Musical Analysis by Computer from Audio Recordings

Hans Fugal
PhD Comprehensive Exam
Computer Science Department
New Mexico State University

What? **analysis** from **recordings**
Why? Have the computer do what we don't want to do, or can't easily do, or do it faster. Also, to learn about our own perception of music.

Crash Course

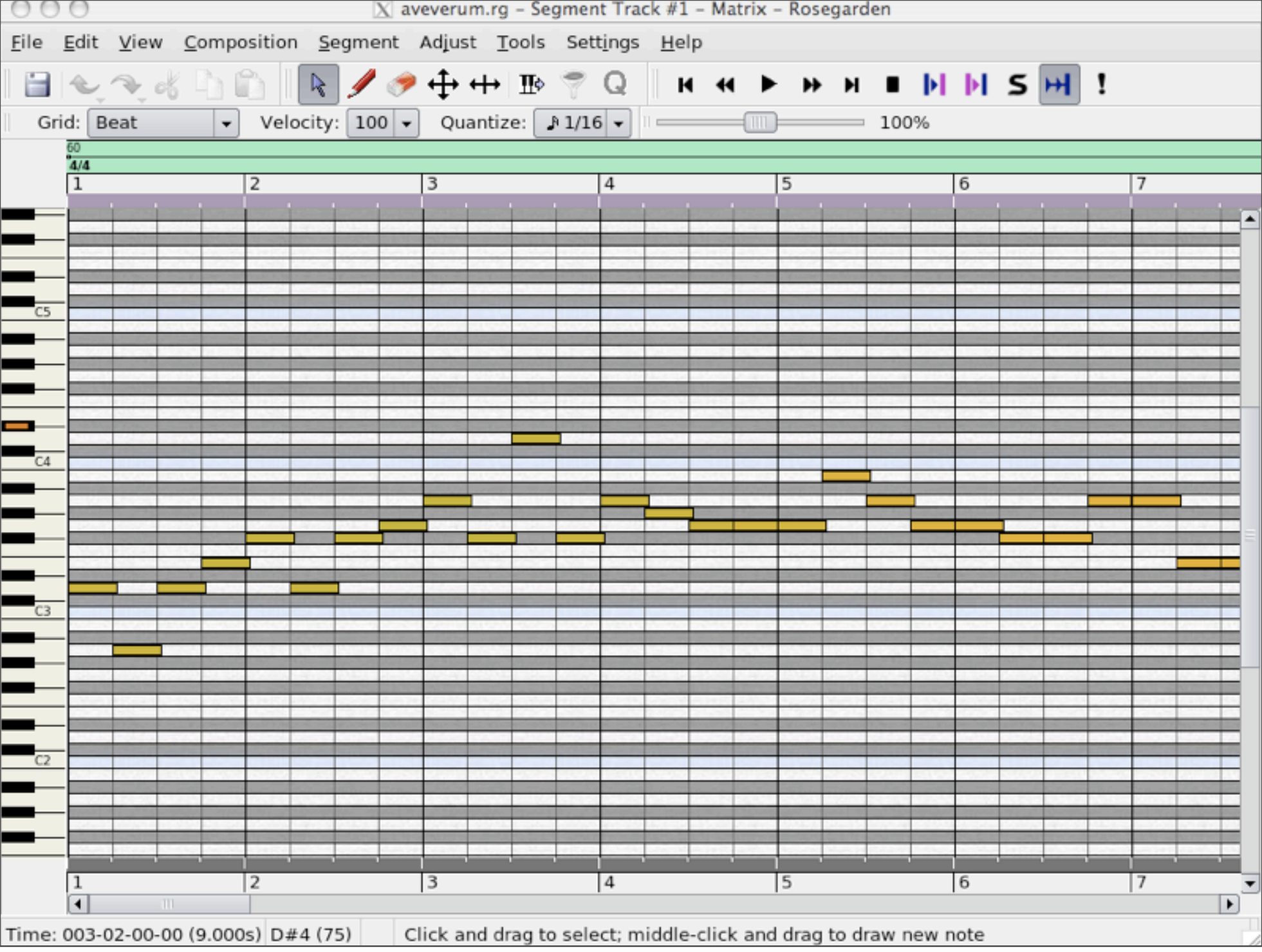
What is Music?

Rhythm

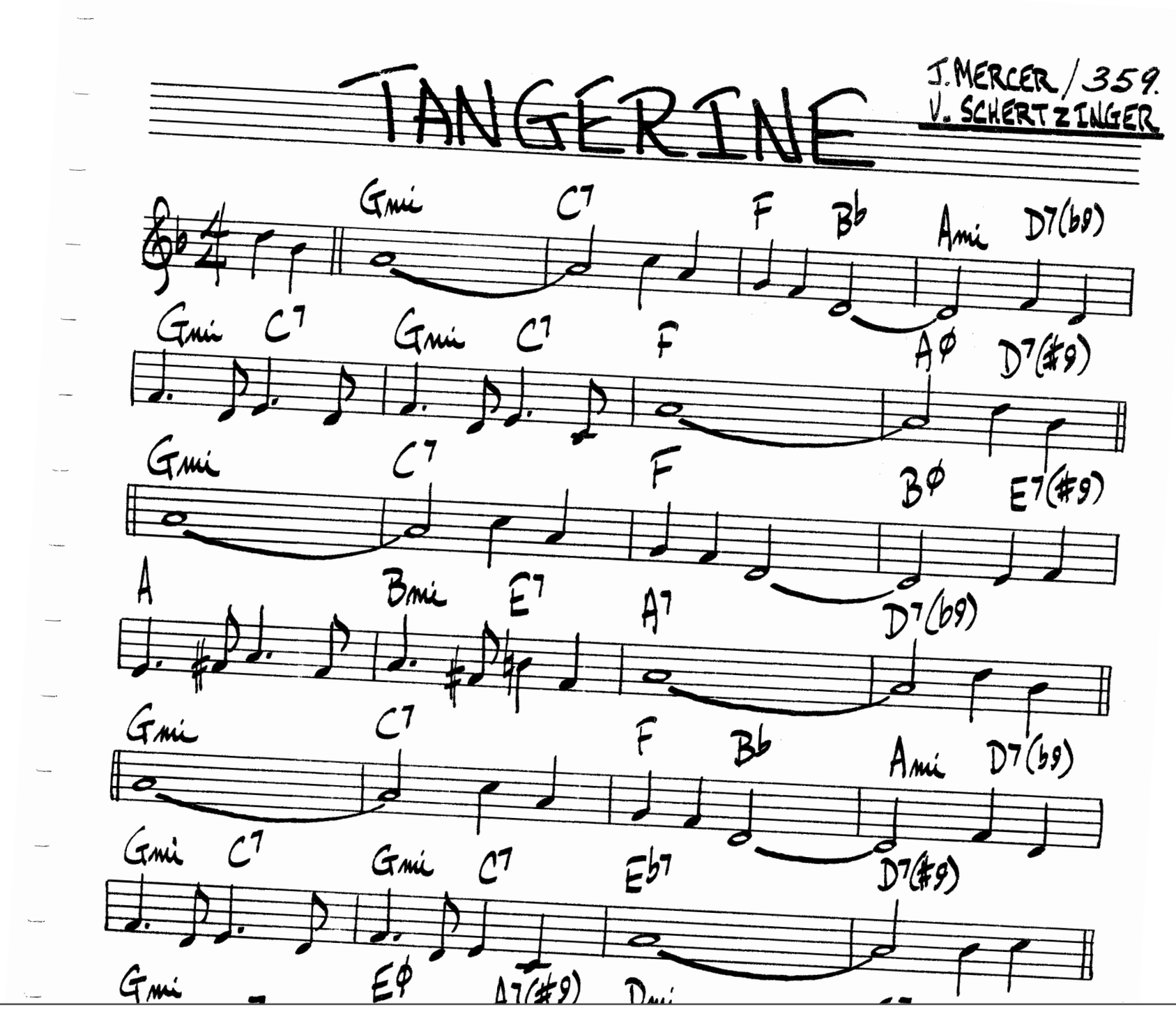
Pitch

Timbre

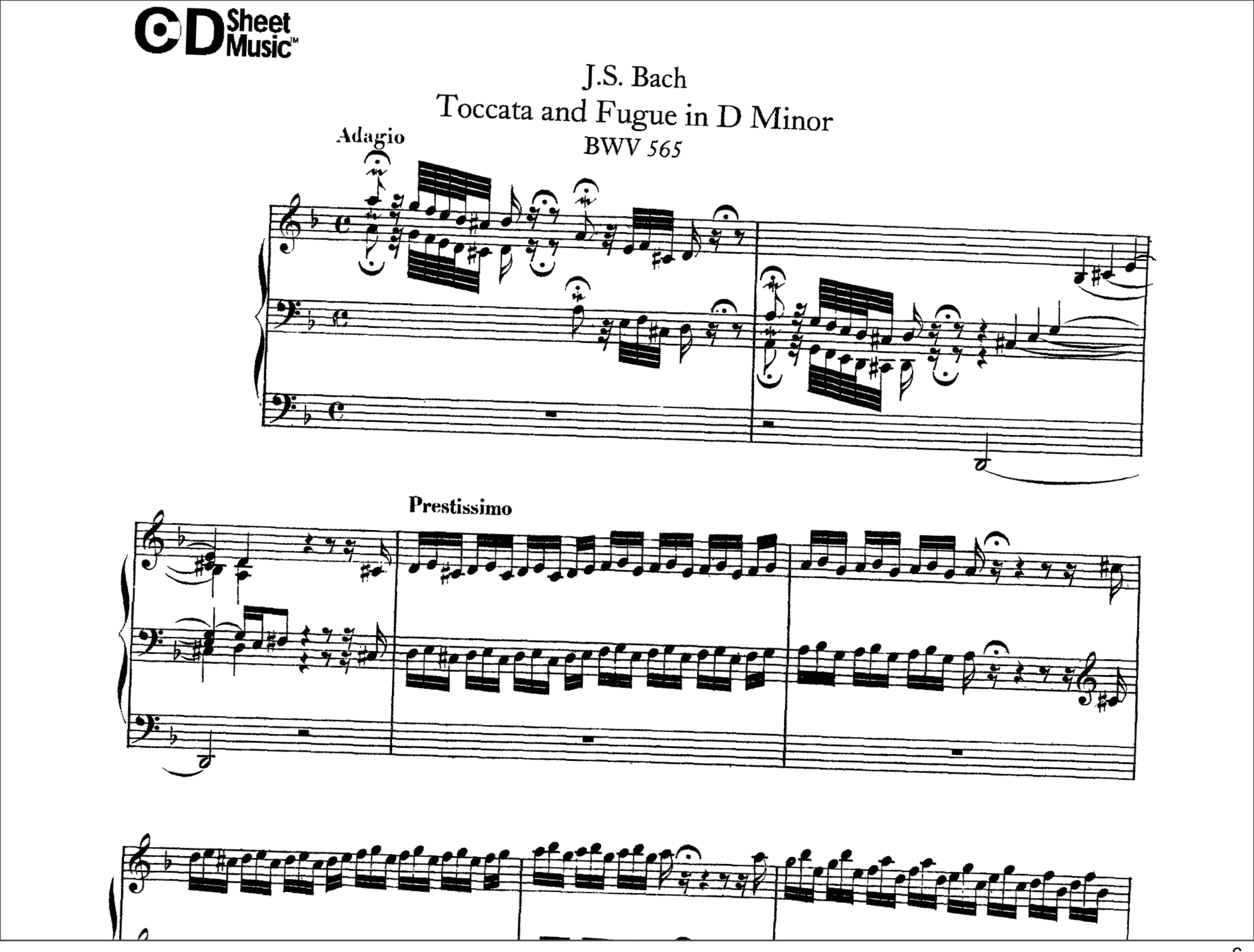
3 axes: time, frequency, texture
mention dynamics in passing



two of the dimensions (time, pitch)
timbre: instrument/performer



Discuss essentials of notation: rhythm (durations, time signature), pitch (notes on staff). Discuss harmony and improvisation. Play the first two lines straight with block chords on the keyboard if possible (even better if you can transpose it down a step to match Brubeck), then play Brubeck clip.



Explicit notation, but still very free-flowing.
Play Herrick clip (first 3 bars).
note timing
Demonstrate timbre by playing the synthesizer (Dorsey) version of the first 3 bars, and compare with organ.

