Part III : Audio Semantics

- What is music?
  - Stylistic Aleatorics
  - Factor Oracle
- Cognitive Model
  - Music as Information Source
  - Listening as Communication Channel
  - Anticipation: description and explanation
  - Emotional Force and Familiarity
  - Information Rate and Signal Recurrence
- Conclusion
  - References and resources

Motivation

- Music is:
  - – Sound ?
  - – Organization ?
  - – Experience ?

Organization of Sounds in order to create an Experience.

Motivation (cont.)

- Many aspects of musical *structure* are impossible to define *formally*
- Nevertheless, music and sound exhibit a great amount of *structure* and *redundancy*
  - Redundancy can be measured statistically
- Musical cognition is guided by *memory and expectation*
  - Guessing the future based on the past

Stylistic Aleatorics

Computational modeling of music that allows reproduction of stylistic music by learning from examples

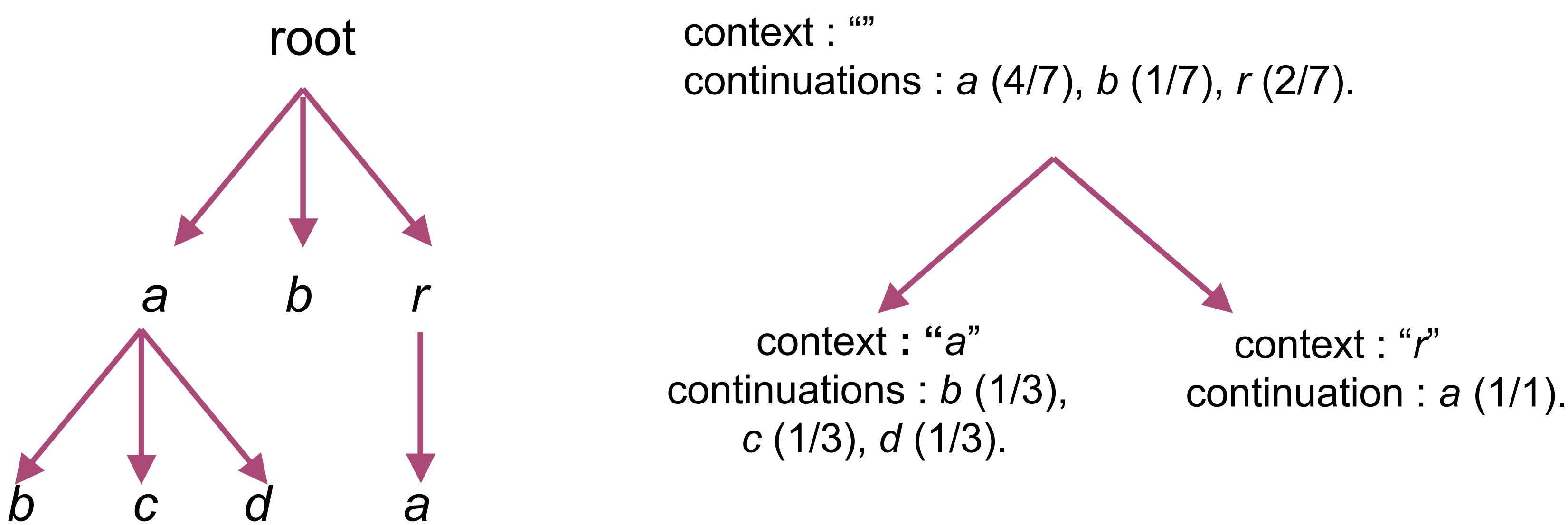
Stochastic Music

Iannis Xenakis “Musiques Formelles” (Formalized Music), published in 1963:

“ Since antiquity the concepts of chance (tyche), disorder (ataxia), disorganization were considered as the opposite and negation of reason (logos), order (taxis), and organization (systasis). [a stochastic process is] . . . an asymptotic evolution towards a stable state, towards a kind of goal, of stochos, whence comes the adjective <stochastic>”.

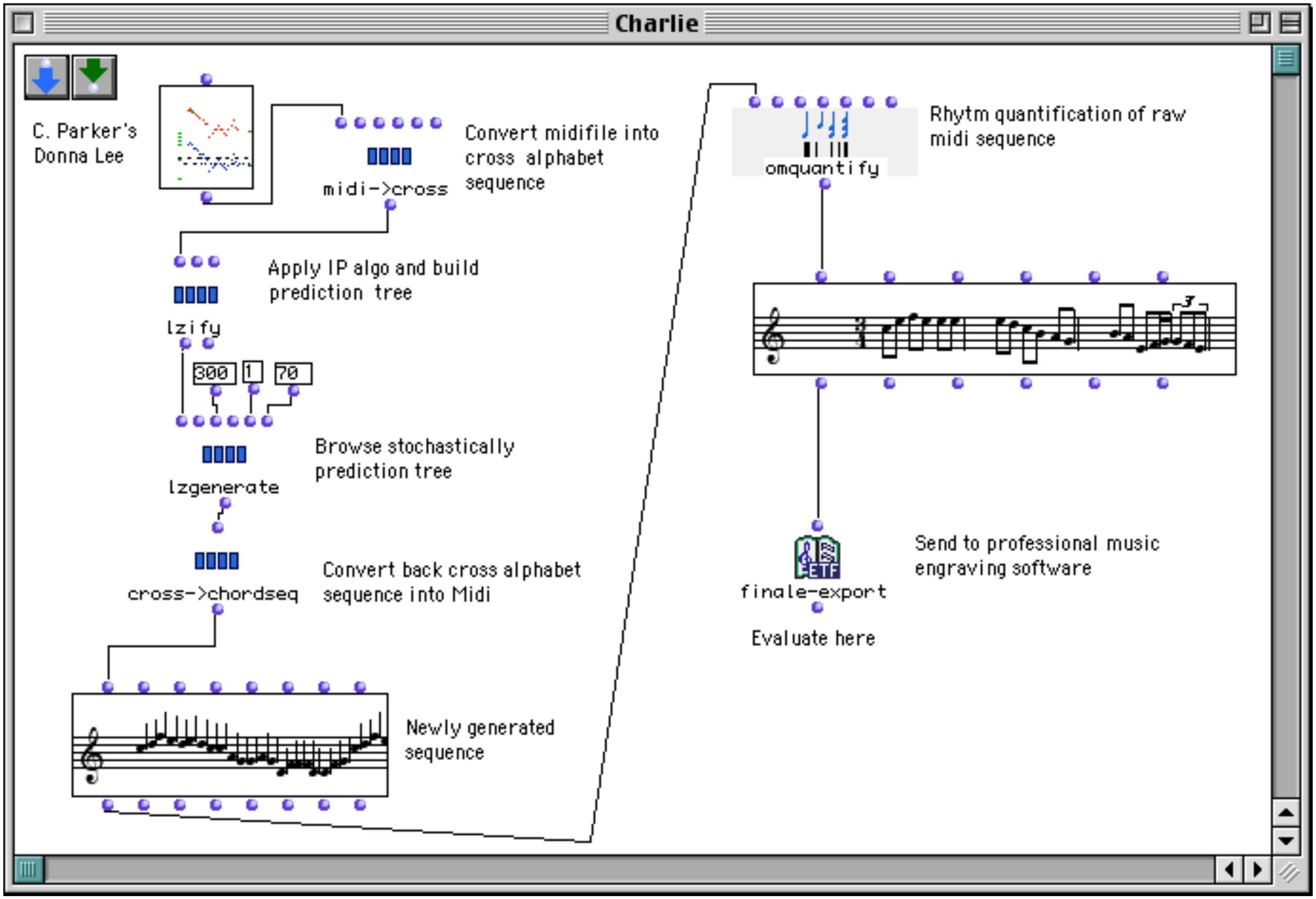
Music as a Prediction Tree

- Analysis of "*abracadabra*".



$$P(\text{generate "}abrac\text{"}) = P(a|\text{""})P(b|\underline{\mathbf{a}})P(r|ab)P(a|ab\underline{\mathbf{r}})P(c|abr\underline{\mathbf{a}}) = 4/7 \cdot 1/3 \cdot 2/7 \cdot 1 \cdot 1/3.$$

lzify



Charlie
C. Parker's
Donna Lee
midi->cross
Convert midifile into
cross alphabet
sequence
Apply IP algo and build
prediction tree
lzify
300
1
70
Browse stochastically
prediction tree
lzgenerate
cross->chordseq
Convert back cross alphabet
sequence into Midi
Newly generated
sequence
omquantify
Rhytm quantification of raw
midi sequence
finale-export
Send to professional music
engraving software
Evaluate here

Examples

http://music.ucsd.edu/~sdubnov/ThoughtsAboutMemex.htm
http://www.ircam.fr/equipes/repmus/MachineImpro/

Chick Corea original

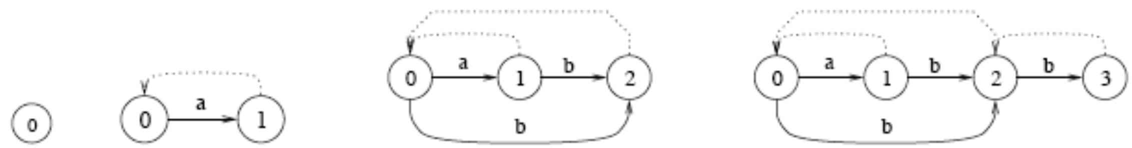
Impro1

Impro2

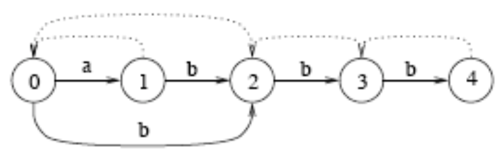
J.S.Bach Ricercar impro

Factor Oracle

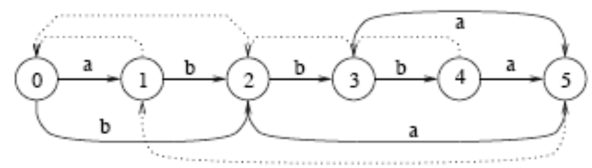
s = “abbbaab”



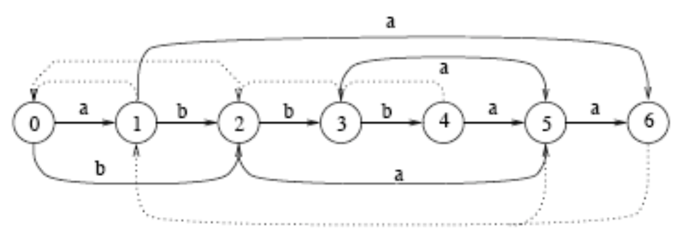
(a) (b) Add a (c) Add b (d) Add b



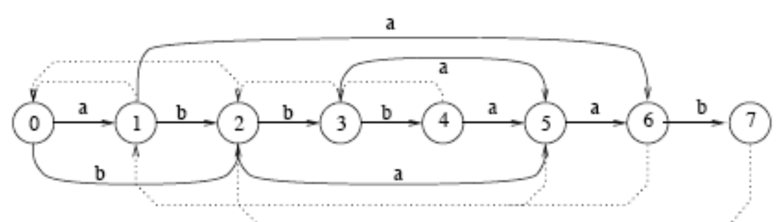
(e) Add b



(f) Add a



(g) Add a



(h) Add b

```
[trn,sfx] = FO(s,a)

% Factor Oracle for sequence s
% input:
% s - string of numbers in range [1,a]
% a - size of the alphabet
% output:
% trm - transition matrix (forward)
% sfx - suffix vector (backward)


[s, kend, ktrace] = FOgen(trn,sfx,n,p,k)
% Generate new sequence using a Factor Oracle
% input:
% trn - transition table
% sfx - suffix vector
% n - length of new string
% p - probability of change
% k - starting point
% output:
% s - new sequence
% kend - end point
```

