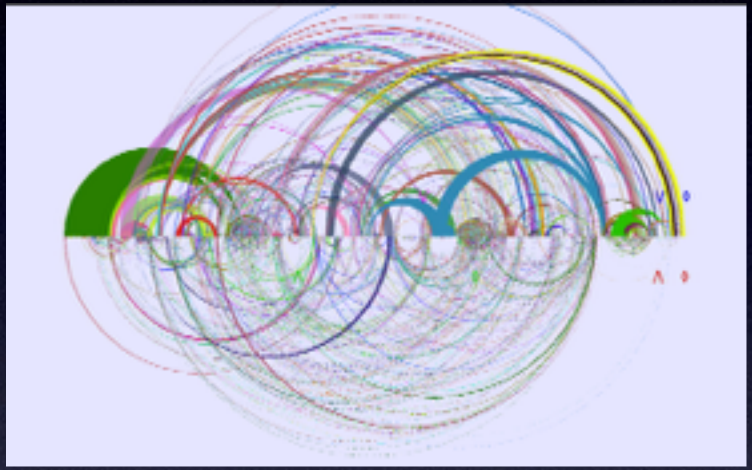
Improvised Symbolic Interaction

A Creative Perspective on Similarity



Gérard Assayag

IRCAM

Sciences and Technologies of Music and Sound (STMS) Lab

Music Representation Team

What are MIR / Sim studies useful to (in a creative perspective)

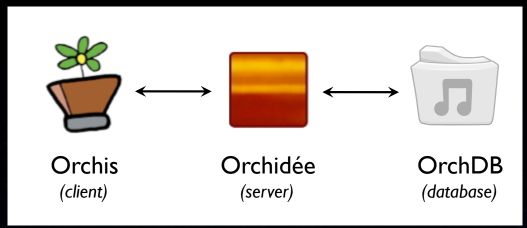
- keywords : find, retrieve, match, identify, align, query / compose, generate, improvise, interact, synthesize, replace, recombine
- in a creative perspective, off-line MIR used for composition, on-line MIR used for direct interaction (Machine Improvisation)
- Going from on line MIR to Machine Audition (artificial listening) to Machine cognition (Machine Musicianship, Rowe 2001)
- Equip autonomous creative agents with Machine Audition and Machine Cognition capabilities
- Escape from the Song (occidental pop song) paradigm : free form, composition, improvisation

Compositional Exemple : orchestral texture derived from a single sound spectral analysis

- Example 1 . *Partiels* by Gerard Grisey, 1975
- Historical example considered at the origin of "Spectral Music"
- The source is the low E note on the trombone. The transformation is an exploration of the spectral content of this note by the whole orchestra
- Litteral reconstruction or transformations inspired by electro-acoustics operators (distortion, ring-modulation, filtering etc.)
- Highlights the need for Computer Assisted Composition Environments

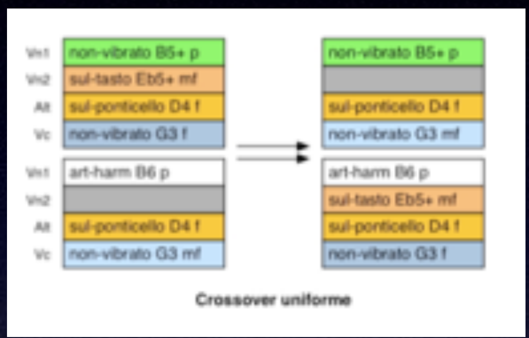
Orchestration

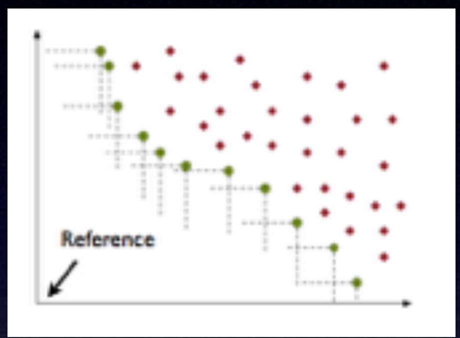
Projet ANR *Sample Orchestrator*

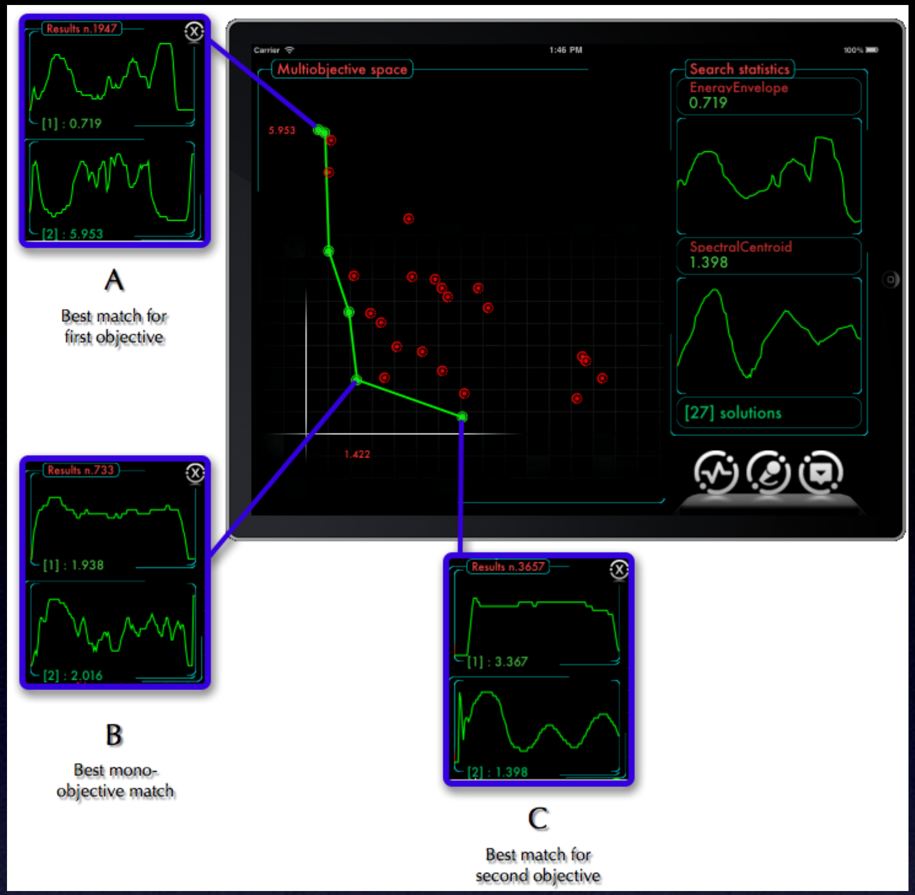


PhD G. Carpentier, Dec. 2008 : Computational Approach of Musical Orchestration

Multi-Objective Optimization





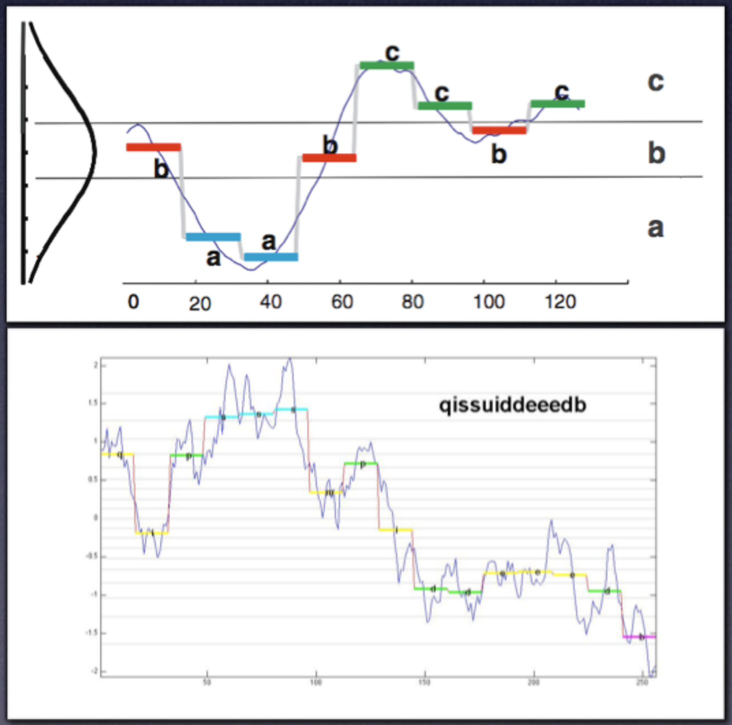


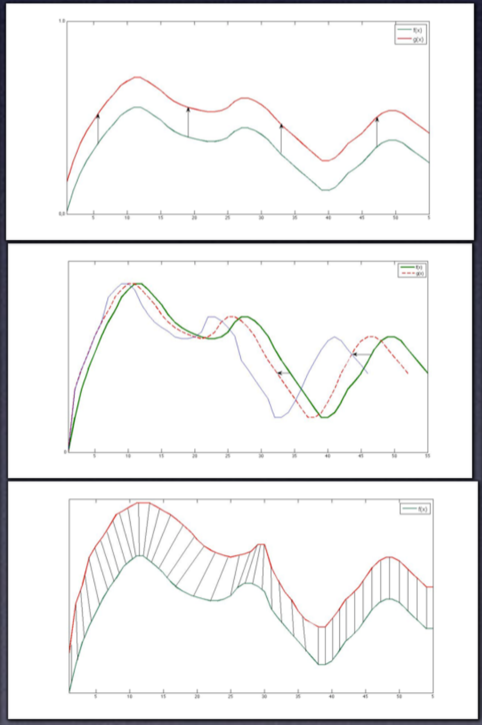
PhD G. Carpentier, Dec. 2008 : Computational Approach of Musical Orchestration