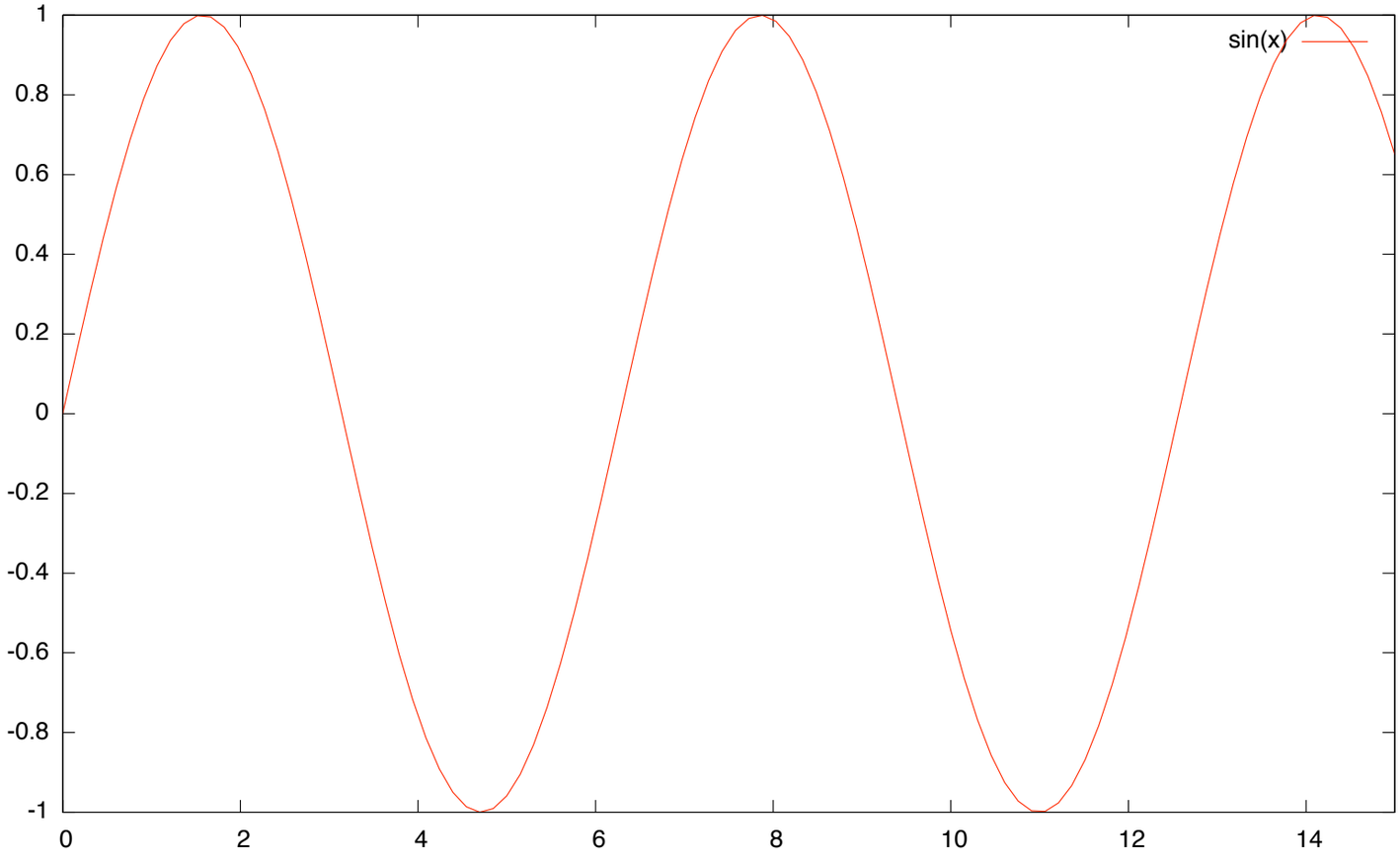
Sound

Vibrations cause Waves

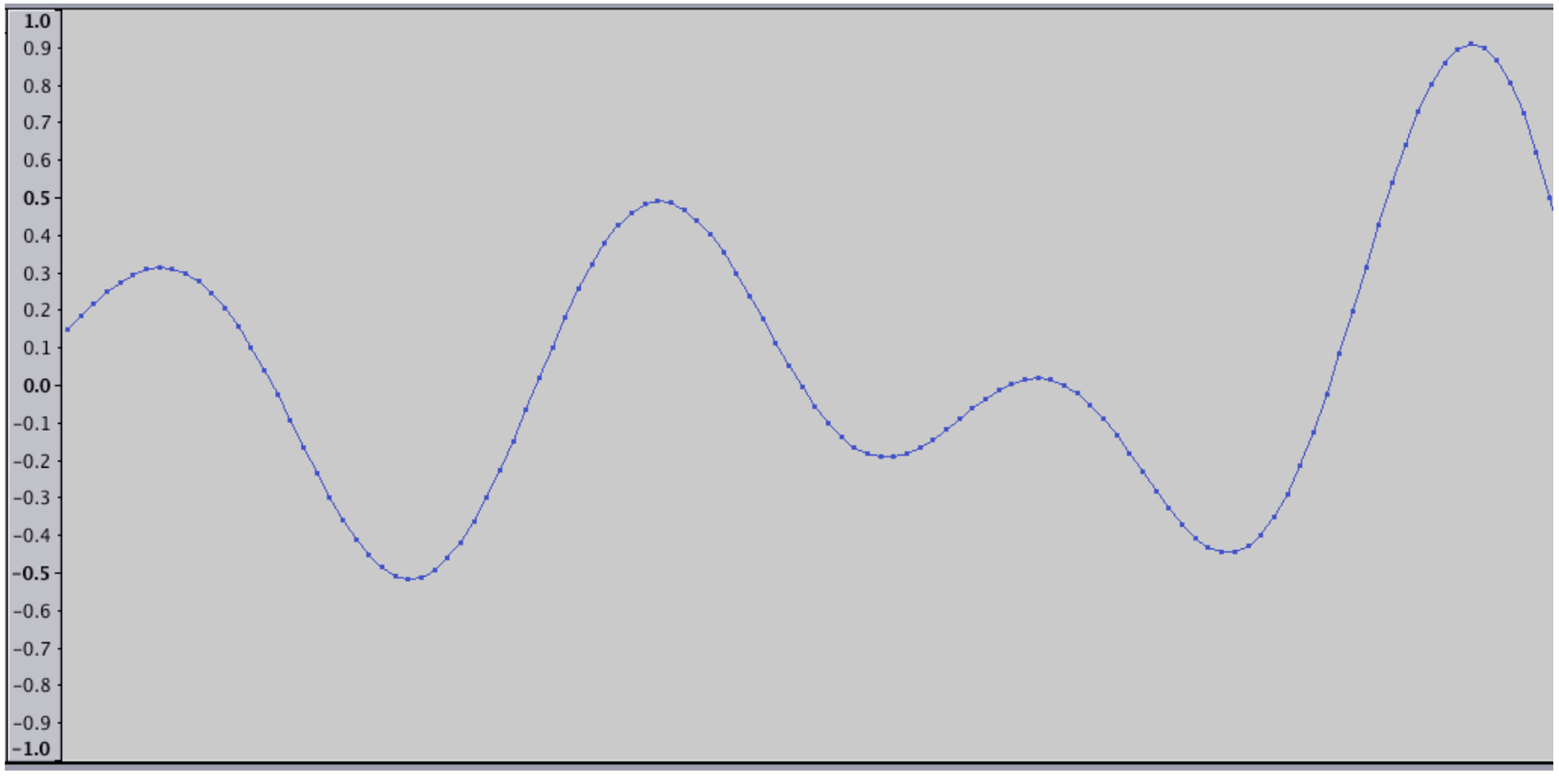
Sinusoidal Composition

Time domain and Frequency domain

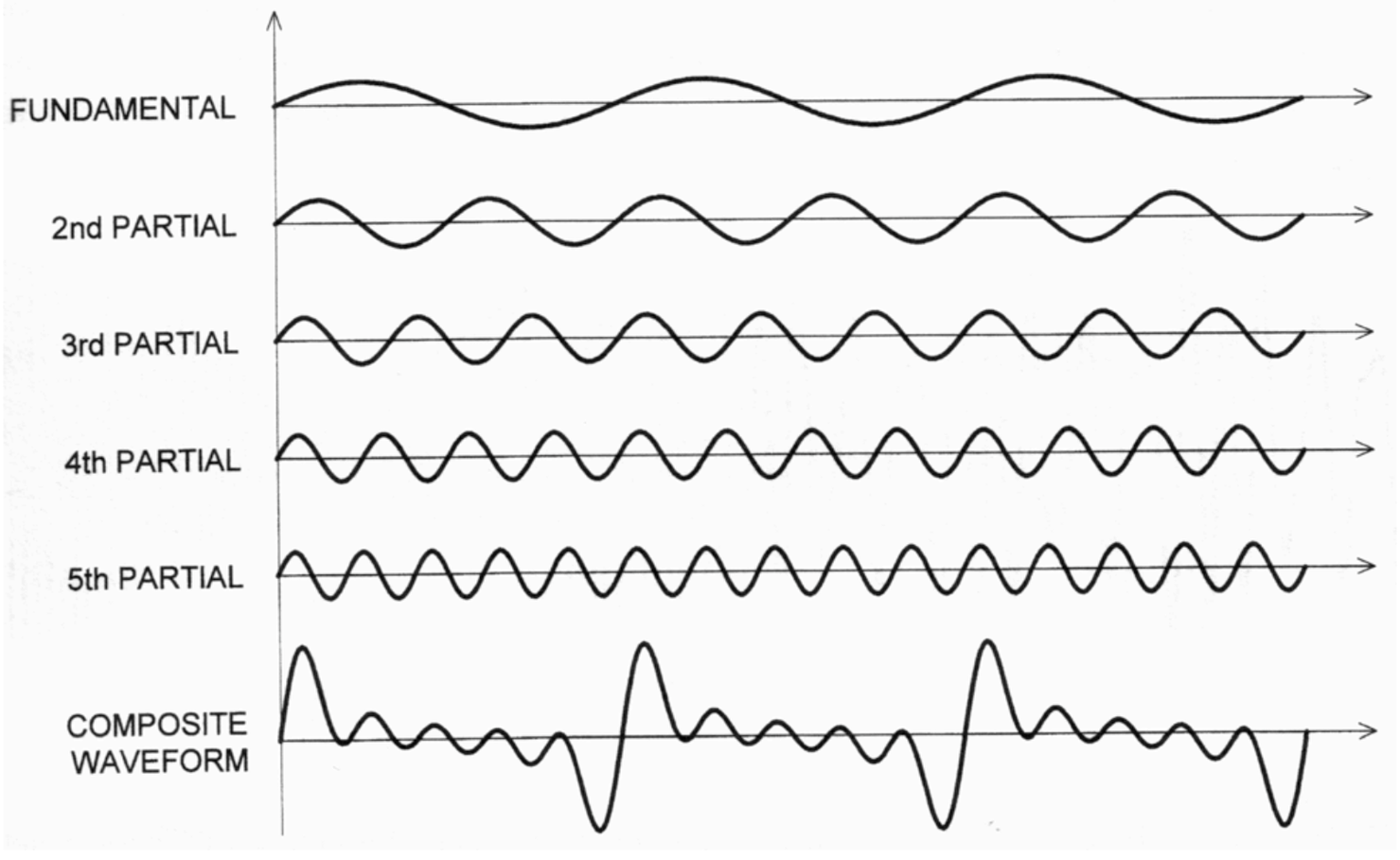
Sound Waves



Sampling



Partials

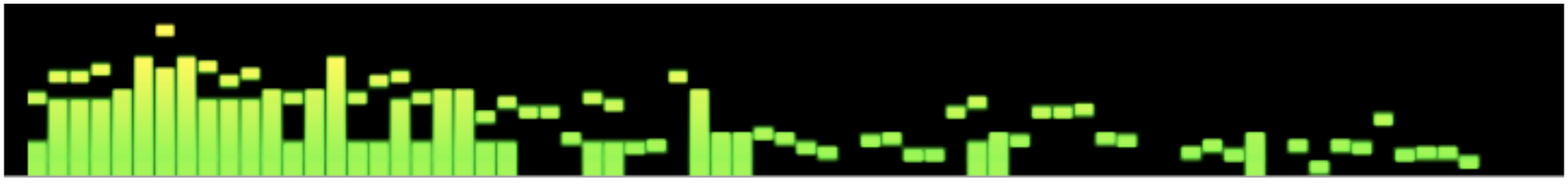


(Dodge and Jerse)

Compare with spectrogram

Discrete Fourier Transform

$$X(\omega_k) = \sum_{n=0}^{N-1} x(t_n) e^{-j\omega_k t_n}$$

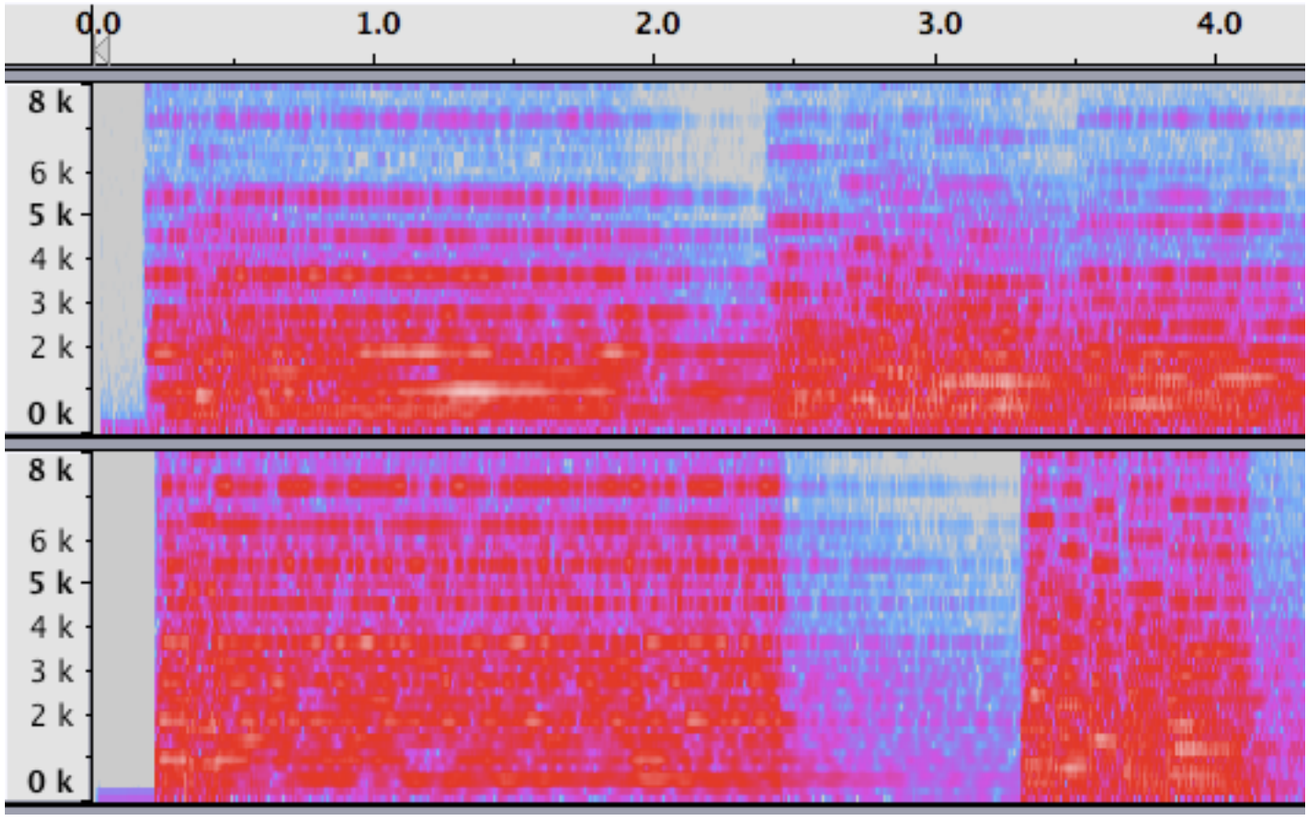


You've seen it before...
Demo Billy Joel, Movin' Out in VLC with Spectrogram.

Spectrograms

Organ

Synthesizer



Herrick (top) and Dorsey (bottom), first few notes.

Rhythm

What?

Tempo

Beat Tracking

Quantization

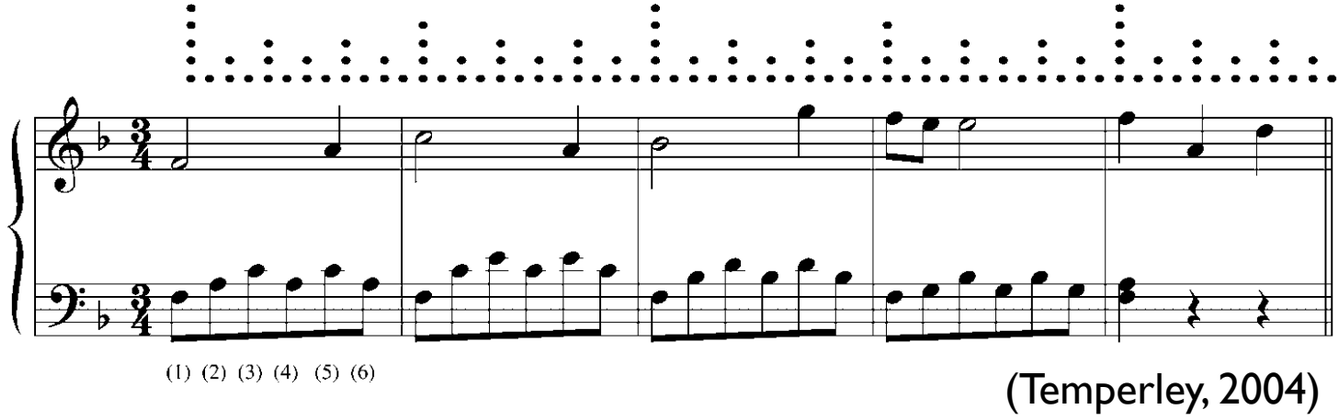
Metrical Structure

Beat and Tempo

Beat: steady pulse

Tempo: beats per minute

Metrical Structure



(Temperley, 2004)

Metrical Levels

Time Signatures

Bars, Downbeats, etc.

Note the hierarchical nature
Which is the primary metrical level? Not always clear-cut
Quantization is the alignment of events to this hierarchical structure

Complications

Tempo Rubato

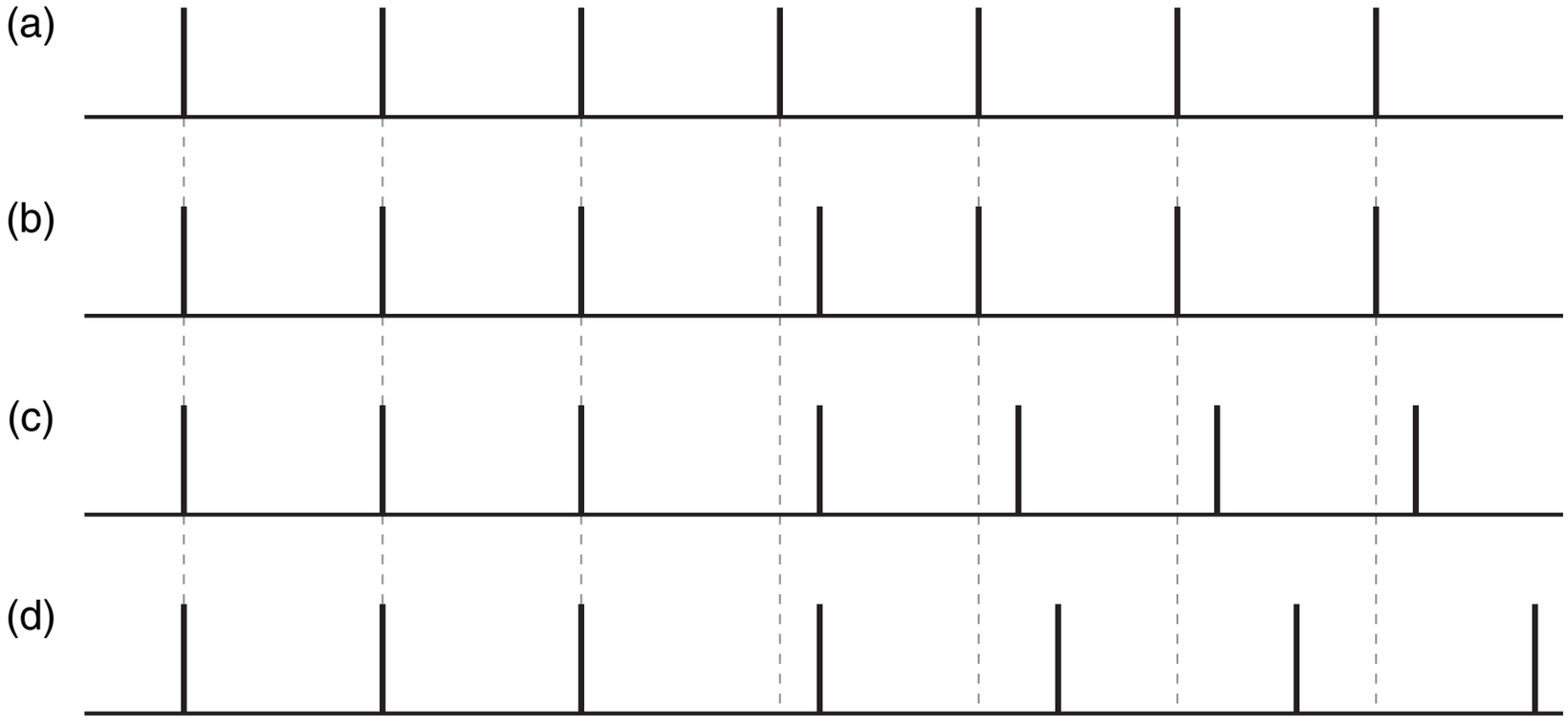
Fermatas, etc.

Syncopation

Articulation

Imperfect Performance

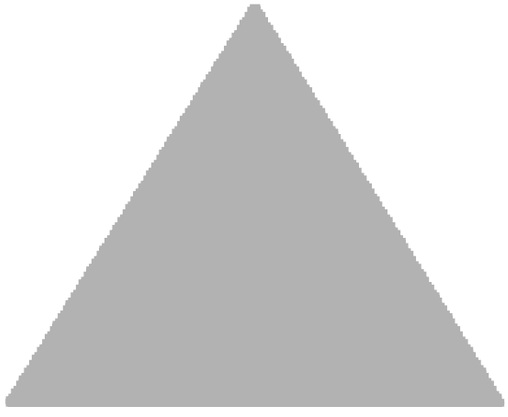
Tempo Deviation



(Gouyon and Dixon 2005)

a – no variation
b – timing of one note
c – offset (e.g. from a fermata)
d – ritardando

Temporal structure



Tempo Timing

(Honing 2001)

naïve algorithms trip up on timing
tempo changes too (rubato)
structure can be very valuable information

How?

Detect Onsets

Construct a Model

the model has hidden state, i.e. the tempo and phase
it predicts and state is adjusted to maximize agreement

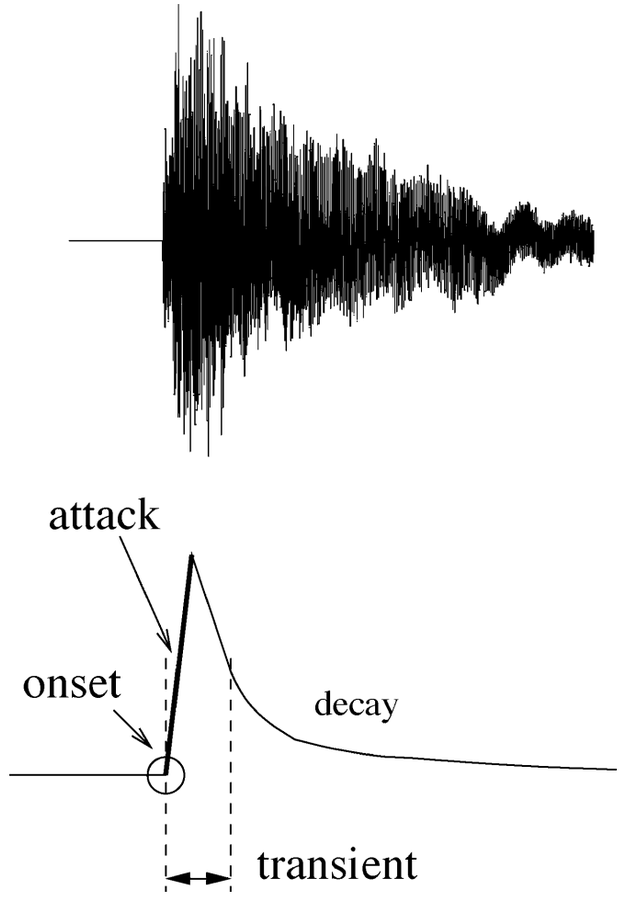
Onsets

Rhythmic events

Points in time

Usually onsets of notes

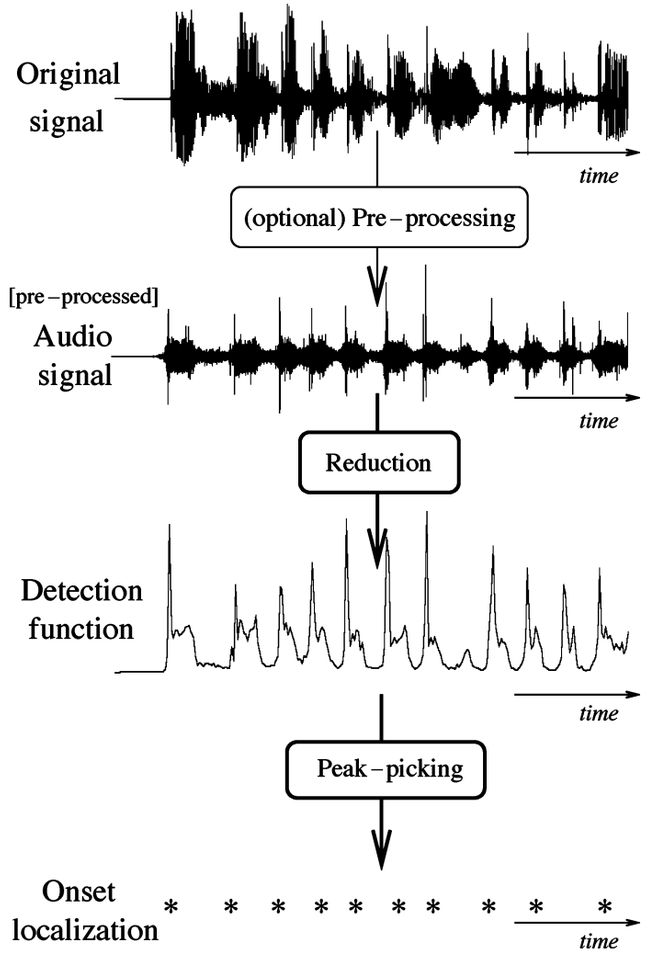
Onset Detection



(Bello et al. 2005)

transients are key

Onset Detection



(Bello et al. 2005)

preprocessing example: high-pass filter
detection functions aka novelty functions

Detecton Functions

Amplitude - beats are louder (or are they?)

