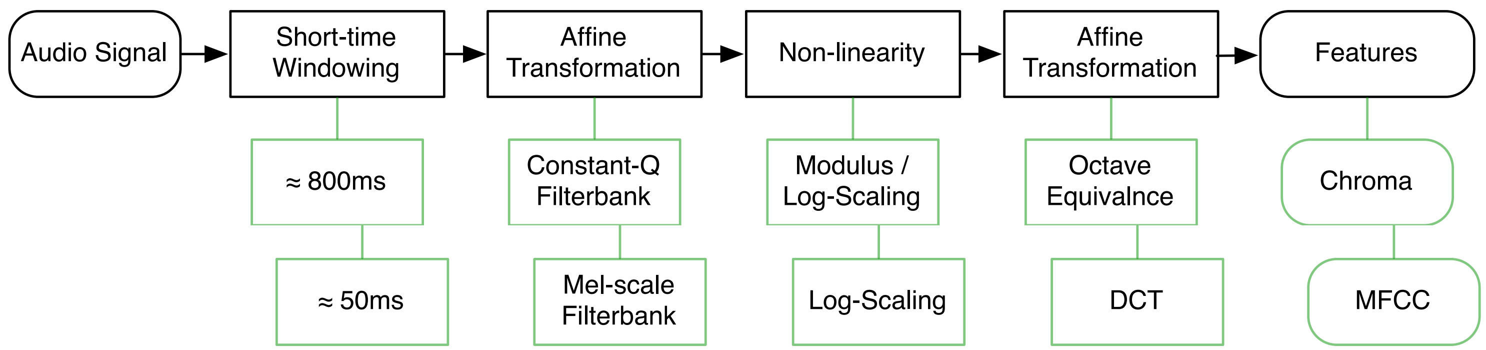


[Humphrey, E.J. // Bello, J.P.]

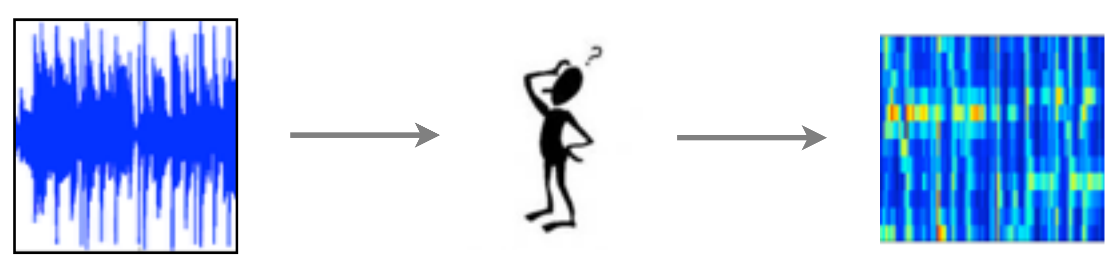
http://marl.smusic.nyu.edu

Deep Learning - A Slightly Different Approach to Design

- The pieces of deep learning are everywhere in feature design:



- What makes feature design so challenging?
  - You have to know what you want
  - You have to know how to do it



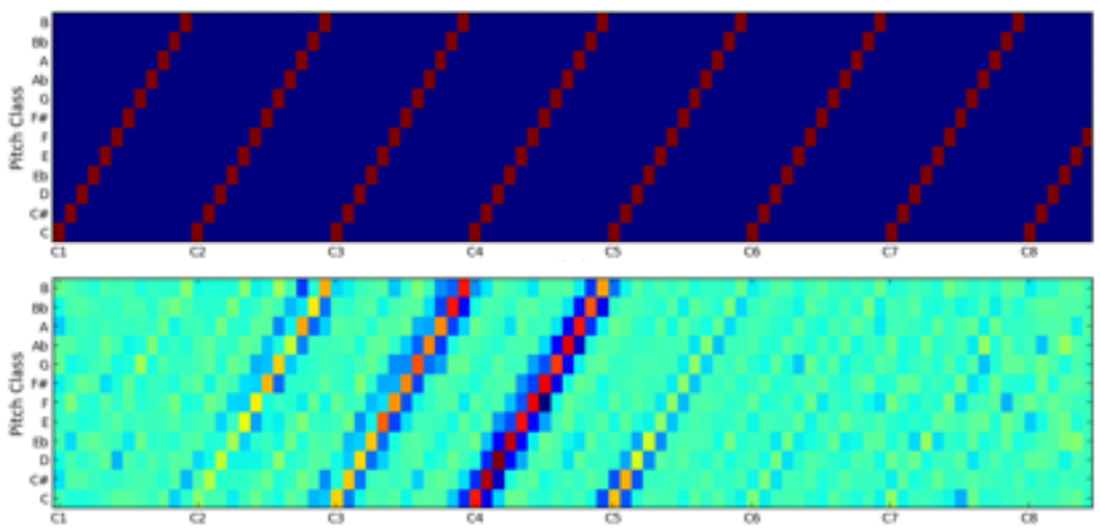
audio function chroma

[Humphrey, E.J. // Bello, J.P.]

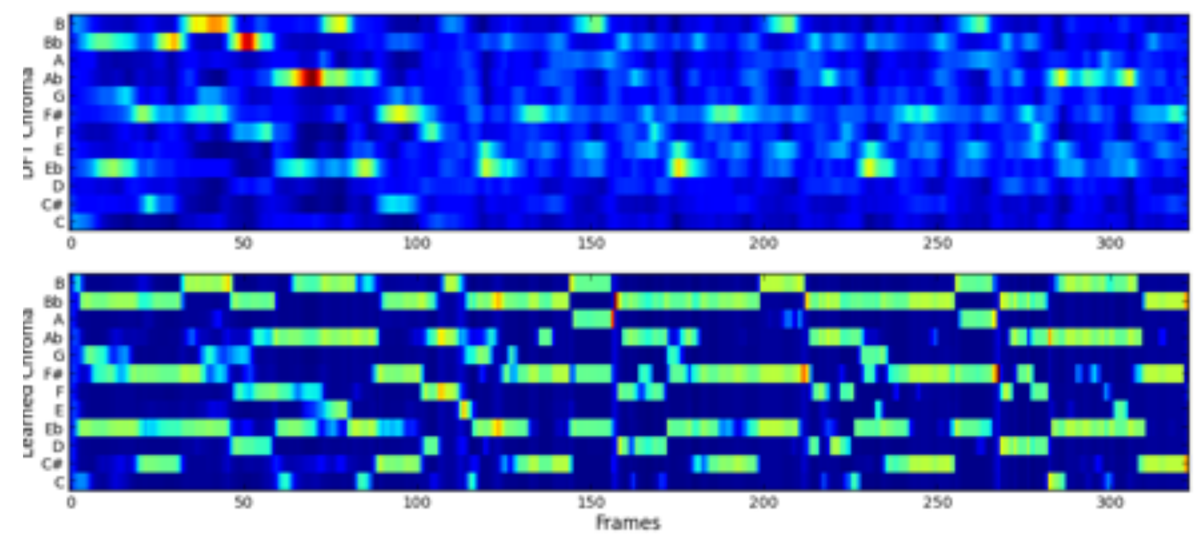
Learning Chroma Features

- Defined versus Learned Features

CQT-to-Chroma Weights



Chroma Features



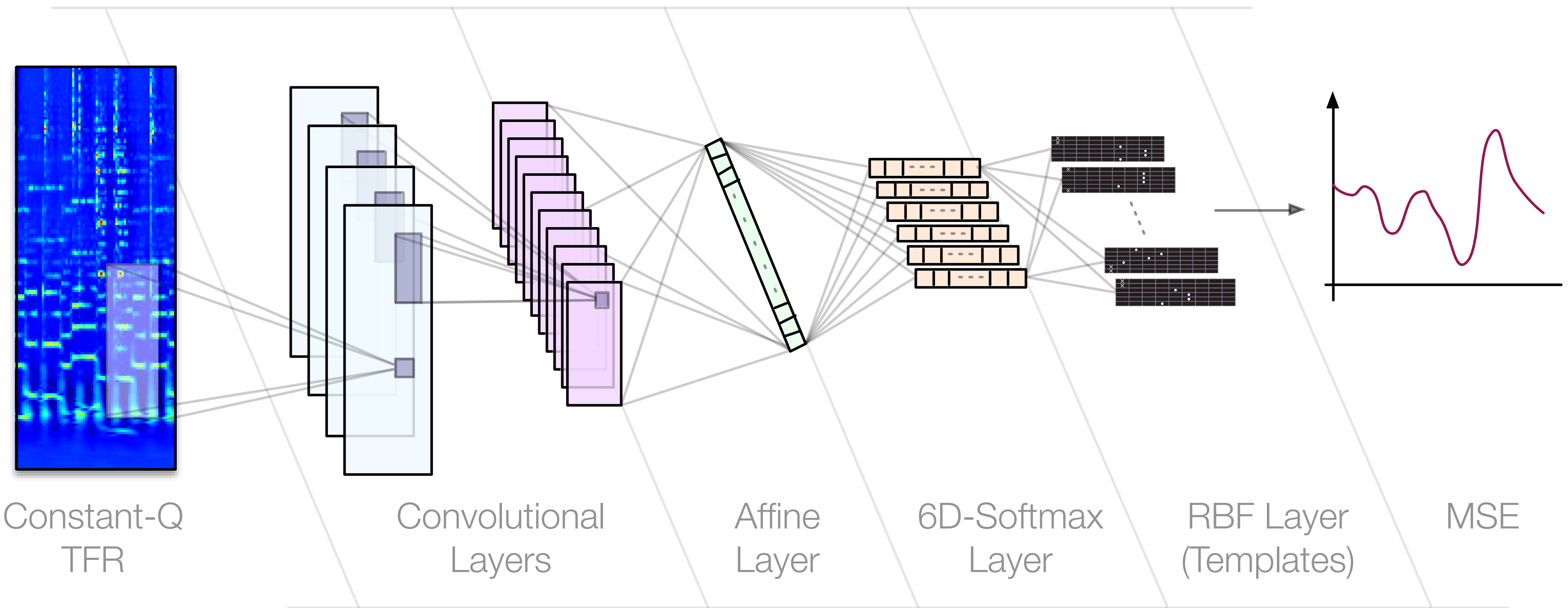
- NB: Tutorial for this is on the MARL website
    - Full Python code + data
    - Could be fun for HAMR time!