Fonction add_letter(Oracle($p = p_1 p_2 \ldots p_m$), σ)

```
Create a new state m+1
Create a new transition from m to m+1 labeled by σ
k ← S_p(m)
While k > −1 and there is no transition from k by σ Do
        Create a new transition from k to m+1 by σ
        k ← S_p(k)
End While
If (k = −1) Then s ← 0
Else s ← where leads the transition from k by σ.
S_pσ(m+1) ← s
Return  Oracle(p = p_1 p_2 ... p_m σ)
```

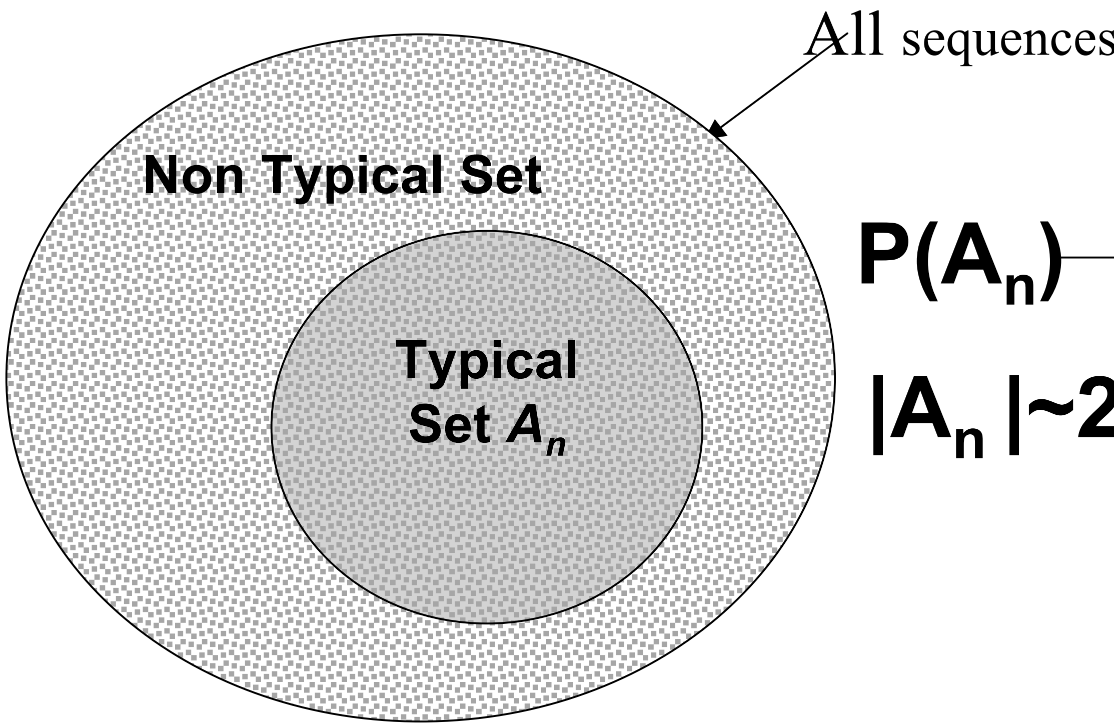
What is style?

- Emergence
  artist’s search for self-expression
- Decision making
  what you do if there is no rational basis for doing it
- Influential
  setting up frameworks of expectations that influence the audience and allow aesthetic planning

Music as Information Source

Typical Set:

$A_n = \{ x_1^n : \quad -\frac{1}{n}\log p(x_1, x_2, .., x_n) \approx H(X) \quad \}$



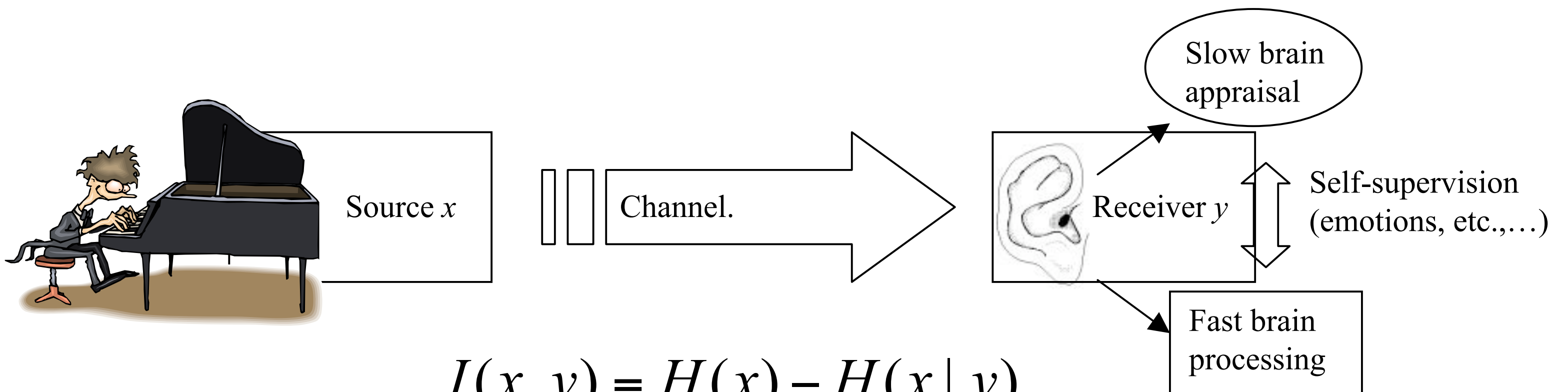
$P(A_n) \longrightarrow 1$

$|A_n| \sim 2^{nH}$

Cog-Comm Model

Aesthetic perception as a communication process:

Information that influences the “cognitive state” of the information receiver

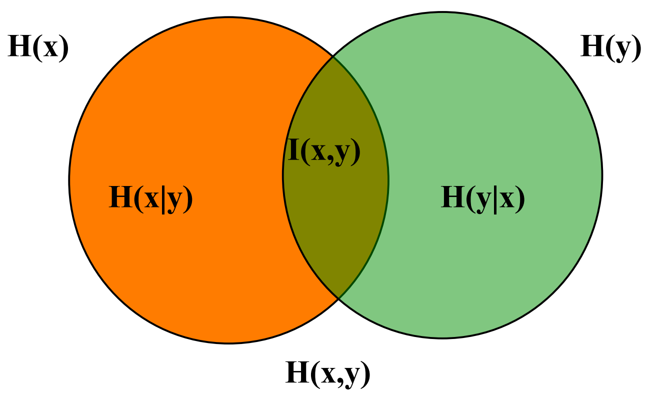


$$I(x,y) = H(x) - H(x \mid y)$$

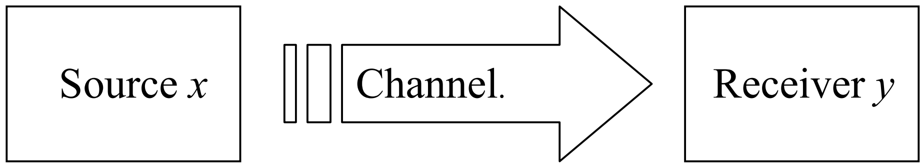
The information paradox:
Discover more by listening more….
(you gain by learning, not get bored!)

Anticipation

- Entropy & Information



- Communication Channel



$$I(x, y) = H(x) - H(x \mid y)$$

y – past experience, what you heard so far

x – new material

H(x) – uncertainty about *x*

H(x|y) – uncertainty about *x* when we know already *y*

I(x,y) – how much the past tells us about the future

“Cognitive” features

Cognitive Measure	**Signal Measure**
Familiarity (Recognition, Categorization)	Similarity Structure, Repetition, Recurrence (Long term)
Emotional Force (Anticipation, Implication-Realization)	Measures of Predictability, Information Rate (Short Term)

Information Rate

Anticipation is proportional to the amount of information that the past "carries" into the present

$$I(x_{past}, x_n) = I(x_1, x_2, \ldots, x_n) - I(x_1, x_2, \ldots, x_{n-1}) = H(x_n) - H(x_n \mid x_1^{n-1})$$

- Increase in information with arrival of new data
- Difference between uncertainty before and after prediction
- Can be estimated from notes or audio (score or recording)
- Psychologically plausible "Inverted U function"

Information Rate (cont.)

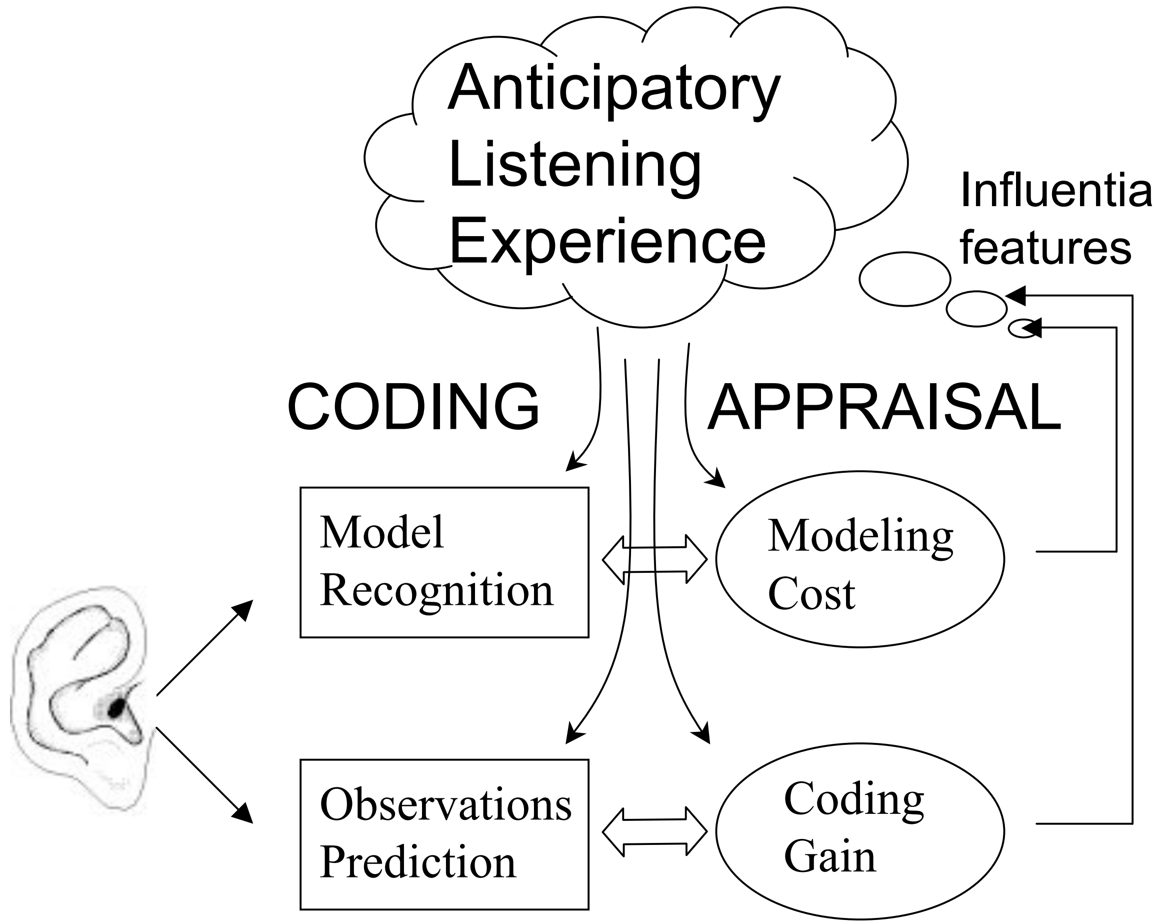
Assuming a model θ we can write

$$I(x_{past}, x_n) \approx \langle I_\theta(x_{past}, x_n) \rangle_{P(\theta)} + \underbrace{H(\theta) + \langle D(\theta \| \theta^*) \rangle_{P(\theta)}}$$