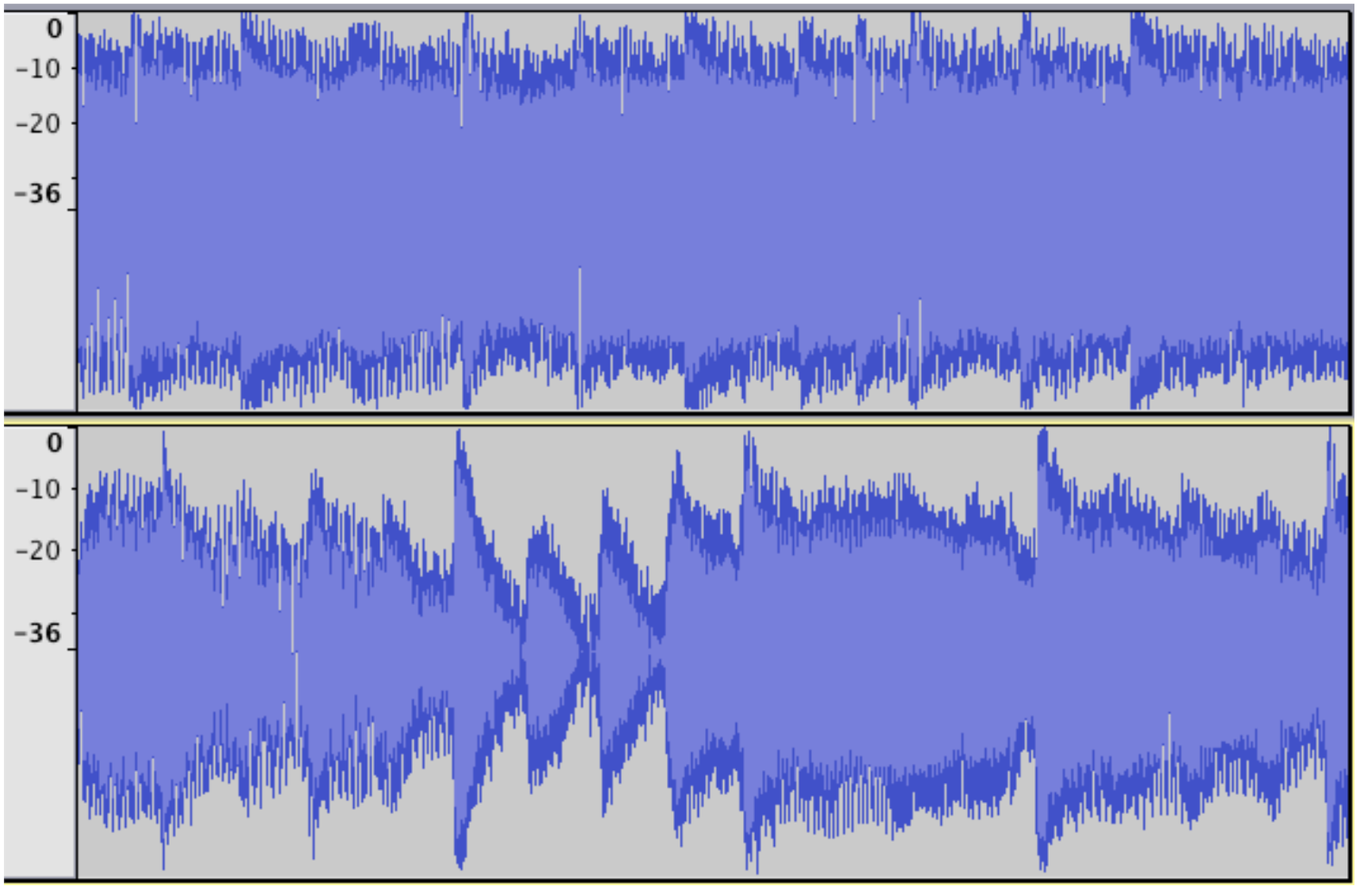
Spectral features

- Look for percussion in high bands
- Spectral Flux

Probabilistic Models

unfortunately many modern recordings have heavy dynamic range compression

Dynamic Range Compression



5 seconds of audio
Top: Keane – Somewhere Only We Know
Bottom: Billy Joel – Movin' Out

Probabilistic Models for Onset Detection

Model the signal

Switching between transient/steady state

Surprise-detecting

modeling the signal with a simplified model works ok if it's a valid model.
surprise: "get used to" the signal then be surprised by an onset.

Tempo Tracking

Stochastic dynamical system

Models beat and period - hidden state

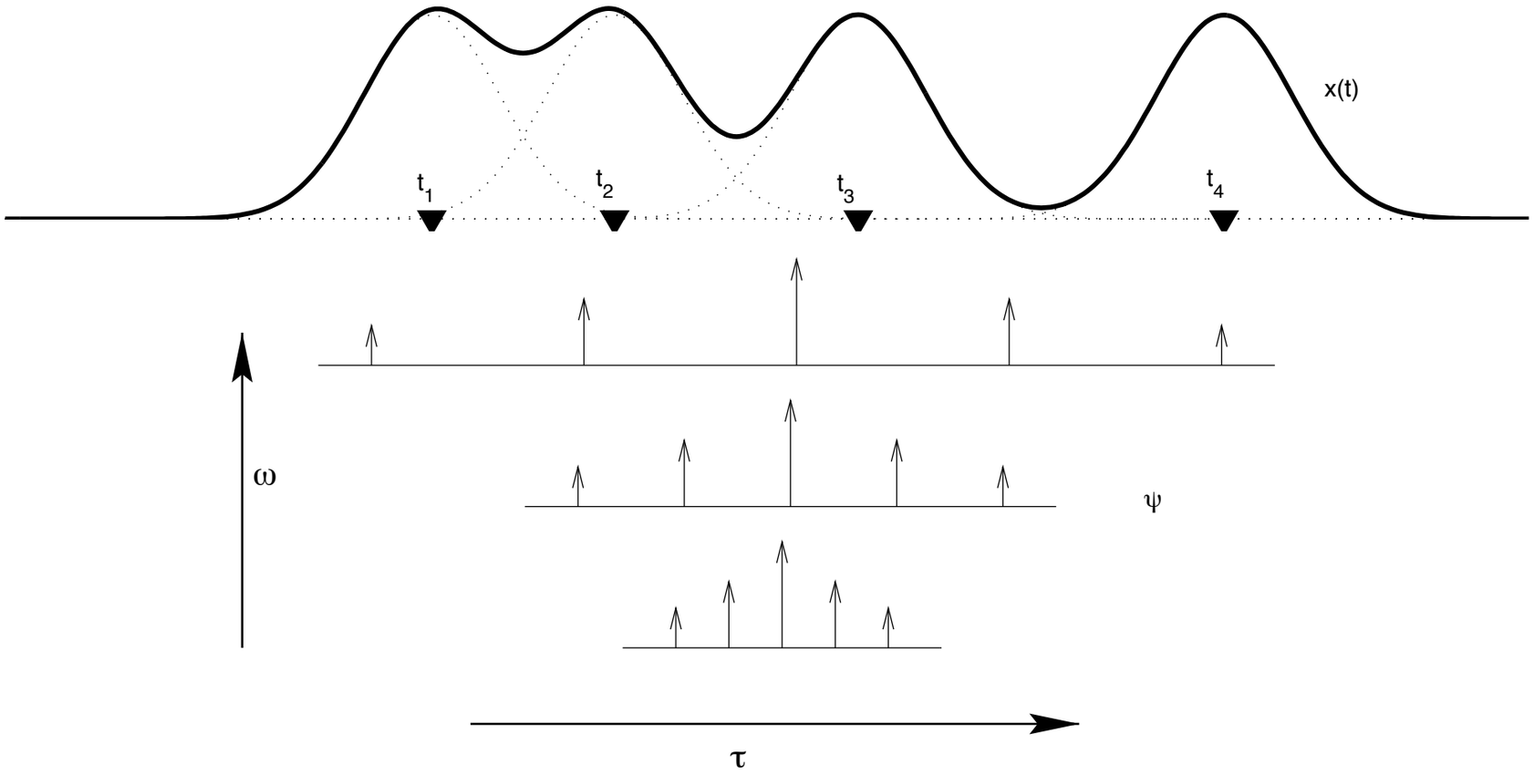
Tempogram

Kalman filter

(Cemgil et al. 2001)

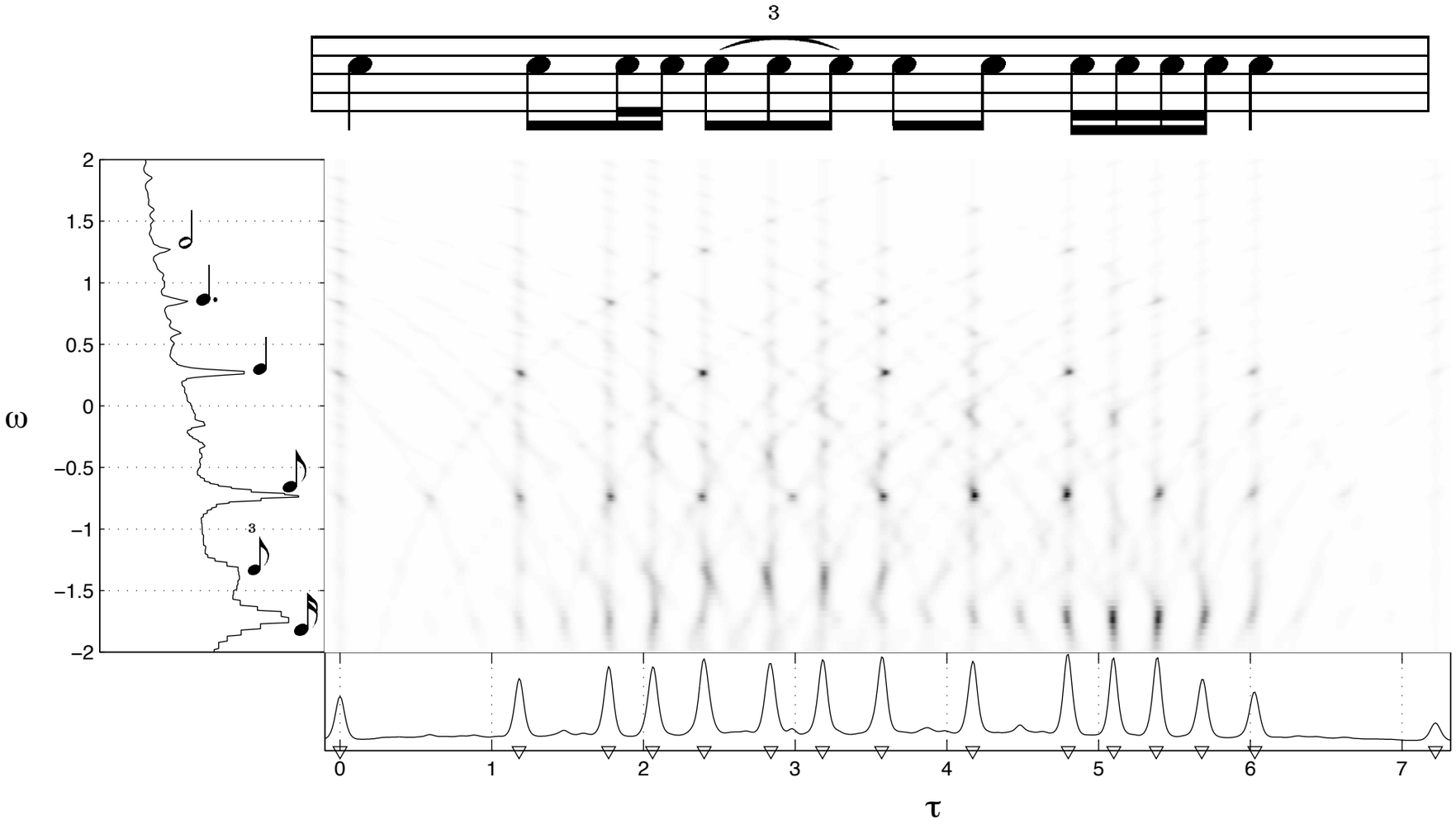
dynamical system: (hidden) state variables and transition rules

Tempogram



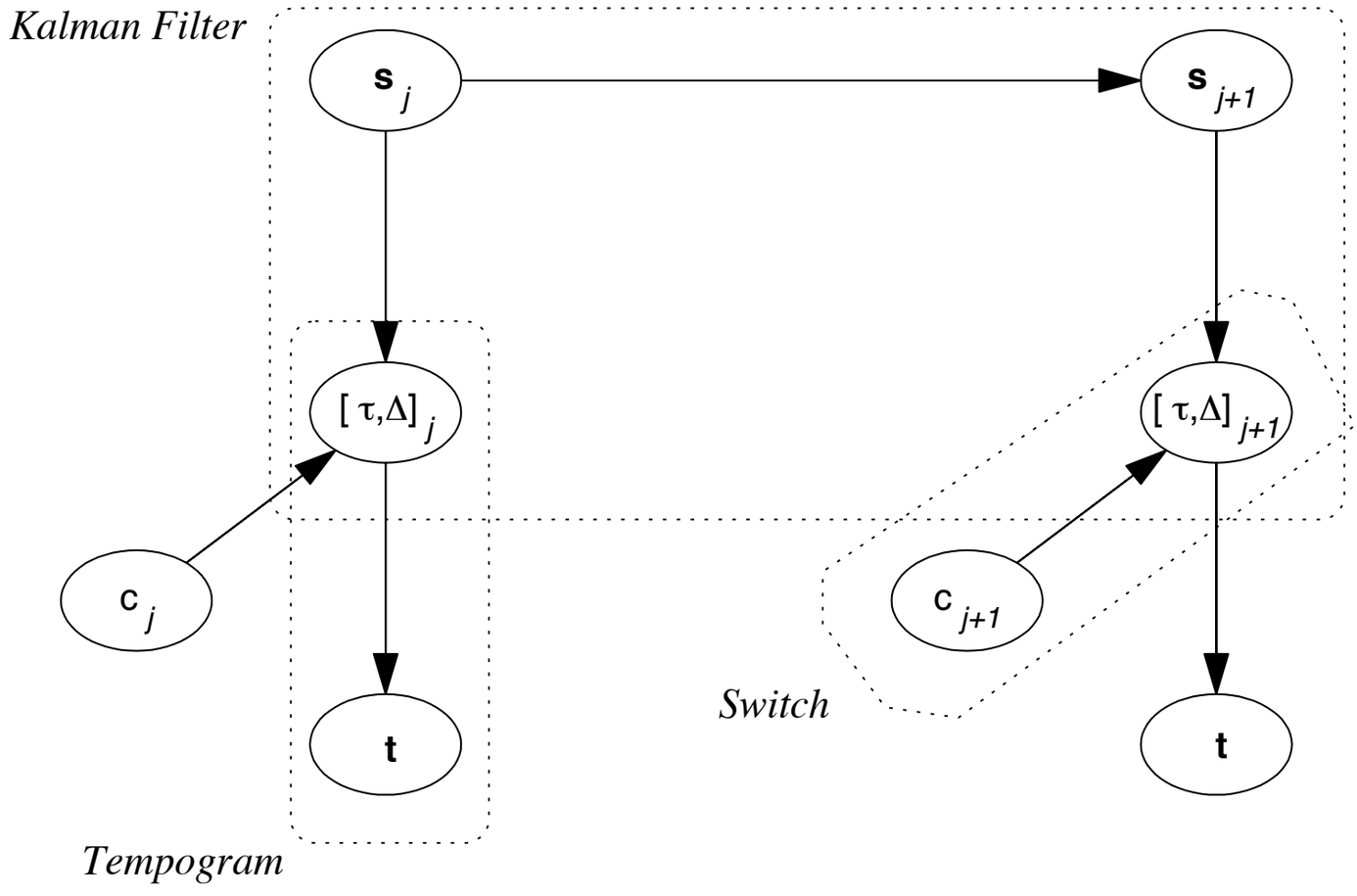
(Cemgil et al. 2001)

Tempogram



(Cemgil et al. 2001)

Kalman Filter



(Cemgil et al. 2001)