[Humphrey, E.J. // Bello, J.P.]

http://marl.smusic.nyu.edu

Learning Human-Readable Representations

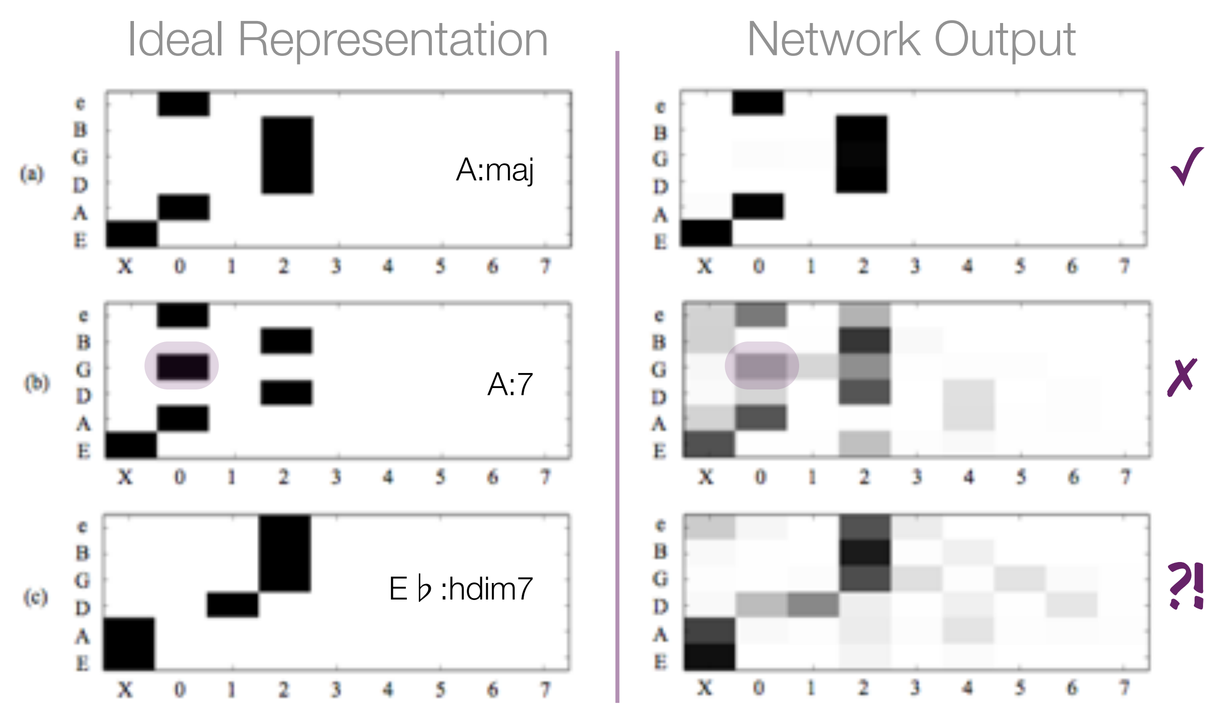
- Can we use chord annotations to directly learn guitar tablature from audio?



[Humphrey, E.J. // Bello, J.P.]

Learning Human-Readable Representations

- Trades slight drop in performance for some notable benefits:
  - representations are directly interpretable by guitarists
  - facilitates large-scale data collection / error correction
  - reduces the degree of time / effort necessary to provide ground truth annotations
  - can generalize to never-before seen chords



	maj	min	maj7	min7	7	N
UC	69.58	57.24	62.08	55.38	49.60	78.21
G	69.52	55.79	63.18	55.52	46.29	77.85

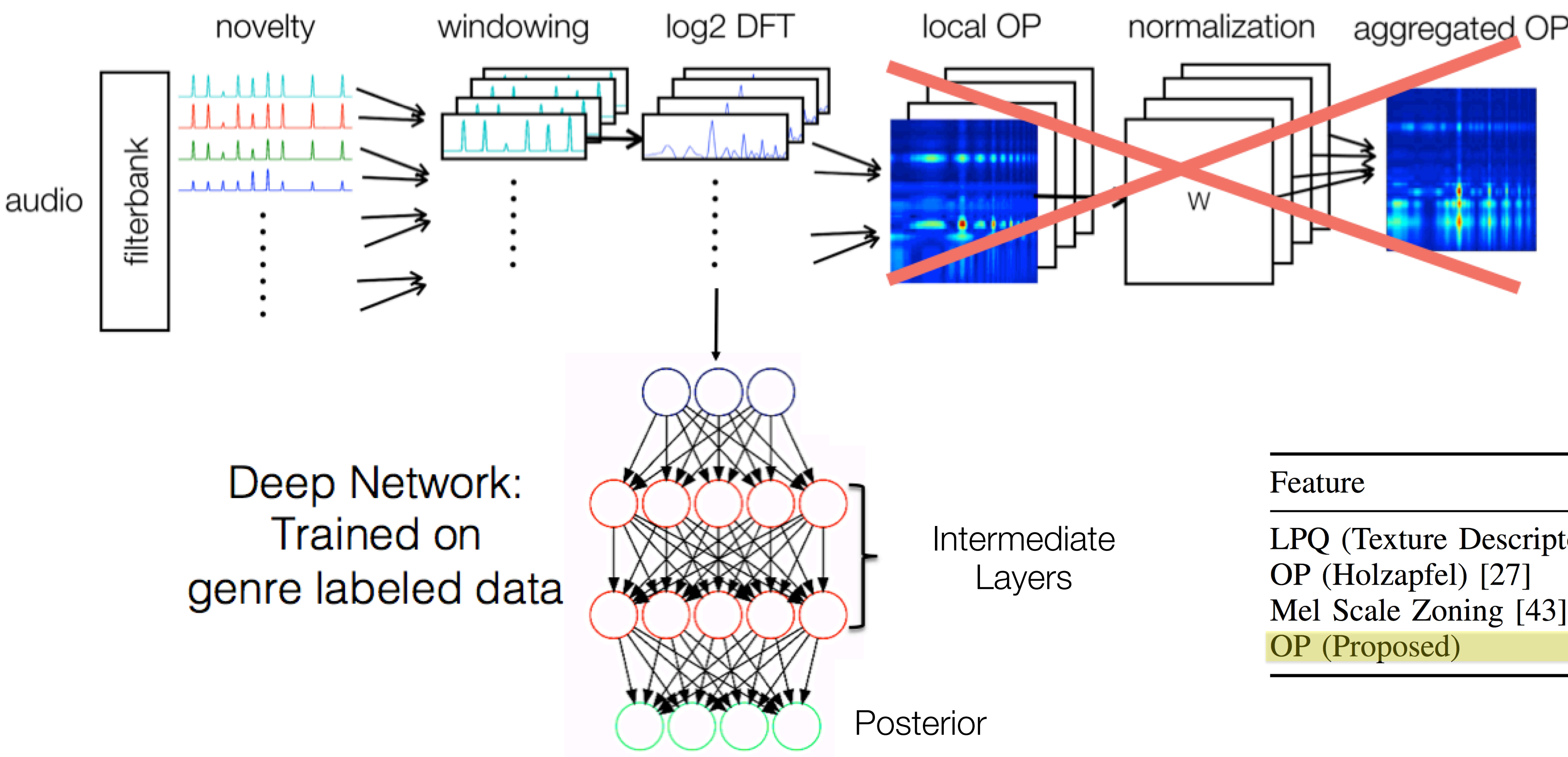
	FWRR	ACQA
UC	58.72	62.02
G	58.26	61.36

[Humphrey, E.J. // Bello, J.P.]