Information rate = Description + Explanation

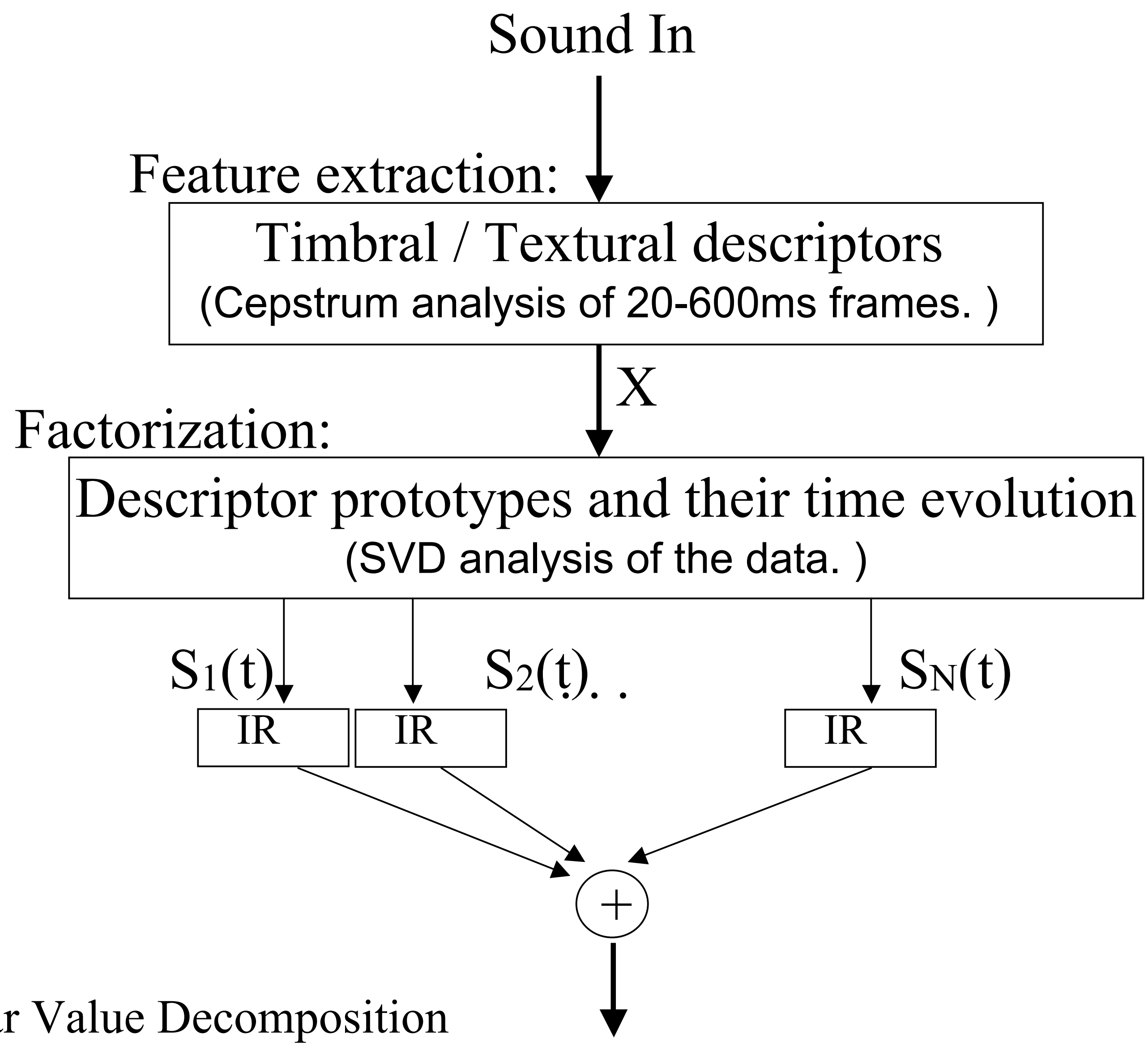
- $I_\theta(x_{past}, x_n)$ Description "within" a model (Coding Gain)
- $H(\theta)$ Model uncertainty (size of model space)
- $D(\theta \| \theta^*)$ Distance between model and "true" distribution

(Model Cost)


Anticipatory
Listening
Experience
Influential
features
CODING
APPRAISAL
Model
Recognition
Modeling
Cost
Observations
Prediction
Coding
Gain

Coding Gain: IR Estimation

IR_Analysis.m



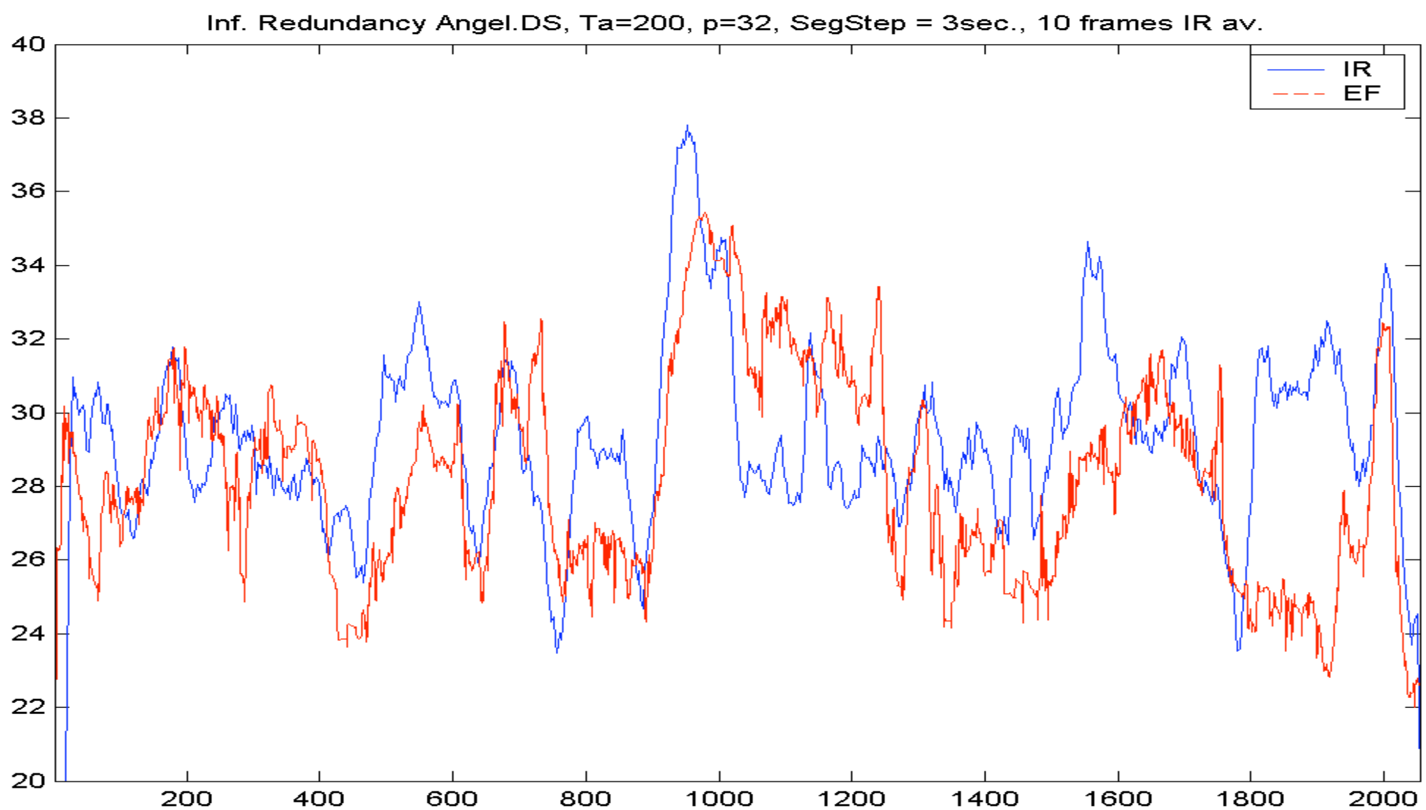
SVD – Singular Value Decomposition

IR – Derived from Spectral Flatness Measure

IR vs. Emotional Force

The Angel of Death by Roger Reynolds

IR using 200 msec cepstral features vs. human judgments of emotional force (EF)



Information Rate (cont.)

How to find $D(\theta_n, \theta^*)$ and $H(\theta_n)$?   Recur_Analysis.m

Idea:

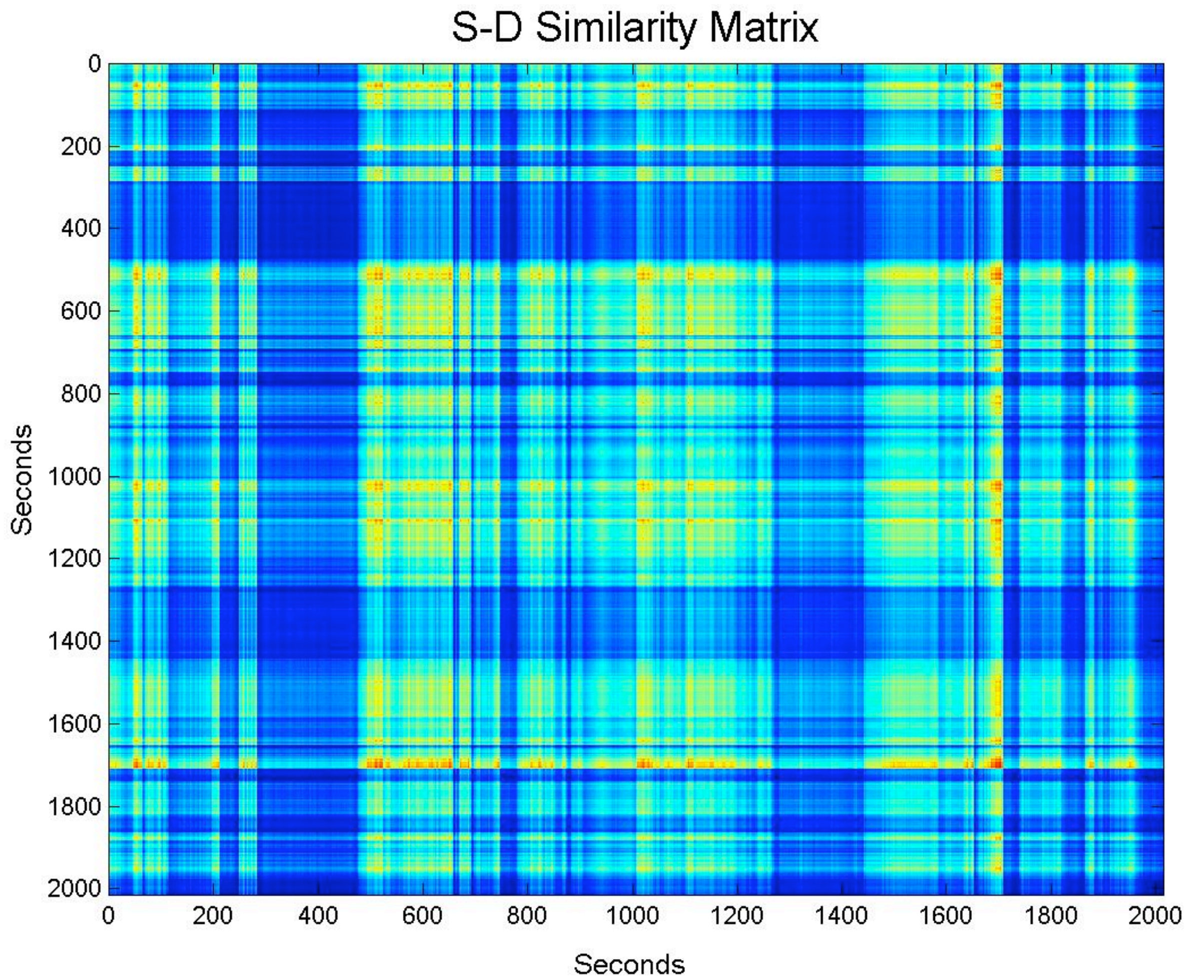
- Transform the similarity matrix D into a Markov transition matrix