Expressive Performances

Onset Salience

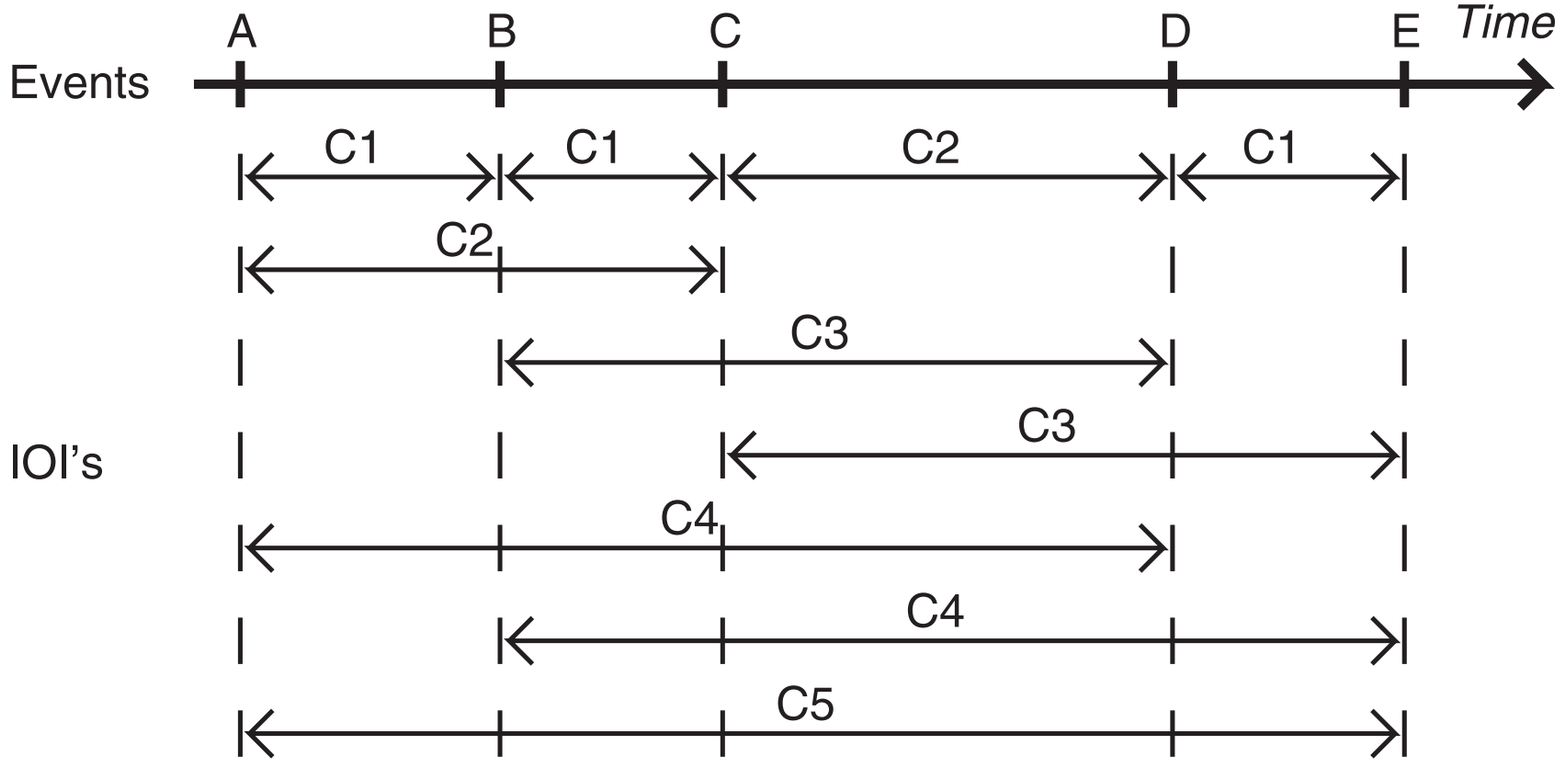
Onset Clustering

Multiple Agents

(Dixon 2001)

imperfect onset detection considered a filter

IOI Clustering



(Dixon 2001)

Define IOI
Overlapping
Clusters C1, C2, C3, C4, C5 of similar IOI length

Agents

Clusters become hypotheses

Agents created for each hypothesis

Predict

Evaluate

(Dixon 2001)

Beat Tracking with and without Drums

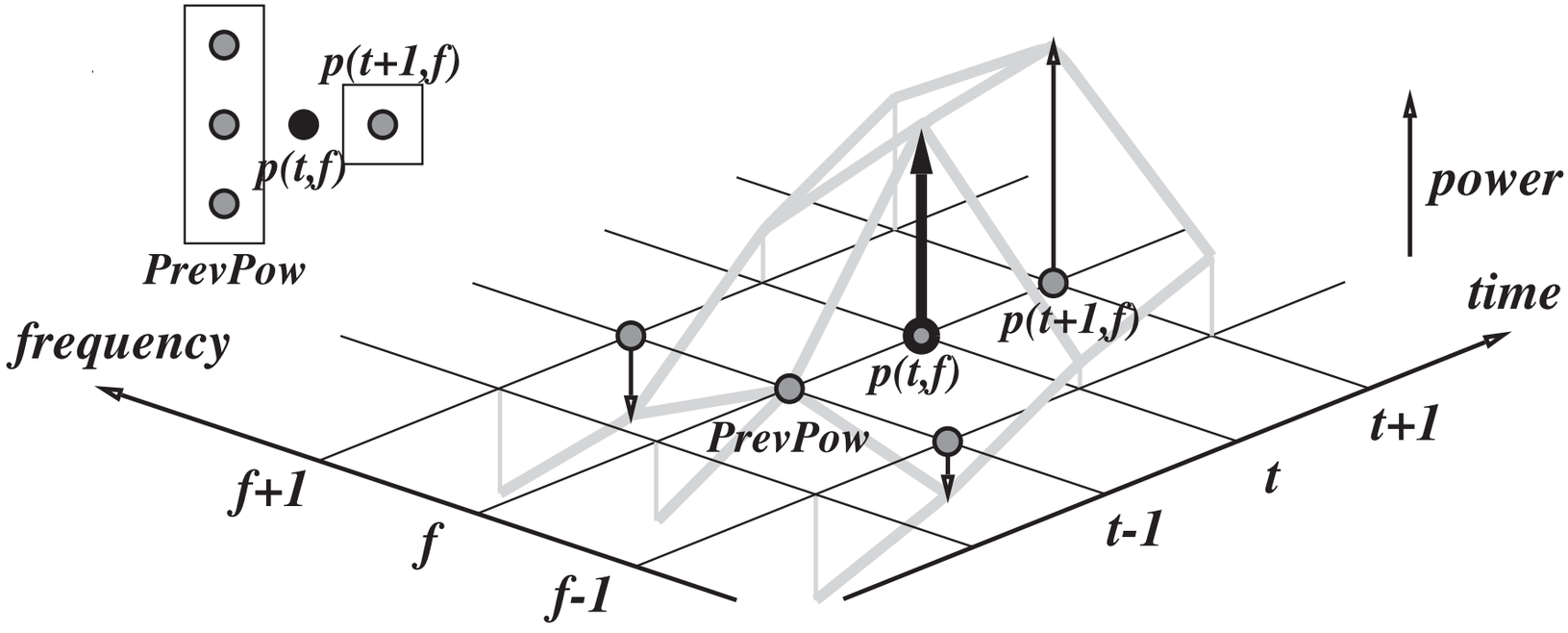
Onsets

Chord Changes

Drum Patterns

(Goto 2001)

Onset Detection in Context



(Goto 2001)

f are bands
t are frames
prevpow – power spectrum of previous frame in same and neighboring bands
p(t+1,f) is the same band in the next frame

Drum Patterns

Bass drum - low frequency

Snare drum - broadband

Find patterns

Metrical structure

(Goto 2001)

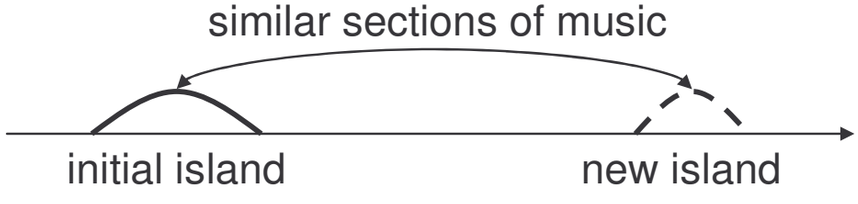
Holistic Beat Tracking

Holistic - metrical structure

Simple initial hypothesis

Self-similarity matrix

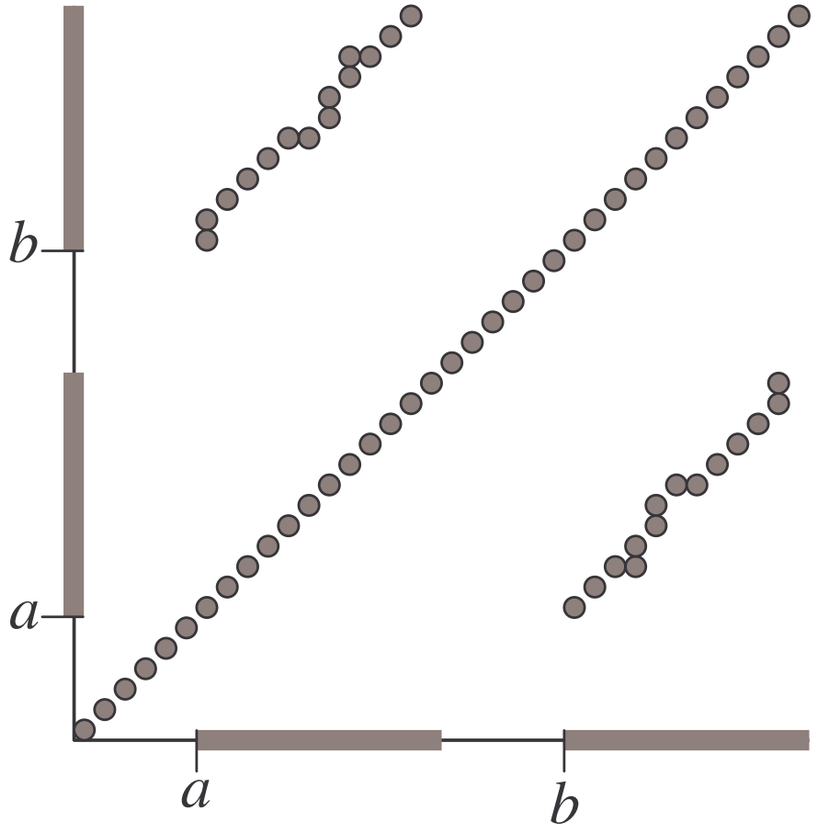
Islands of Tempo



(Dannenberg 2005)

He implemented a basic beat tracker, then added holistic information and found improvement

Self-Similarity Matrix



(Dannenberg 2005)

song segments a, b
diagonal where moment x is similar to itself

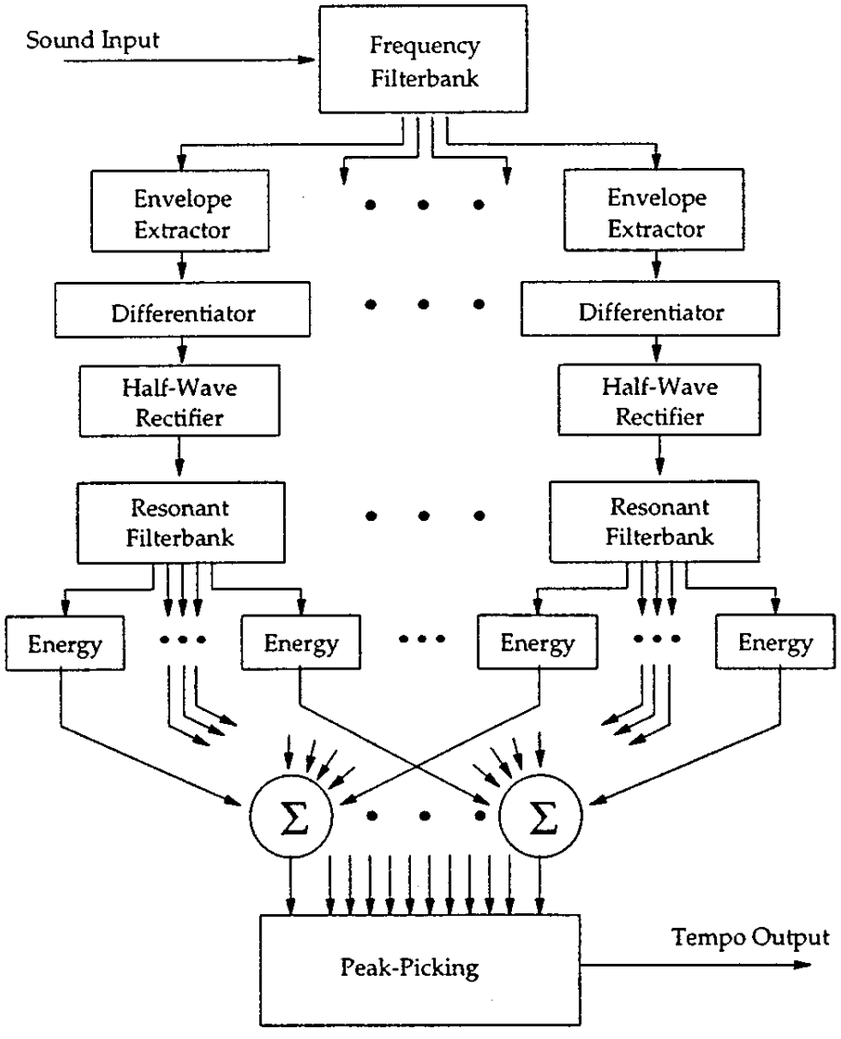
Perceptual vs. Transcriptive

Onsets: detect or not?

Implicit Rhythmic Events

scheirer's AM noise

Perceptual Beat Analysis



(Scheirer 1998)

The resonators are comb filters
6 bandpass filters

Perceptual Bayesian

Rhythm Tracks

Bayesian Analysis or Gradient Descent

(Sethares et al. 2005)

Particle Filters used for approximating intractable bayesian analysis

Evaluation

No common corpus

No ground truth

Confounding comparisons, e.g. viola/violin

(Temperley 2004)

Pitch

What?

Transcription

Chords

Inform holistic approaches

Holistic approaches like Dannenberg 2005 which we have just seen

How?

Rule-based

Probabalistic/Learning

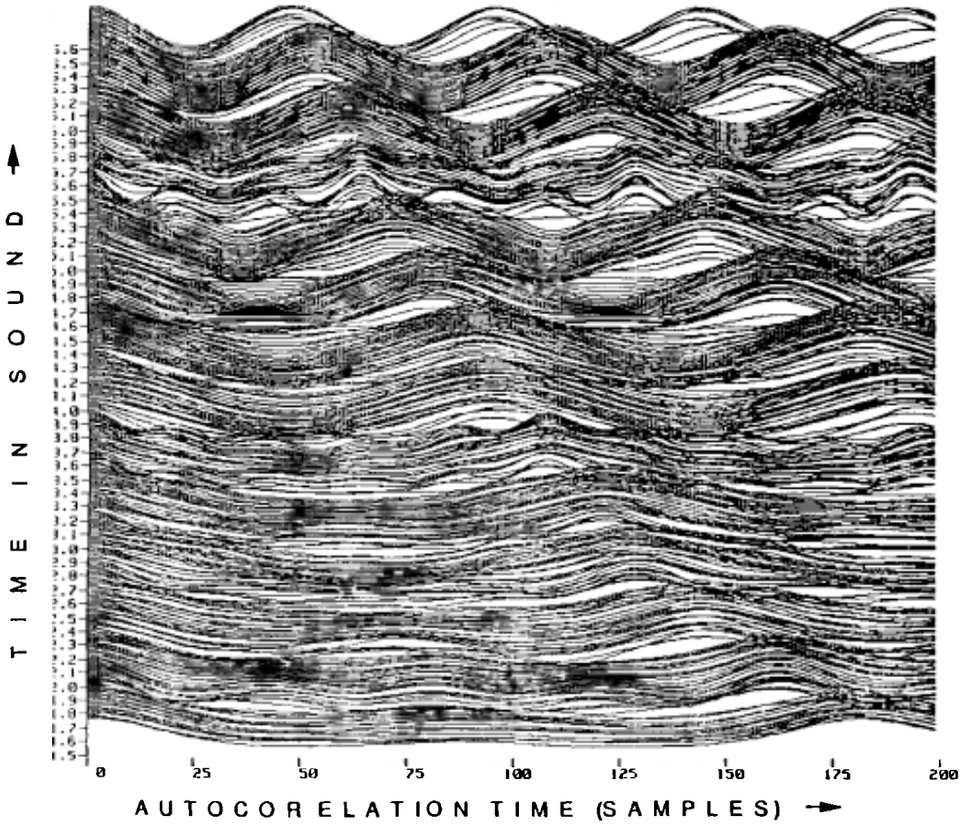
Transcription

Just pick the frequencies from the spectrum, right?