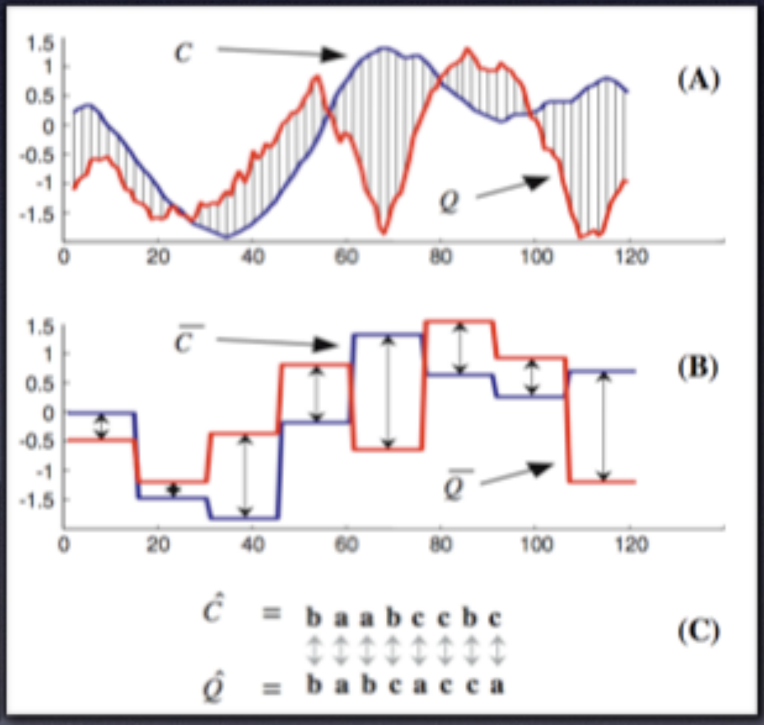
PhD P. Esling 2013

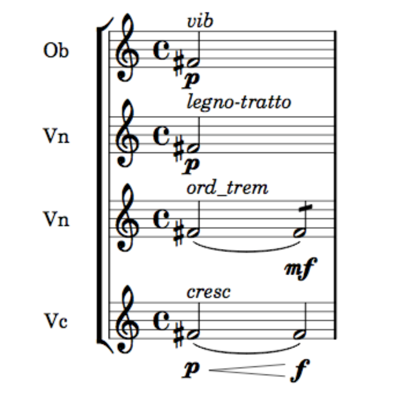
Multi-Objective Time Series Matching

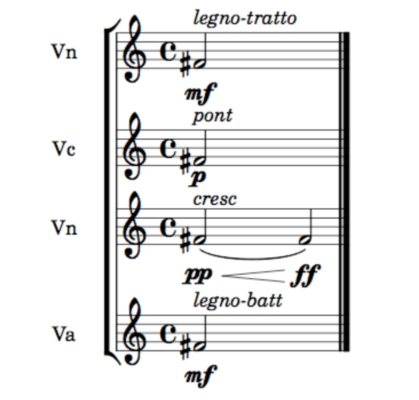
A complex sound target











Dynamic Musical Orchestration Using Genetic Algorithms and a Spectro-Temporal Description of Musical Instruments, **Appl. Of Evolutionary Apps., LNCS 2010** Vol 6025/2010

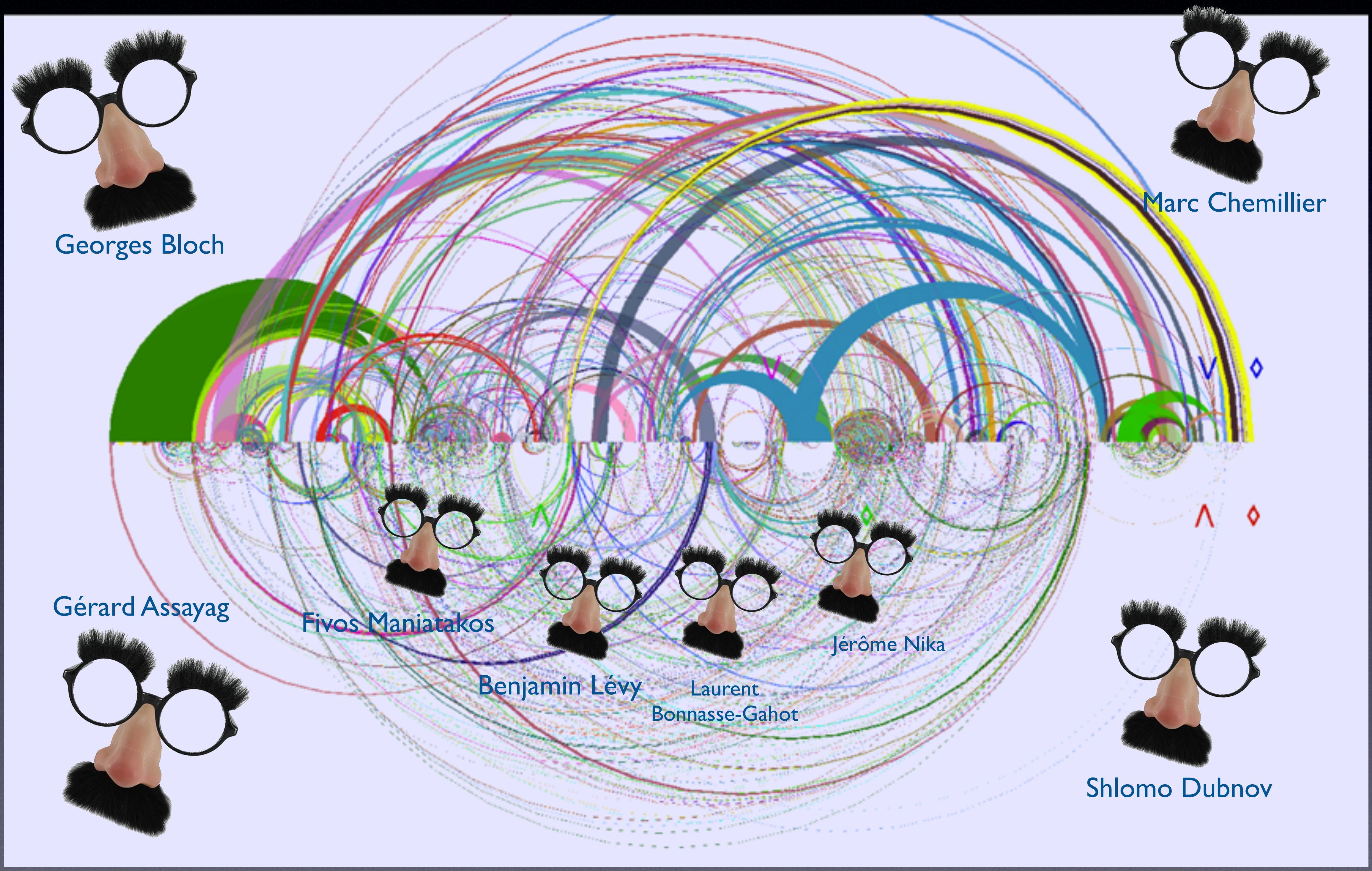
Time series data mining and analysis, **ACM Computing Surveys 2011** (Accepted proof).

Intelligent sound samples database with multiobjective time series matching", **IEEE Transactions on Speech Audio and Language Processing 2012**

Getting it from scratch

The OMax Project

featuring the OMax Brothers



The OMax Paradigm
& family

Style Modeling → Formal grammars for jazz impro Chemillier 2000

Markov

LZ-IP/PST
Dubnov, Assayag al. [1998-2003]

Factor Oracle

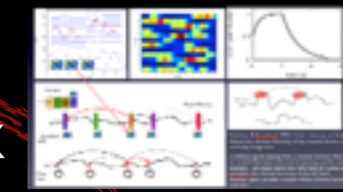
OM-LZ
Assayag 2000

Dubnov, Assayag, Cont
2007 - 2012
Audio Oracle / Information Rate
Information Geometry

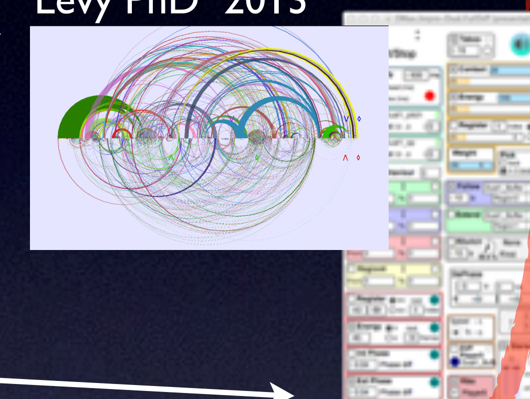
OMax Paradigm

- Listen, learn, model, generate loop
- Assayag, Chemillier, Bloch (2003)

SoMax
Bonnasse-Gahot, Assayag, Bloch
2013



OMax
Assayag, Chemillier, Bloch, 2004-2013
Lévy PhD 2013

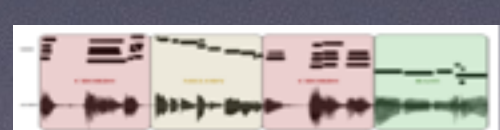


ImproTeK
Chemillier & Nika 2013

....
Conklin [2003]

Continuator
Pachet [2002-2013]

....



