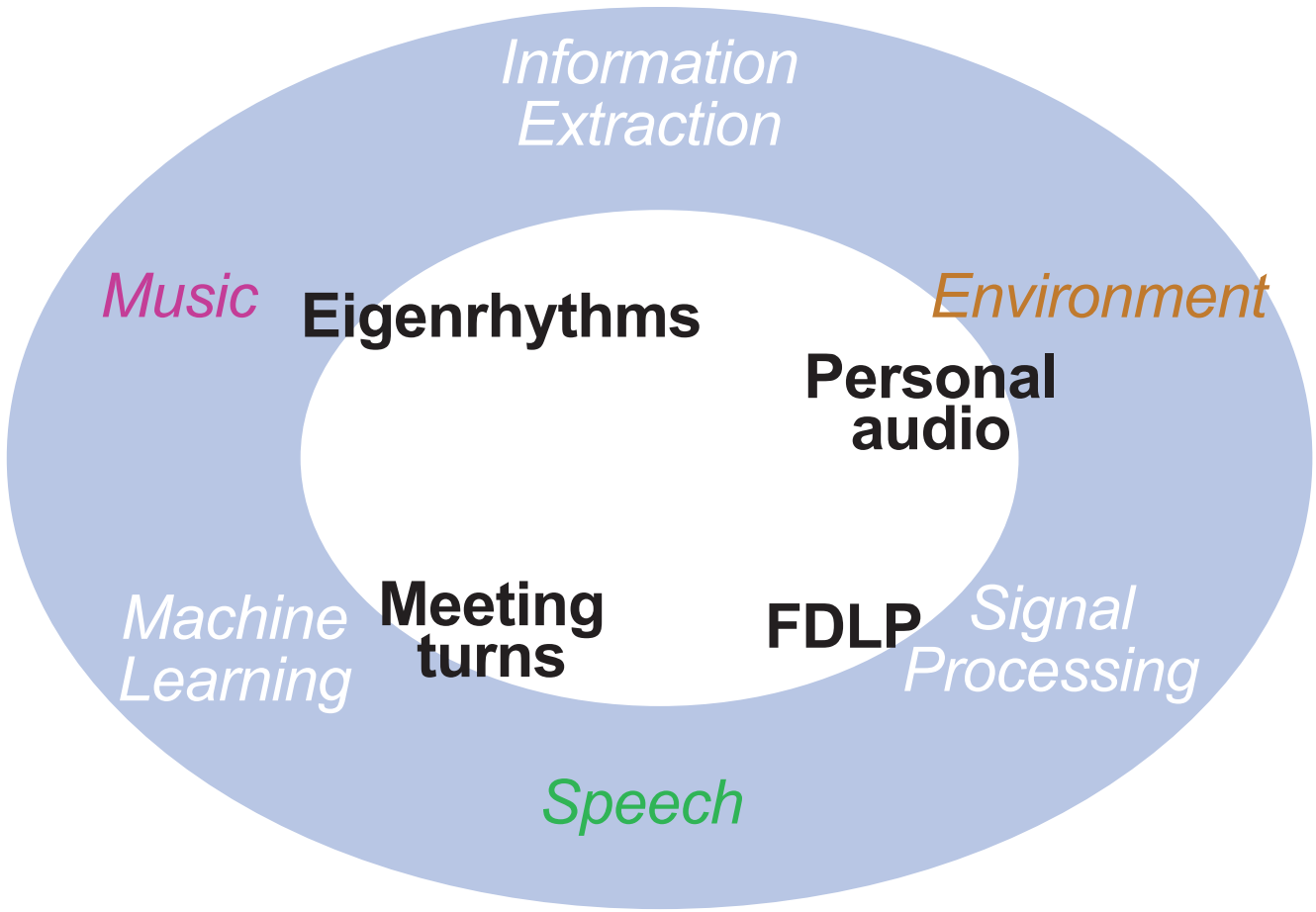
Sound Analysis Research at LabROSA

Dan Ellis
Laboratory for Recognition and Organization of Speech and Audio
Dept. Electrical Eng., Columbia Univ., NY USA
dpwe@ee.columbia.edu http://labrosa.ee.columbia.edu/

