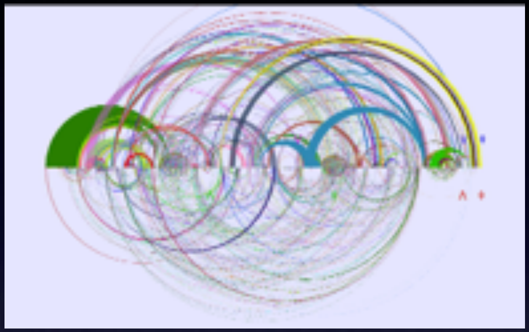
Creative Dynamics of Improvised Musical Interaction



IRCAM

Sciences and Technologies of Music and Sound (STMS) Lab

Music Representation Team

Creative Dynamics of Improvised Interaction between Humans and Digital Agents

- At a current time, a creative agent's improvised decision process relies on
  - evaluation of past history
  - analysis of incoming events
  - anticipation strategies
- it takes time to come to a decision, and part of this decision can also be to act later
- Involves time and memory at different scales, just as in music composition
- Cannot be fully apprehended by mainstream signal / event processing technologies (machine musicianship)

Creative Dynamics of Improvised Interaction between Humans and Digital Agents

- Bring into synergy a combination of means
  - **machine listening**—extracting high level features from the signal and turning them into significant symbolic units;
  - **machine learning**— discovering and assimilating on the fly intelligent schemes by listening to actual performers;
  - **stylistic simulation**—elaborating a consistent model of style;