http://marl.smusic.nyu.edu

From Genre Classification to Rhythmic Similarity

- Leverage feature learning to optimize onset patterns

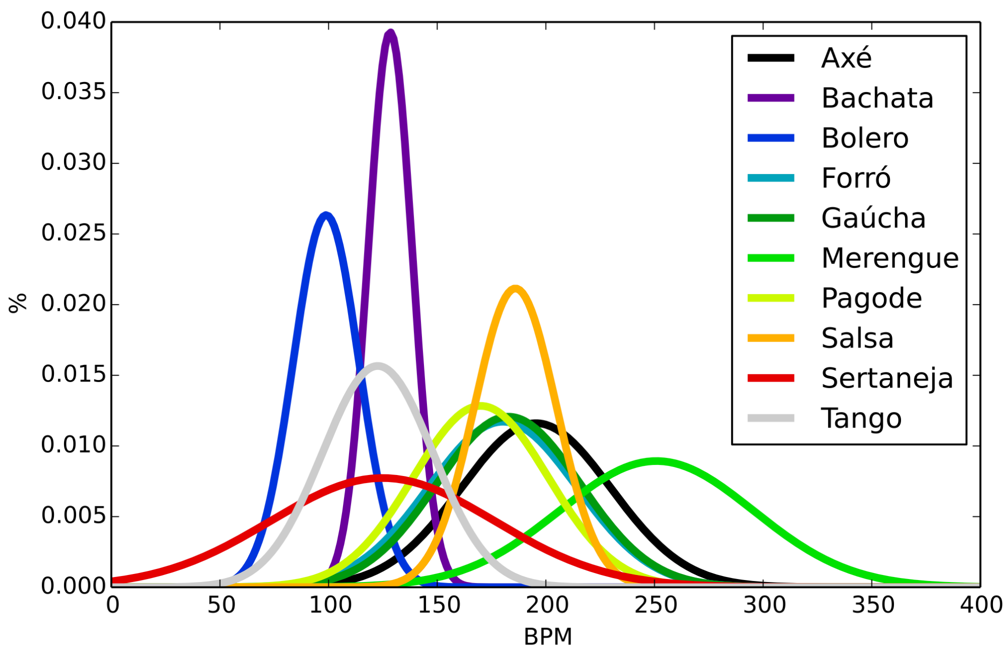


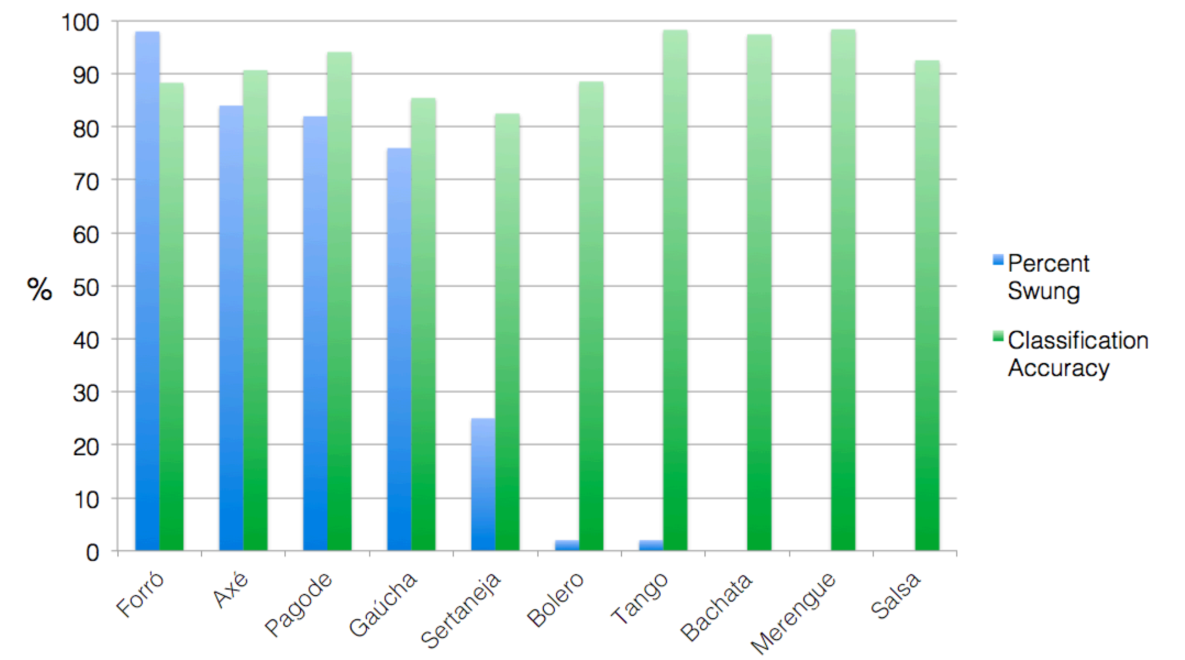
Feature	Accuracy (%)
LPQ (Texture Descriptors) [42]	80.78
OP (Holzapfel) [27]	81.80
Mel Scale Zoning [43]	82.33
OP (Proposed)	91.32

[Esparza, T. // Bello, J.P. // Humphrey, E.J.]

From Genre Classification to Rhythmic Similarity

- Approach demonstrates sensitivity to certain rhythmic nuances
  - Tempo dependence shows no significant effect
  - Fine-grained changes (swung rhythms) affect classification accuracy



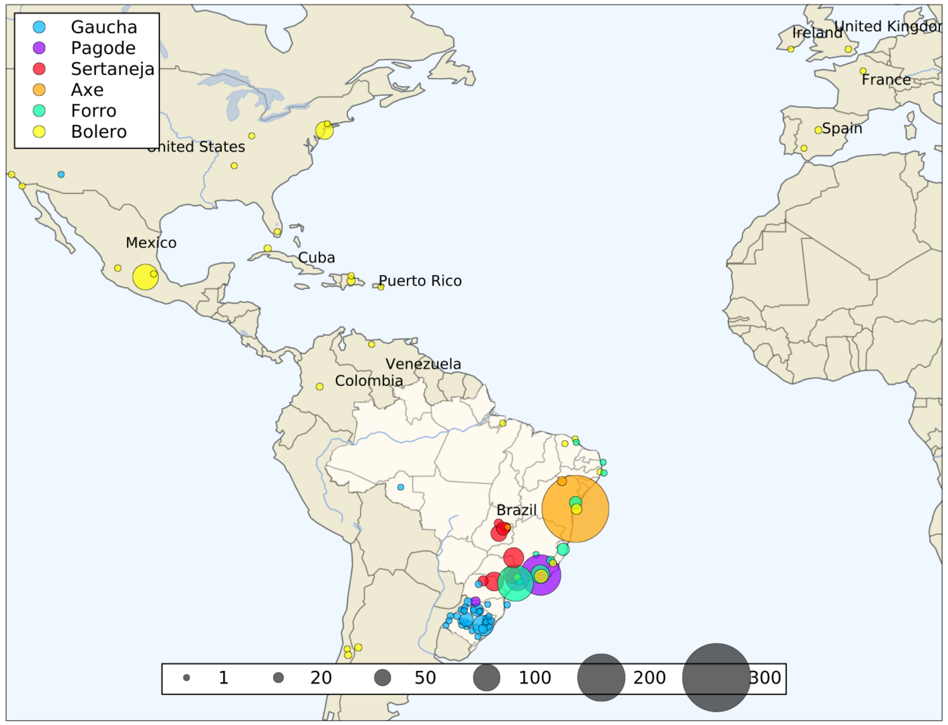


[Esparza, T. // Bello, J.P. // Humphrey, E.J.]

http://marl.smusic.nyu.edu

From Genre Classification to Rhythmic Similarity

- Nuances of the Latin Music Dataset undermine rhythmic similarity evaluation
  - Annotation: Use trumps content
  - Selection: Brazilian bias skews discrimination
  - Unintended correlations: Tango exhibits unique signal-level qualities (bandwidth)
- Genre is a poor proxy for rhythmic similarity
  - Sertaneja is better defined by lyrical themes
  - Global pop influence flattens rhythm content



[Esparza, T. // Bello, J.P. // Humphrey, E.J.]