1. Speech
2. Music
3. Environmental Sound

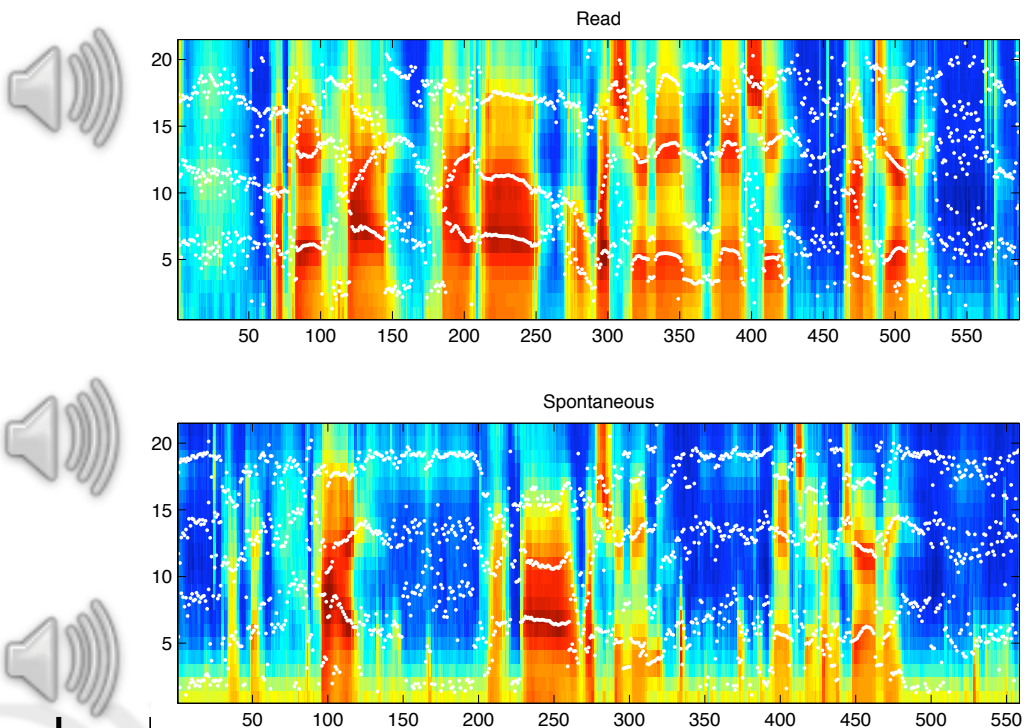
LabROSA Overview

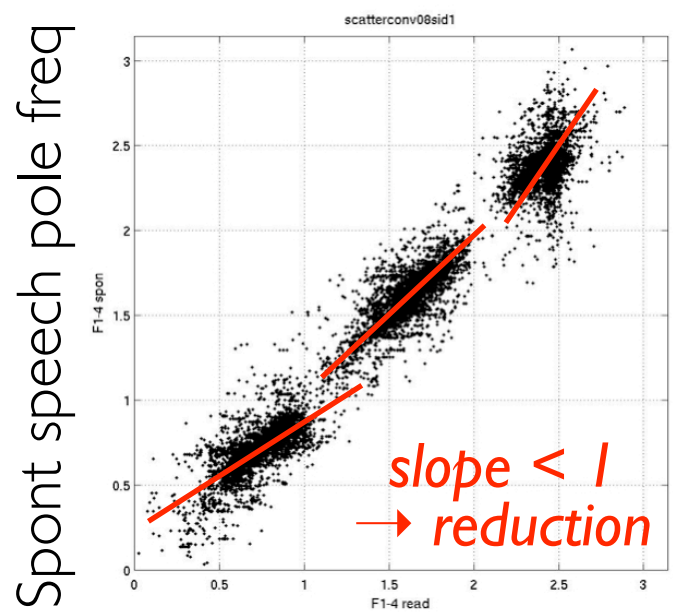


1. Speech Analysis / Recognition

with Sambarta Bhattacharjee

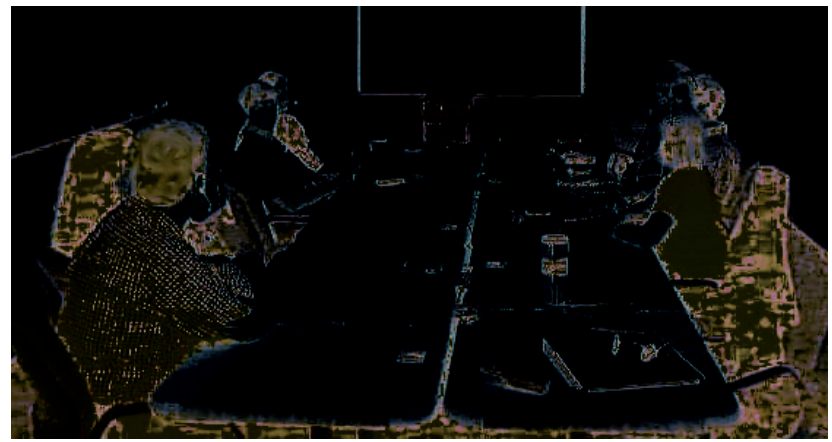
- Speech recognizers work for read speech poorly for spontaneous
  - e.g. 5% errors → 30%
- Transform spontaneous speech to read?