  - **symbolic music representation construction** —formalized representations connecting to organized musical thinking, analysis and composition.
- These means are parallel processes that **cooperate** and/or **compete**
- Multi-level **machine musicianship** / **smart musical memory** with discovery and generative skills contributing to the emergence of **creative digital musical agents**

Creative Dynamics of Improvised Interaction between Humans and Digital Agents

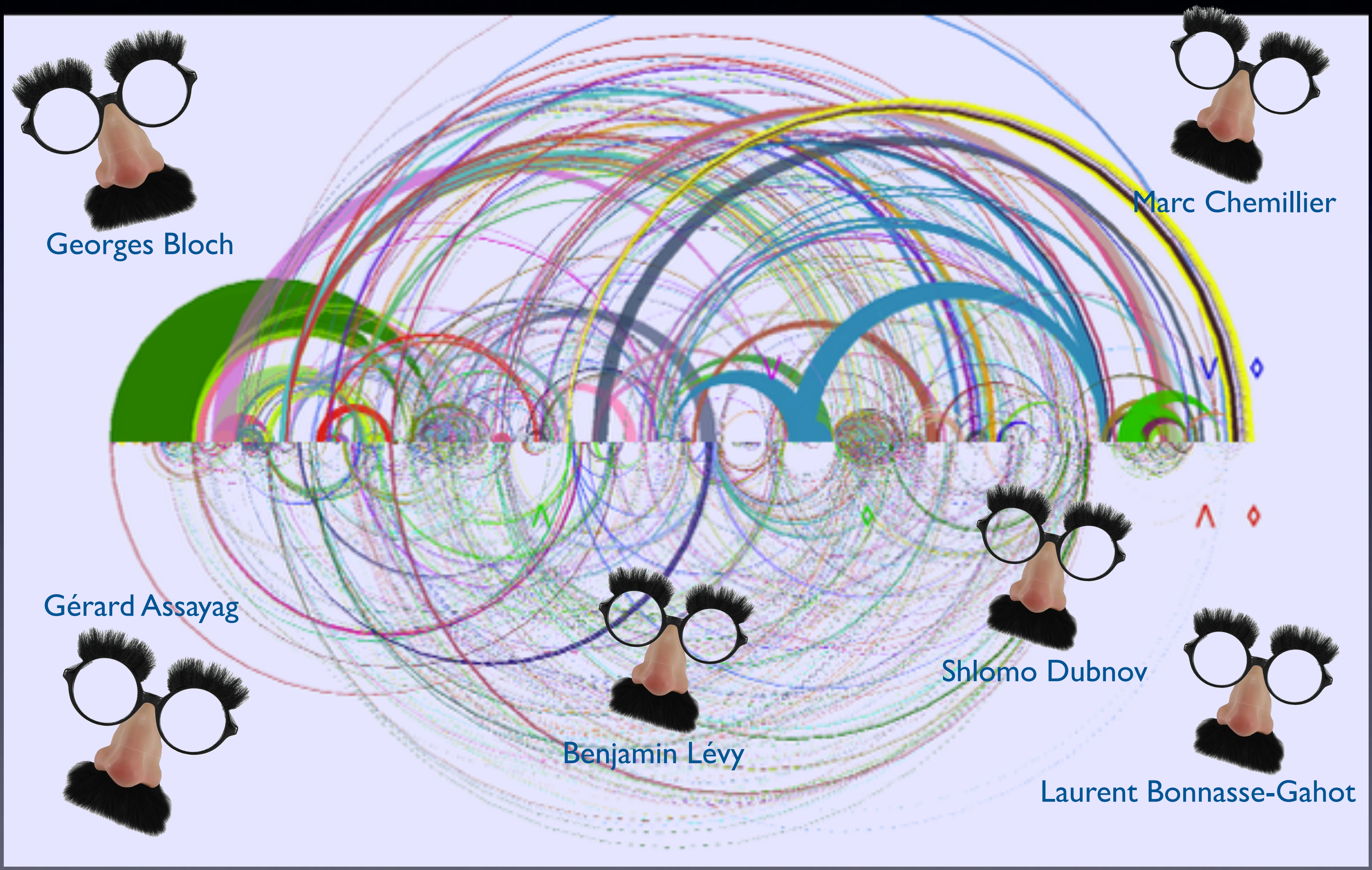
- An instance of a creative agent : **OMax** and its siblings (SoMax, ImproTek, MiMi, Native Alien etc.)
- Cooperation between heterogeneous components specialized in **real-time audio** signal processing, high level **music representations** and **formal knowledge** structures (symbolic interaction).
- **Learns and plays on the fly** in live setups
- Connect instant **contextual listening** to **corpus based knowledge,** with longer term investigation and decision processes allowing the system to **refer to large-scale structures and scenarios** while following the human.