MIMI
François [2007]

Native Alien
Bhagwati [2011]

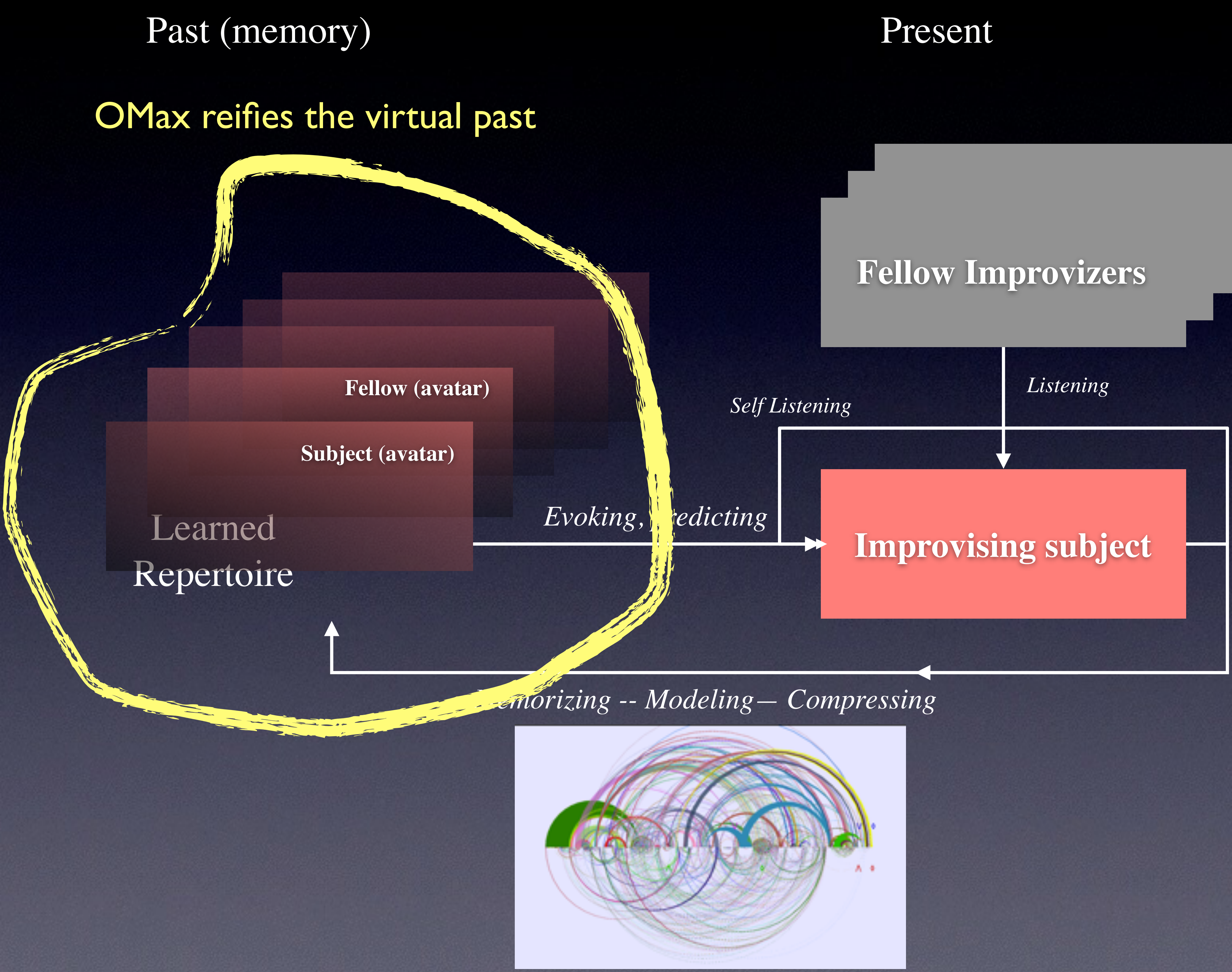
PyOracle
Surges & Dubnov [2013]

Impro Control
Donzé, Seshia, Wesse
(2014)

A Virtual Musician who Learns on the fly

- Before learning one must ***listen*** : efficient machine listening
  - **perception aware** signal segmentation into ***musical units*** distributed on some *geometry*
  - **cognition aware** discovery of an ***alphabet structure*** on the musical units delivering a *symbolic stream*
- ***Learn*** incrementally a ***stylistic sequence model*** into a formal representation directly from the stream of symbols
- ***Generate*** and ***render*** new sequences by *navigating* the model
- These 3 processes (Listen, Learn, Generate) are real-time and concurrent *(competitive and cooperative)* : a unit played by the musician is recognized and integrated in the model after a few milliseconds (close to human performance)

Modeling Improvised Interaction as ***Stylistic Reinjection***



Variable Memory (adaptive) Markov Models

Context-based methods in statistical learning

IPG based on [Lempel,Ziv,78]

Dict = {} ; S : Sequence

While S ≠ ε

S = *pu* with *p* = shortest prefix(S) ∉ Dict

Dict = Dict U {p}

S = u

add shortest prefix not already in dict and move forward

a b a b a c a b a a b c

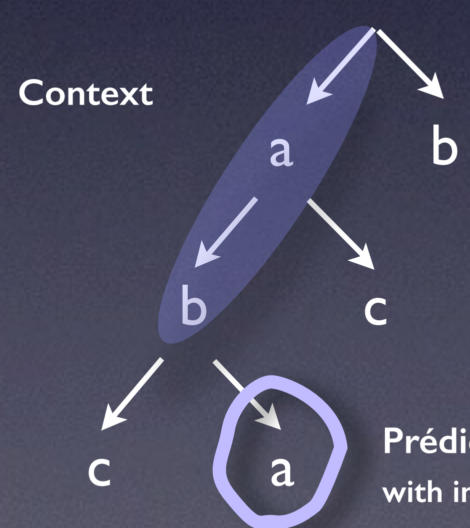
Dict = {a, b, ab, ac, aba, abc}

Optimal coding: average code length for contextes converges to entropy of the source

$$\mathrm{Lim}_{\infty}\ c(n)\ log_2(c(n)) / n = H$$

n : length, c(n) nb of contextes, $log_2(c(n))$ average length of code words

Universal Predictor : adaptively combines predictability of Markovian Models with increasing orders, converges to an **optimal coding** without knowing a-priori the statistical model of the source
Outperforms any fixed-order Markov predictor.



(Papentin L2 level complexity)
The lower the L2 complexity, the more generative

Prédiction (continuation)
with implicit probability distr.

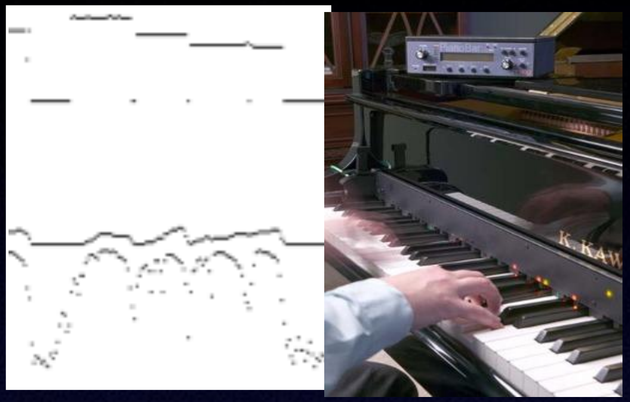
PST : Retain continuation **t** for **abc** iff
P(t | *abc*) > *Pmin*
P(t | *abc*) *sign. different from* P(t | *bc*)

Shape Creation through Context - Prediction
equivalent to navigation into compressed representation

Geraint : Similarity related to Prediction
Daniel Muellensiefen : compression distance performs well in evaluating similarity

Listening Strategies

Monodic signal : pitch follower (Yin~), or Midi Helper



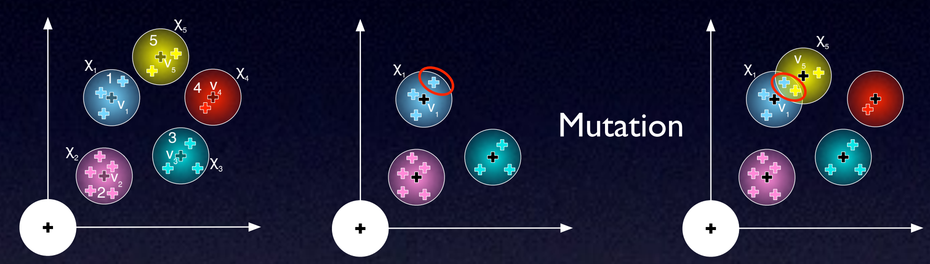
Yin~ quality function

Moog pianobar
Key detection

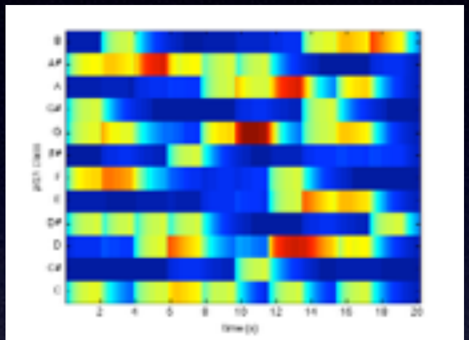
Arbitrary complex signal : audio descriptors

- stream of frame-wise perceptual spectral descriptors (MFCC)
- adaptive quantization of descriptor vectors
- local grouping / averaging -> musical tokens
- adaptive KNN -> symbolic alphabet