http://marl.smusic.nyu.edu

Melody Extraction from Polyphonic Audio

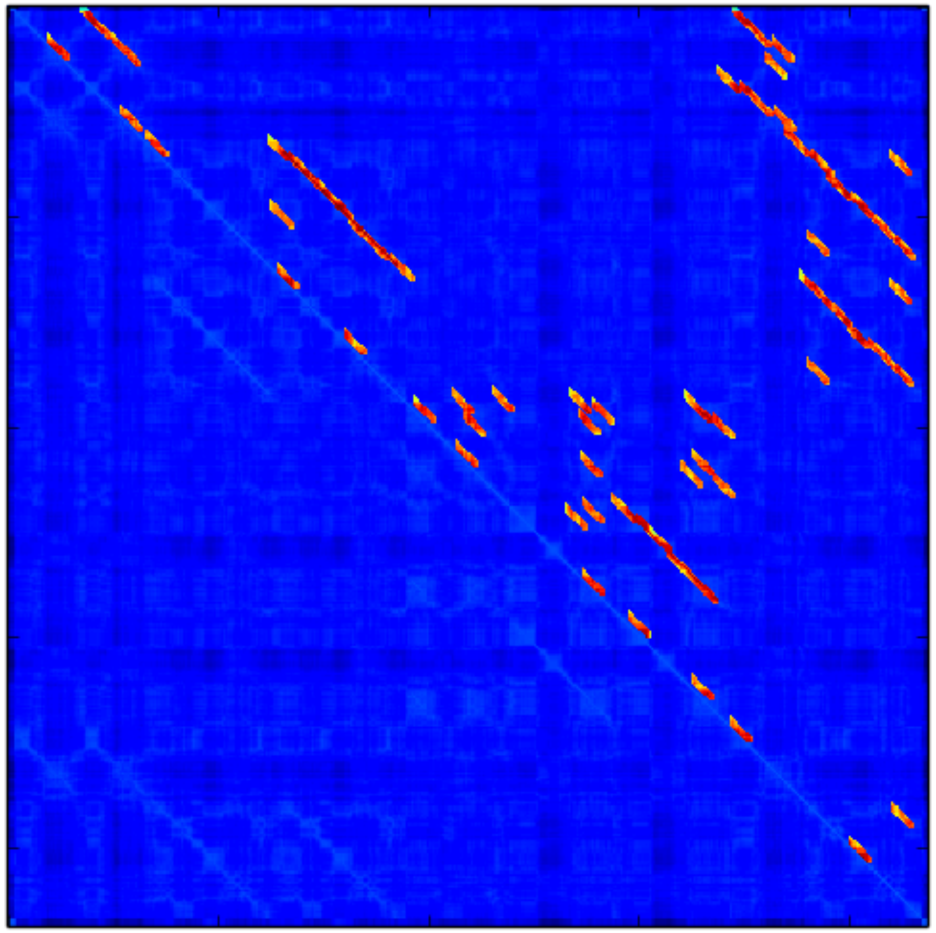
- We're curating a dataset!
- Goals:
  - A few hundred full-length pieces
  - Annotations:
    - Predominant f0
    - Time-aligned Instruments / Sources
    - Genre
- Developing tools for monophonic f0 annotation (collaboration w/C4DM)
- Targeting a May / ISMIR release
- Let us know if you'd like to help!

[Bittner, R. // Salamon, J. // Bello, J.P.]

Pattern Discovery via Segmentation Methods

- Motives are short melodic/harmonic ideas that occur at least twice in a piece
- Idea: Use tools from music segmentation to discover these patterns
- Approach:
  - key-invariant self-similarity matrix (SSM)
  - novel path finding algorithm
- Works both on symbolic and audio representations
- Best MIREX results using audio as input
  - (Also worst results :-D)

Patterns found in key-invariant SSM of Beethoven Op. 2 No.1



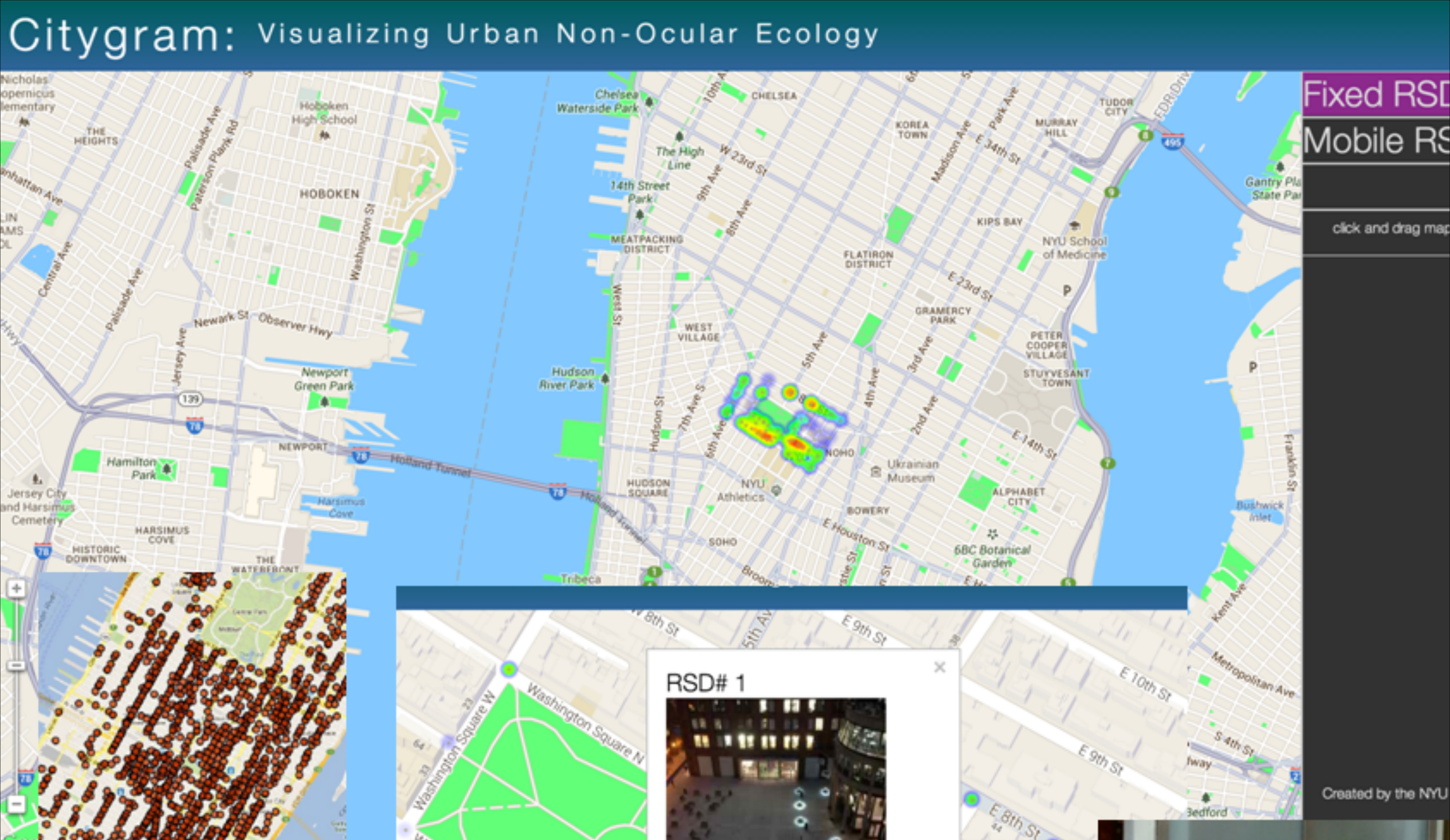
[Nieto, O. // Farbood, M.]

Perceptually-Based Evaluation of Music Boundaries

- Goal: Explore the relevance of the Precision and Recall values when evaluating the boundaries of music segmentation algorithms
- Method: Three experiments where subjects rate the quality of various boundaries.
- Take-aways:
  - Precision is more perceptually relevant than Recall
  - Proposed an F_α measure instead of F_1 score (with $\alpha < 1$)

$$F_\alpha = (1 + \alpha^2)\frac{P \cdot R}{\alpha^2 P + R}$$

[Nieto, O. // Jehan, T. // Farbood, M. // Bello, J.P.]