Creative Dynamics of Improvised Interaction between Humans and Digital Agents

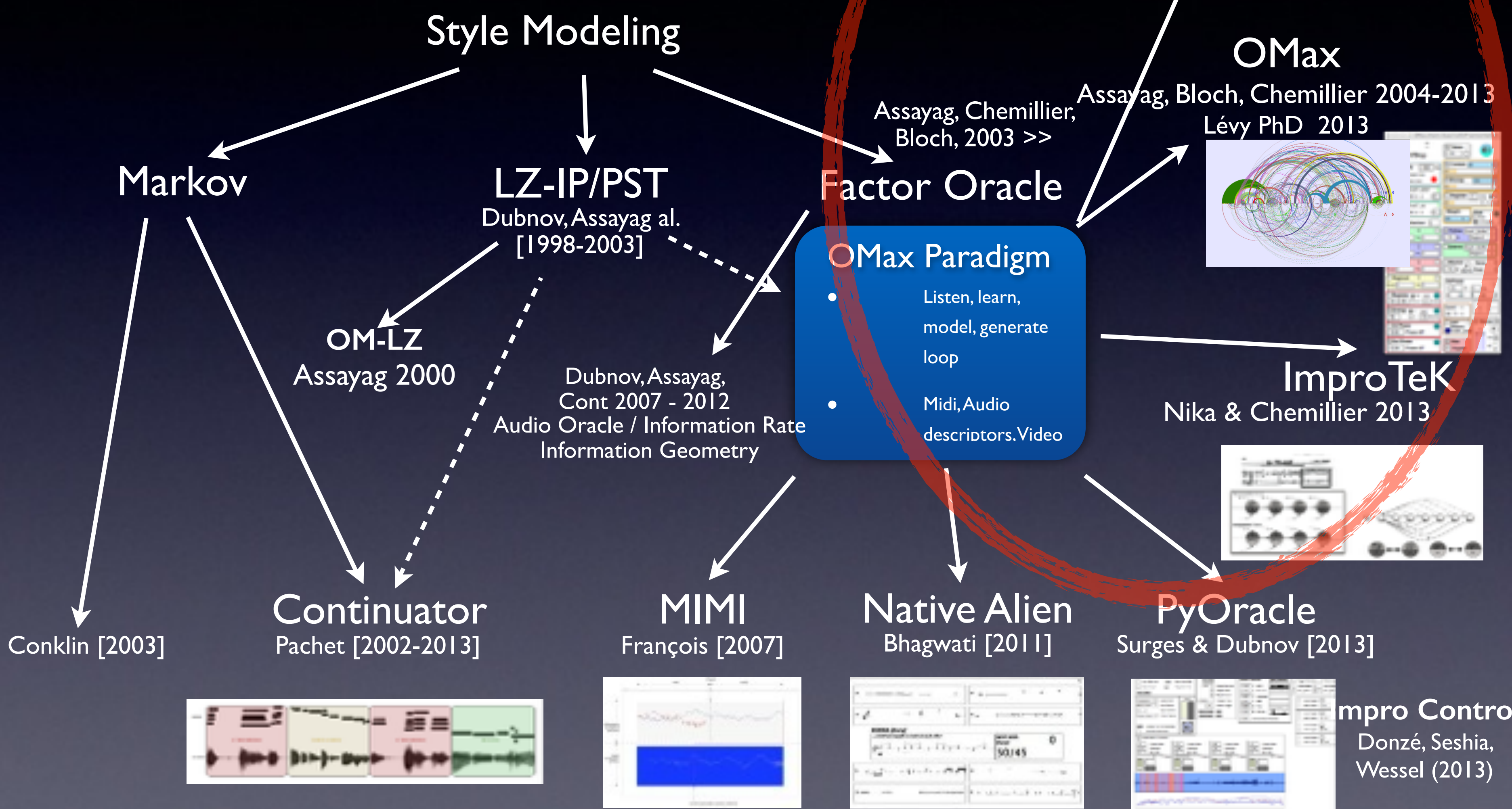
- **Symbolic Interaction** defines a new artificial creativity paradigm in computer music, and extends to other fields as well.
- **Symbolic Interaction** brings together the advantages of the worlds of interactive real-time computing and of intelligent, content-level analysis and processing
- **Interact** with an agent which develops freely in its own ways while keeping in style with the user.
- B**ring together composition and improvisation** through modeling cognitive structures and processes
- A **decision-making paradigm**, weave decisions step after step either by deciding to **relate** to an **overall structural determinism** / music memory or to **jump** in an improvised way so as to generate surprise.
- **Liveliness** is precisely a mixture of **deterministic** and **unexpected** (**improvised**) **behaviors** as in living organisms
- Music is a synthesis of time from the dialectics of immanent forms and transcendent causality (*Platonic Rhizomes 16/9 @ Plato Academia*)