$$\mathbf{P}_{ij} = P(j \mid i) = \frac{d(X_i, X_j)}{\sum_j d(X_i, X_j)}$$

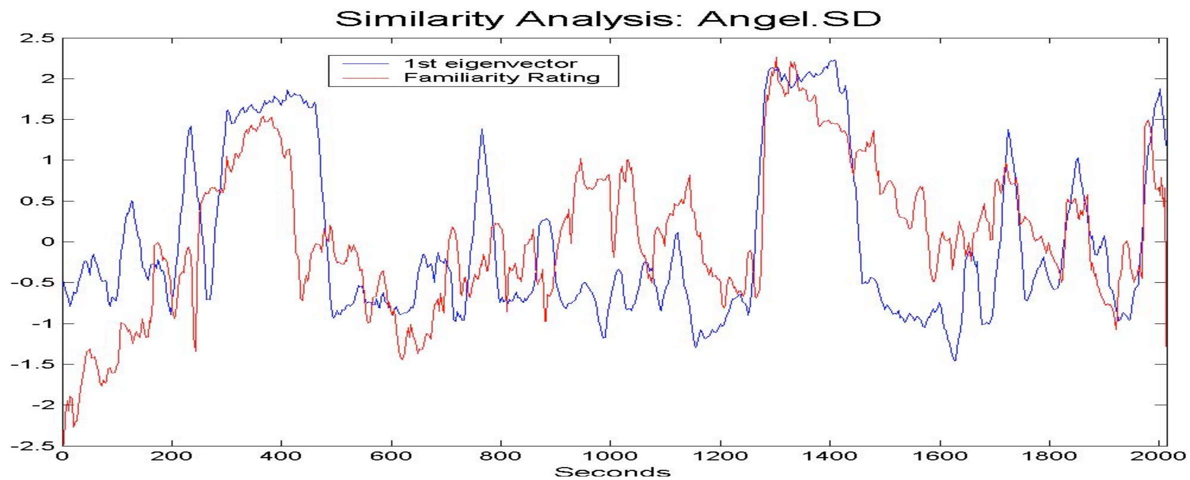
- Approximate $P(\theta^*)$ by eigenvectors of $\mathbf{P}$
- Consider it as a measure of "explanation"

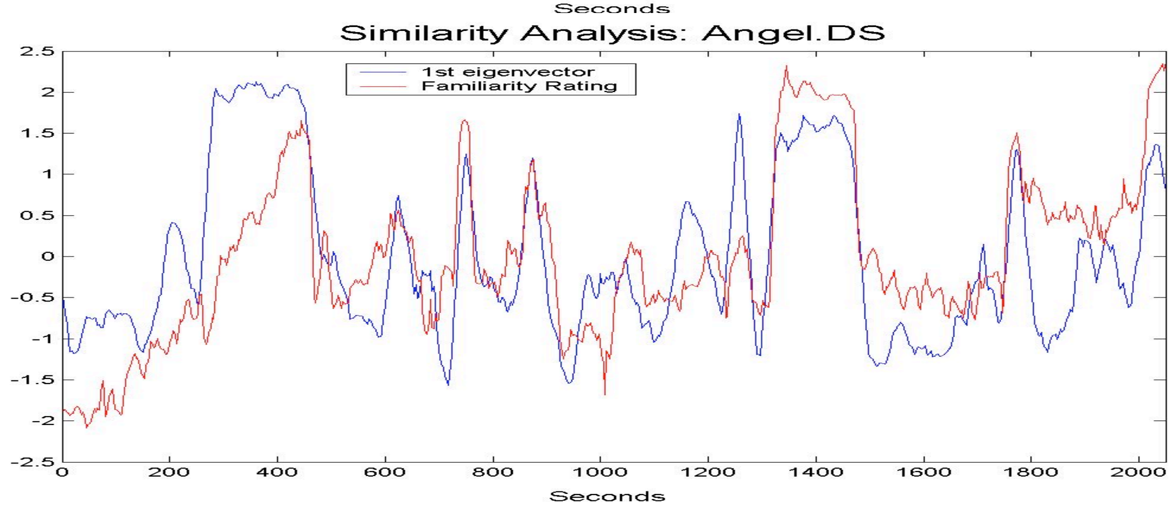
$$H(\theta_n) + D(\theta_n, \theta^*) = <\log P(\theta^*)>_{\theta_n} \sim \log P_n(\theta^*)$$

The Angel of Death by Roger Reynolds



“Explanation” Profile vs. Familiarity





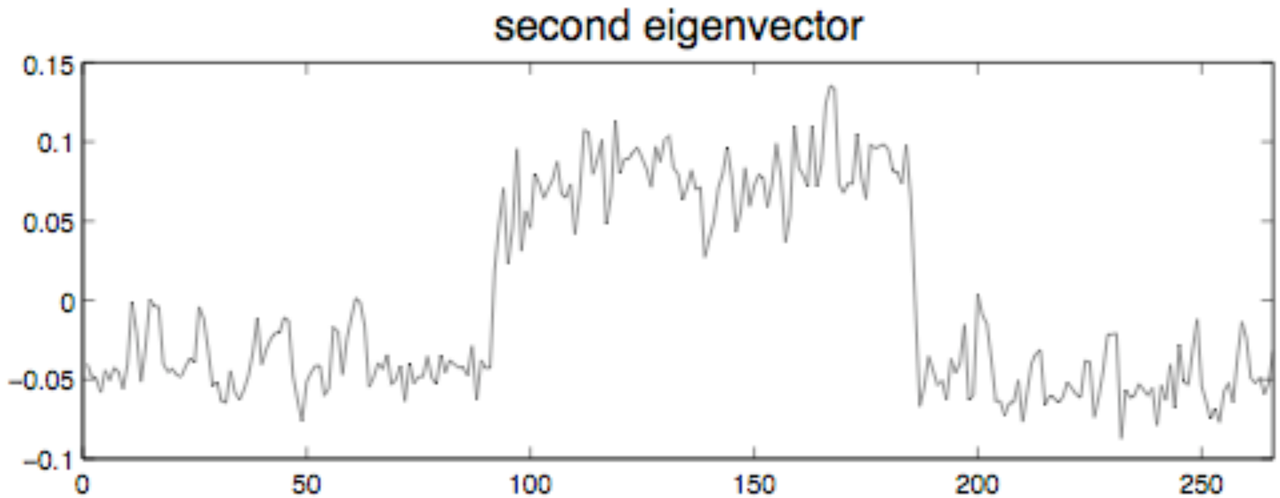
JASIST, Special Issue on Style, Sep. 2006

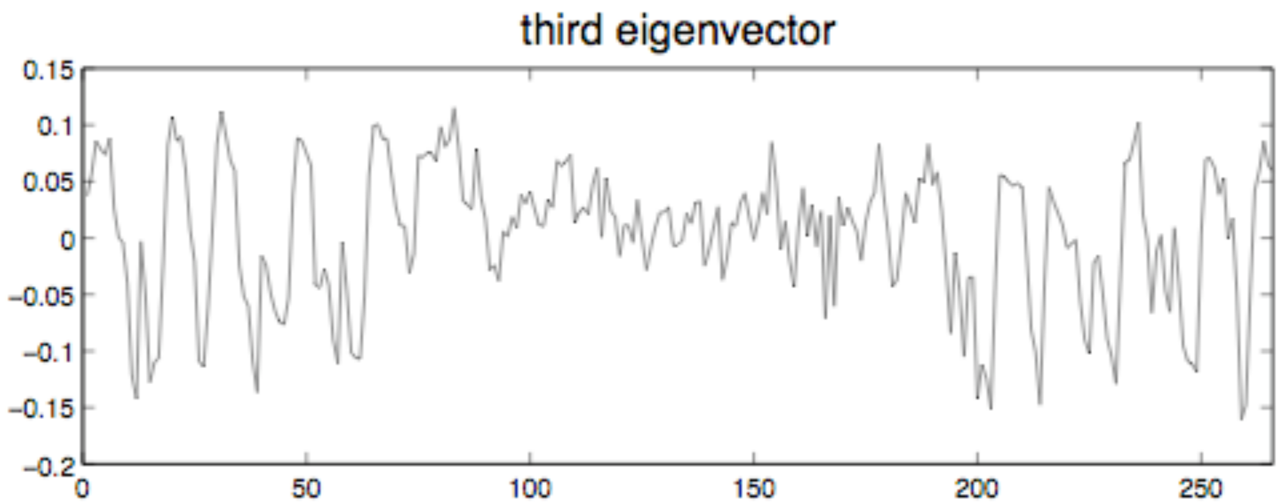
Segmentation Application

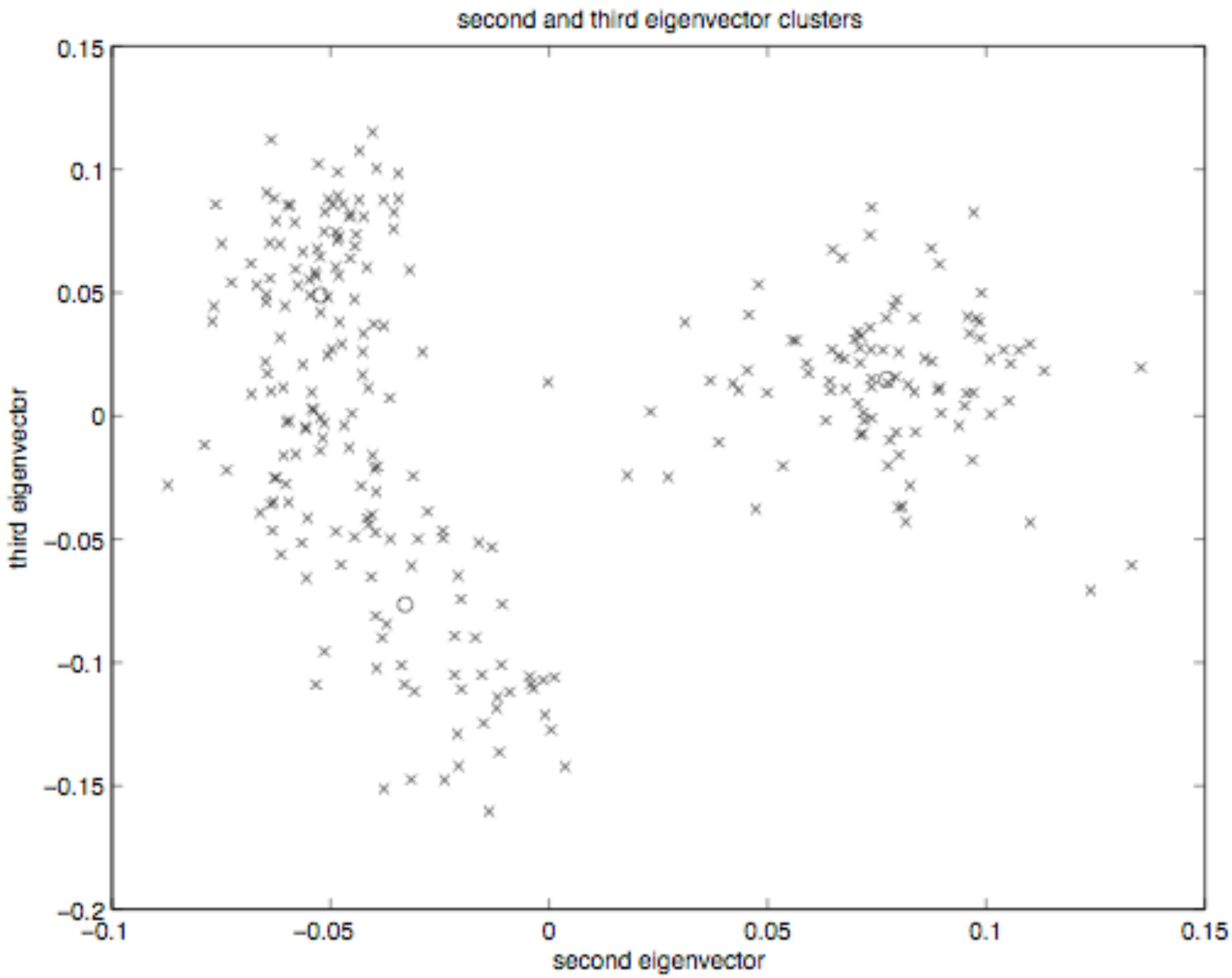
Apel & Dubnov, ICMC 04

Original   Segmented

- Segmentation is done by grouping the values of 2 largest eigenvectors into 3 clusters
- Sounds are associated to their nearest cluster


second eigenvector
0.15
0.1
0.05
0
−0.05
−0.1
0
50
100
150
200
250


third eigenvector
0.15
0.1
0.05
0
−0.05
−0.1
−0.15
−0.2
0
50
100
150
200
250



second and third eigenvector clusters
0.15
0.1
0.05
0
−0.05
−0.1
−0.15
−0.2
third eigenvector
−0.1
−0.05
0
0.05
0.1
0.15
second eigenvector