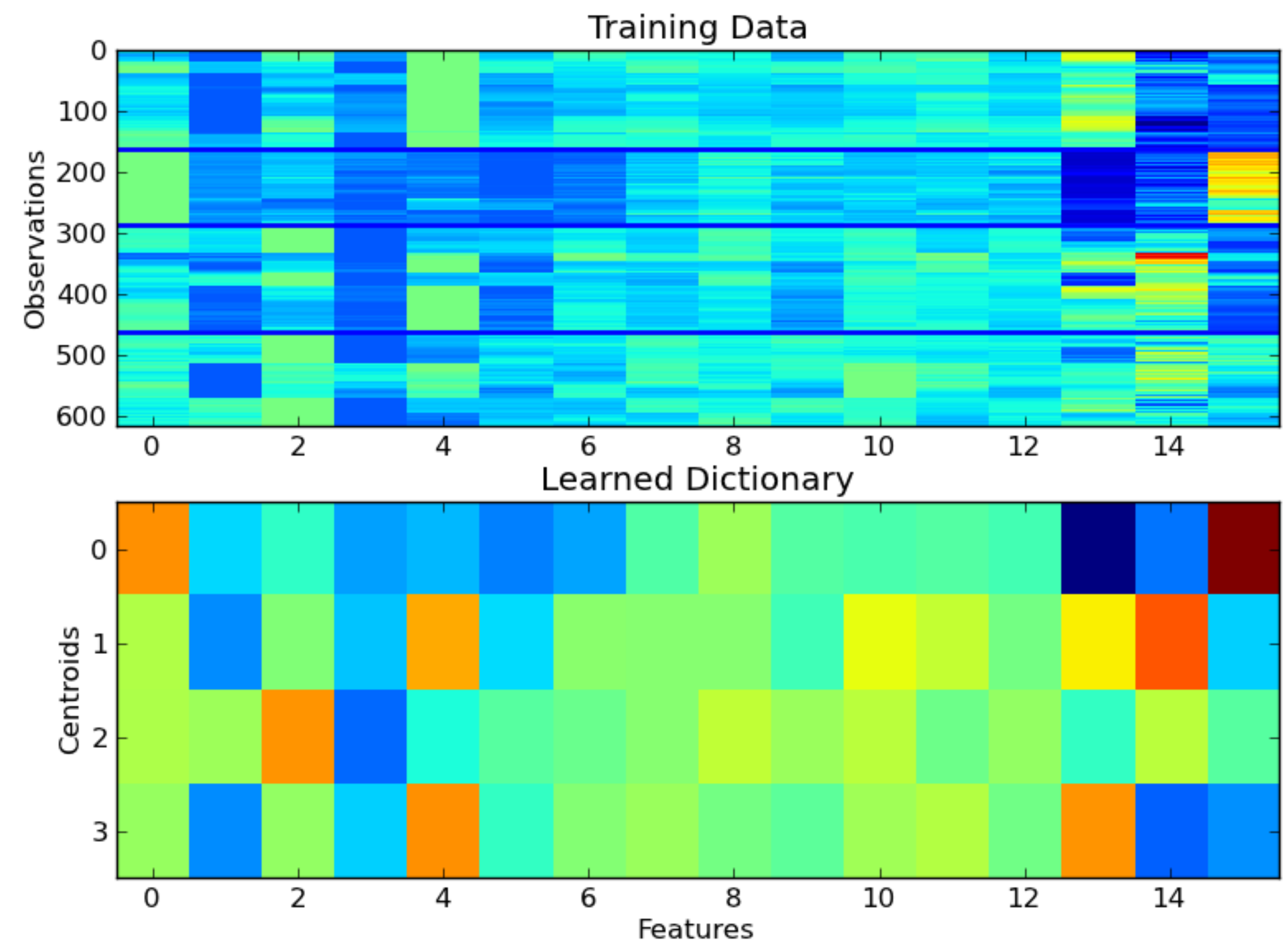


Non-Music Audio Research

Extreme Vocal Effects
Citygram One
Acoustic Ecology

Extreme Vocal Effects

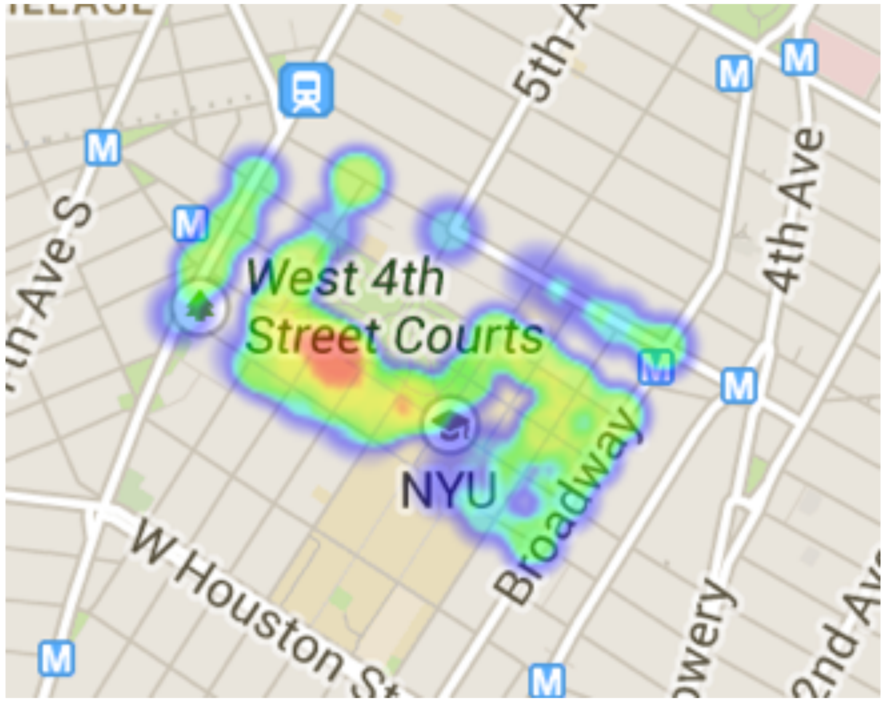
- Automatic classification of EVEs
- EVE Types:
  - Growl
  - Fry Scream
  - Roughness
- Features:
  - MFCC
  - Spectral Contrast
  - Zero Crossings in TD
  - Loudness (RMS)
  - K-means



[Nieto, O.]

Citygram One - Mapping Acoustic Ecology

- Mapping non-ocular spatio-acoustic energy
    - Dynamic, quasi-real-time sound maps
    - Publicly accessible and as open as possible
    - Exploration portal for the public, artists, policy-makers, and researchers
- Soundmaps are valuable, but non-existent
    - Invisible energies such as sound underrepresented
    - Accurately quantify and measure “noise pollution”
    - Richer representation of urban landscapes

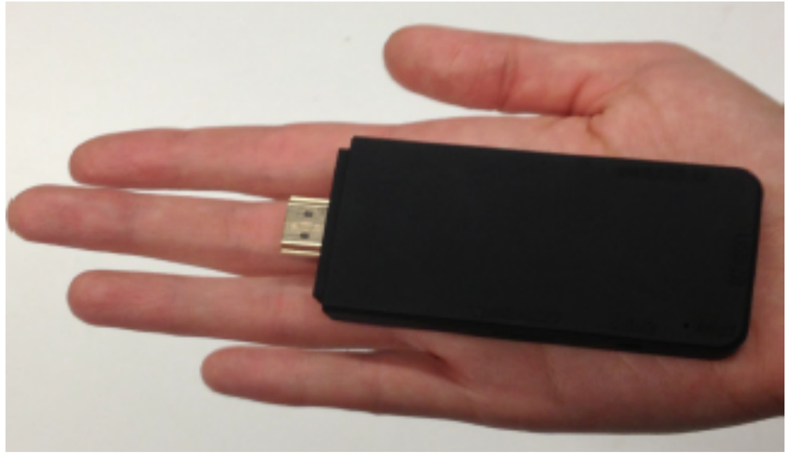


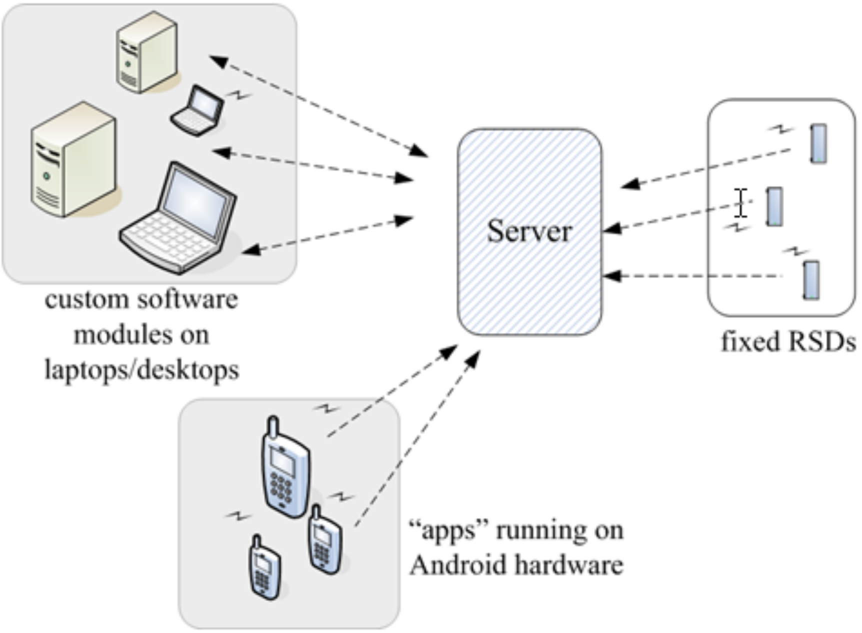
[Jacoby, C. // Salamon, J. // Park, T.H. (PI)]

http://marl.smusic.nyu.edu

Citygram One - Mapping Acoustic Ecology

- Goal: Create and deploy a cyber-physical system
  - Acquisition – build/deploy remote sensor network
  - Analysis – content-based + context-based
  - Visualization – map overlays, multiple features
  - Citizen science – sound recording / annotation