Transcription

- Complications
  - Complex and changing timbre
  - Noise
  - Polyphonic Confusion
  - Rhythm

Autocorrelation

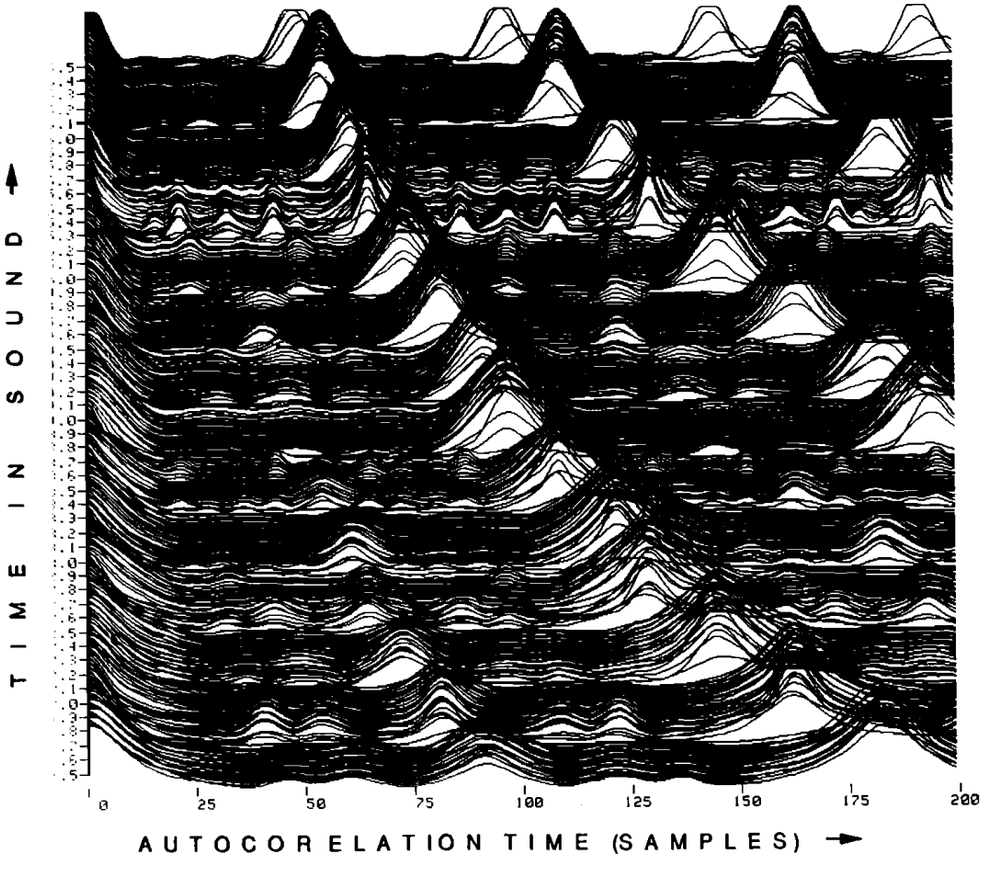


$$r_{xx}(n) = \frac{1}{N} \sum_{t=1}^{N-n-1} x(t)x(t+n)$$

(Brown and Zhang 1991)

time-domain
c major scale on piano

Narrowed Autocorrelation



$$S_N(\tau)^2 = \langle |f(t) + f(t-\tau) + f(t-2\tau) + \cdots + f(t-(N-1)\tau)|^2 \rangle$$

(Brown and Zhang 1991)

Where <|x|^2> seems to mean <x,x> (inner product)
same scale
note logarithmic nature
Neither is suitable for polyphony

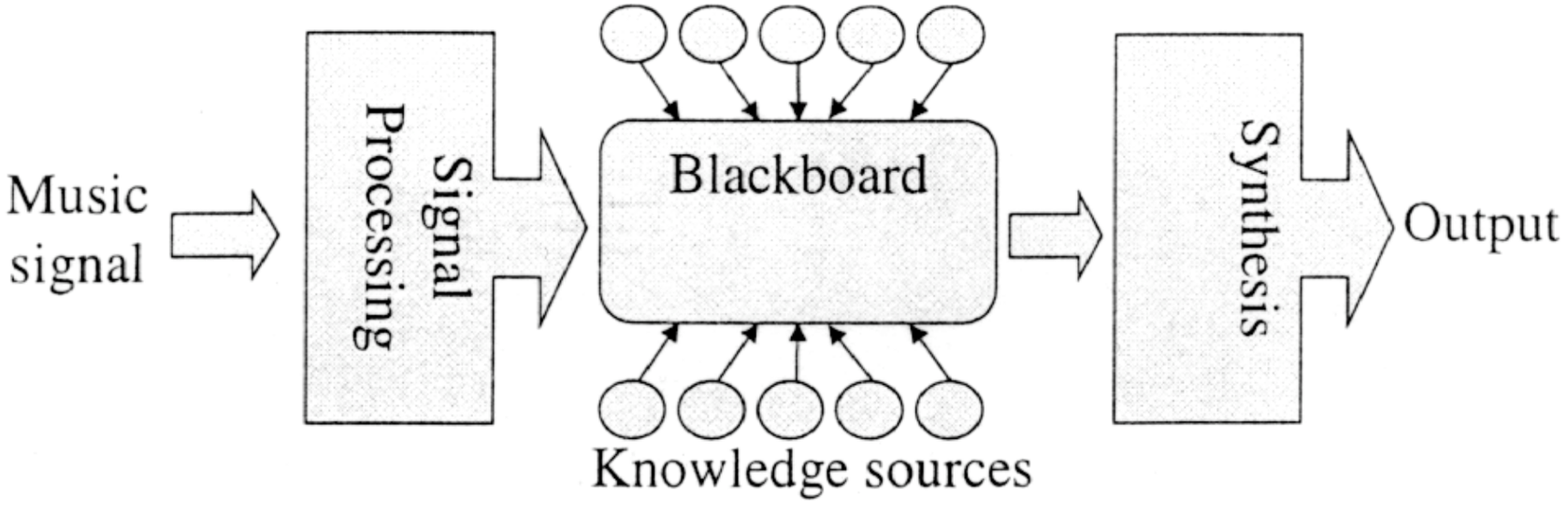
Transcription

approaches:

- Knowledge-based Blackboard
- Multiple-Cause Model
- Independent Component Analysis

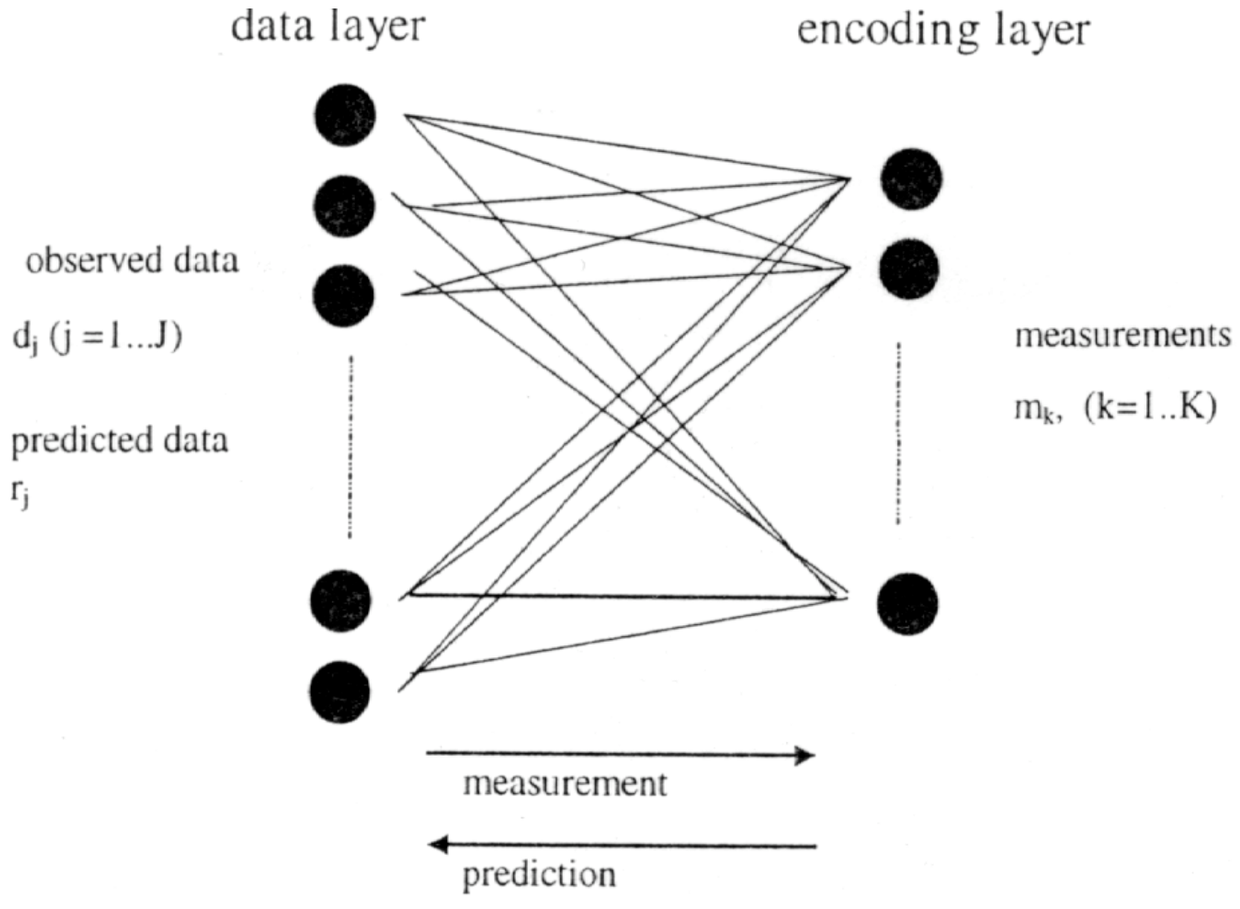
(Plumbley et al. 2002)

Blackboard



(Plumbley et al. 2002)

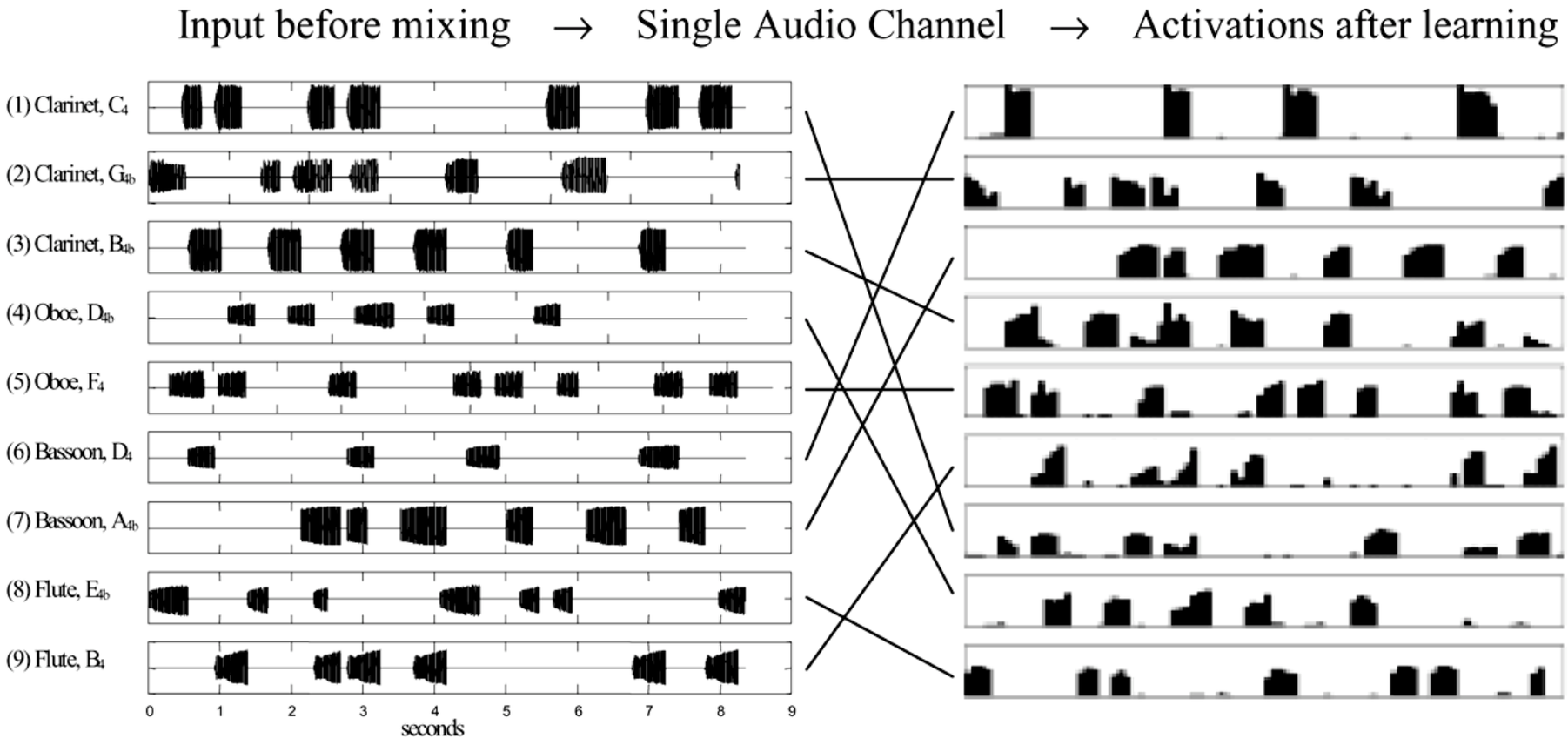
Multiple-Cause Model



(Plumbley et al. 2002)

neural network
feeds forward to give measurements, then backward to predict
weights adjusted by minimizing prediction error

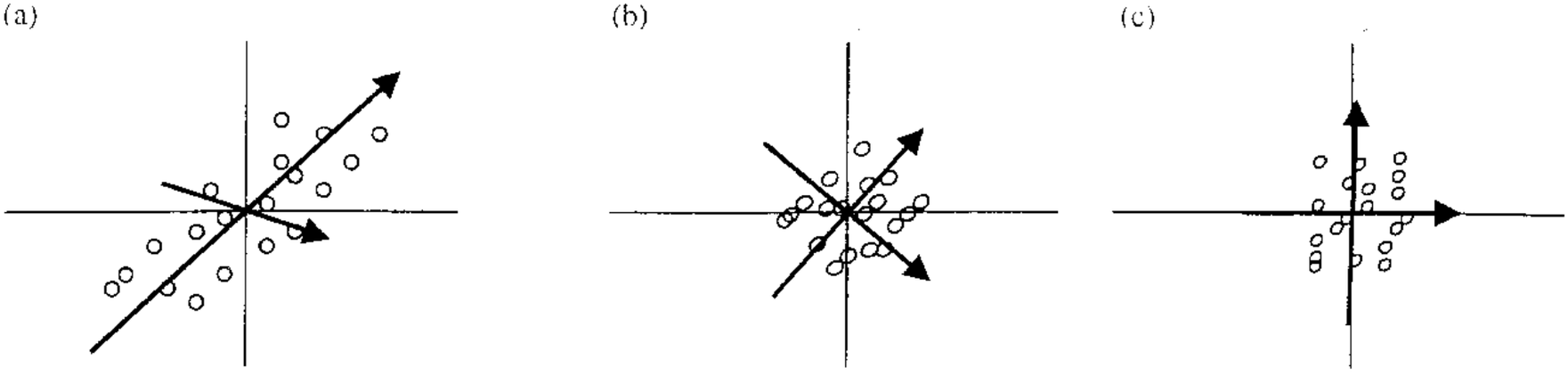
Multiple-Cause Model



(Plumbley et al. 2002)

Mapping unknown
flute/clarinet confusion, because of similarities?

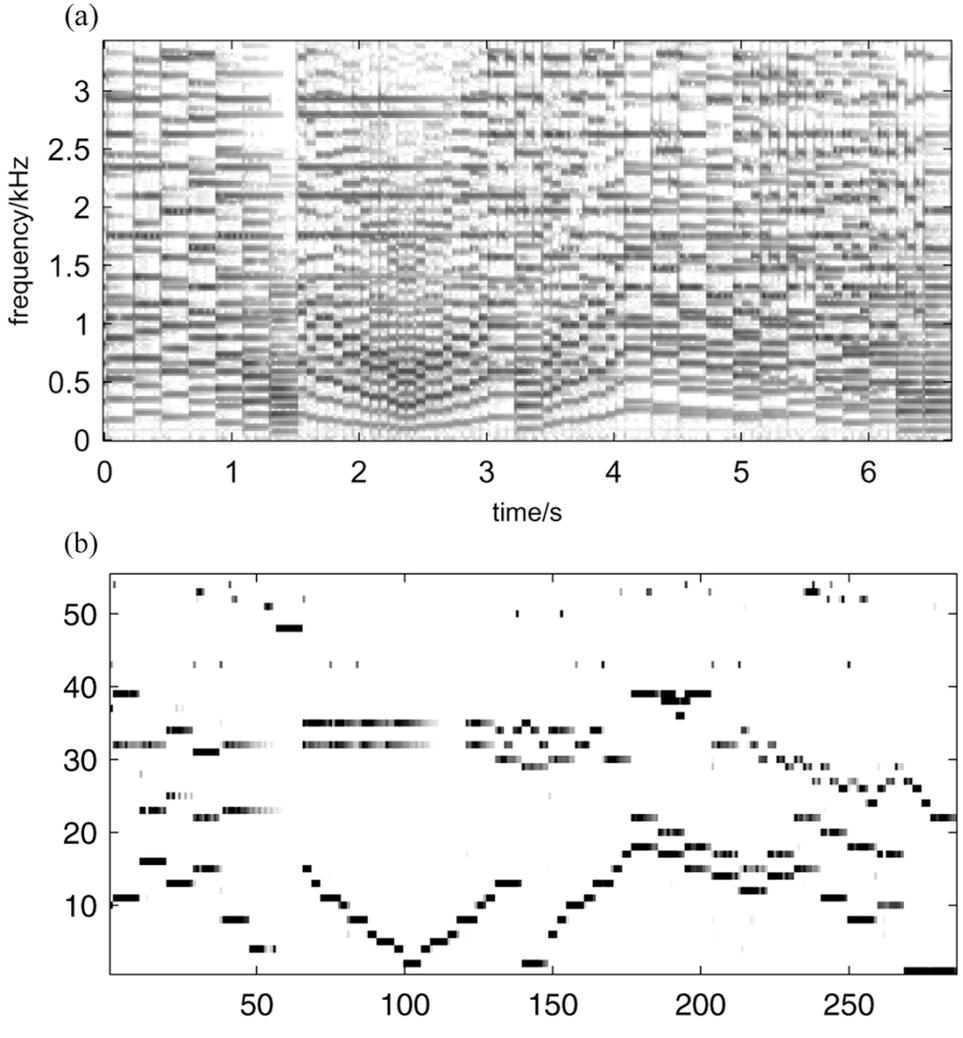
ICA



(Plumbley et al. 2002)

Independent Component Analysis
a) original data
b) after whitening
c) after rotation
Often used in source separation

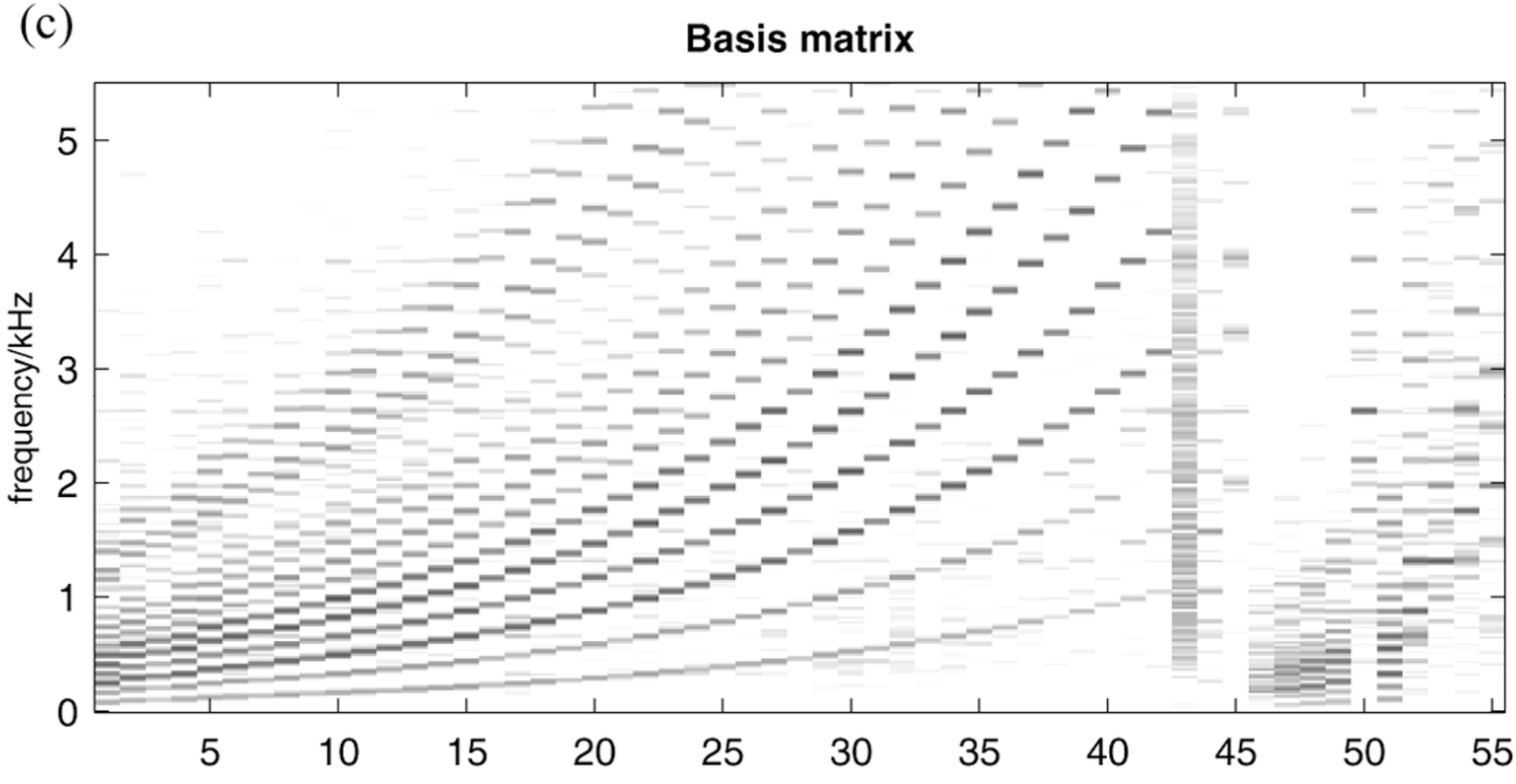
ICA



(Plumbley et al. 2002)