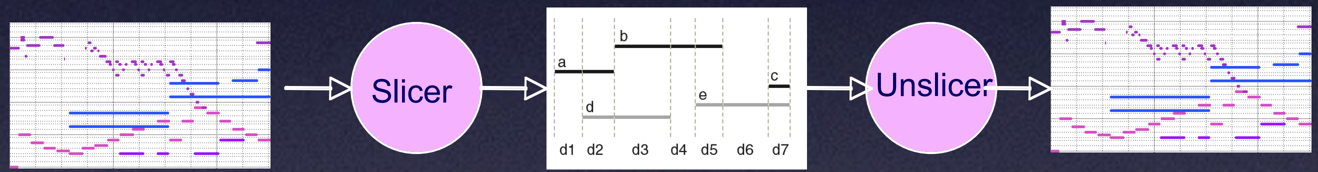
Euclidean Space



Chromagrams

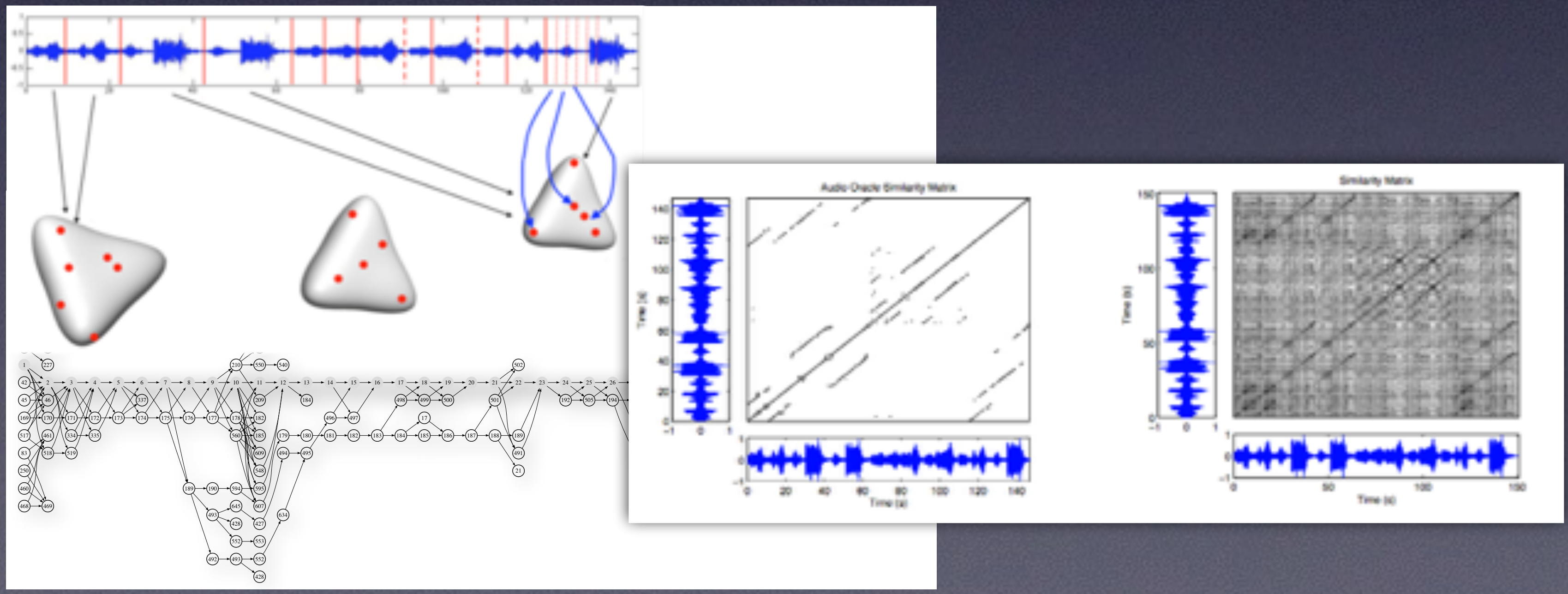


Polyphonic Midi Processing



Information Geometry

- **Machine Listening in an Information geometry** Framework : Distributing Sound Frames over **Riemannian Manifolds** with information metrics over parametric exponential probability spaces
- **Points** are probability distributions (approxed as **frequency domain descriptors**)
- **Distances** are **Bregman Divergences** : amount to relative entropy between perceptual descriptors
- Bregman Balls cluster points into stable musical units abstracted as **symbols** in a formal language alphabet

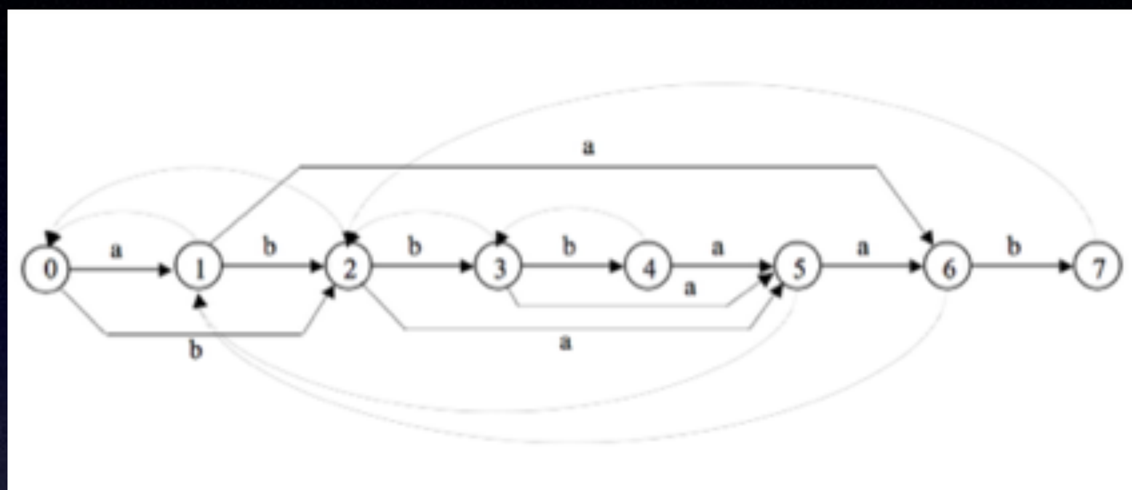


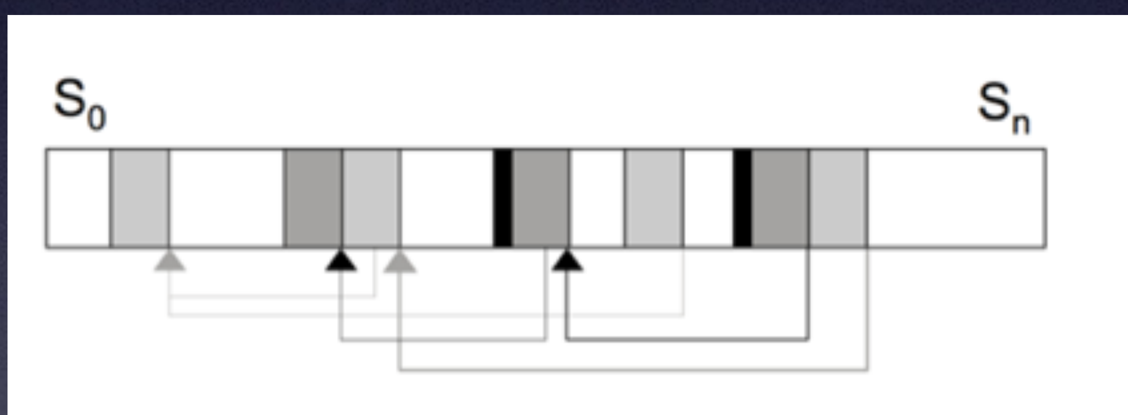
Cont A., Dubnov S., and Assayag G., *On the Information Geometry of Audio Streams with Applications to Similarity Computing*, **IEEE Transactions on Audio, Speech and Language Processing**, Vol. 19, no. 4, Pp. 837-846, May 2011.

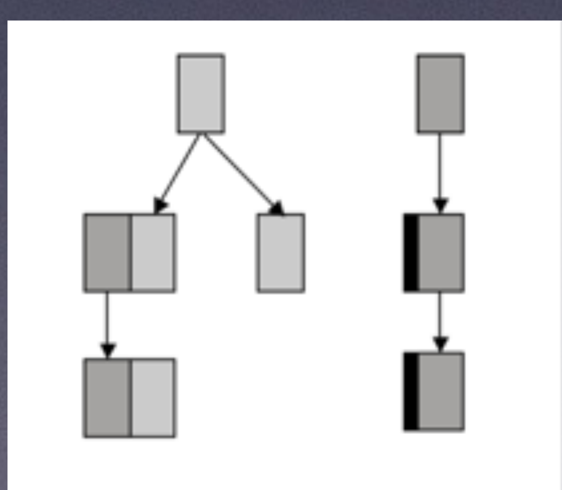
Listen + Learn : Factor Oracle and Suffix Link Tree