The OMax Project

featuring the OMax Brothers



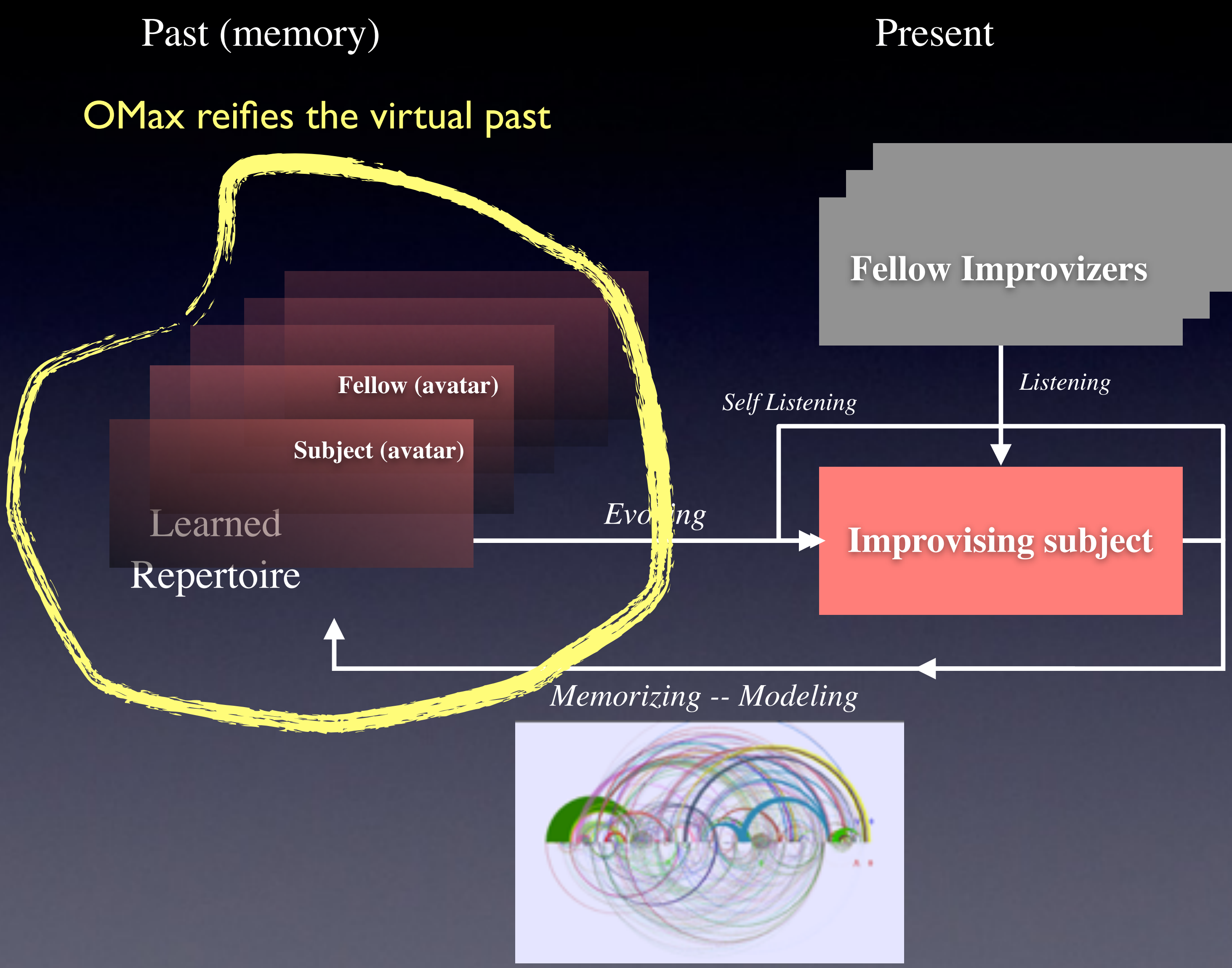
The OMax Paradigm
& family



Virtualizing a Musician

- **Context** : a live music performance
- **Data** : audio streams arriving in in real time
- **Model** : a style model capturing the musical logics underlying the signal stream
- **Output** : a virtual musician agent that plays with the same musical logics and with the same sound (musical behavioural simulation)
- Resulting situation : the human musician plays freely along with the virtual one (creative clone) in an improvisation setup

Modeling Improvised Interaction as stylistic reinjection



A Virtual Musician that Learns

- Before learning one must ***listen*** : efficient machine listening
  - **perception aware** stream 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*** using formal languages methods on the symbolic stream
- ***Generate*** and ***render*** new sequences by navigating the model
- These 3 processes (Listen, Learn, Generate) are <u>real-time</u> and <u>concurrent</u> : a unit played by the musician is recognized and integrated in the model after a few milliseconds (close to human performance)

Stylistic Modeling of Musical Sequences seen as compression

- In 2000, Hutter showed that **finding the optimal behavior of a rational agent is equivalent to compressing its observations.** The proof applies to any goal-seeking agent in any unknown environment which can be simulated by a computer.
- There is a close connection between machine learning and compression: **a system that predicts** the posterior probabilities of a sequence given its entire history can be used for **optimal data compression** ... , while an **optimal compressor can be used for universal prediction** (by finding the symbol that compresses best, given the previous history). This equivalence has been used as justification for data compression as a benchmark for "general intelligence"

Variable Memory (adaptive) Markov Models

Context-based methods in statistical learning

IPG based on [Lempel,Ziv,78]