Spectral Clustering

The above clustering method is closely related to spectral clustering

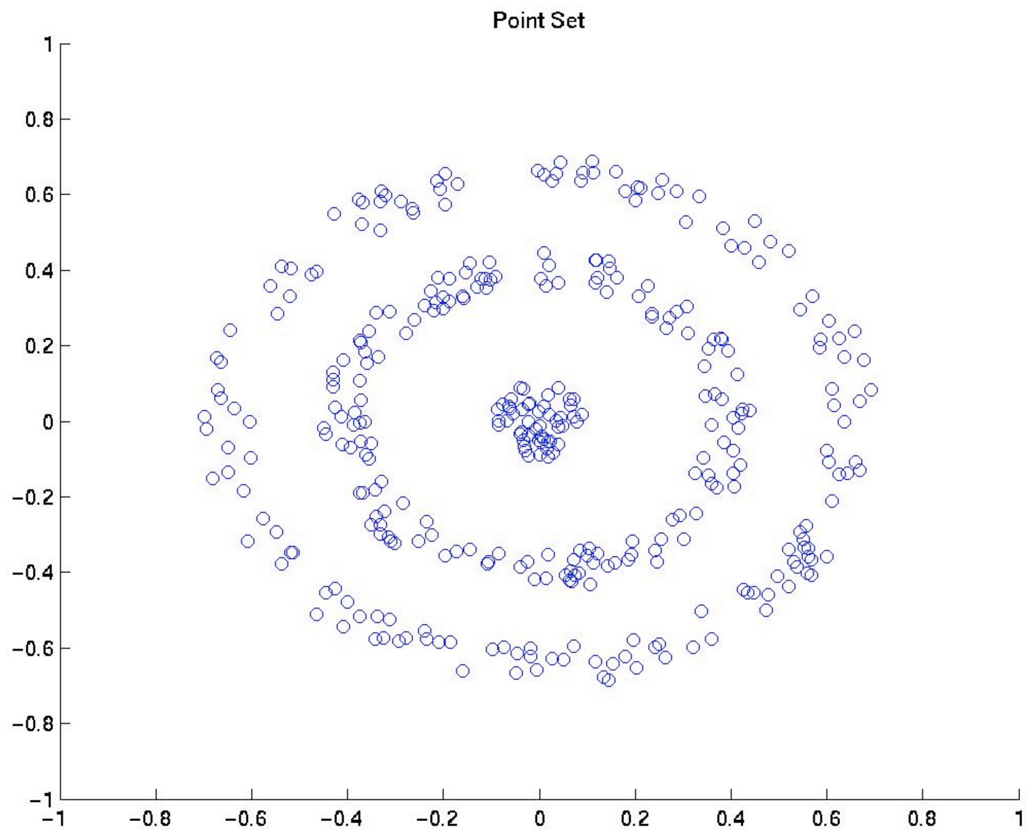
Shi & Malik Algorithm:

- Construct the matrices D and Z. $Z = diag(\sum_j d(X_i, X_j))$
- Find the second smallest generalized eigenvector of (Z-D) i.e.

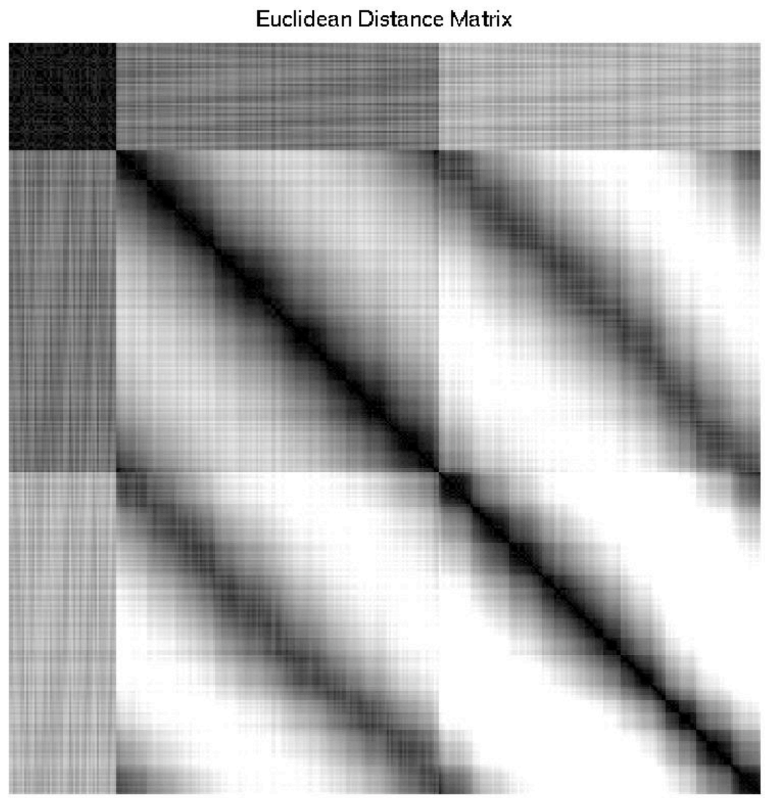
$$(Z - D)y = \lambda Zy$$

- Threshold *y* to get a partitioning of the graph.