Signal / Symbolic articulation

Factor Oracle + Suffix Link Trees

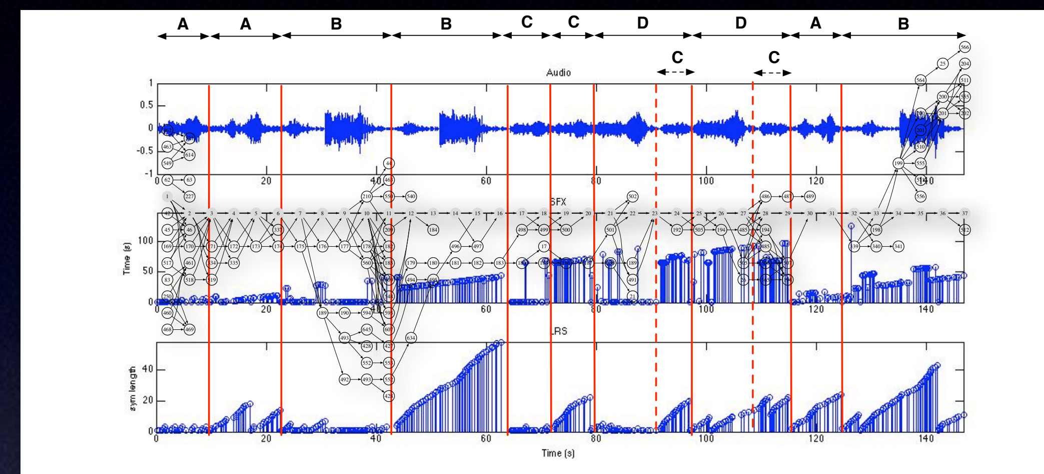


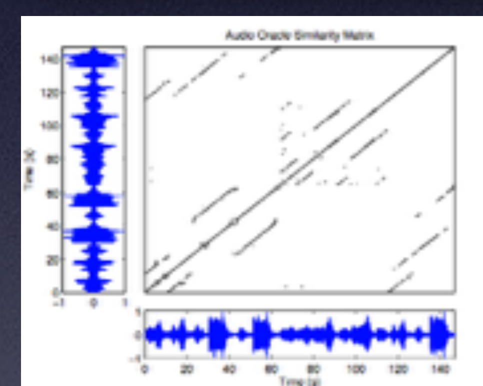


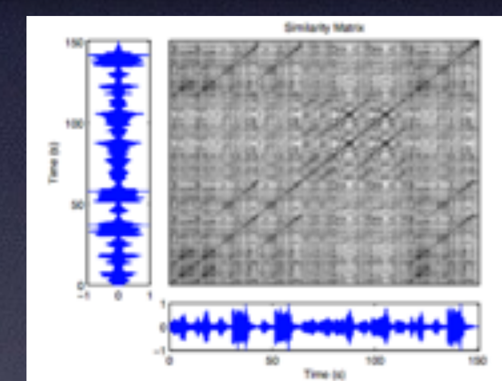


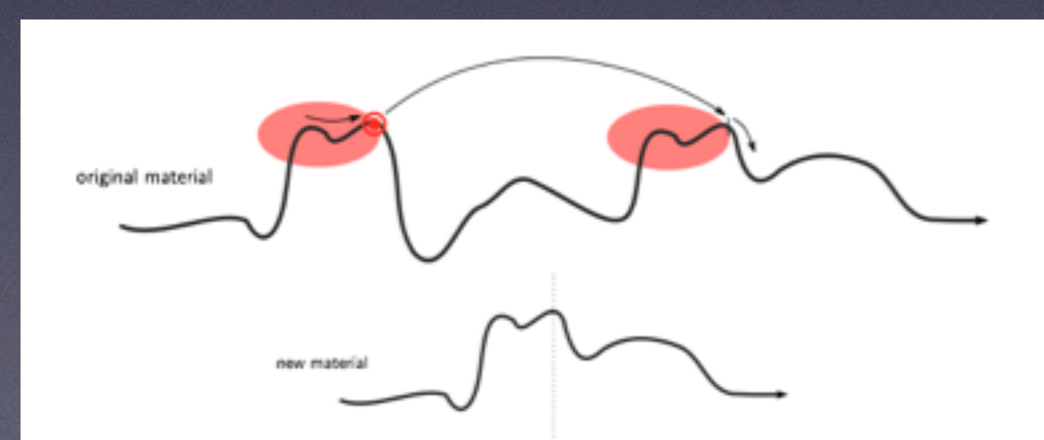
Suffix Link Trees form
a forest of trees fully
explaining
the algebraic partial
order of patterns
in learned sequence

A. Cont, S. Dubnov Info-Geo + Similarity
on Beethoven's First Piano Sonata played by Gulda in 1958









OMax versatility

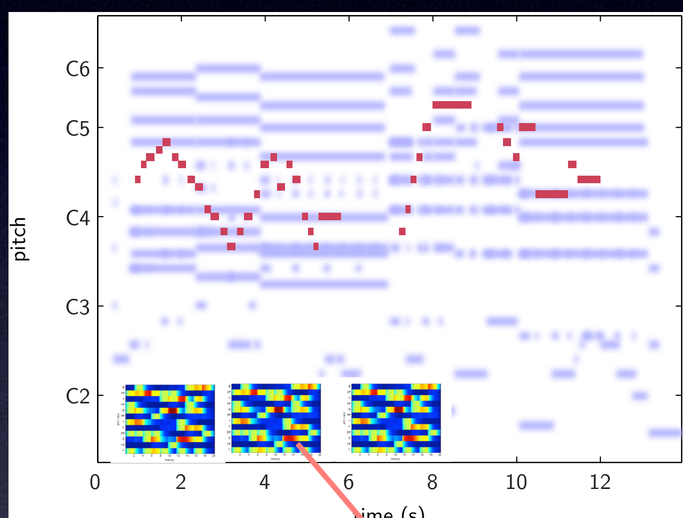
Rhythm and pulse derive from off-line learning of the multi-track (beat aligned) studio sessions

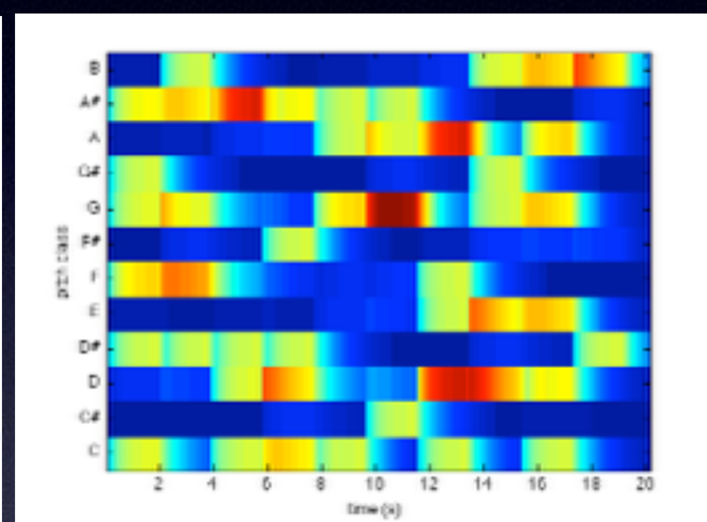
Rhythm and pulse derive from the syllable and prosodic level analysis of similarities

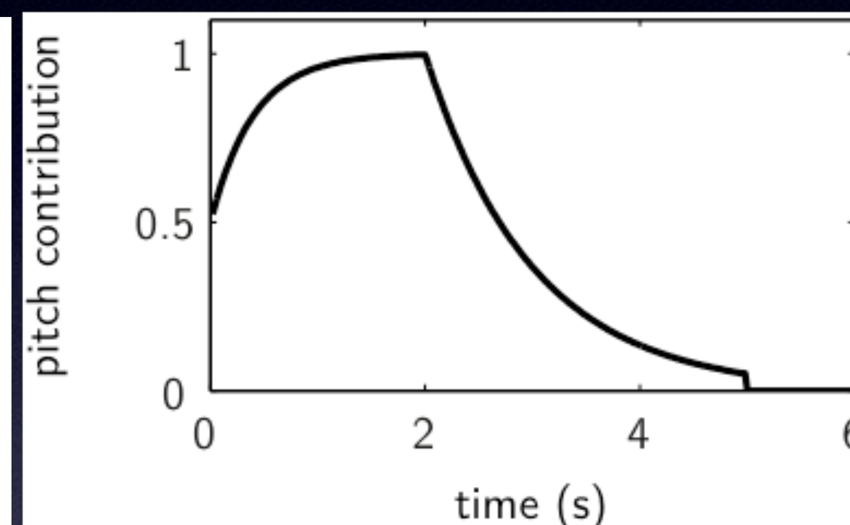
Going SoMax : listen carefully

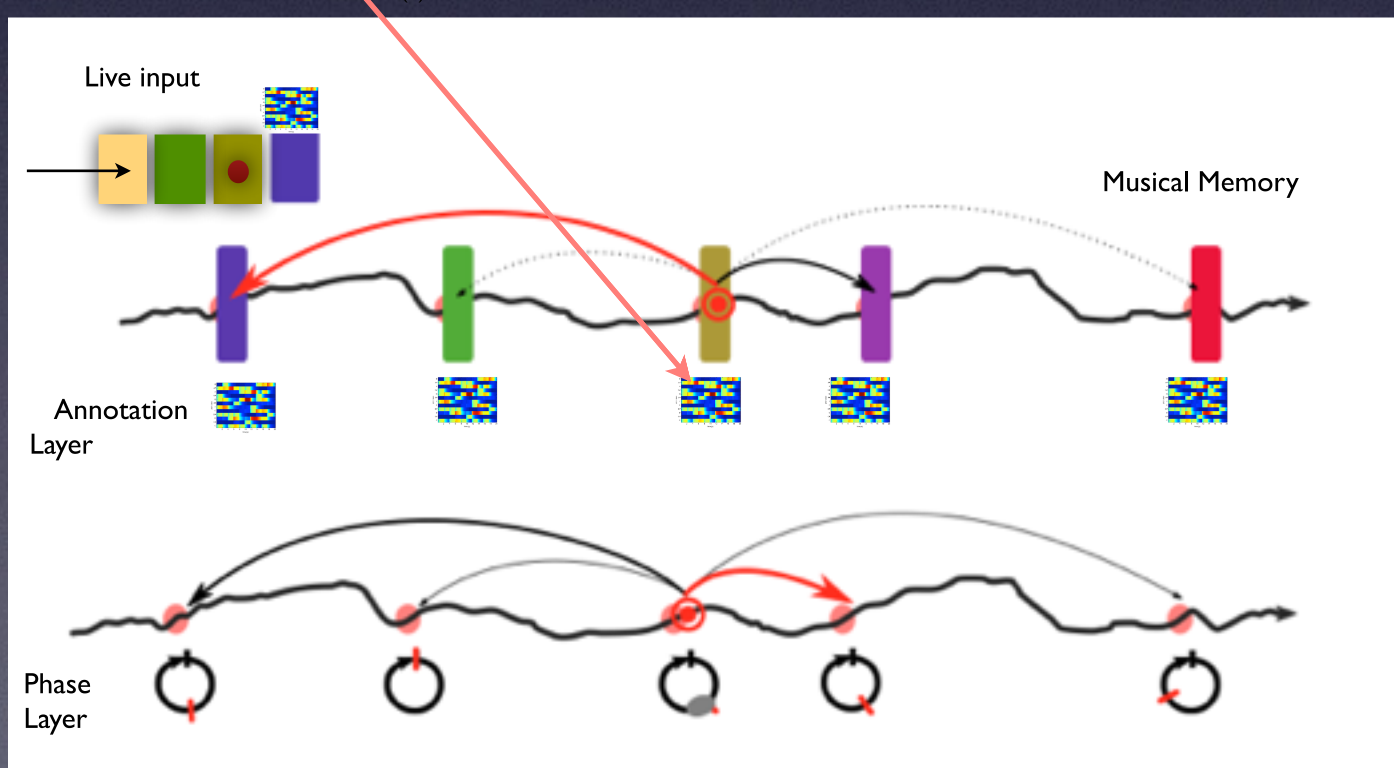
(ANR SOR2, Post-doc L. Bonnasse-Gahot)

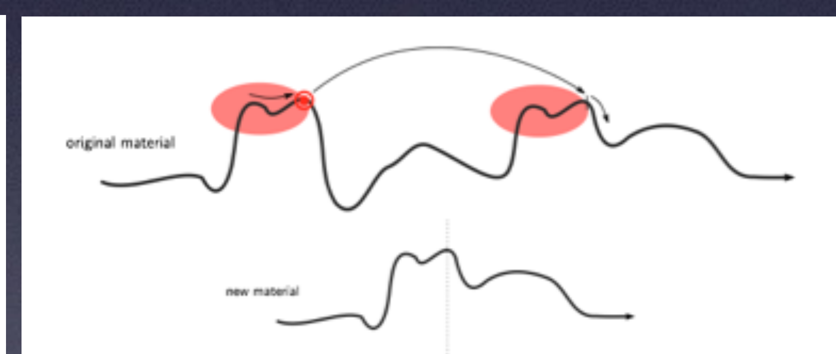
Listen to several channels (foreground, background, voices etc.), One for the main memory model, the others for automatic annotation
Create e.g. a solo memory model plus loose harmonic / textural annotation, or the other way round, or both.
At generation time, match annotation with features extracted from the input