Meeting Recordings

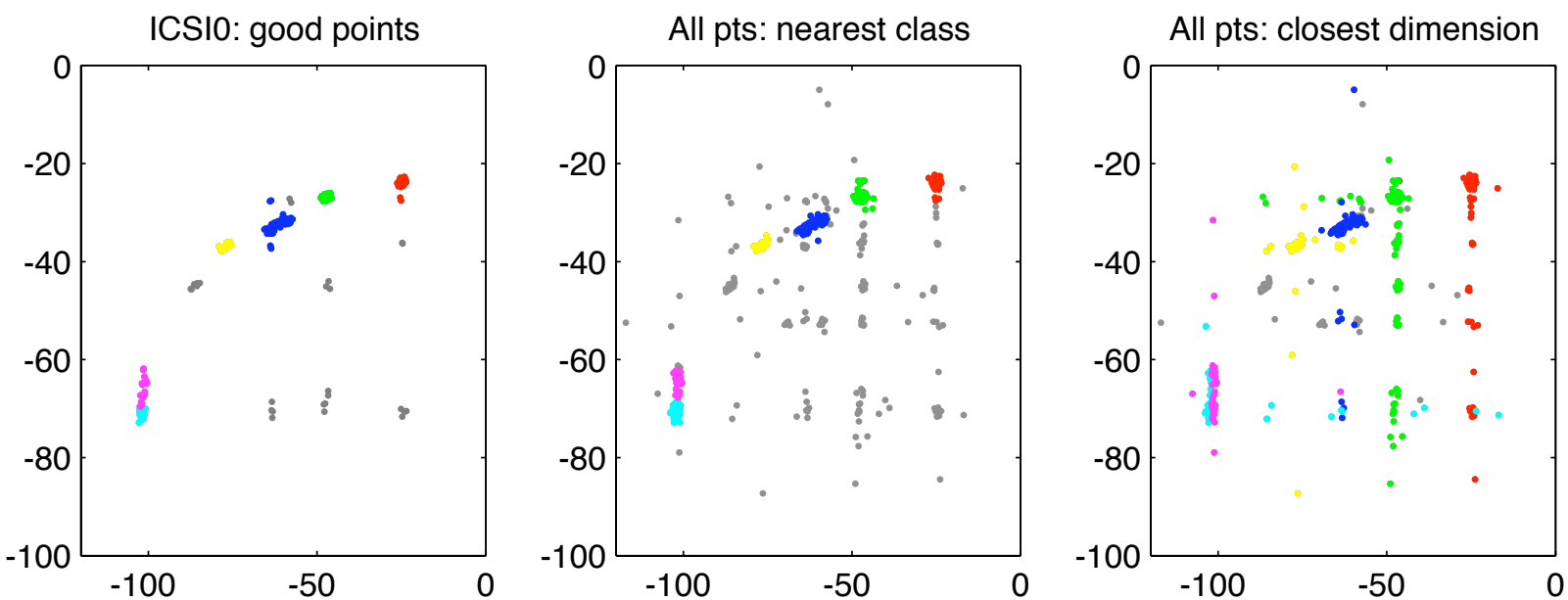
with Jerry Liu and ICSI



- Multi-mic recordings for speaker turns
  - every voice reaches every mic... (?)
  - ... but with differing coupling filters (delays, gains)
- Find turns with minimal assumptions
  - e.g. ad-hoc sensor setups (multiple PDAs)
  - differences to remove effect of source signal
    - no spectral models, < 1xRT

Speaker Turns from Timing Diffs

- Find best timing skew between mic pairs
- Find clusters in high-confidence points
- Fit Gaussians to each cluster,
  assign that class to all frames within radius



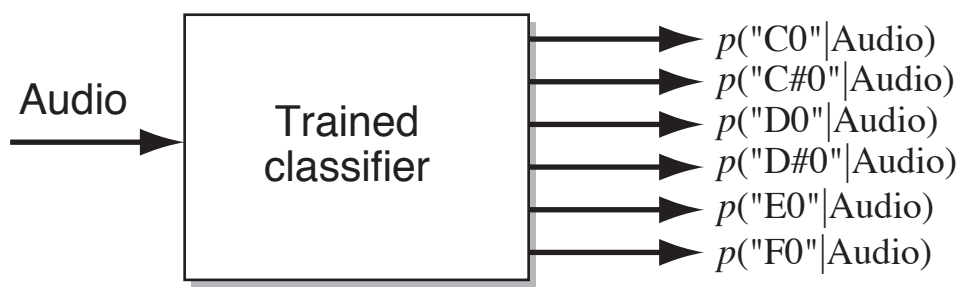
2. Music Signal Analysis

- A lot of music data available
  - e.g. 60G of MP3
    ≈ 1000 hr of audio/15k tracks
- What can we do with it?
  - implicit definition of 'music'
- Quality vs. quantity
  - Speech recognition lesson:
    10x data, 1/10th annotation, twice as useful
- Motivating Applications
  - music similarity / classification
  - computer (assisted) music generation
  - insight into music

Transcription as Classification

with Graham Poliner

- Signal models typically used for transcription
  - harmonic spectrum, superposition
- But ... trade domain knowledge for data
  - transcription as pure classification problem:



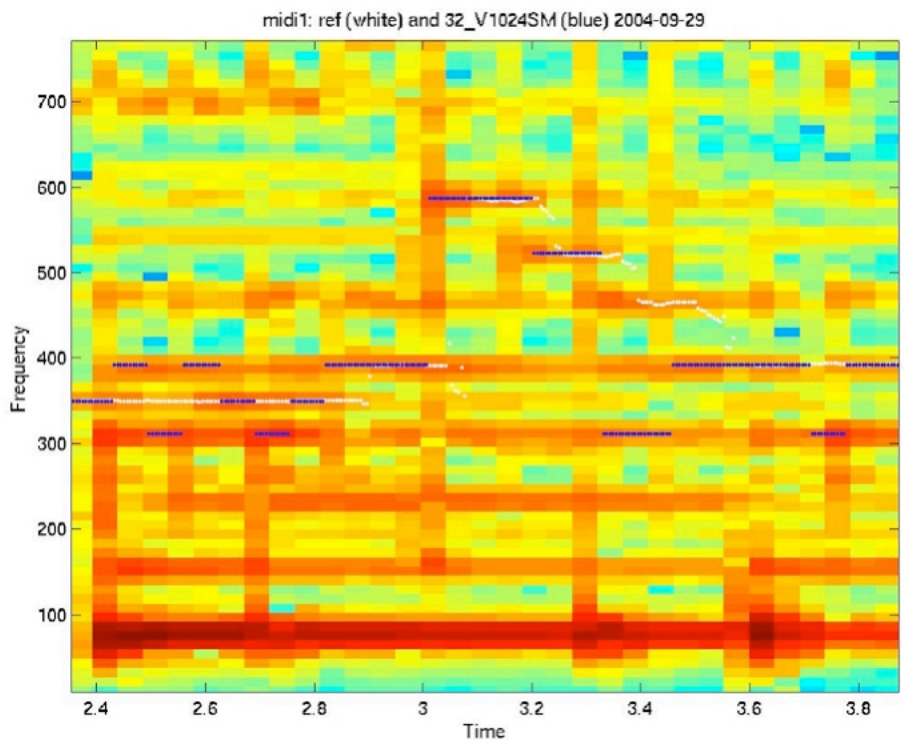
  - single N-way discrimination for "melody"
  - per-note classifiers for polyphonic transcription

Classifier Transcription Results

- Trained on MIDI syntheses (32 songs)
  - SMO SVM (Weka)
- Tested on ISMIR MIREX 2003 set
  - foreground/background separation

Frame-level pitch concordance

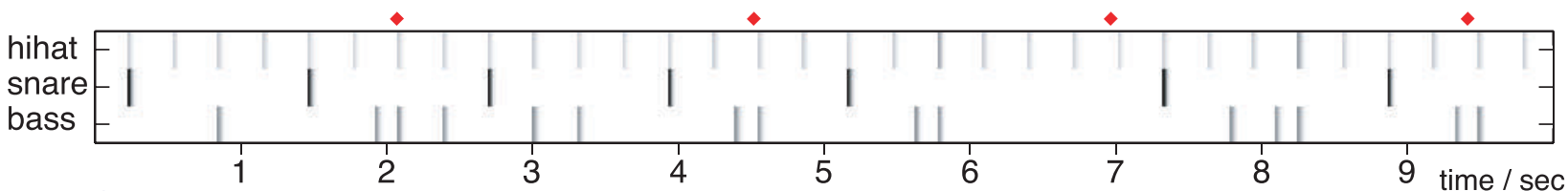
system	*"jazz3"*	*overall*
fg+bg	71.5%	44.3%
just fg	56.1%	45.4%