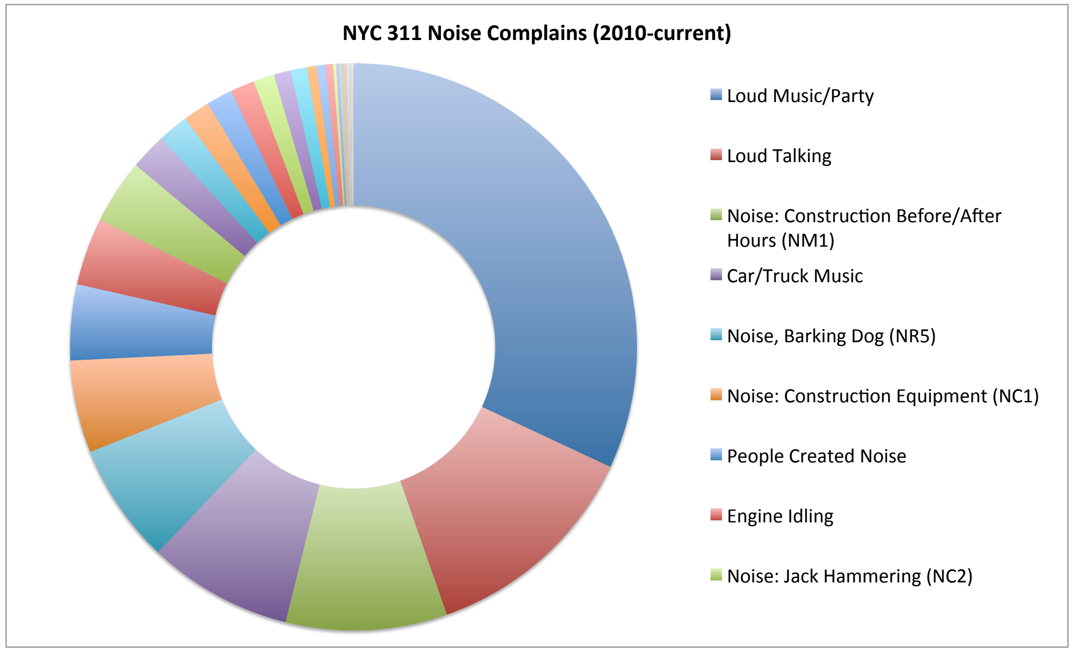


[Jacoby, C. // Salamon, J. // Park, T.H. (PI)]

http://marl.smusic.nyu.edu

Urban Auditory Scene Analysis

- Phase I: Source ID (siren, jackhammer, gunshot…)
  - Curate dataset (annotated urban sound collections are scarce!)
  - Train/test ML algorithms for source ID
- Phase 2: Content + Context
  - Explore relation with other sources of city data (311 noise complaints, crime stats, etc.)



[Salamon, J. c/o CUSP]

Acoustic Ecology

MARINEXPLORE WHALE DETECTION

The Marinexplore and Cornell University Whale Detection Challenge

Finished

Friday, February 8, 2013 $10,000 • 249 teams Monday, April 8, 2013

Dashboard

Public Leaderboard • **Private Leaderboard**

This competition has completed. This leaderboard reflects the final standings.

See someone using multiple accounts? Let us know.

#	Δ1w	Team Name * in the money	Score	Entries	Last Submission UTC (Best – Last Submission)
1	↑3	**SluiceBox** *	0.98384	70	Sun, 07 Apr 2013 18:58:34
2	↑3	**alfnie** *	0.98379	27	Sun, 07 Apr 2013 22:47:36 (-1.2h)
3	↑11	**RBM**	0.98226	32	Sun, 07 Apr 2013 23:22:16 (-1.3h)
4	↓3	**Free Willzyx**	0.98210	38	Sun, 07 Apr 2013 23:52:09 (-1.8h)
5	↓3	**Jure Zbontar**	0.98080	24	Mon, 01 Apr 2013 15:52:11 (-5.1h)

[Humphrey, E.J. // Cheung, B. (UC-Berkeley)]

Computer Music Systems

Automatic Accompaniment
AirJam
Audio Continuators
Interactive Performance

Automatic Musical Accompaniment

- Goal: Given a melody, automatically generate accompaniment
- Applications:
  - Automated composition tools
  - Automated real-time accompaniment
  - Algorithmic composition
- Given a melody note sequence find the most likely sequence of chords:

$$\hat{c} = \underset{m \in \Sigma^*}{\arg\max} \; \Pr\left[c \mid m\right] = \underset{m \in \Sigma^*}{\arg\max} \; \Pr\left[m \mid c\right] \Pr[c]$$

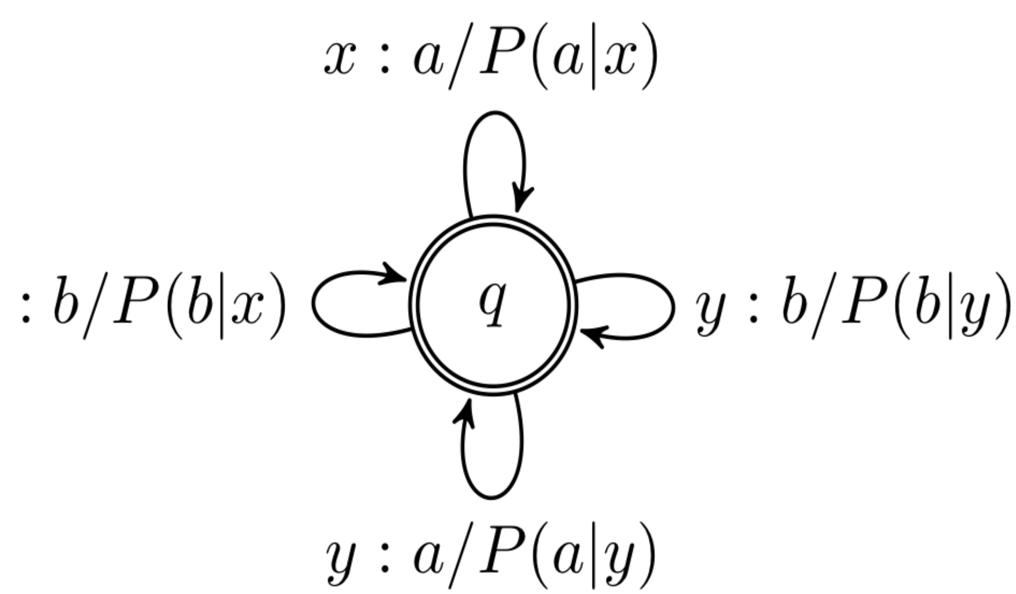
c: chord sequence
m: melody sequence
Σ*: set of all possible melody sequences

- Model Pr[c|m] and Pr[c] using separate finite state machines, which can then be combined

[Forsyth, J.]

Automatic Musical Accompaniment

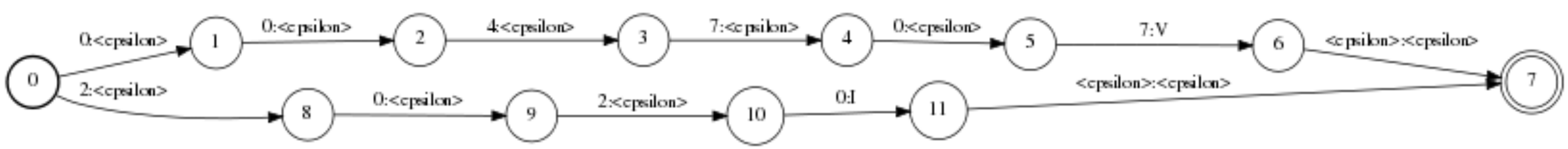
- Methodology:
    - train on Bach four-voice chorales (MIDI)
    - use different n-gram orders, chord quantization strategies, key normalization
    - evaluate using cross-fold validation
    - compute accuracy, average Euclidean distance between ground truth and generated sequences
    - key normalization, n-gram order improve performance



model Pr[c] using n-gram
model Pr[m|c] using one-state FST

Automatic Musical Accompaniment

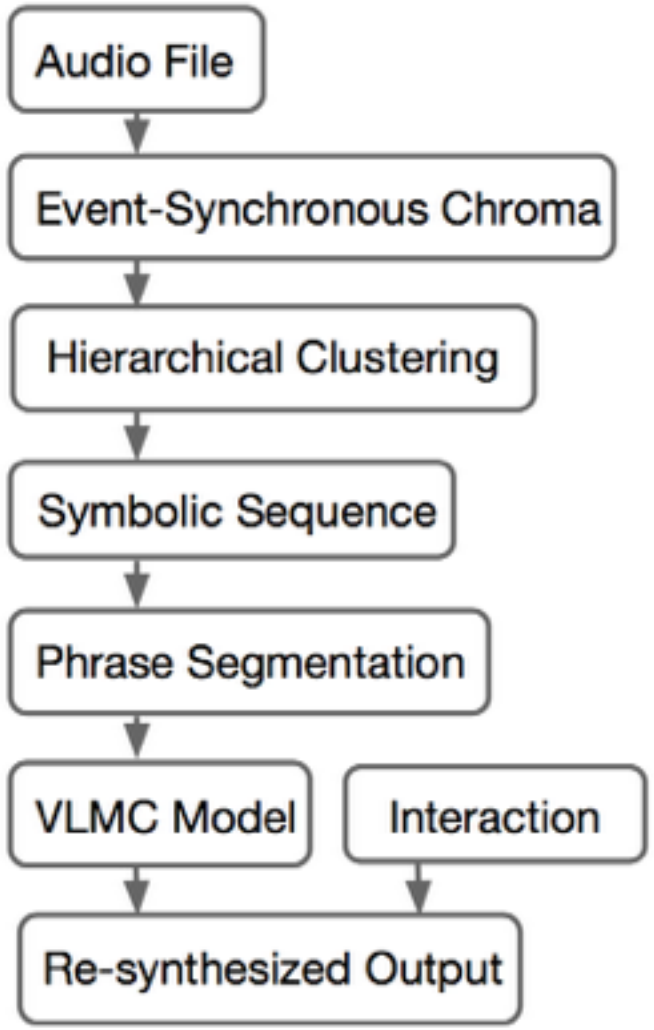
- Extending this approach with a speech recognition framework:
  - melody notes —> “phonemes”, chords —> “words”
  - FST maps sequences of notes (melody) to chords