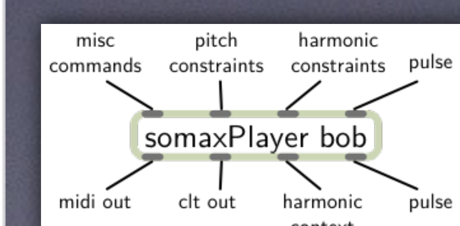




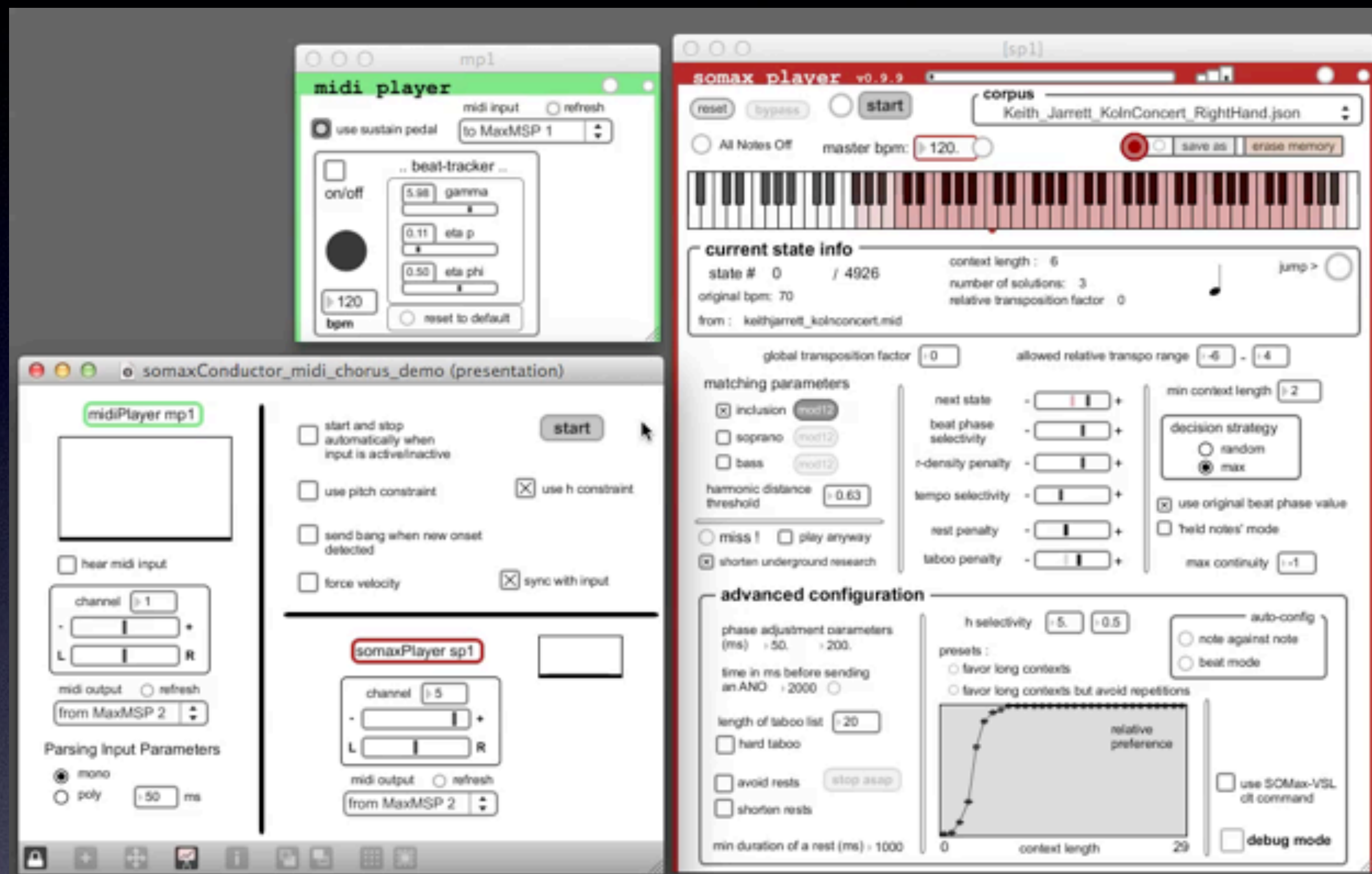


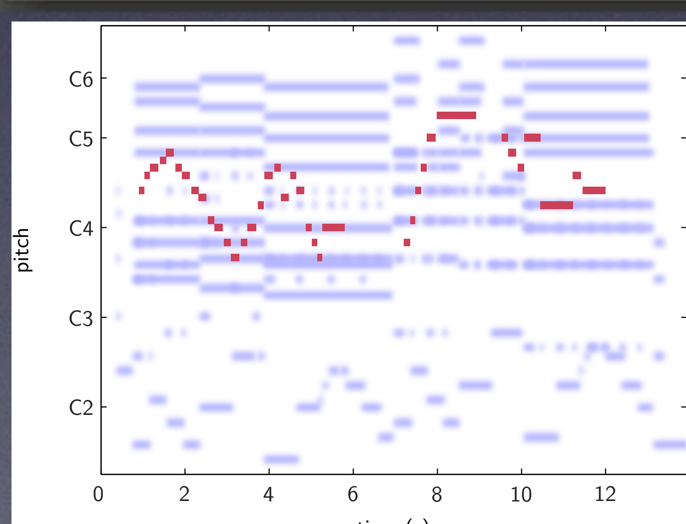
Toiviainen & Krumhansl, 2003, Echoic memory and leaky integration