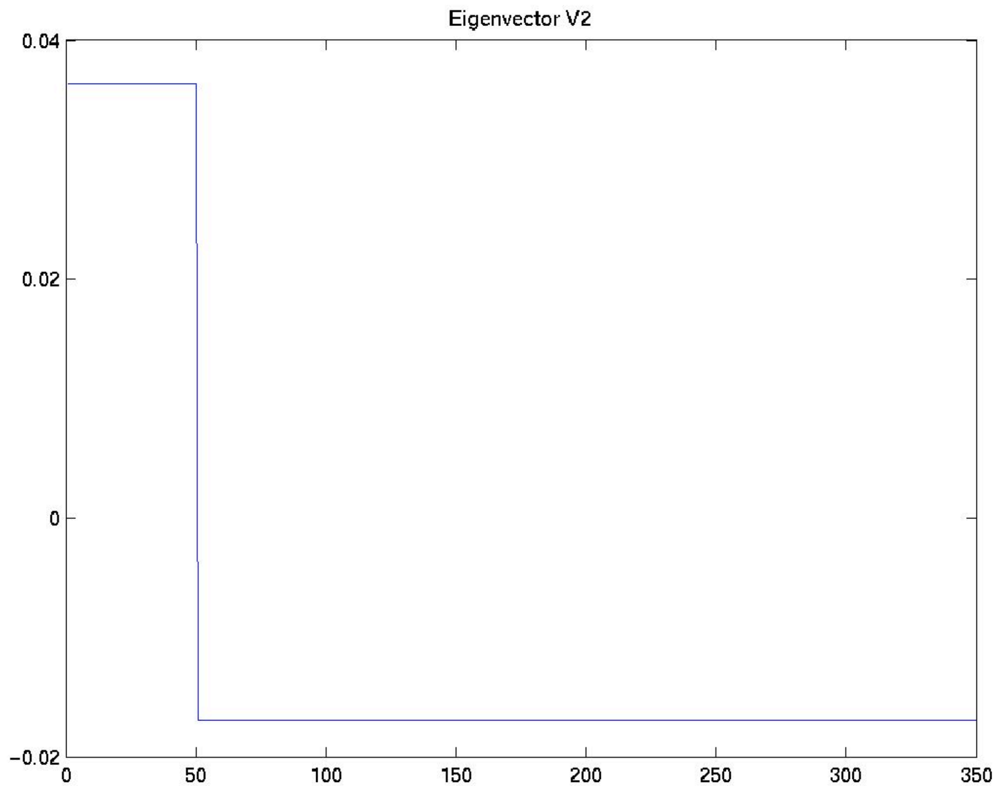
Example



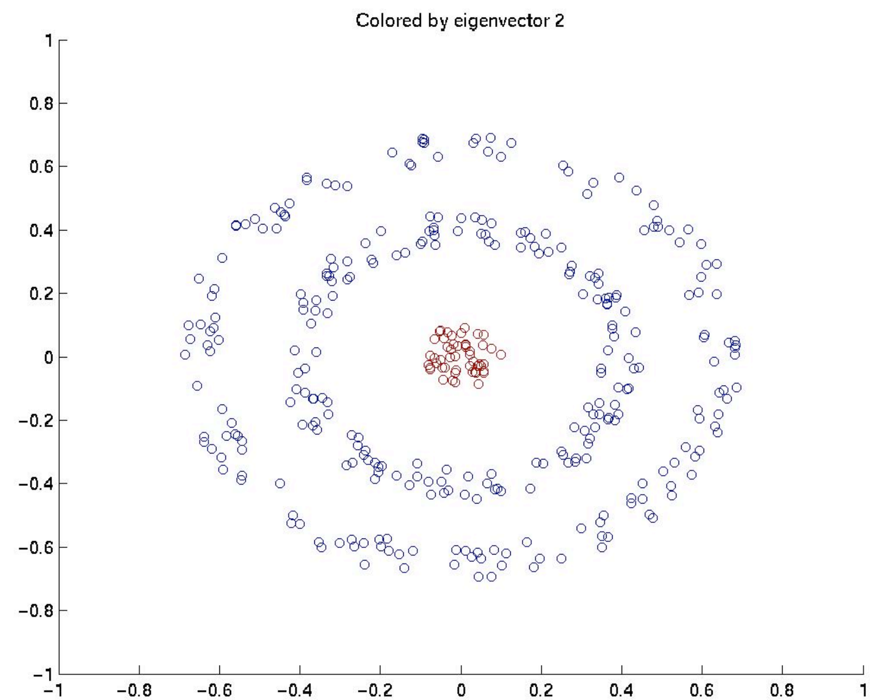
Distance Matrix



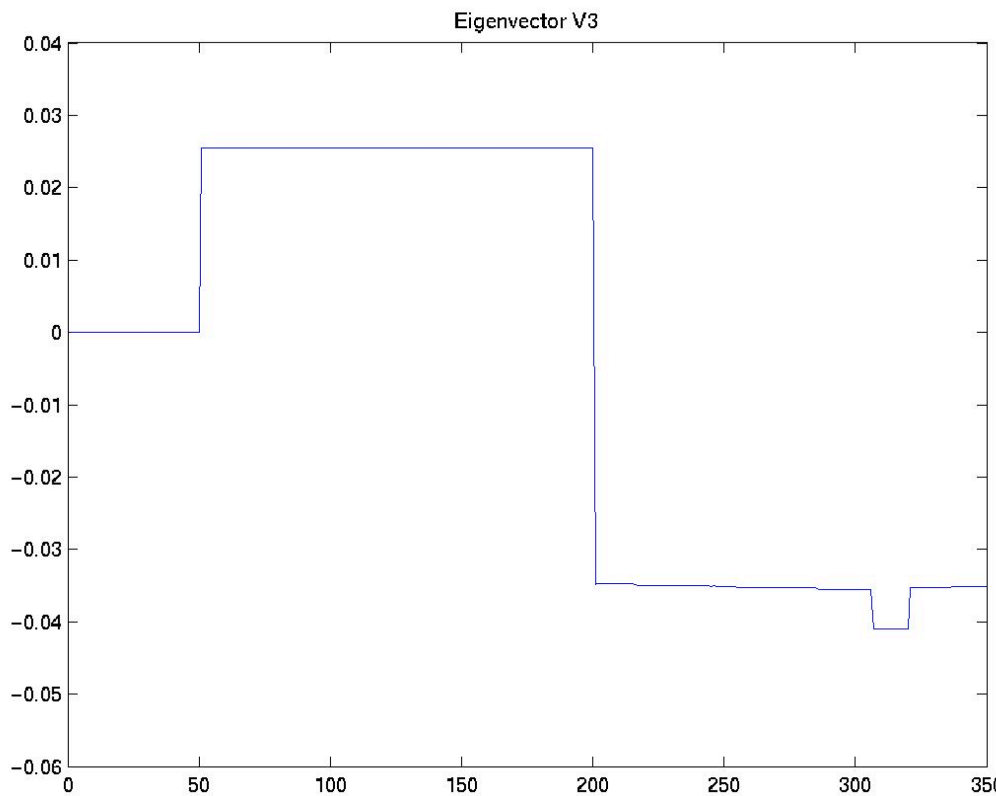
Second generalized eigenvector



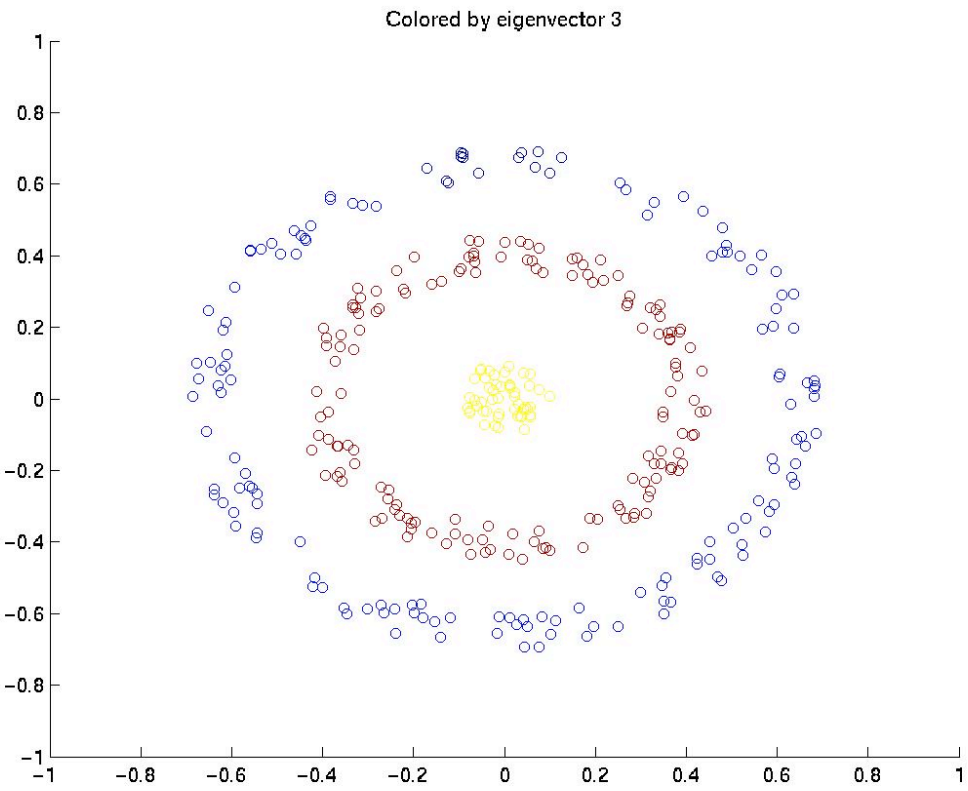
The first partition



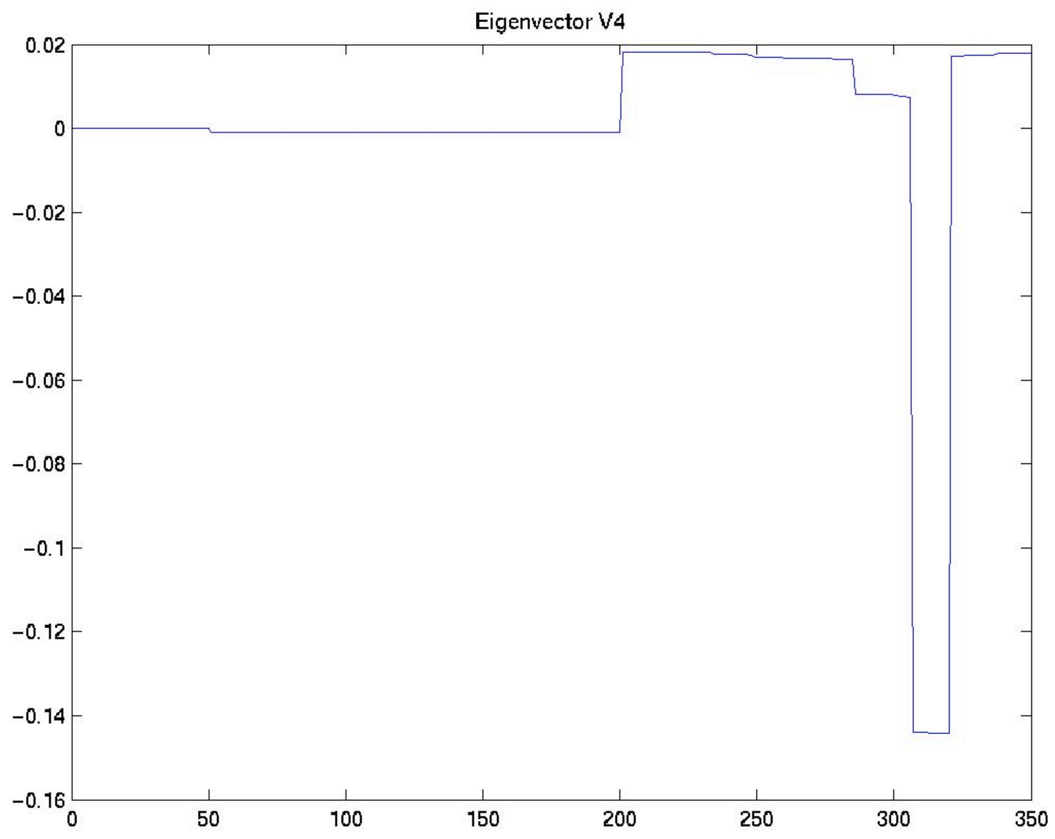
Third generalized eigenvector



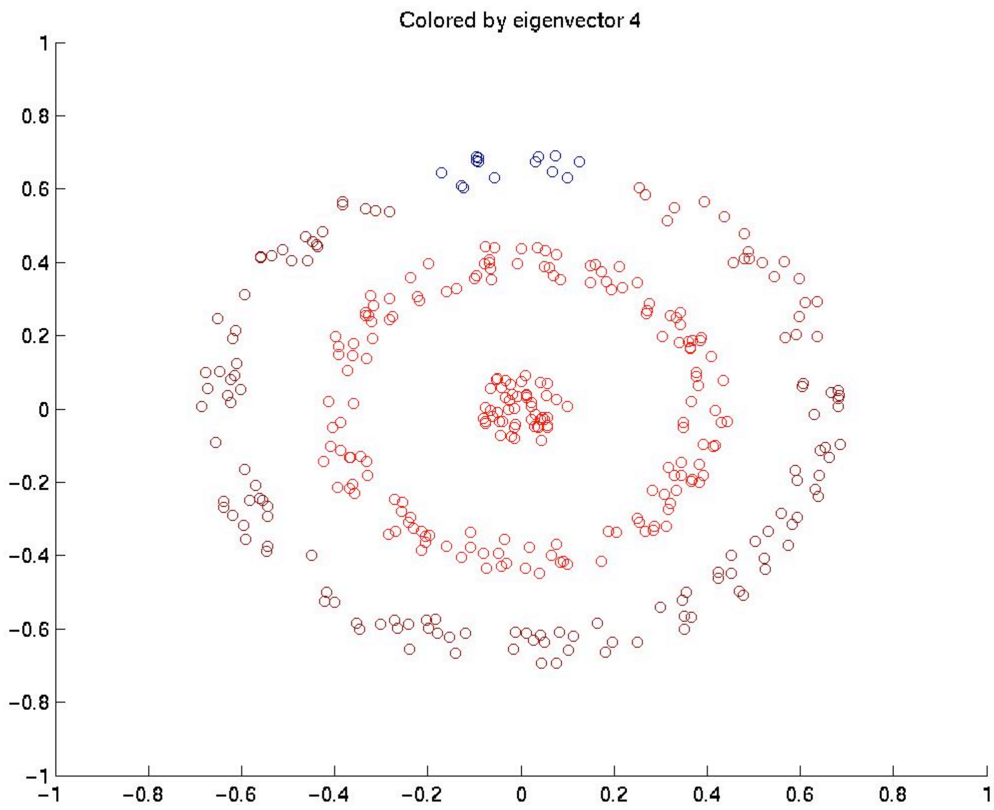
The second partition



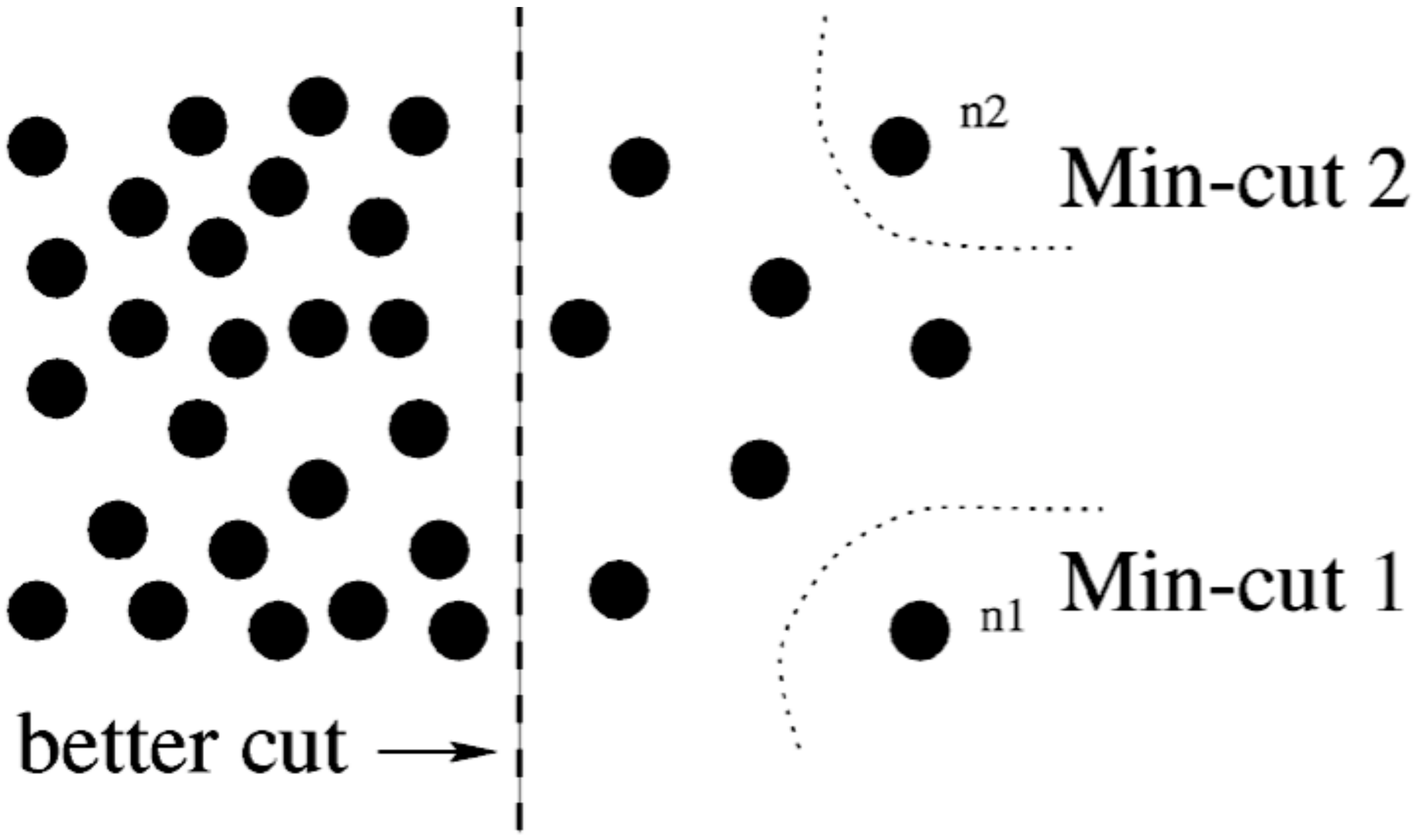
Fourth generalized eigenvector



The third partition



Normalized Cut



From Shi and Malik, 2000

Conclusion

http://cosmal.ucsd.edu/cal/

1. **Modeling Music and words**, Douglas Turnbull, Luke Barrington, and Gert Lanckriet - ISMIR 06

2. **Musical Boundary Detection using Boosting**, Douglas Turnbull, Gert Lankriet, Elias Pampalk, and Masataka Goto - Submitted to ICASSP 07

Modeling music and words [TBL06]

Design a **statistical system** that **learns a relationship between music and words.**

Applications:

1. **Annotation:** Given a audio-content of a song, we can **'annotate'** the song with semantically meaningful words.