Plumbley et al., each note a source
Bach Partita
Spectrum and output note activations

ICA



(Plumbley et al. 2002)

Note Shapes
Ordered Manually
Clear harmonic structure below 43
above 43 probably transients

Generative Model for Transcription

Graphical (dynamical) model

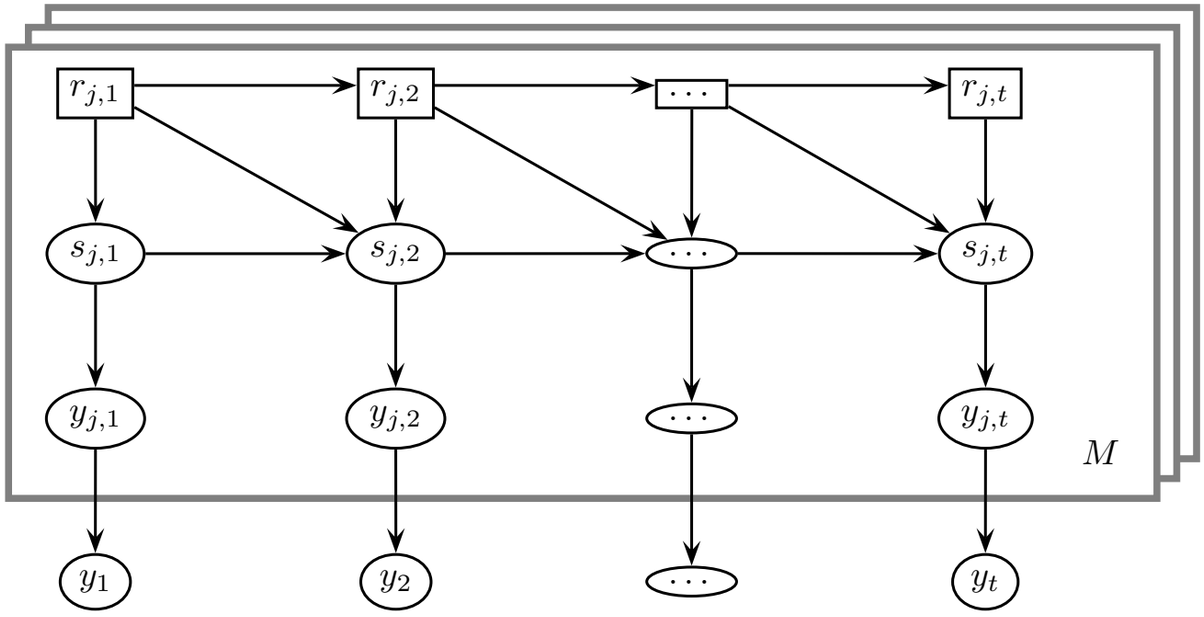
Bayesian Network

Models sound generation

Kalman

(Cemgil et al. 2004)

Bayesian Network



$$P(r|y) \propto P(y|r)P(r)$$

(Cemgil et al. 2004)

Predict from state + piano roll
M plates (j=1..M), one for each sound generator
time t
r for roll, s for state
sum sound generators together at each **sample** to get y
P(y|r) from the generative model (Kalman filter used to find hidden states in box)
P(r) prior can incorporate musical knowledge

Graphical Model

For sung melodies

Joint model of pitch, rhythm, tempo, segmentation

Event List as a Markov Chain

Optimal parameters of model via dynamic programming

Branch and Prune

(Raphael 2005)

Markov Chain

Sequence of Random Variables

Given present, future is independent of past

i.e. only need the last observation

(Raphael 2005)

Audio thumbnails

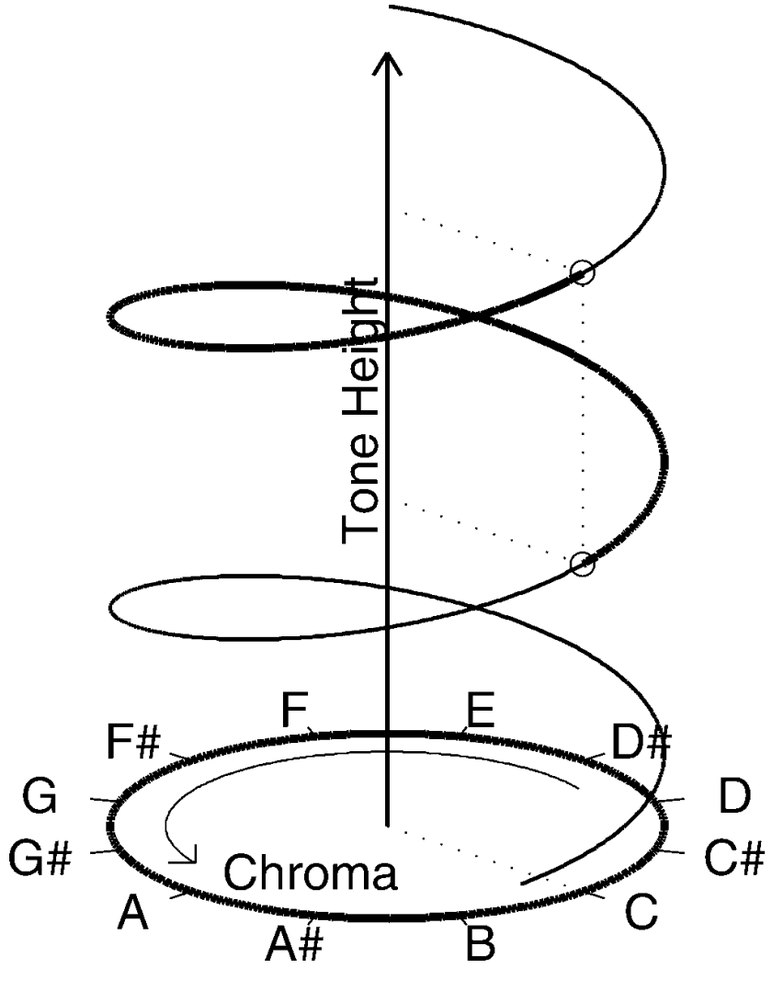
Chromagram

Similarity Matrix

Most Strongly Repeated section

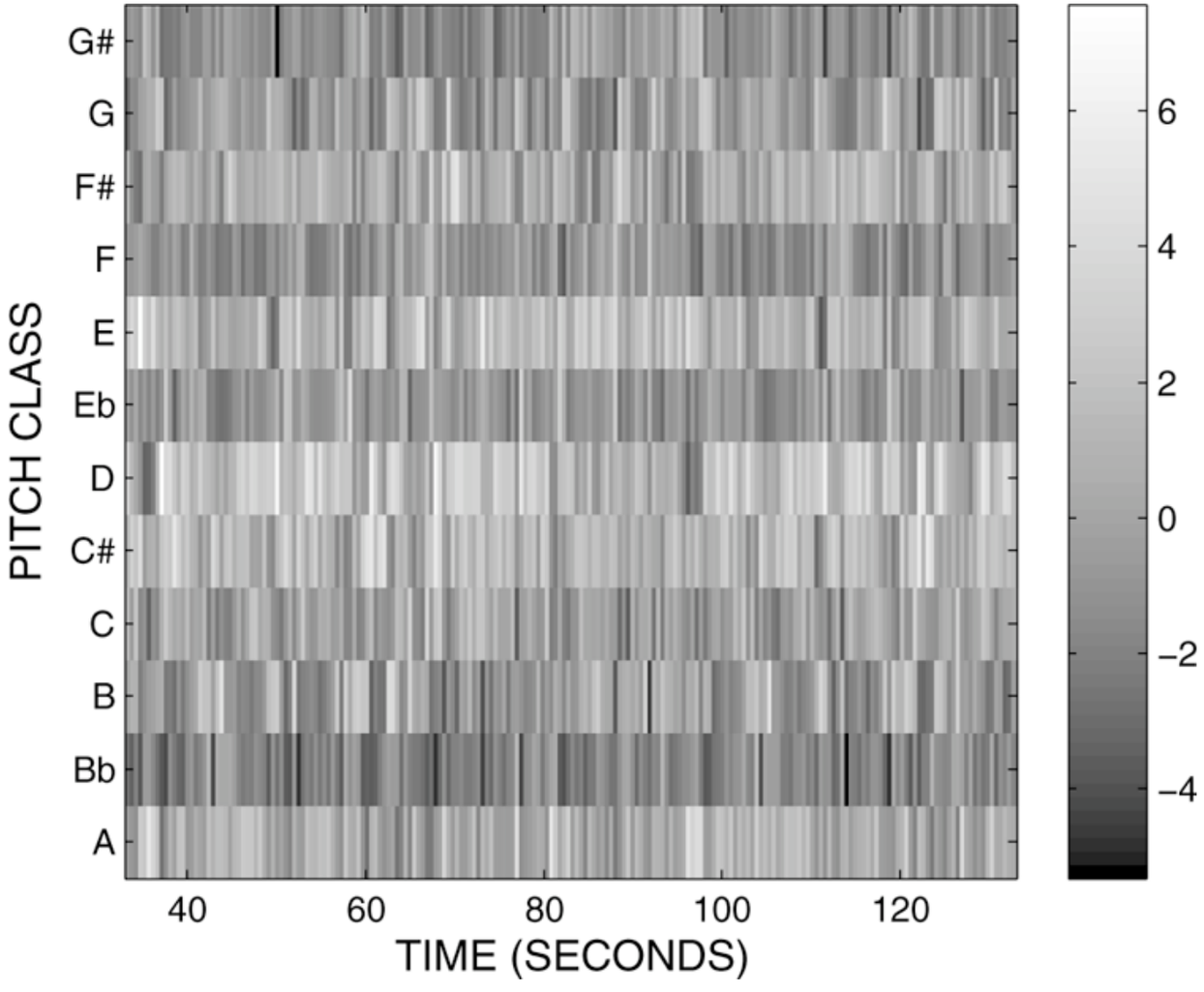
(Bartsch 2005)

Chroma



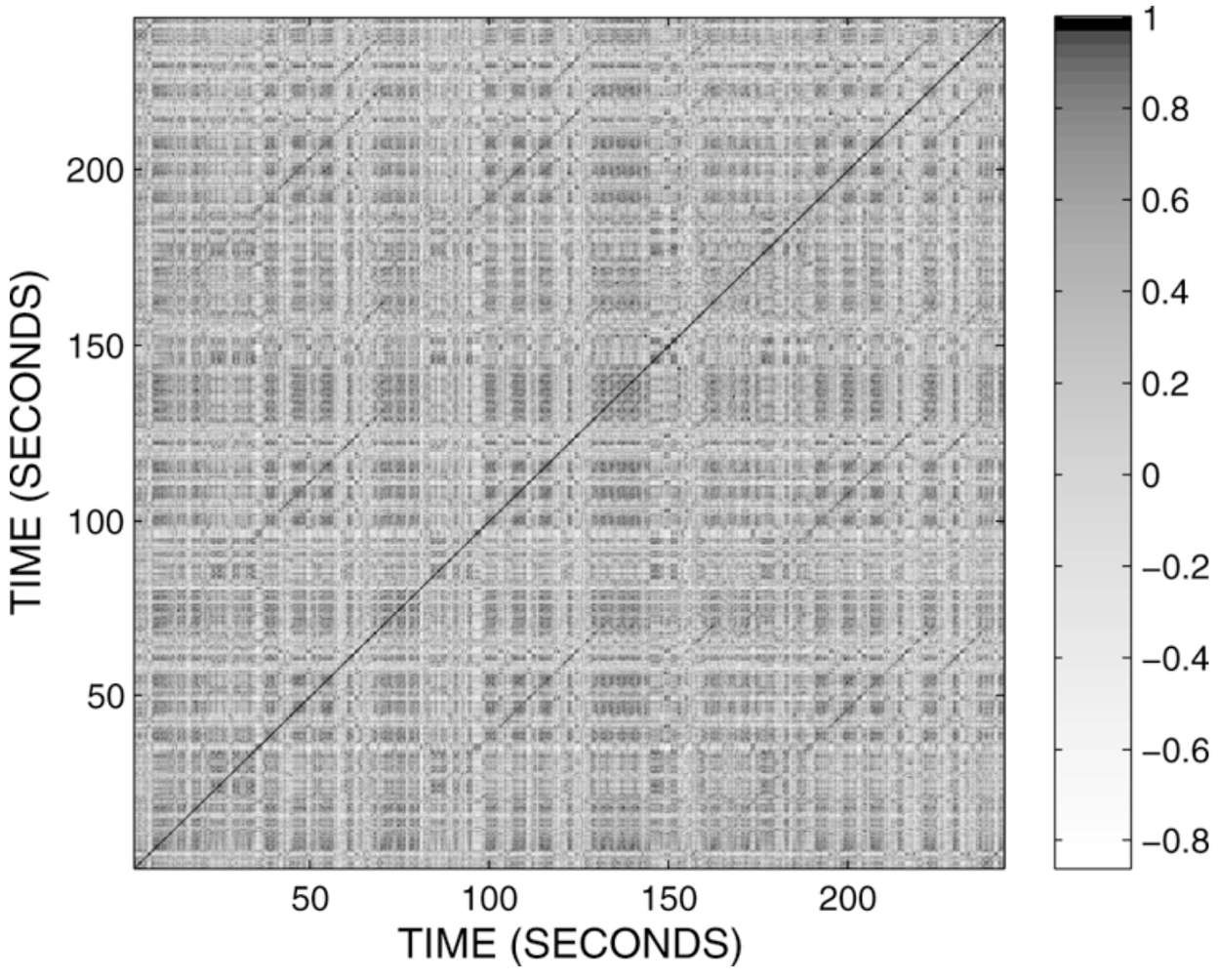
(Bartsch 2005)

Chromagram



(Bartsch 2005)

Similarity matrix



(Bartsch 2005)

Key, Chord, and Rhythm

Holistic: Key, Chords, Rhythm

Rule-based

(Shenoy and Wang 2005)

Steps

Beat Tracking

Audio Segmentation

Chroma Features Extracted

Chord and Key

Eliminate non-key chords

Inconsistent/Missing Chords

Rhythm/Metrical considerations

(Shenoy and Wang 2005)

Many simplifying assumptions made (4/4, replacing chords willy-nilly), but performs promisingly

Conclusion

Monophonic is workable

Polyphonic is hard

Probabalistic/Learning methods more robust

Timbre

What?

Instrument Identification/Classification

Instrument Separation

Genre Classification

How?

Overwhelmingly by Machine Learning

Some rule-based systems

Machine Learning

Learning Cycle

Observation

Action

Evaluation

Adjustment

Problem Formulation

Important!