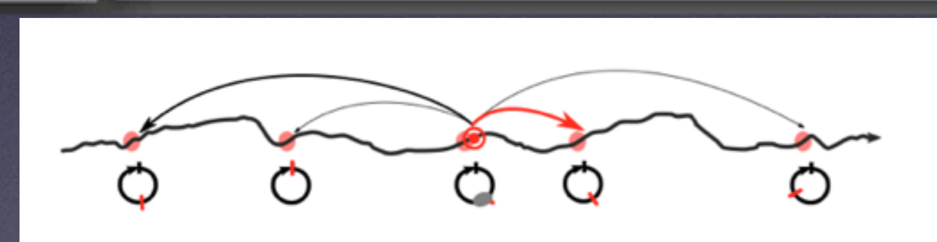
Durational accent; see Parncutt, 1994

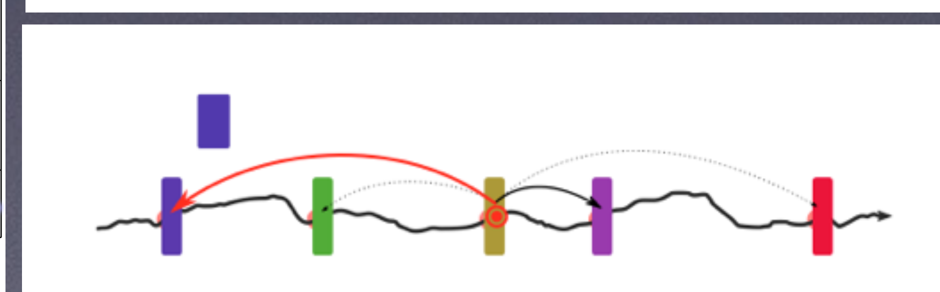


Going SoMax : Autumn in Köln





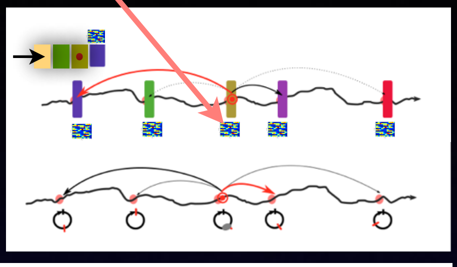


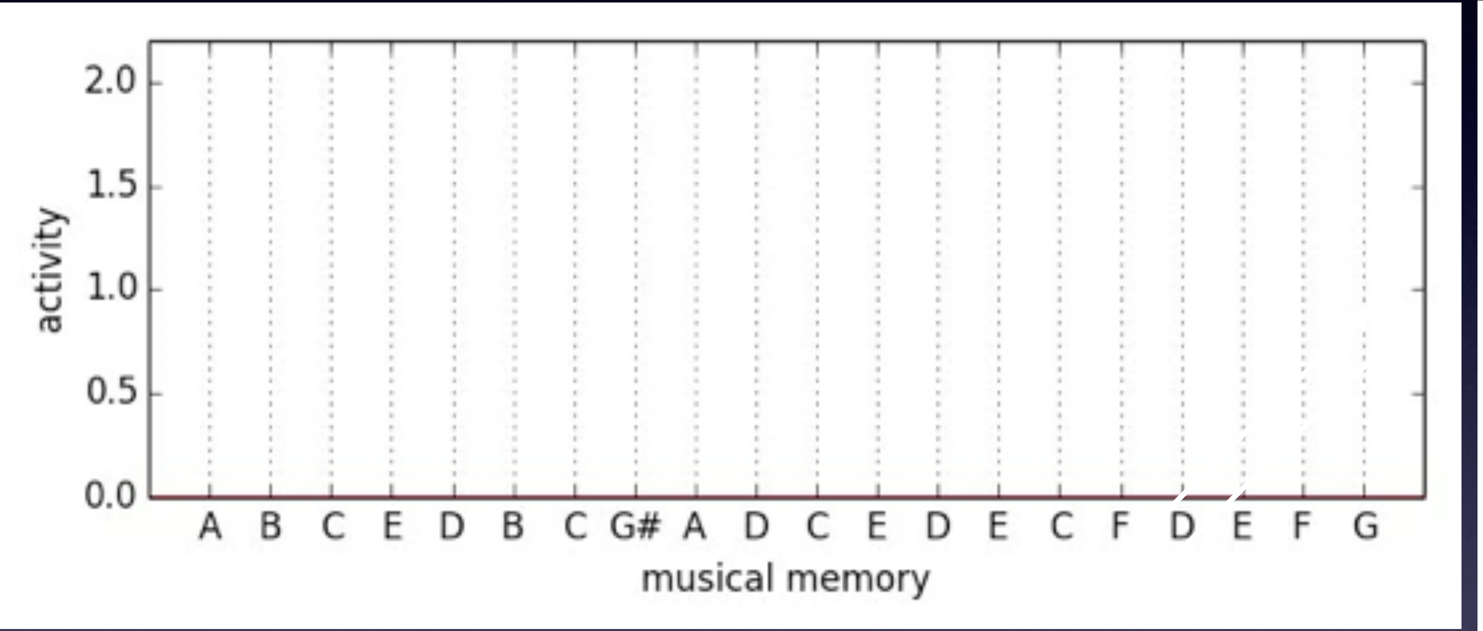


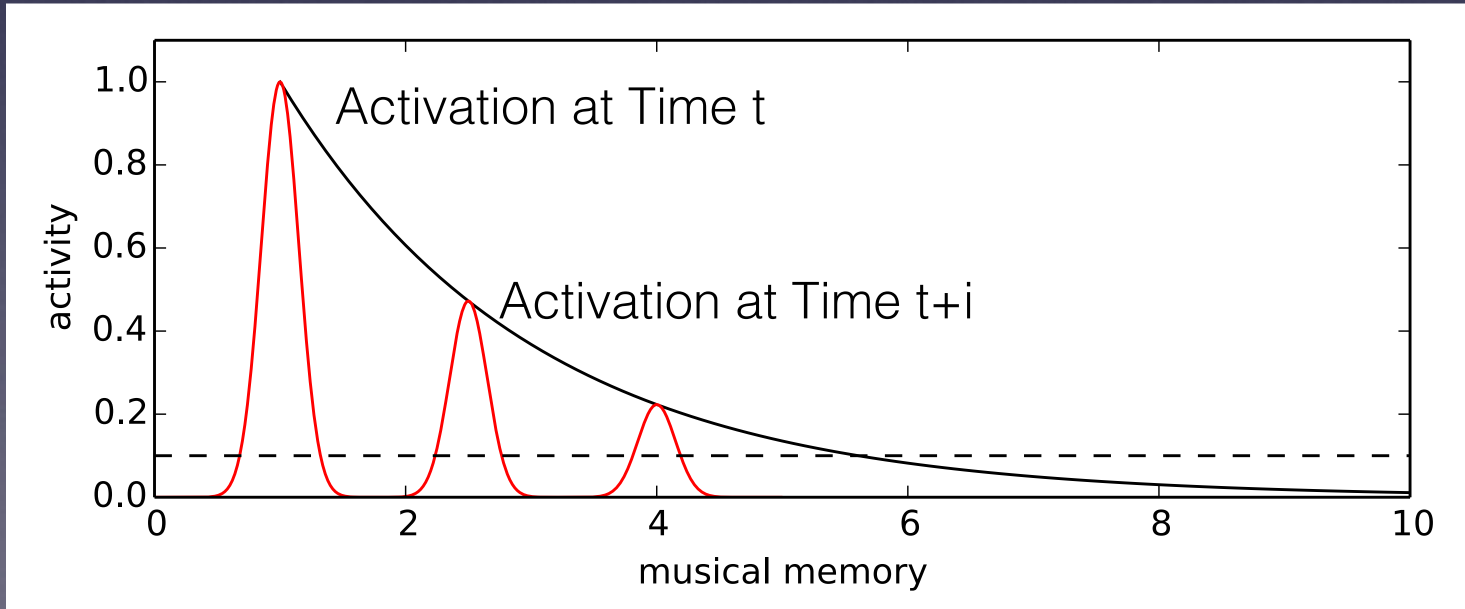
SoMax

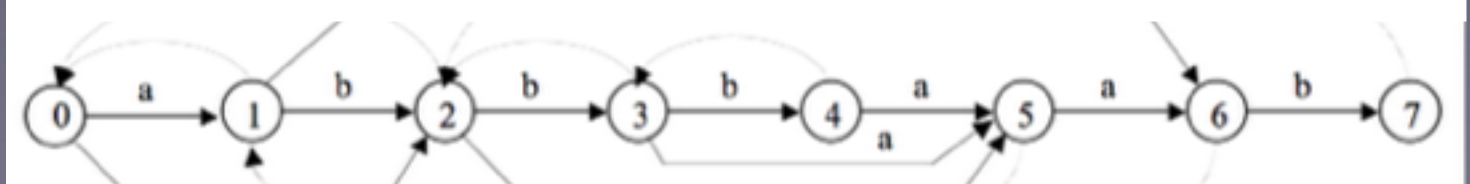
Memory Activation Scheme

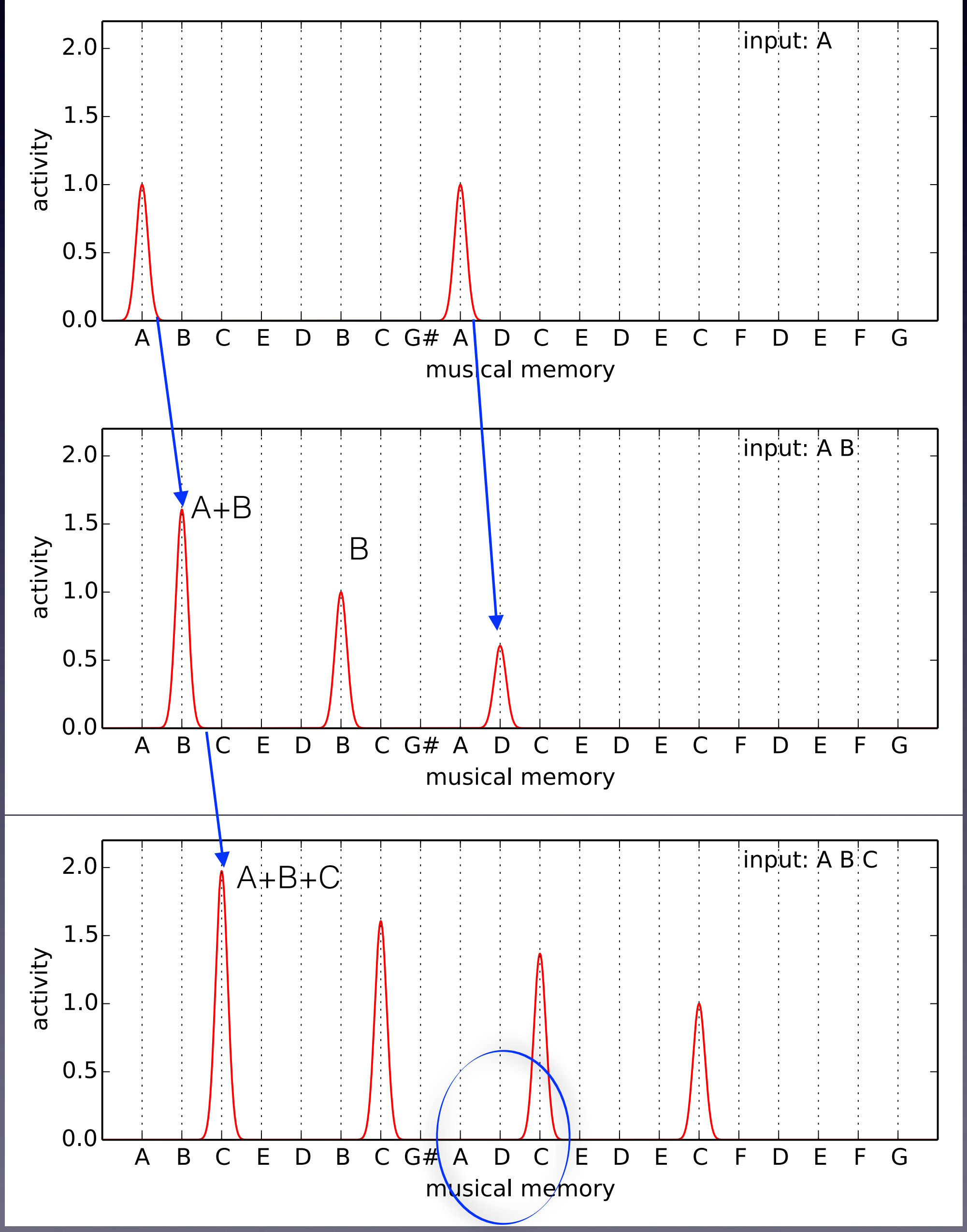
Addressing the **cartographical blindness**, the **evidence accumulation**, and the **cognitive persistence** questions