Eigenrhythms: Drum Pattern Space

with John Arroyo

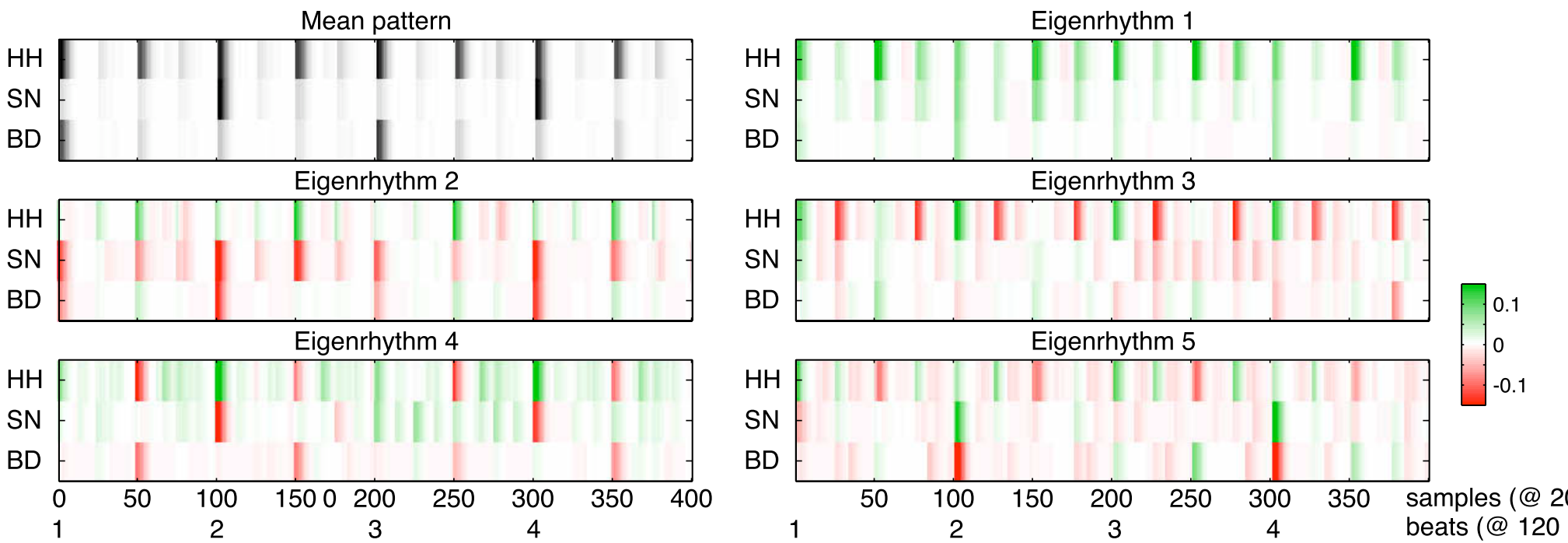
- Pop songs built on repeating "drum loop"
  - bass drum, snare, hi-hat
  - small variations on a few basic patterns



- Eigen-analysis (PCA) to capture variations?
  - by analyzing lots of (MIDI) data
- Applications
  - music categorization
  - "beat box" synthesis

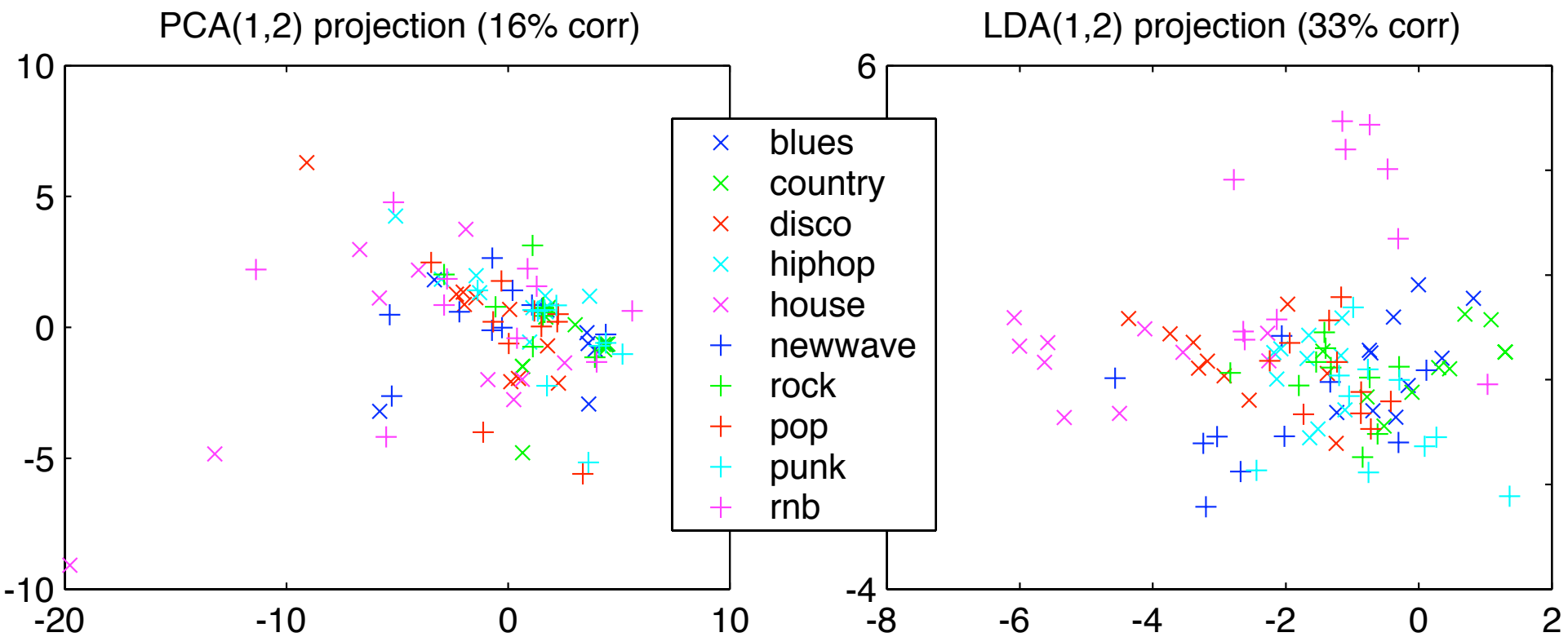
Eigenrhythms

- Need 20+ Eigenvectors for good coverage of 100 training patterns (1200 dims)
- Top patterns:



Eigenrhythms for Classification

- Projections in Eigenspace / LDA space

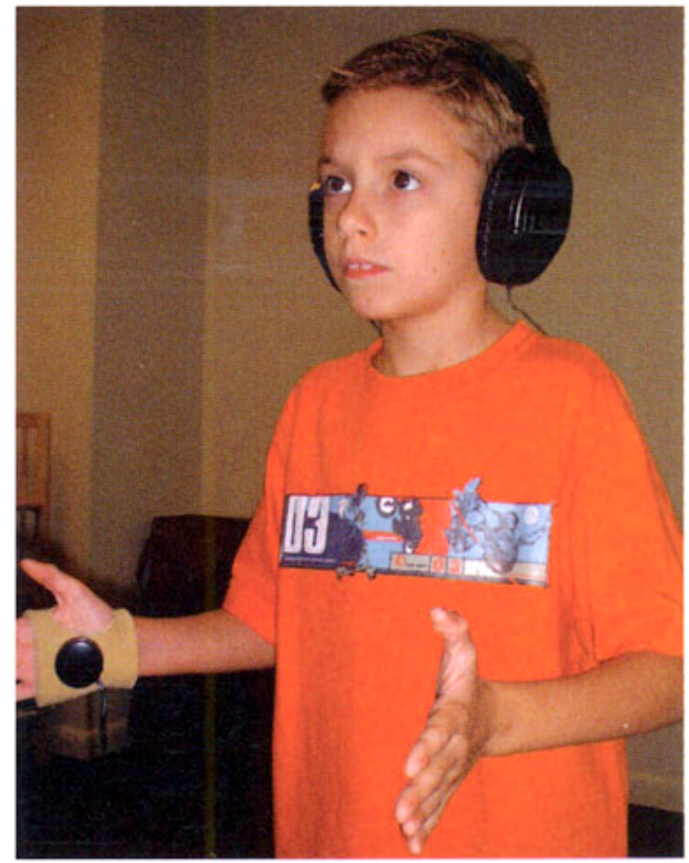


- 10-way Genre classification (nearest nbr):
  - PCA3: 20% correct
  - LDA4: 36% correct

3. Other Sounds: Clap Detection

with Nathan Lesser

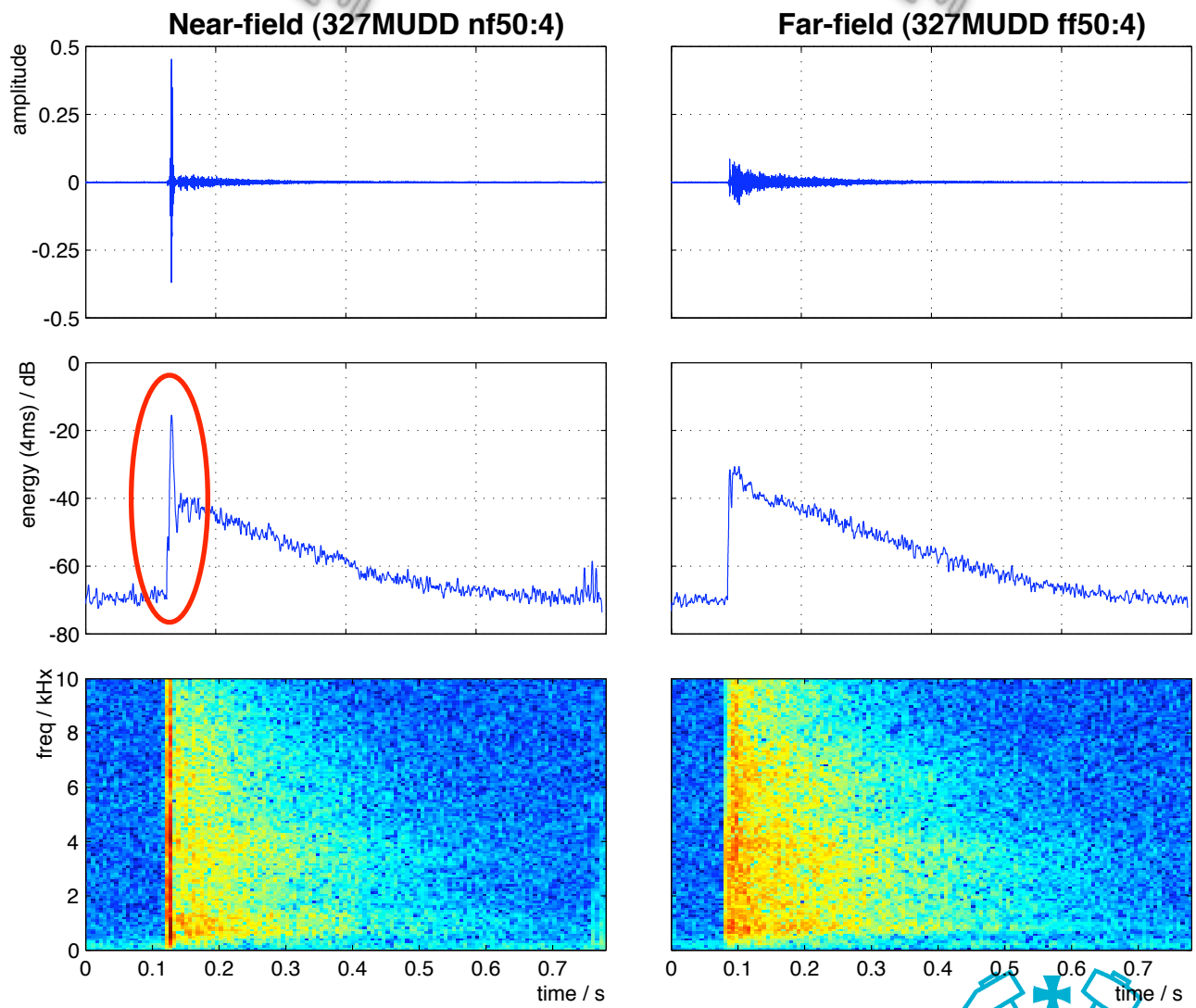
- Rhythmic clapping may help neural development
  - sensori-motor planning
  - focus and attention
- "Interactive metronome" devices
  - give feedback on synchrony
  - sensor-based
- Classroom deployment?
  - acoustic-based?
  - for multiple simultaneous users??