Sufficient and representative data

Evaluable and useful task

Proper Evaluation, ensure generalization

Instrument Identification

Monophonic or Polyphonic

Single note or whole passages

Synonyms recognition and classification, though the latter implies categorization or may be unsupervised

Steps

Extract Features

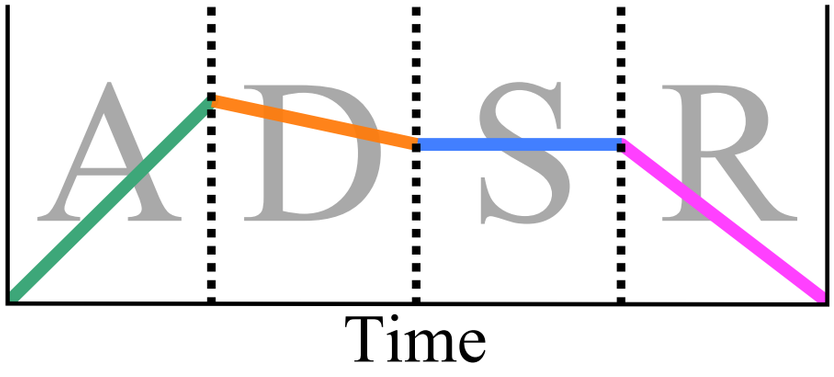
Apply Classifiers

Select Relevant Features/Classifiers

Features

- Amplitude
- Energy
- Zero-Crossing Rate
- Frequency
- Spectral Shape
- Harmonics
- Inharmonics
- Cepstral Coefficients
- Bark, ERB
- Correlogram

Amplitude



Attack, Decay, Sustain, Release

Volume

Frequency

Fundamental (F0)

Range

Vibrato

Spectral Shape

Centroid

Spread

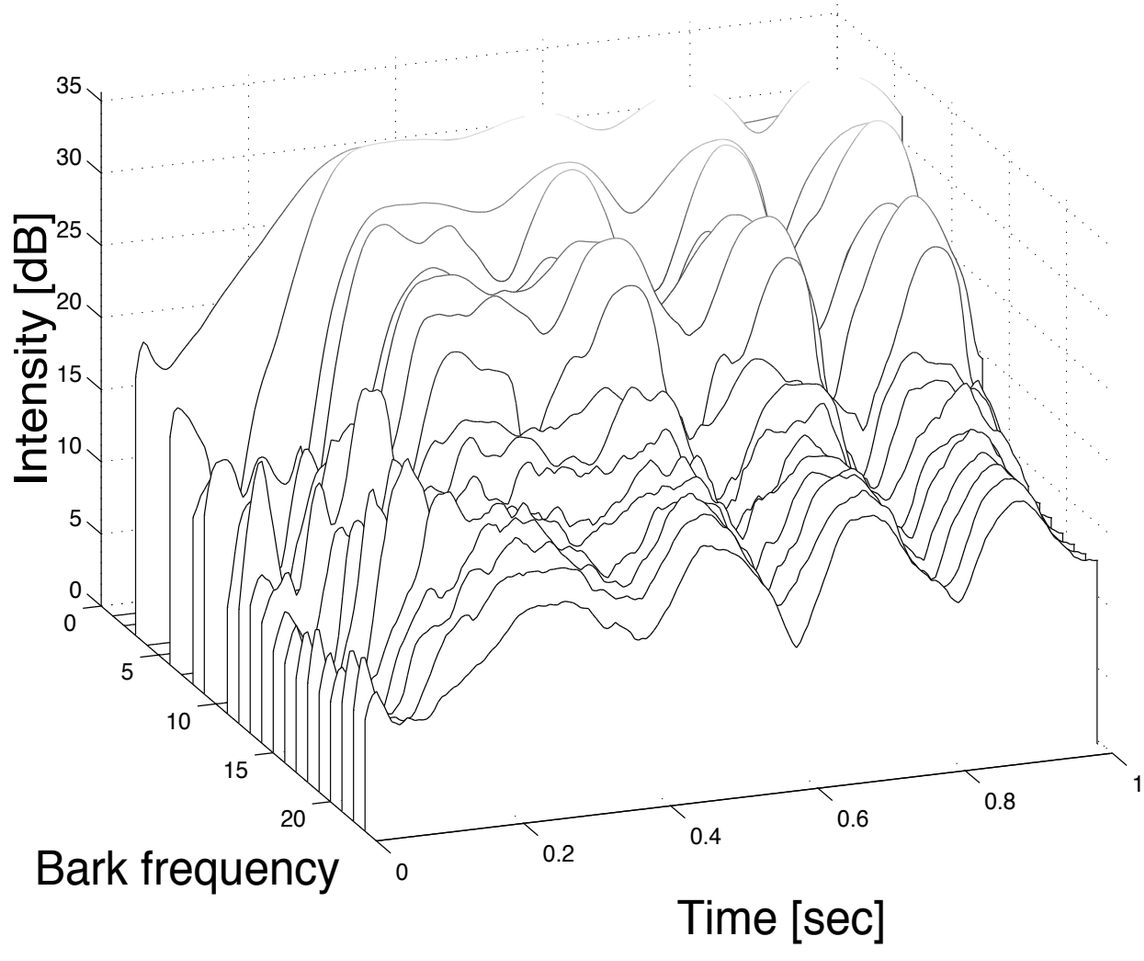
Flatness

Flux

Skew

centroid like mean of normal, most popular

Sinusoid Envelopes



(Eronen 2001)

Like Fourier, but bins calculated independently
sample-accurate
useful for examining spectral evolution
That's a flute

Cepstral Coefficients

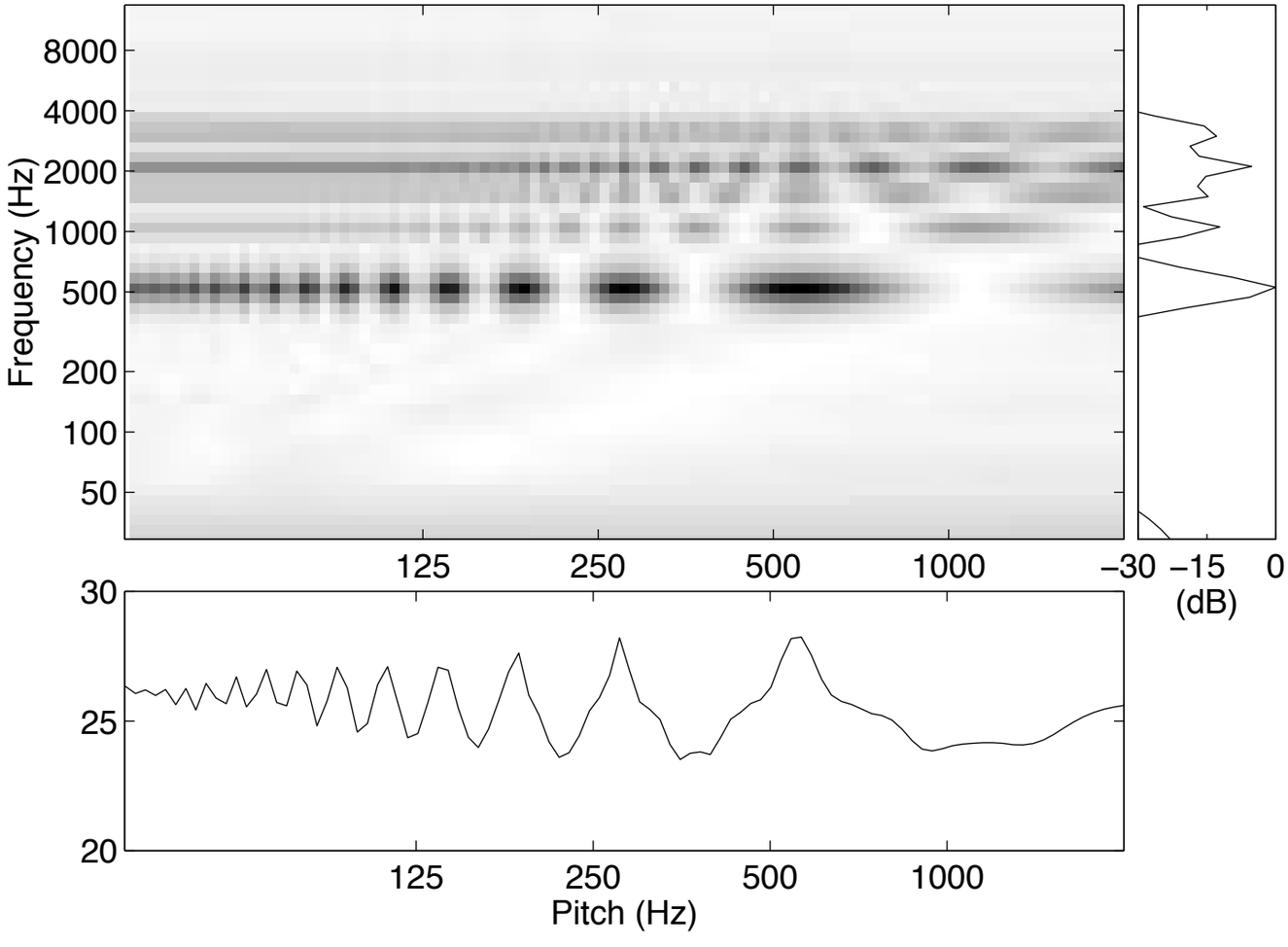
$$\boldsymbol{c} = \mathcal{F}^{-1}\left(\log\left|\mathcal{F}(\boldsymbol{x})\right|\right)$$

Usually in conjunction with the mel scale (MFCC)

$$m(f) = 2595\log(1 + f/700)$$

very popular
conceptually, information on the rate of change of the spectrum

Correlogram



(Martin and Kim 1998)

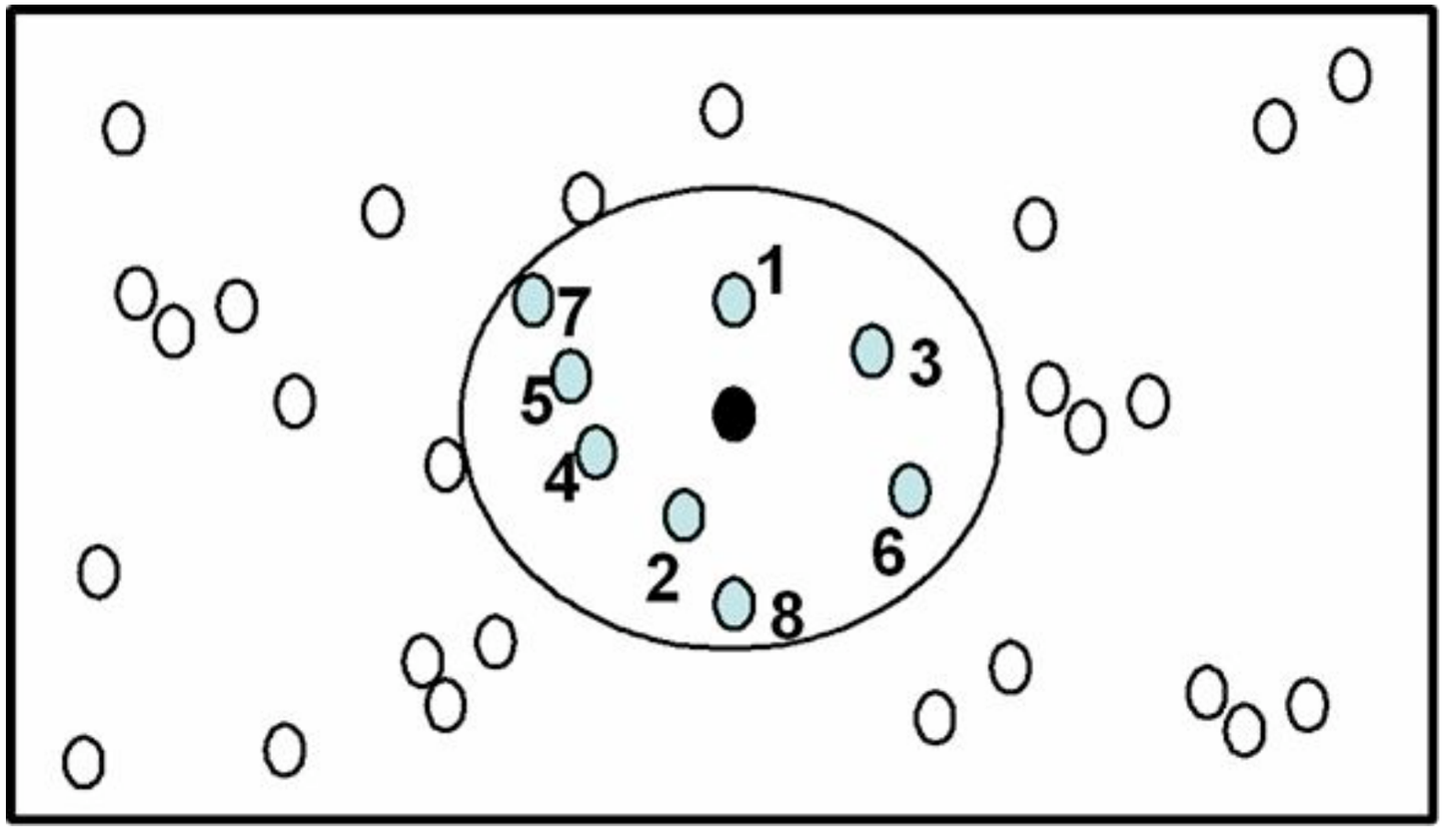
"not based on the assumption that the signal is periodic"
modeled after cochlea and inner hair cells
filtered into channels and autocorrelation
that's a violin

Classifiers

- k-NN
- Naive Bayes
- Discriminant Analysis
- Decision Trees
- Support Vector Machines
- Artificial Neural Networks
- Rough Sets
- Hidden Markov Models
- Gaussian Mixture Model
- Independent Subspace Analysis

GMM: linear combination of gaussians
SVM: combination of typical spectra and noise

K-Nearest Neighbors



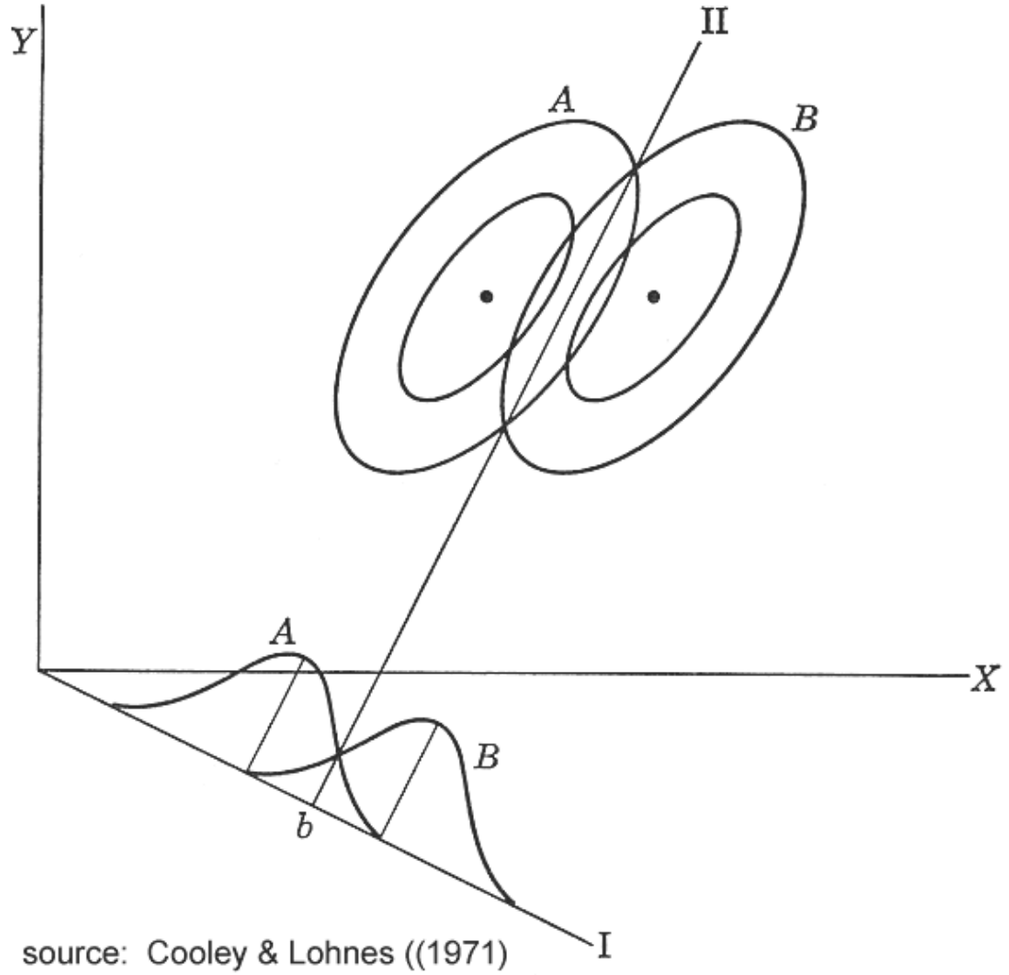
(University College Dublin)

Naive Bayes

V is set of classes, a are features

$$v_{\mathrm{NB}} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \prod_i P(a_i | v_j)$$

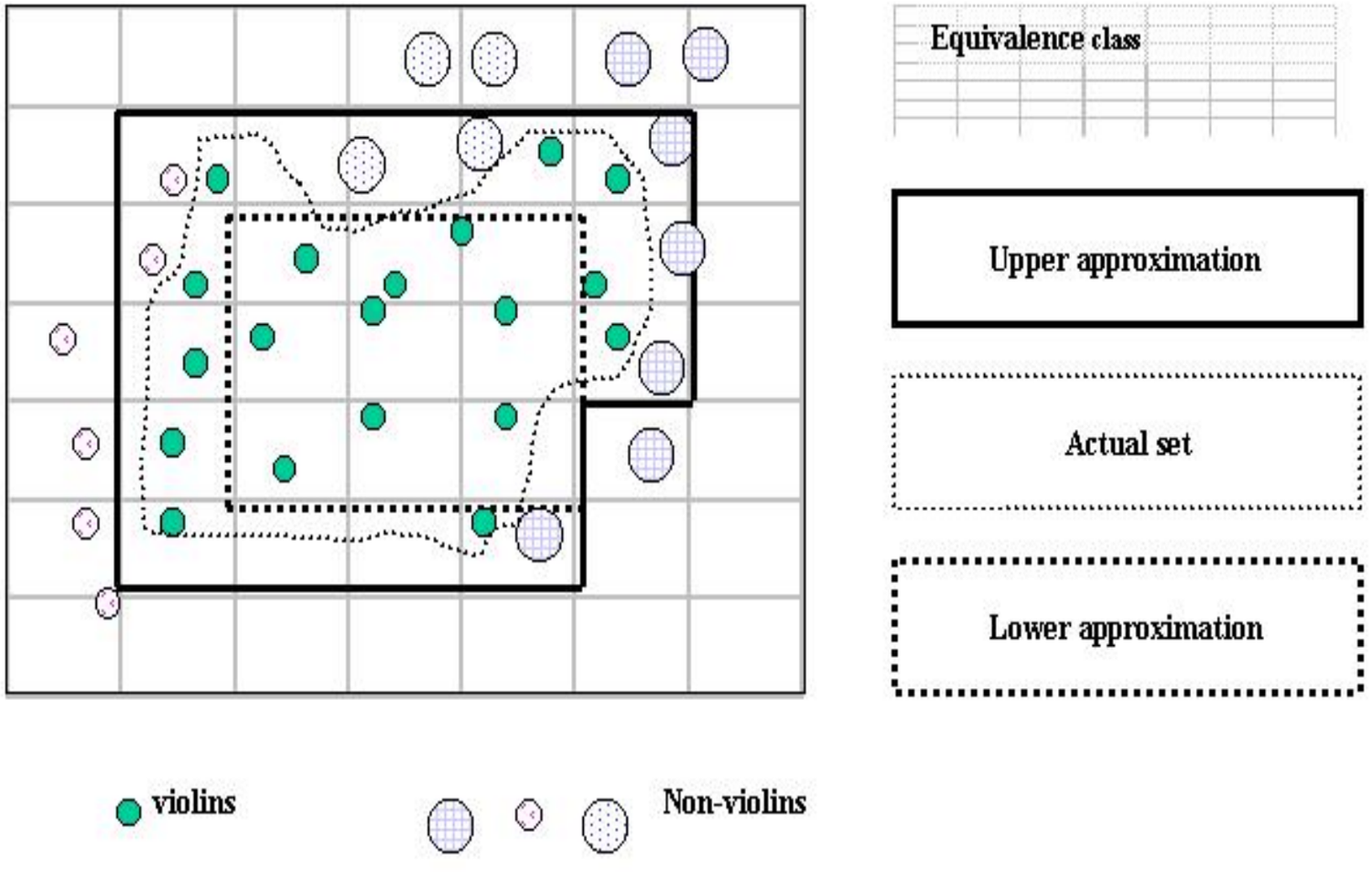
Discriminant Analysis



source: Cooley & Lohnes ((1971)

More frequently applied to selection

Rough Sets



(Herrera-Boyer et al. 2003)

Hidden Markov Model

Bayesian Network

State Transition Probabilities

Observation Probabilities

States, transition probabilities, observation probabilities hidden - only get observations