- Method:
  - Trained chord model and chord-melody map using the Rock Corpus Dataset (de Clerc, Temperly at U Rochester)
  - Models built using openFST and openGRM-ngram

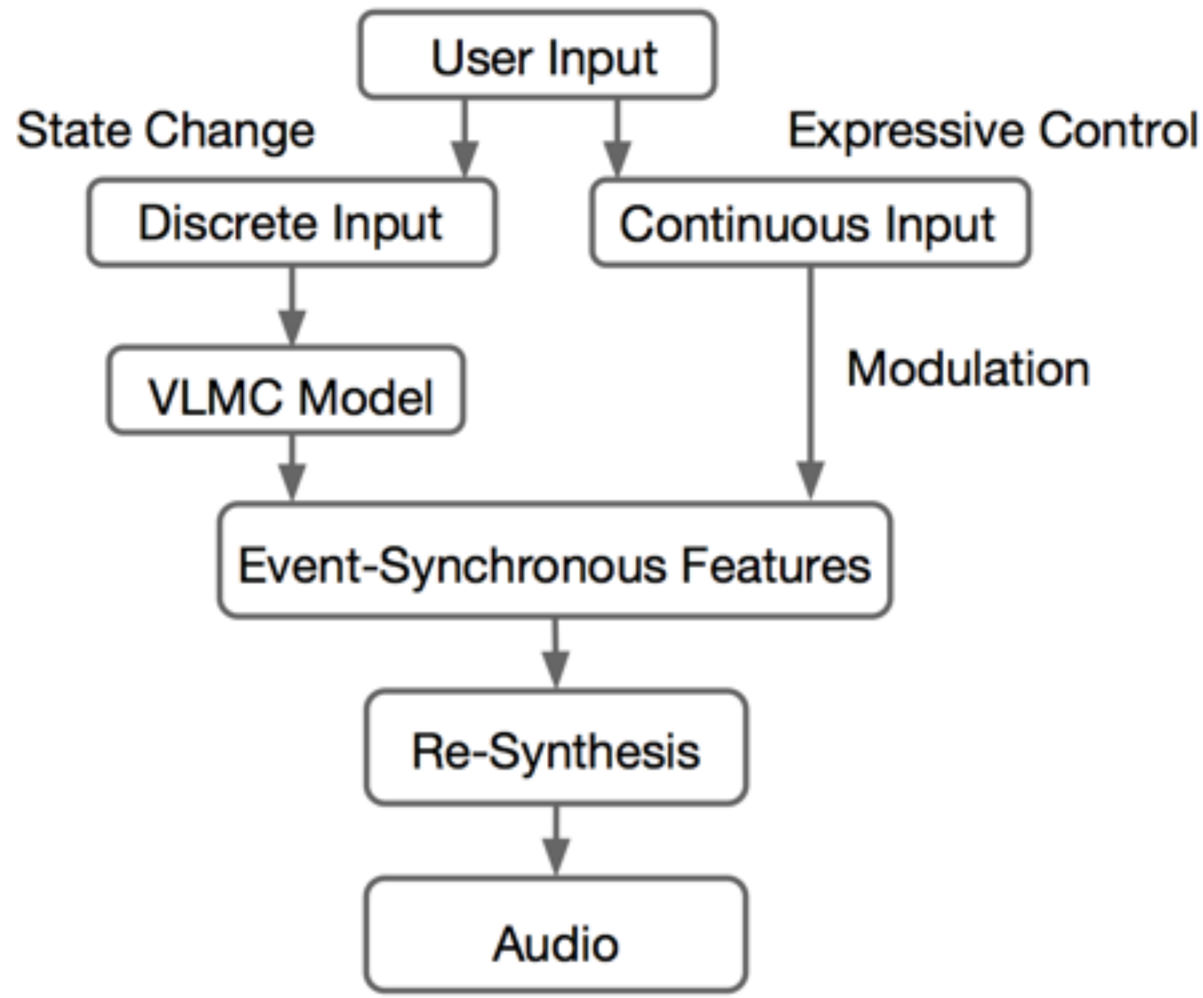
Audio Continuators

- Builds upon and extends previous work with Continuators (Pachet, Marchini, Kosta)
    - Improved clustering methods
    - Phrase segmentation
    - Introduces interaction paradigm
- Developed objective evaluation metrics of recurrence and novelty for the system's output
- Python Continuator implementation found at http://github.com/amlal/vlmc



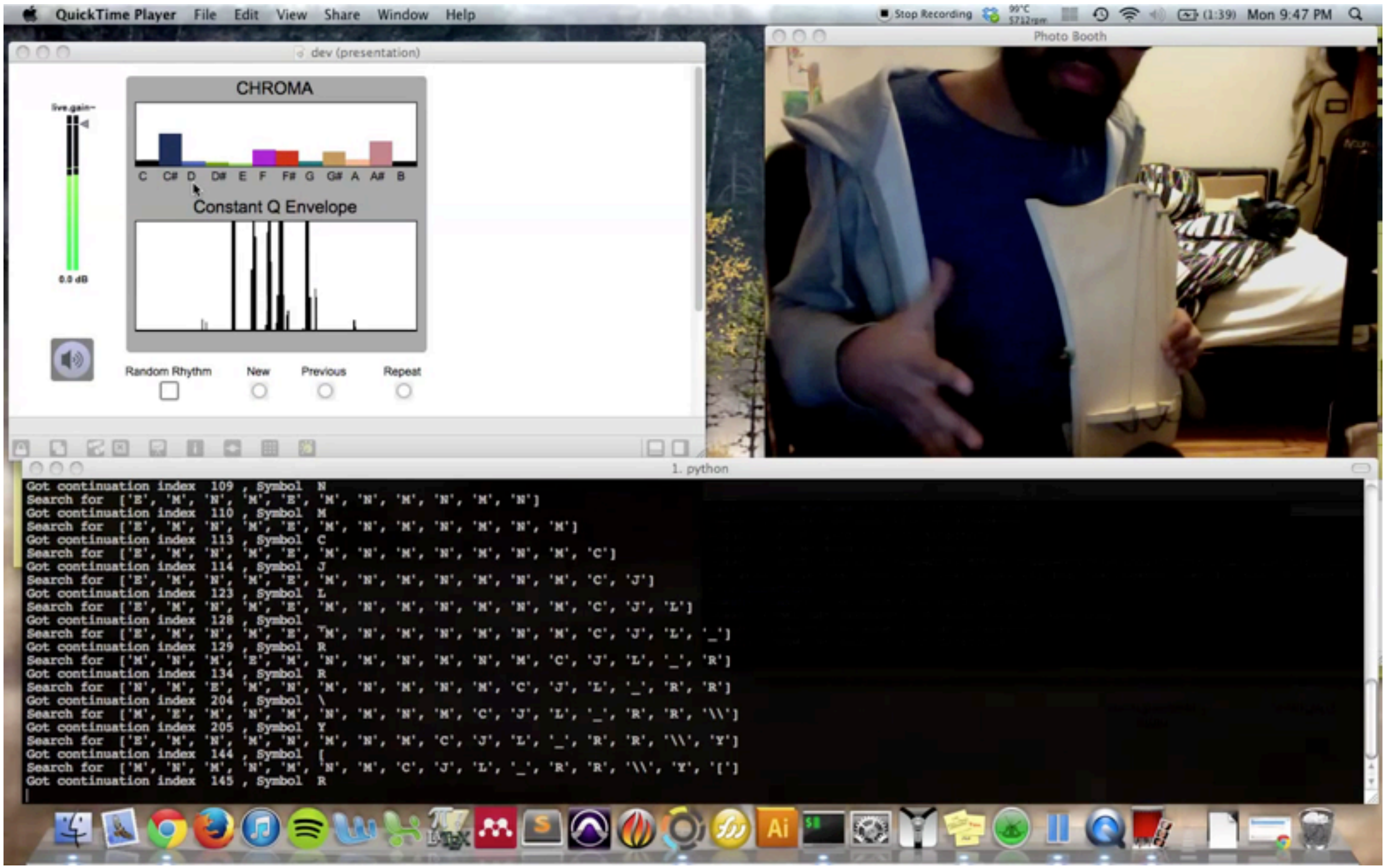
Audio Continuators

- Once trained, the interaction paradigm operates in a feed-forward manner:



[Lal, A. // Bello, J.P.]