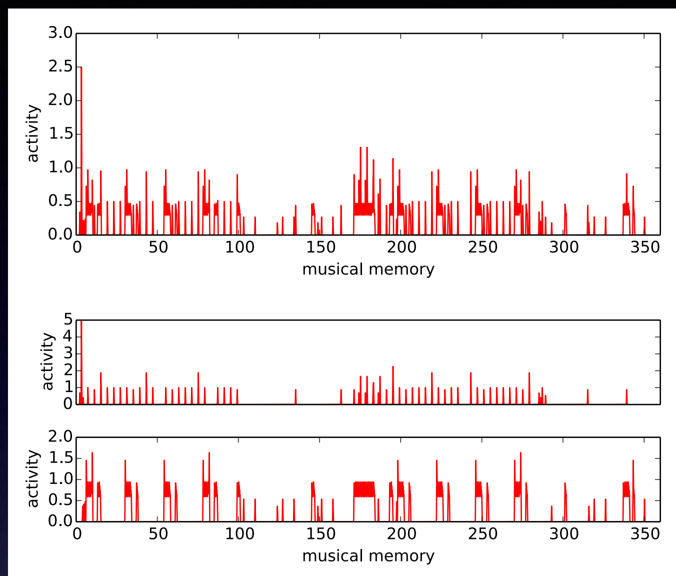


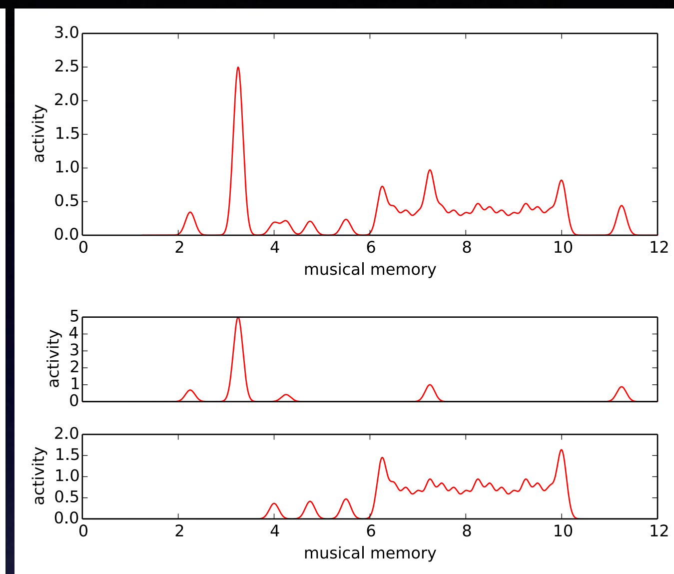


Fuzzy pattern
escape from the purely markovian sequence logic

SoMax

Summing up activation profiles of parallel annotation views including self listening





Beat/Phase Profile

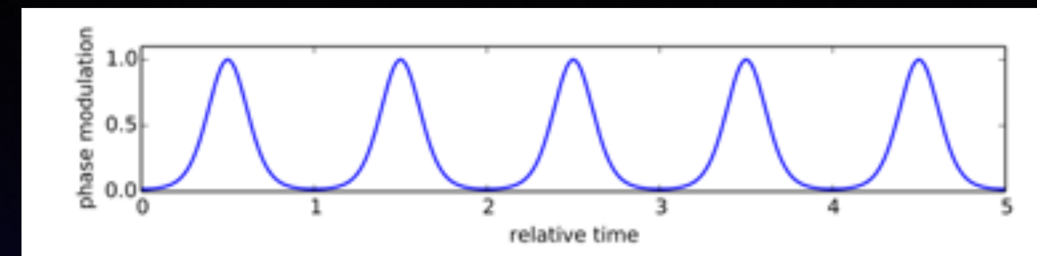
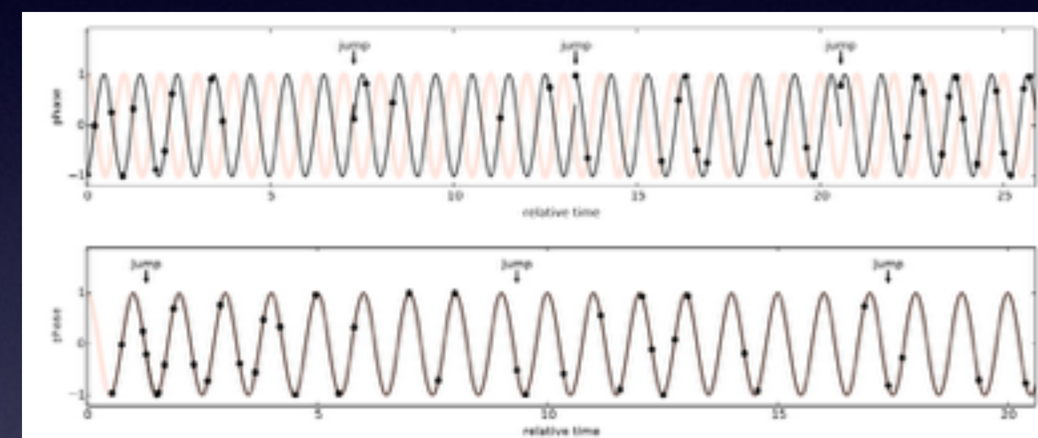
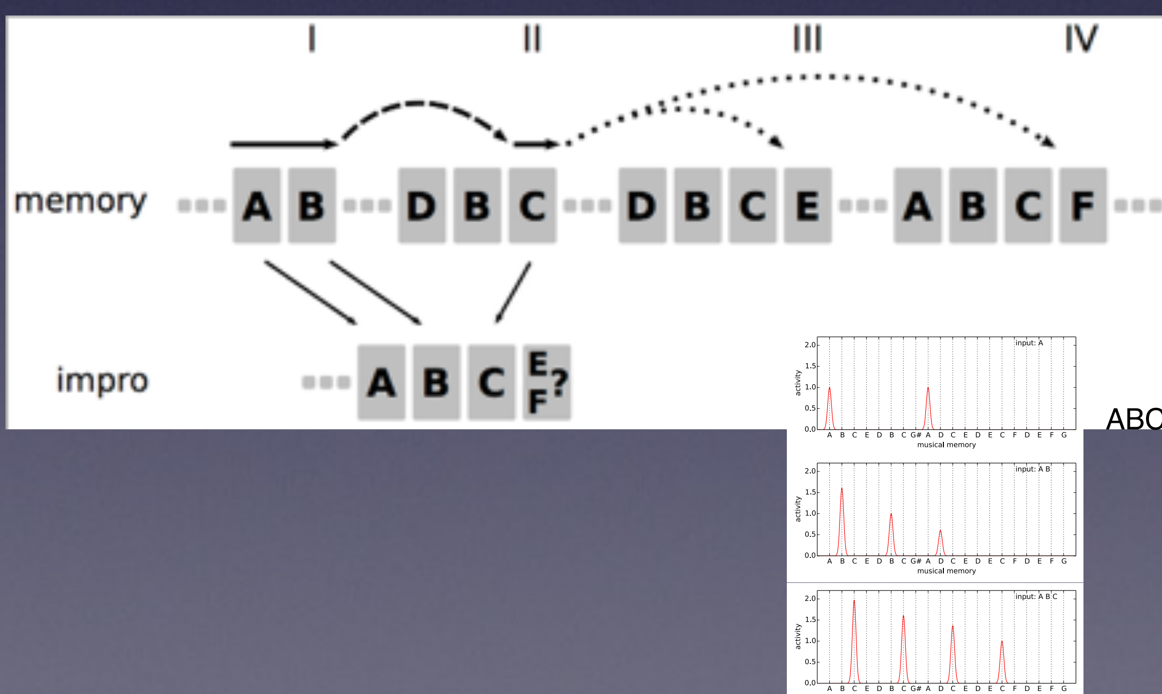
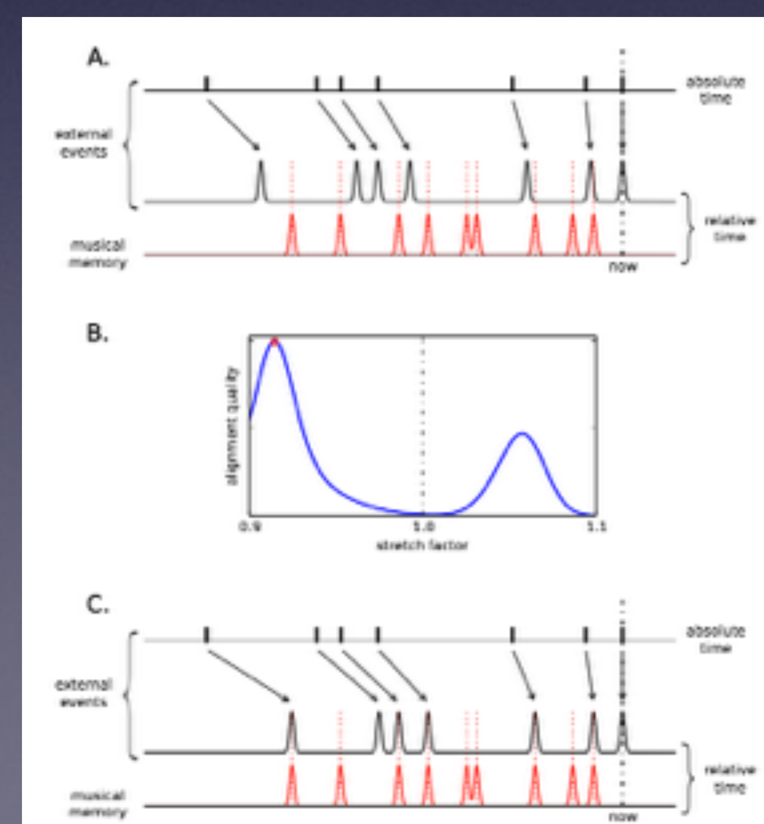


Figure 14: Phase modulation function, $\exp(\eta[\cos(2\pi(\xi - \xi_{\text{target}})) - 1])$. In this illustration, $\xi_{\text{target}} = 0.5$ and $\eta = 2.0$.



Self Listening : listening to the memory, or to the generation

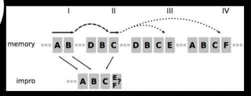


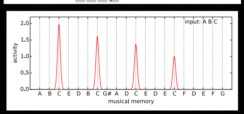


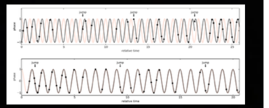
Flexible Time

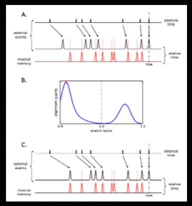
In the mood of Time Remembered (Rémi Fox + Bill Evans)

1 agent trained on Bill Evans music. note/note, flexible time adjustment, melodic listening
Right anticipation at 01:20









Schoenberg revisited

1 agent A1 trained off-line on corpus of Schoenberg's Drei Klavierstücke, Op. 11. A1 improvises with melodic listening to Rémi's Impro. 2nd agent A2 learns on the fly from Rémi.s impro audio stream, with additional harmonic view coming from A1's impro. In second part, (2:36)A1 and A2 improvise together :A2 listens to A1's harmony, A1 listens to A2's melody. Rémi and Laurent get back into the game.