Support Vector Machines

Map data with kernel function

Find linear hyperplane to minimize error

Selection

Principal Components Analysis

Discriminant Analysis

Sequential Forward/Backward Generation

Gradual Descriptor Elimination

features and/or classifiers
SFG/SBG: add/remove most/least relevant features one at a time
GDE: choose most irrelevant feature using LDA and remove, estimate at every step

Status

Much monophonic work done

Much polyphonic work to be done

Shotgun approach most successful

Understanding

Spectral Anticipations

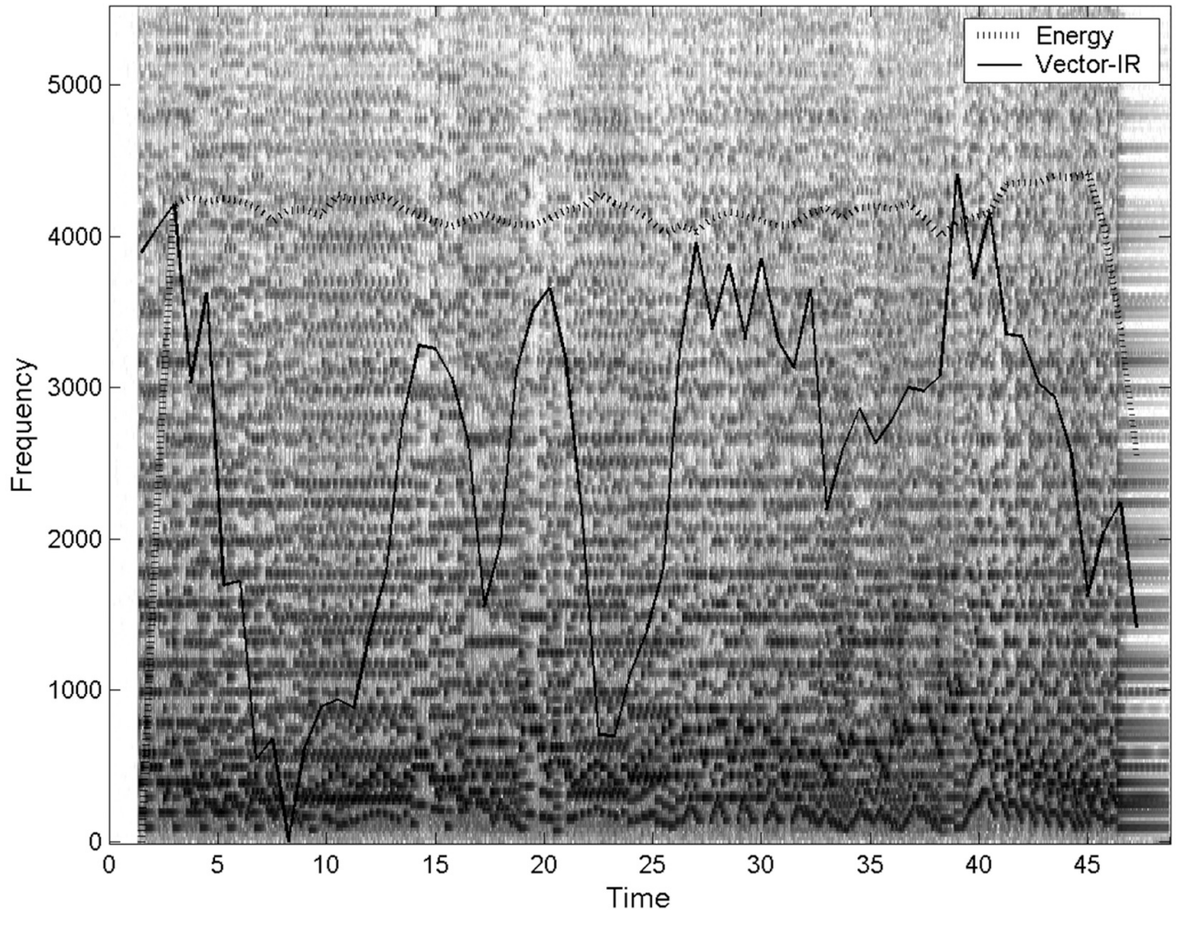
Inverted U: silence and noise are both boring

Structure carries information

Anticipation and Surprise

(Dubnov 2006)

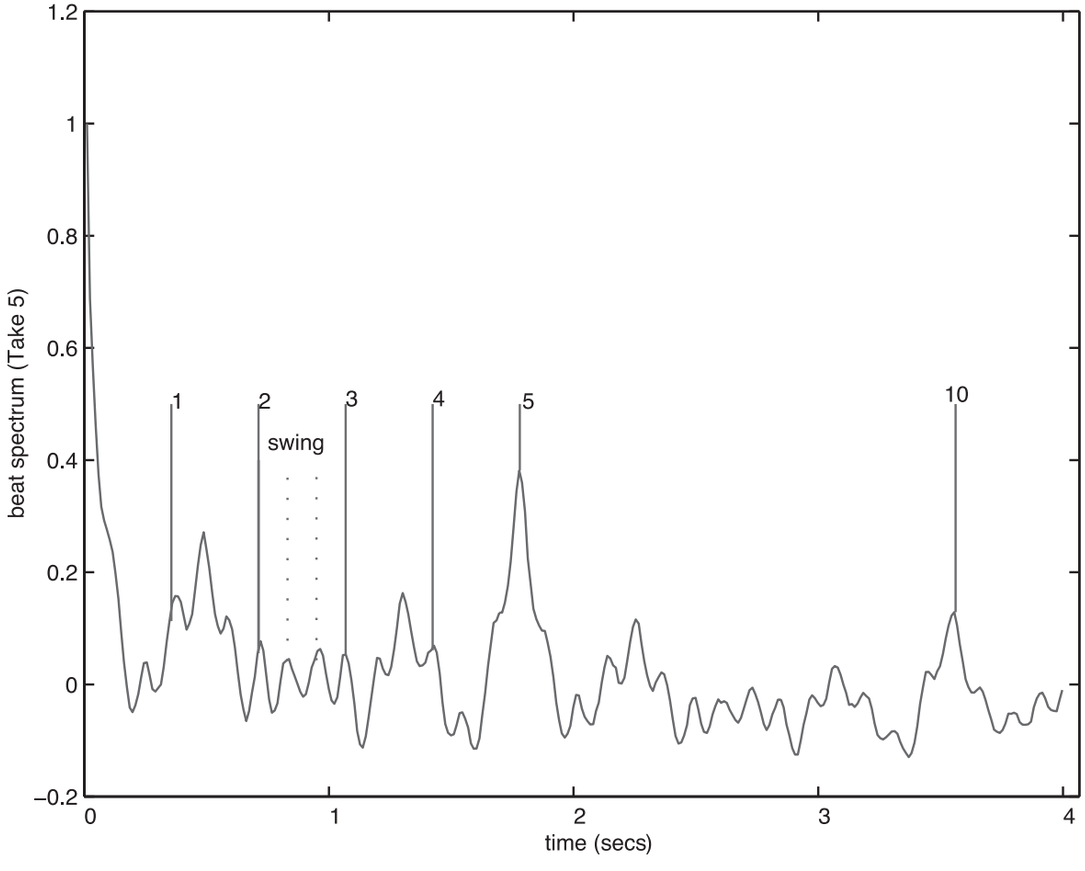
Spectral Anticipations



(Dubnov 2006)

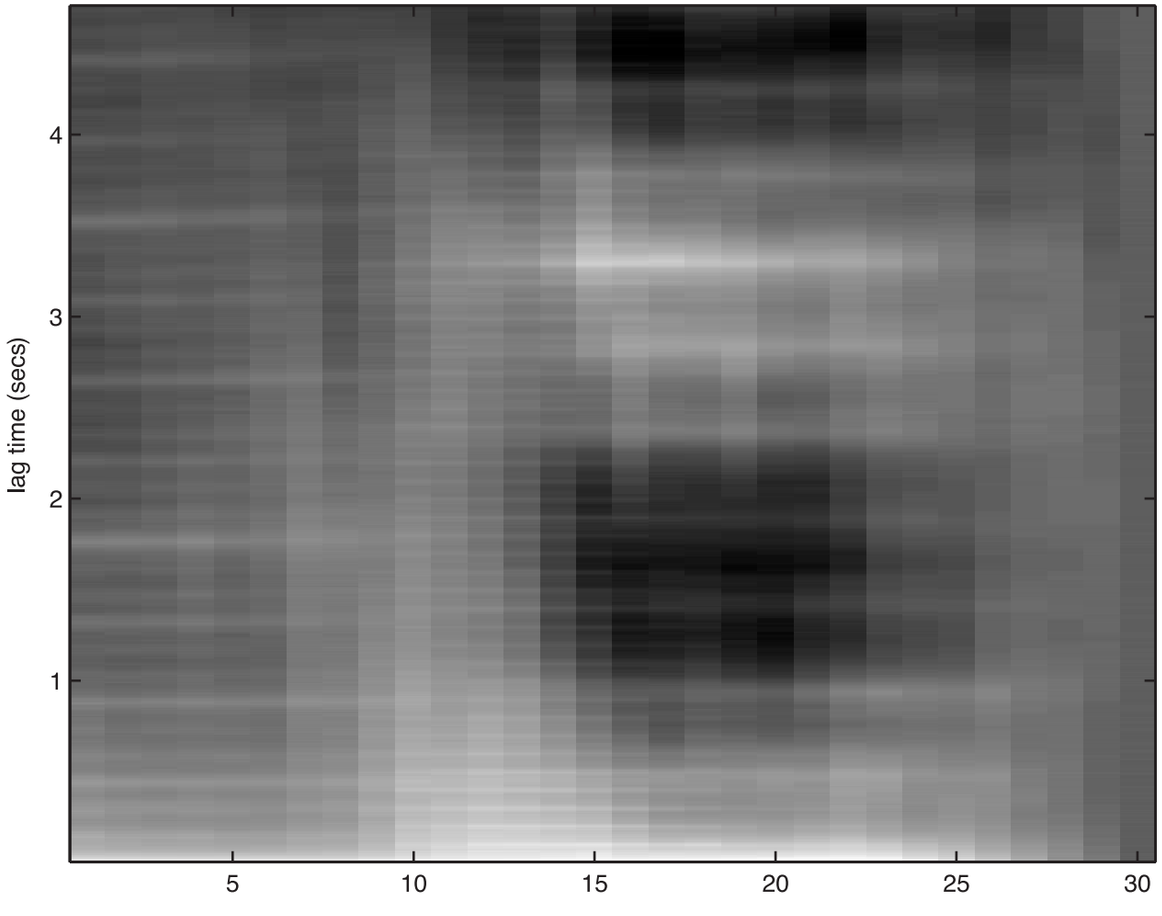
Energy doesn't say much, anticipation profile has more information

Beat Spectrum



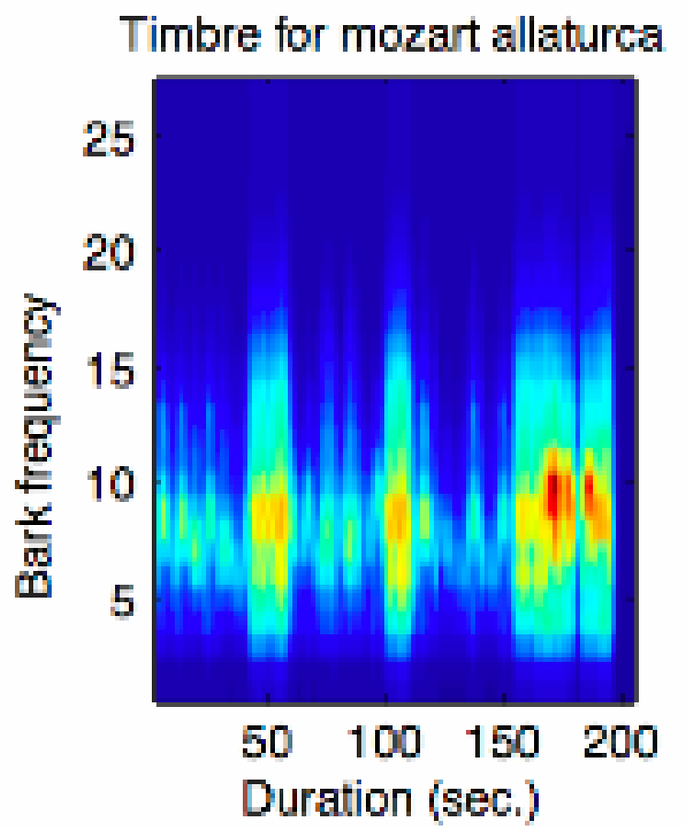
(Cooper et al. 2006)

Beat Spectrogram



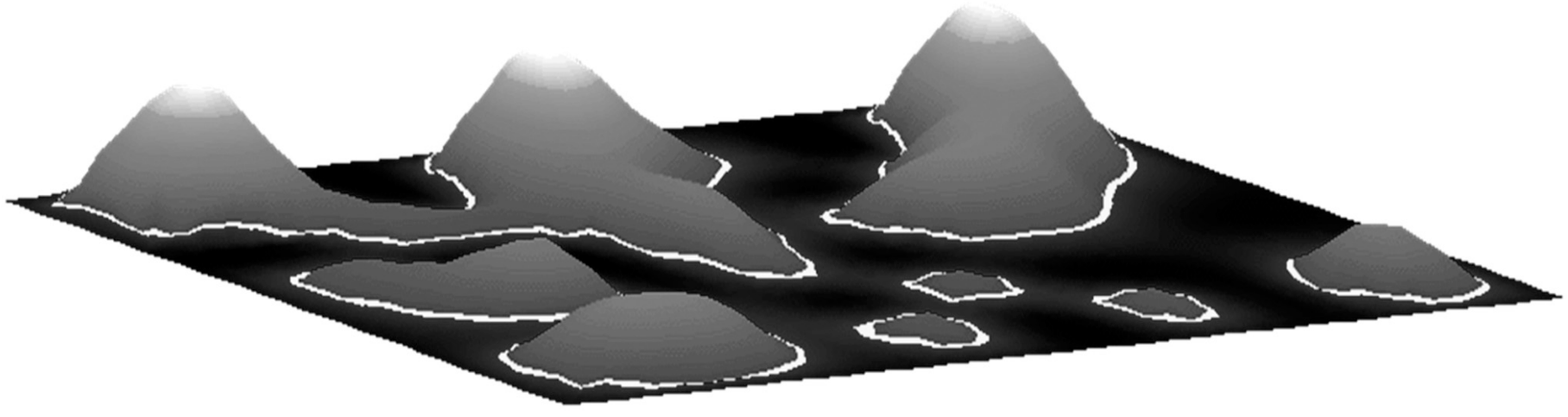
(Cooper et al. 2006)

Timbregram



(Kuhl and Jensen 2008)

Islands of Music



(Cooper et al. 2006)

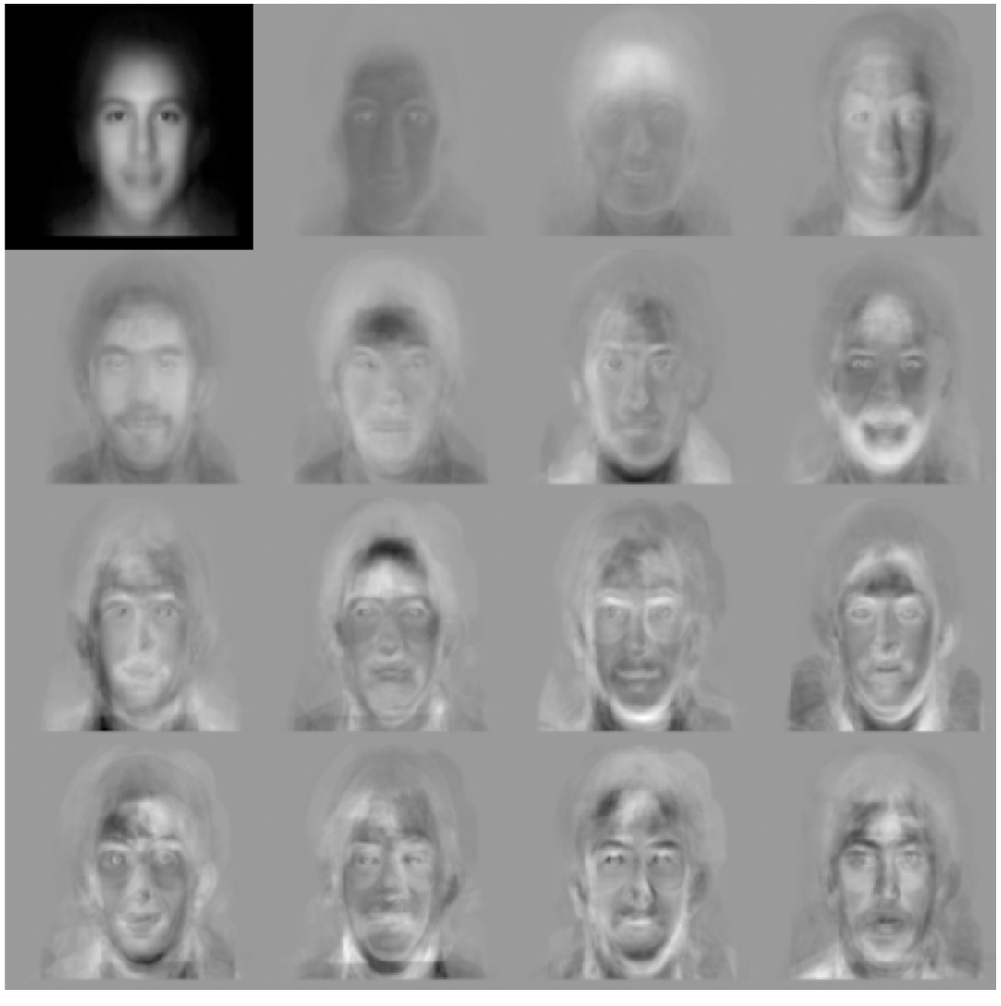
Classification via Visualization

Spectrogram and MFCC

Texture-of-textures

(Deshpande et al. 2001)

Eigenfaces



(Zhao and Chellappa 2006)

Eigensound

Abstract Away auditory events

PCA

A Generalized Spectral Template

Whitman's Eigenradio

(Recht and Whitman 2003)

Play snippet from Singular Christmas - Grand hotel pout twice

Style Classification

MIDI

Lyrical or Energetic or ...

13 MIDI-based Features

Naive Bayes, GMM, ANN

(Dannenberg et al. 1997)

Speech/Music Discriminator

Features: 4 Hz modulation energy, low-energy frames, skew, centroid, flux, ZCR, Cepstral, "Rhythmicness"

GMM

(Scheirer and Slaney 1997)

Conclusion

Probabalistic and Machine Learning over direct programming
Wide application of many disciplines of computer science and signal processing
Complexity and vagueness of perception
Learn more about our own perception as we find what does and doesn't work