Audio Continuators

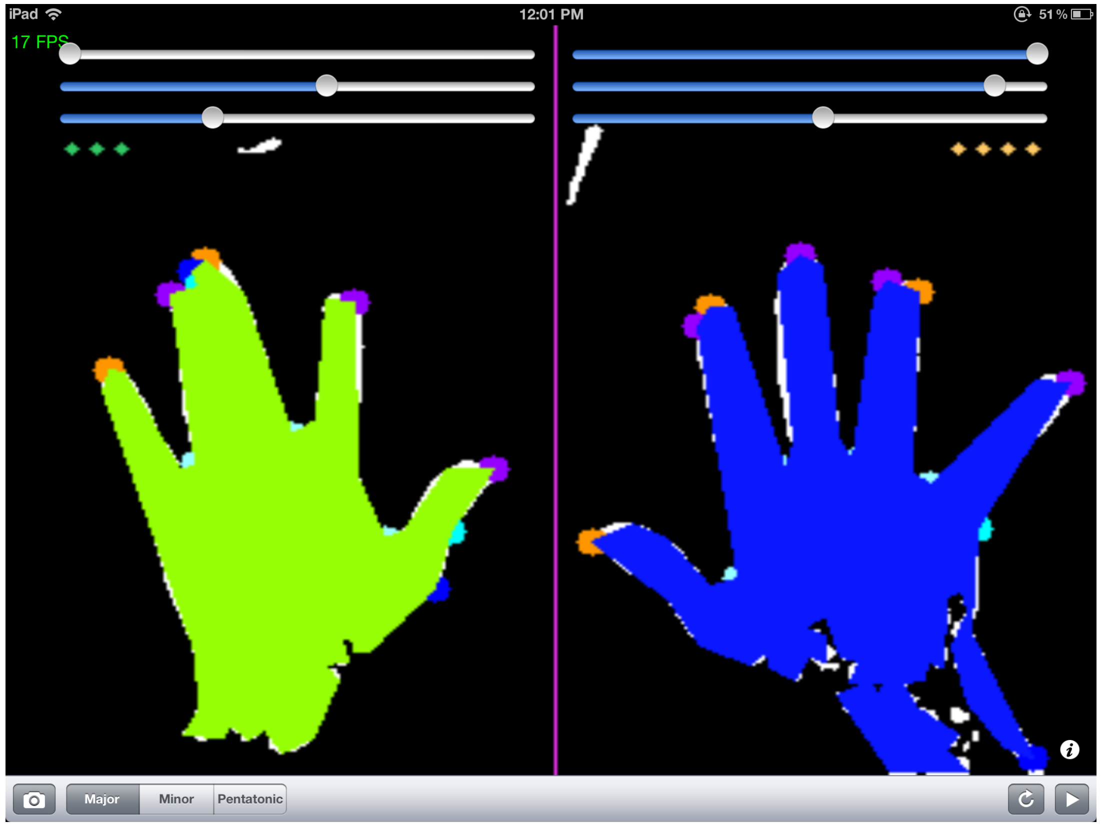


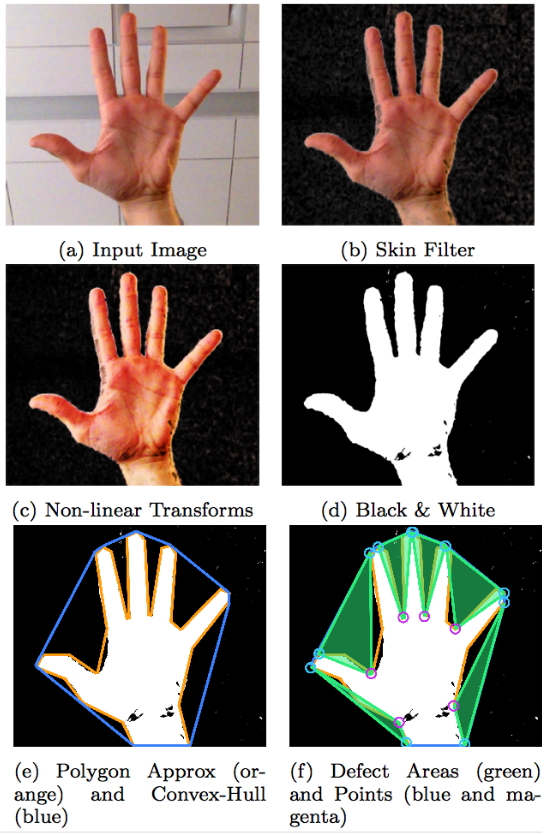
[Lal, A. // Bello, J.P.]

http://marl.smusic.nyu.edu

AirJam - Pose Recognition for Instrument Control

- Real-time computer vision on mobile devices using built-in front camera
- Convex-hull + heuristics to detect gestures
- AirJam published on the AppStore for iPad!





[Nieto, O. // Shasha, D.]

http://marl.smusic.nyu.edu

AirJam

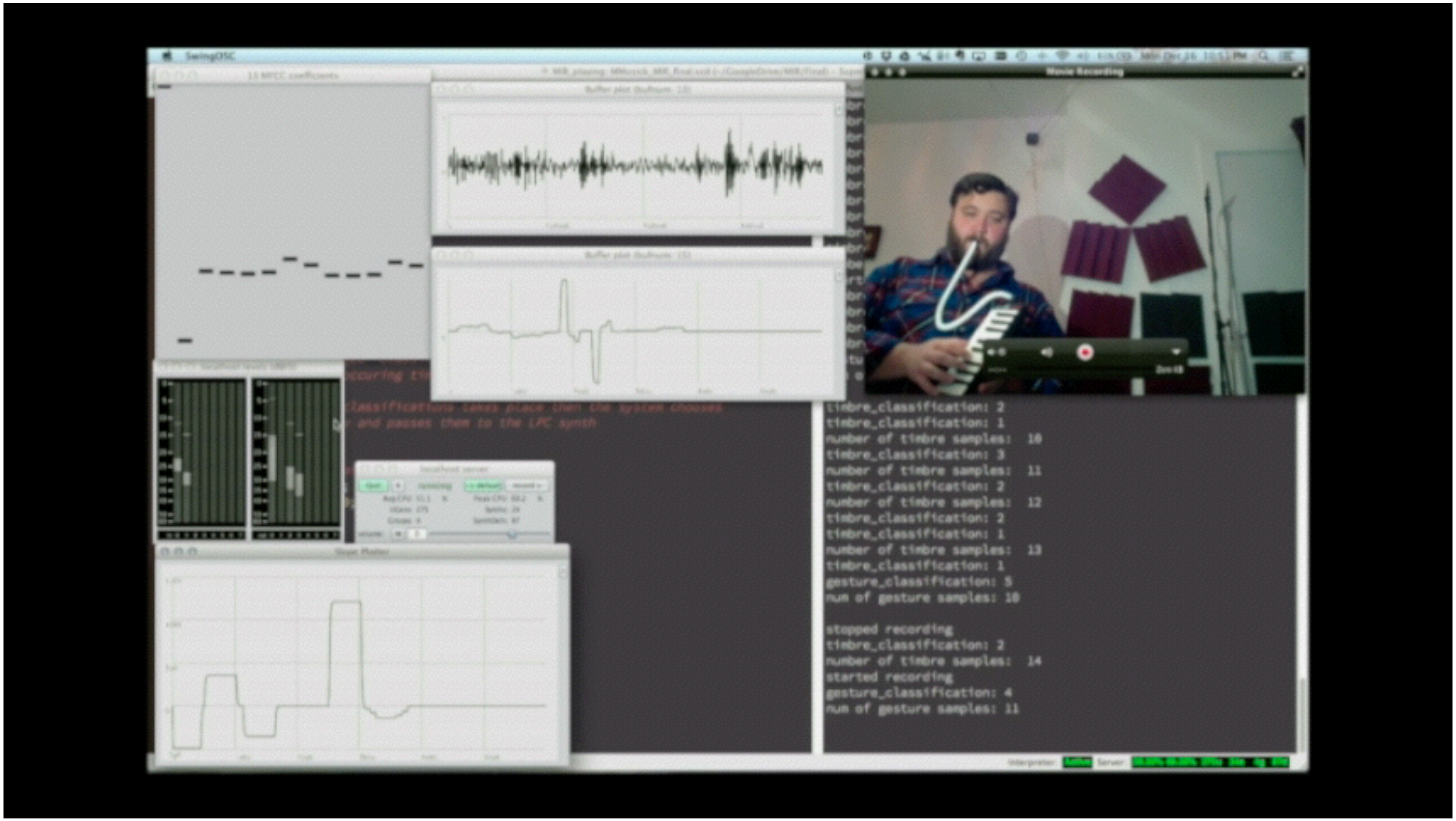
[Nieto, O. // Shasha, D.]

Interactive Performance Systems

- Objective: develop systems for
  - live-performance (e.g. concert), with musicians who improvise with the system
  - gallery installations, where participants are free to explore / play the system
- Emphasizes:
  - Composition of a set of interactions, not a single state (i.e., a score)
  - The system interface is sonic, rather than physical or tactile controllers
- Use feature extraction and classification to expand the range of interaction
  - Composer maps detected events to musical responses
  - Feature design embodies a creative element (what sonic behaviors are encoded?)

[Musick, M.]

Interactive Performance Systems



[Musick, M.]

http://marl.smusic.nyu.edu

Interactive Performance Systems

[Musick, M.]

Thanks! // Questions?