Dict = {} ; S : Sequence

While S ≠ ε

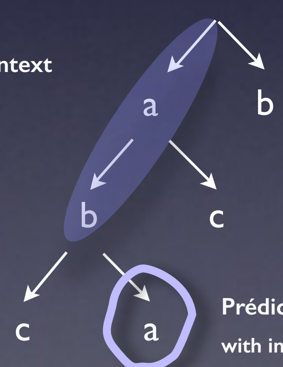
- S = *pu* with *p* = shortest prefix(S) ∉ Dict
- Dict = Dict U {p}
- S = u

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}

Context



(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*)

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

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

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.

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

Sequence Modeling and stylistic imitation

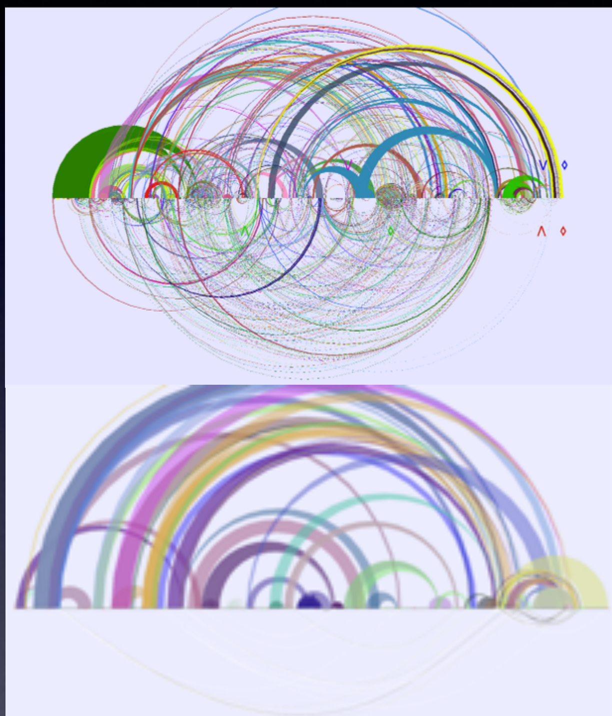
- Generation in the style of Ricercar J.S. Bach
- A Chorus in the style of Charlie Parker

Bernard Lubat and the bebop style

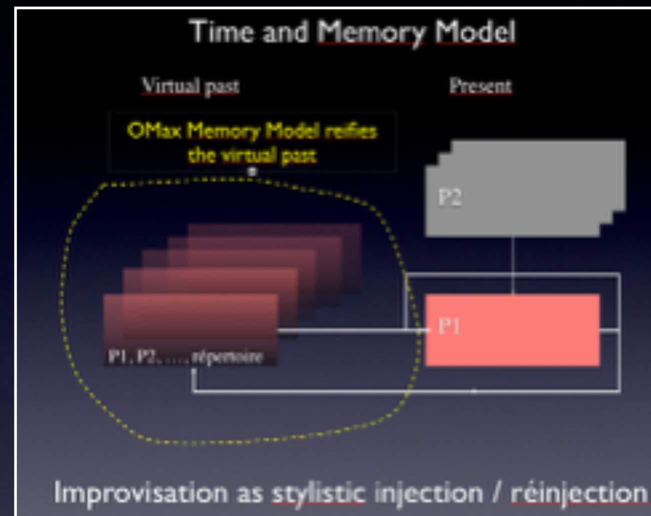
From Midi to Audio : a chorus in the style of Jaco Pastorius