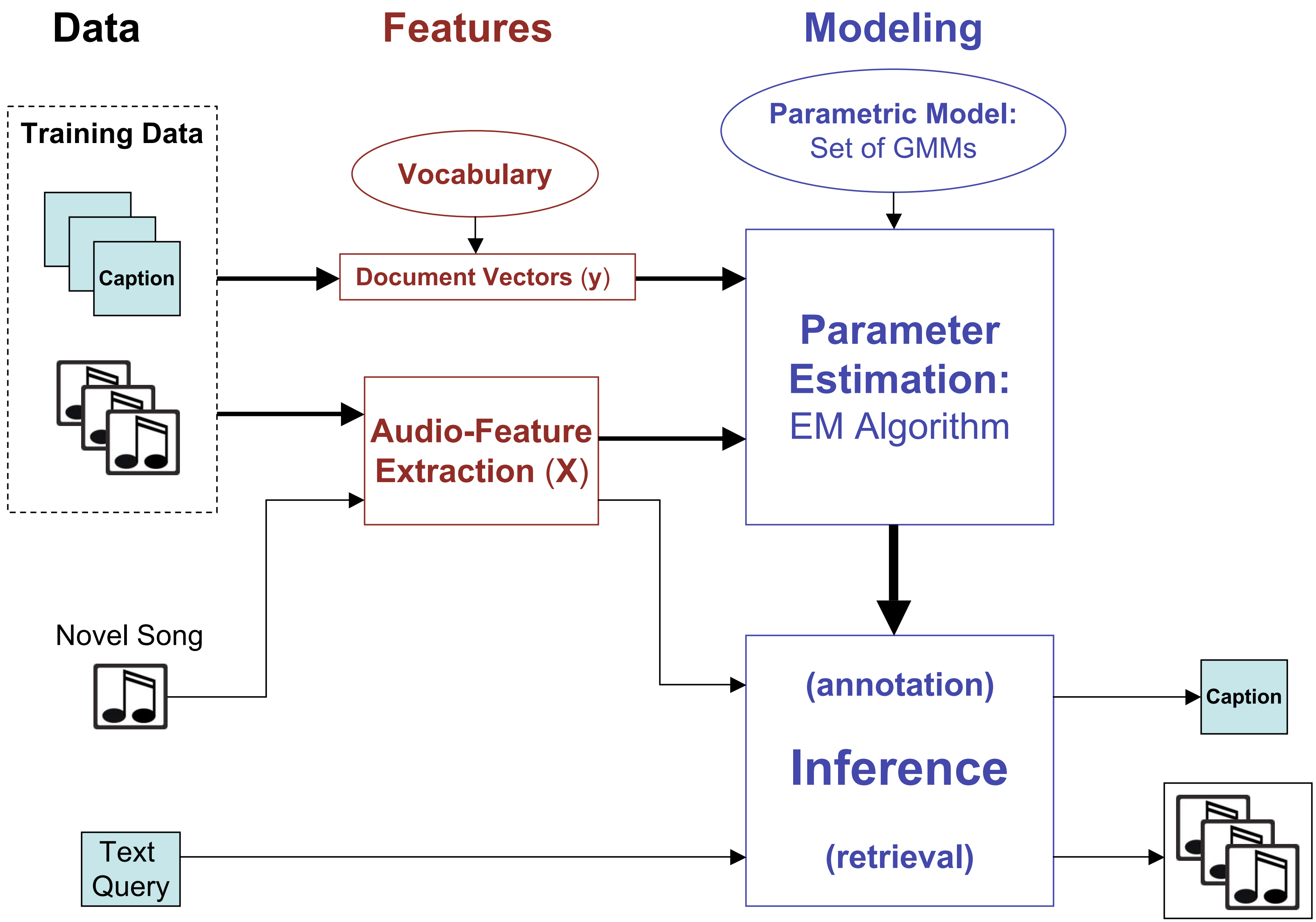
   song → words

2. **Retrieval:** Given a text-based query, we can **'retrieve'** relevant songs based on the audio content of the songs.

   words → songs

The parameter for the model are learned using a heterogeneous data set of song and song reviews.

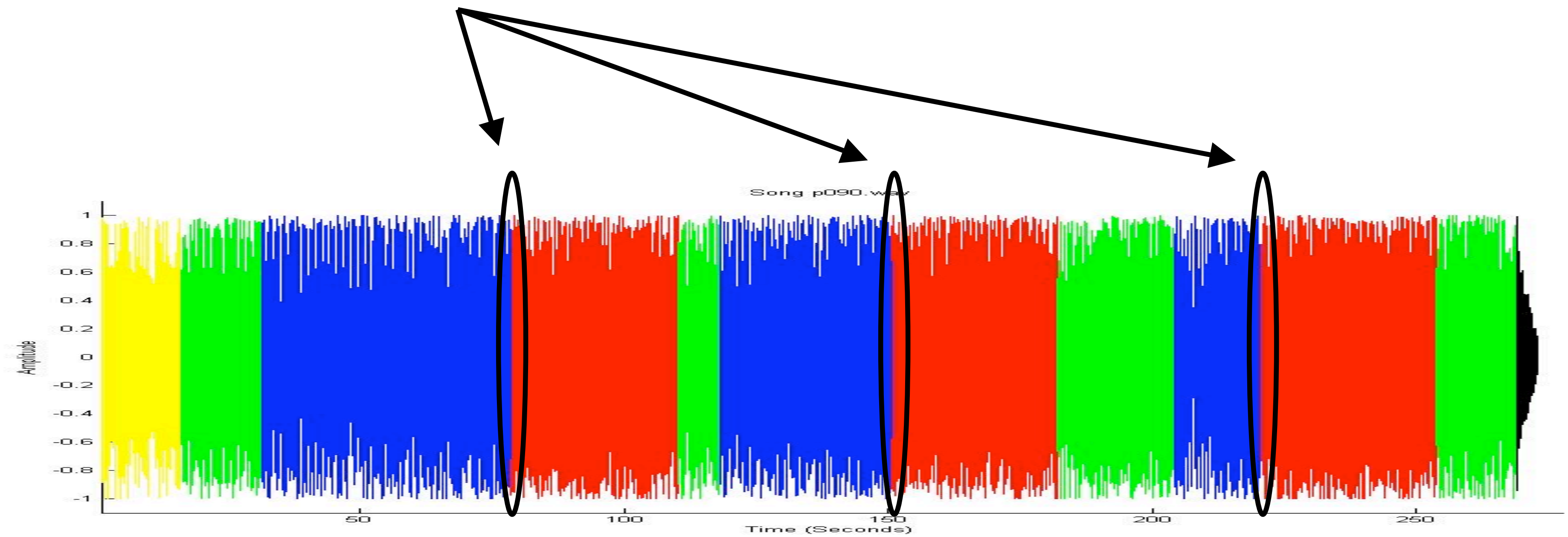
Modeling music and words [TBL06]



Supervised Musical Boundary Detection [TLPG06]

Consider the **structure** of a song:

- **Musical Segment**: a song is composed of segments:
  - **Introduction, Bridges, Verses, Choruses, Outro**
- **Musical Boundary**: a boundary between two musical segments.
  - e.g., the end of a **verse** and the beginning of a **chorus**



Framework for Anticipatory Machine Improvisation and Style Imitation

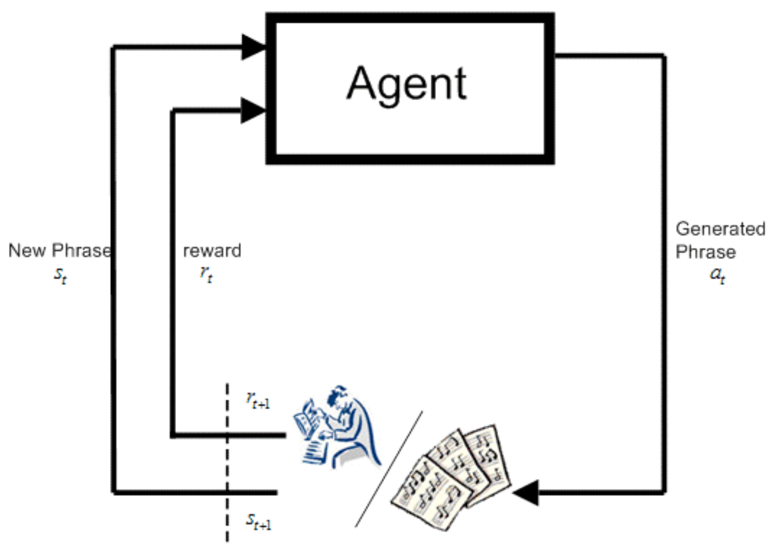
Arshia Cont[1,2], Shlomo Dubnov[1] and Gérard Assayag[2]

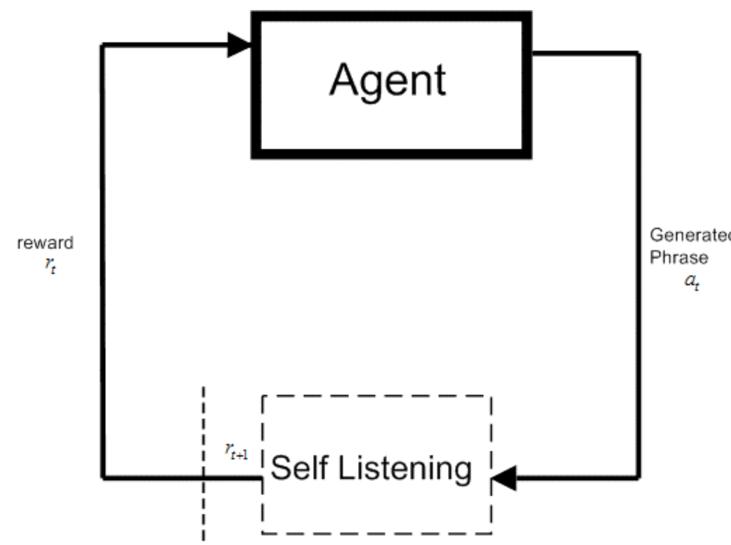
[1] Center for Research in Computing and the Arts (CRCA), University of California in San Diego (UCSD).
[2] Ircam-Centre Pompidou, Paris, France.
{Cont,assayag}@ircam.fr, sdubnov@ucsd.edu