from interactivemetronome.com

Clap Range Discrimination

- Absolute level varies
- Decay slopes ~ same
  - reverberation
  - (RT_{60} ~ 900ms)
- **Initial burst** for near-field
  - "direct sound"



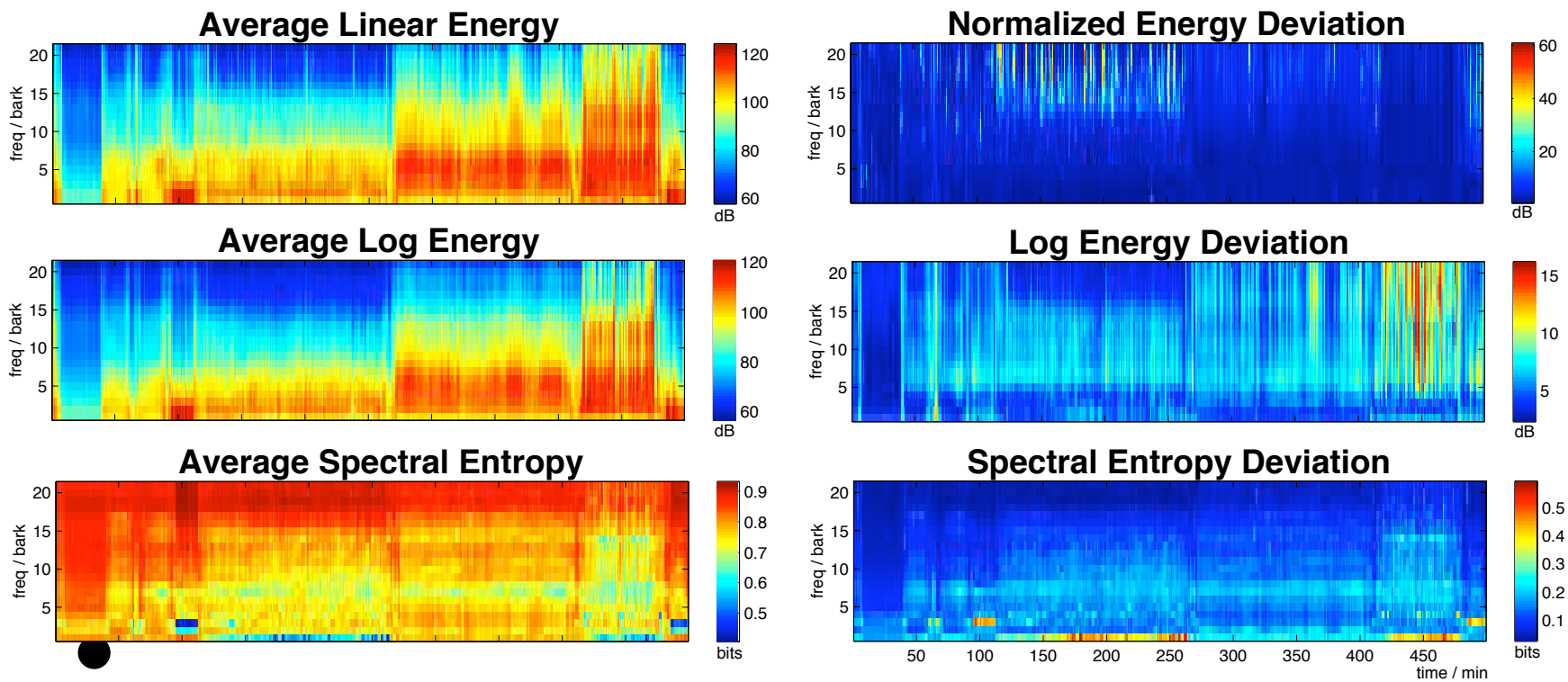
"Personal Audio"

with Keansub Lee

- Easy to record everything you hear
  - ~100GB / year @ 64 kbps
- Very hard to find anything
  - how to scan?
  - how to visualize?
  - how to index?
- Starting point: Collect data
  - ~ 60 hours (8 days, ~7.5 hr/day)
  - hand-mark 139 segments (26 min/seg avg.)
  - assign to 16 classes (8 have multiple instances)

Features for Long Recordings

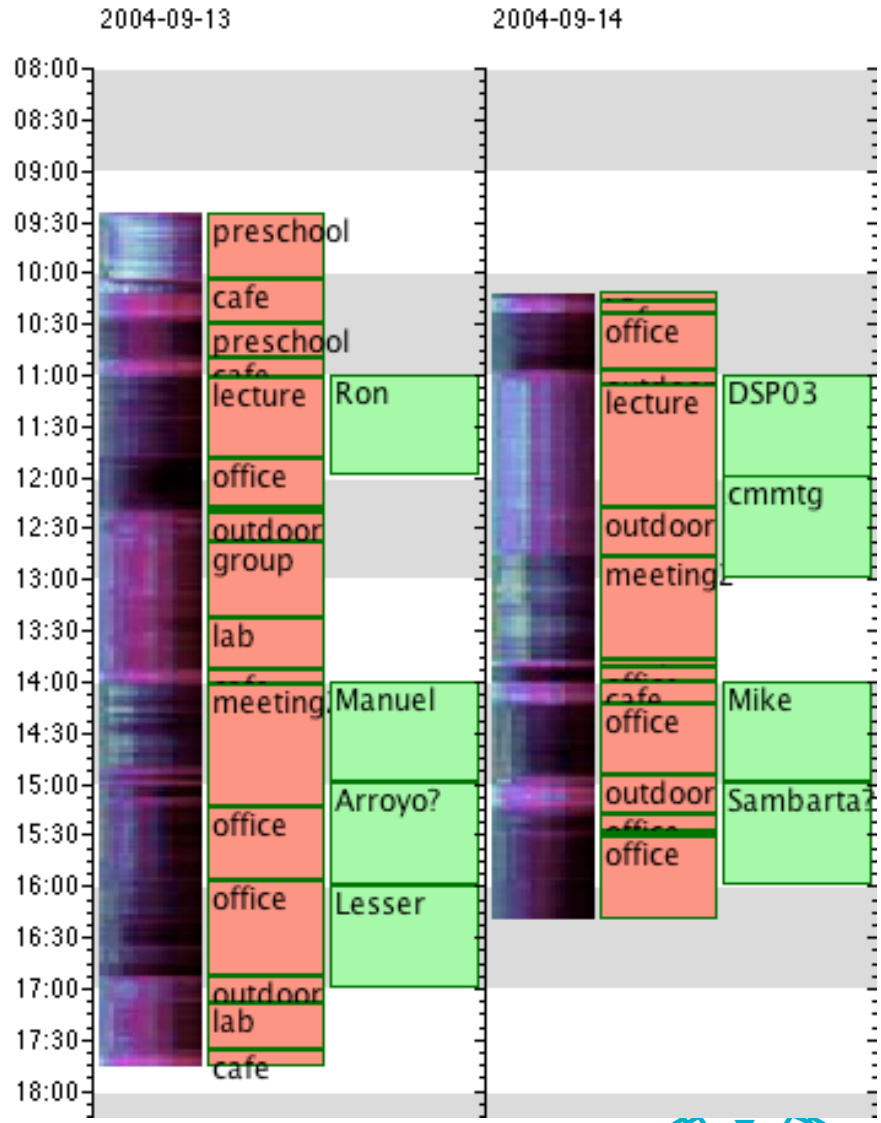
- Feature frames = 1 min (not 25 ms!)
- Characterize variation within each frame...



- and structure within coarse auditory bands

Personal Audio Applications

- Visualization / browsing / diary inference
  - link in other information sources
    - diary
    - email
- NoteTaker interface:
  - "what was I hearing?"



LabROSA Summary

- LabROSA
  - signal processing
    + machine learning
    + information extraction
- Applications
  - Speech: Recognition, Organization
  - Music: Transcription, Recommendation
  - Environment: Detection, Description
- Also...
  - signal separation, compression, dolphins...