Framework for Anticipatory Machine Improvisation and Style Imitation [CAD06]

Approach

- AI:
  - Interactive Reinforcement Learning with an environment
  - Multiple-agents with collaborative and competitive learning
  - Memory-based learning
  - Main schema for interactive modeling:

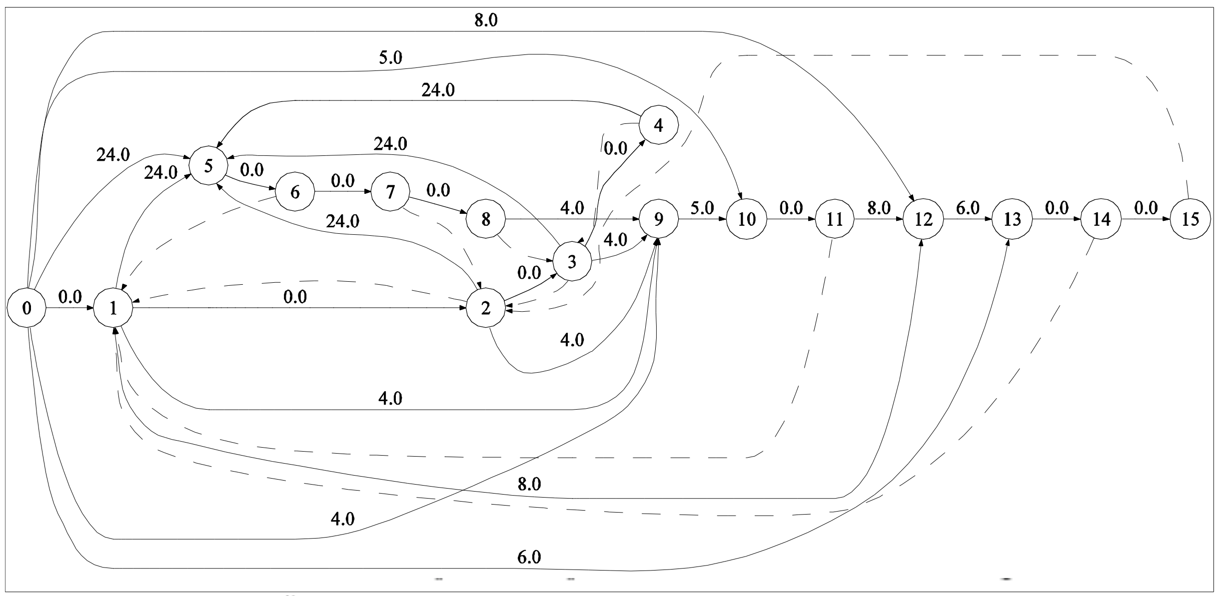




Framework for Anticipatory Machine Improvisation and Style Imitation [CAD06]

Memory Models

- Using Factor Oracles:
  - Linear time and space complexity
  - Suffix links: give the maximum repeated suffix given the context.
  - Forward Transitions: give maximal length context.



Event Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Pitch (MIDI)	0	51	63	62	63	0	65	0	67	67	63	68	68	58	60
Harmonic Int.	0	0	0	0	24	0	0	0	4	5	0	8	6	0	0
Duration	4	4	4	4	4	4	4	8	4	4	4	4	4	4	4
Pitch Deriv.	0	0	12	-1	1	0	0	0	0	0	1	-3	0	-4	2
Harm. Deriv.	0	0	0	0	0	0	0	0	0	1	0	0	-2	0	0
Dur. Deriv.	1	1	1	1	1	1	1	2	0.5	1	1	1	1	1	1

Framework for Anticipatory Machine Improvisation and Style Imitation [CAD06]

Long-term planning/formal shapes

Real-time Multiple Pitch Observation using Sparse Non-Negative Constraints

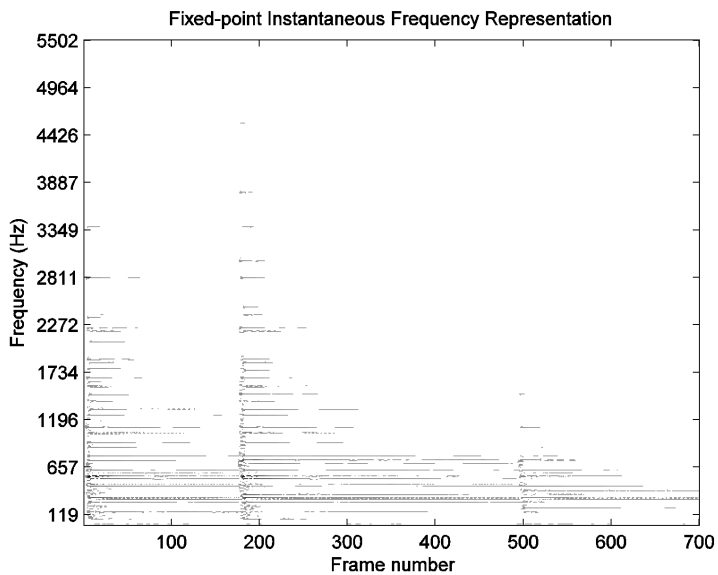
Arshia Cont

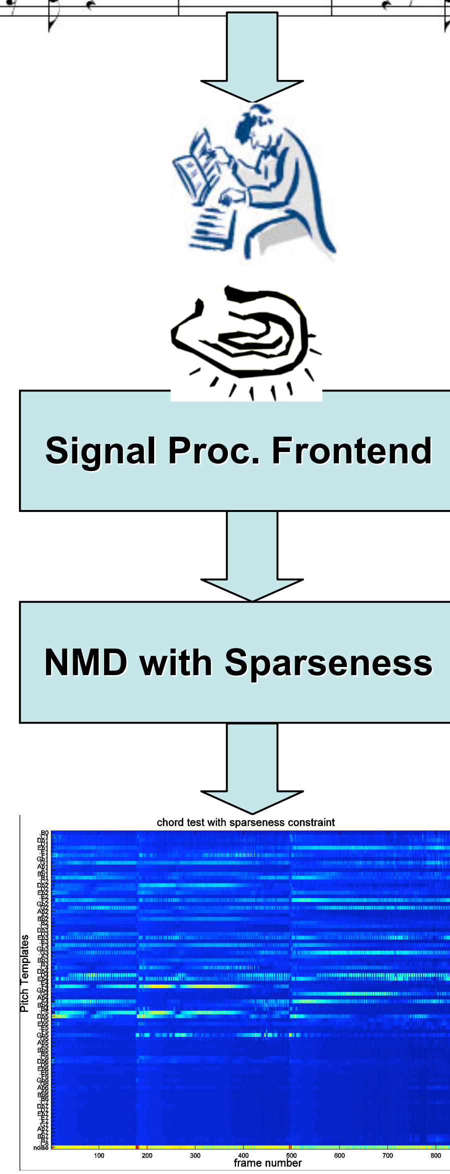
IRCAM, Realtime Applications Team, Paris.
Center for Research in Computing and the Arts (CRCA),
University of California in San Diego (UCSD).

http://crca.ucsd.edu/arshia/

Realtime Multiple-pitch Observation [CONT06]

Implementation



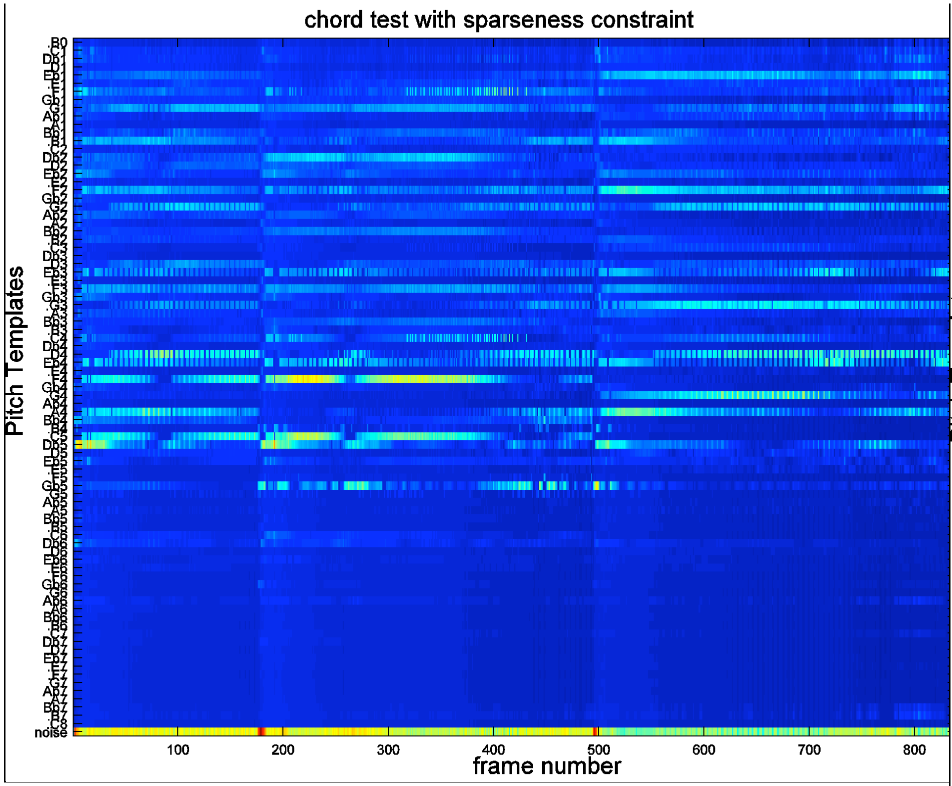


Algorithm *Non-Negative Matrix Deconvolution with Sparseness Constraint*
Given V and W, find the non-negative vector H with a given ℓ_ε norm and ℓ_2 norm.

1. Initialize H to random positive matrices
2. Iterate
   (a) Set $H = H - \mu_H W^T(WH - V)$
   (b) Set $s_i = h_i + (\ell_\varepsilon - \sum tanh(h_i^2))/N$ and $m_i = \ell_\varepsilon / N$
   (c) Set $s = m + \alpha(s - m)$
   where $\alpha = \frac{-(s-m)^T m + \sqrt{((s-m)^T m)^2 - \sum(s-m)^2(\sum m^2 - \ell_2^2)}}{\sum(s-m)^2}$
   (d) Set negative components of s to zero and set $H = s$

Realtime Multiple-pitch Observation [CONT06]

Sparseness Comparison



Conferences

- ICMC, International Computer Music Conference
- ISMIR, Int. Conf. on Music Information Retrieval
- DAFX, Int. Conf. on Digital Audio Effects
- NIME, New Interfaces for Musical Expression
- ICAD, Int. Conf. on Auditory Display
- ICASSP, Int. Conf. on Audio, Speech and Signal Processing
- ACM Multimedia
- AES, Audio Engineering Society Conferences
- ASA, Acoustical Society of America Meetings

Some Web Resources

- Organizations
  - Computer Music Association http://www.notam02.no/icma/
  - Music information retrieval research http://www.music-ir.org/
  - AUDITORY list home page http://www.auditory.org/
  - Acoustic Society of America http://www.acoustics.org/
- This
  - http://music.ucsd.edu/~sdubnov/ComputerAudition/
  - http://cosmal.ucsd.edu/cal/