From Style Model to Creative Agents

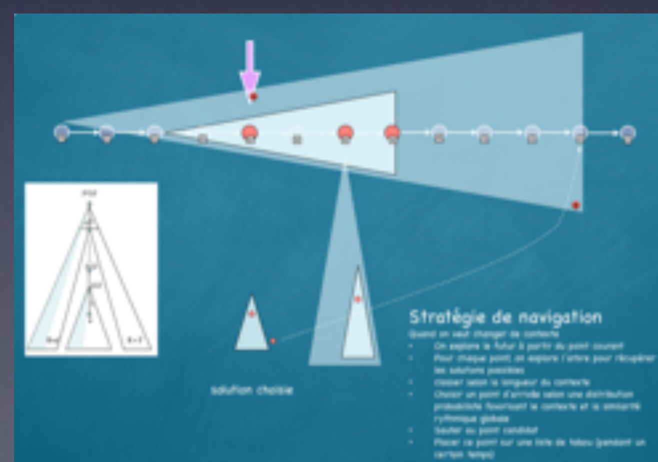
Representation of knowledge



General epistemological framework



Perceptual and Generative strategies

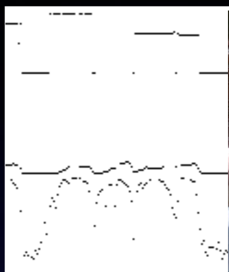


A Virtual Musician that Learns

- Before learning one must ***listen*** : efficient machine listening
  - **perception aware** stream 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*** using formal languages methods on the symbolic stream
- *Generate* and *render* new sequences by navigating the model
- These 3 processes (Listen, Learn, Generate) are <u>real-time</u> and <u>concurrent</u> : a unit played by the musician is recognized and integrated in the model after a few milliseconds (close to human performance)

Listening Strategies

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



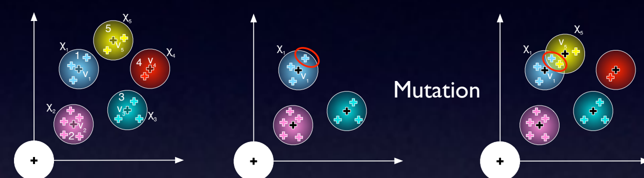
Yin~ quality function

Moog pianobar
Key detection

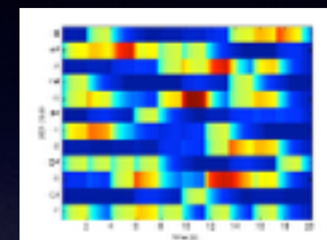
Arbitrary complex signal : audio descriptors

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

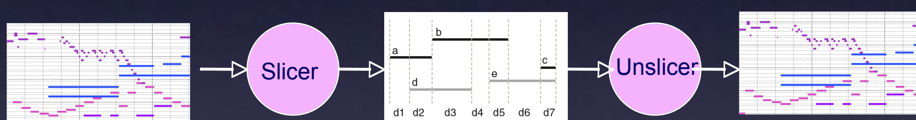
Euclidean Space



Chromagrams



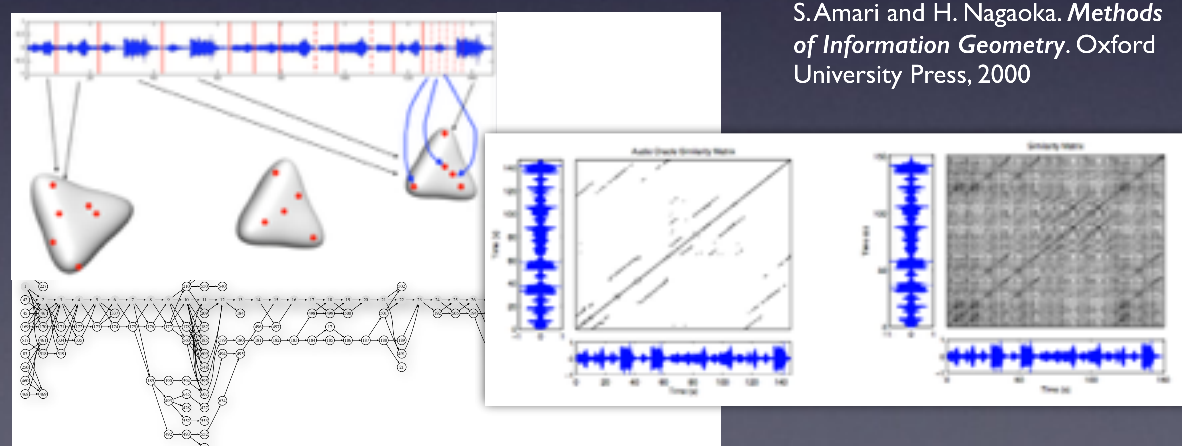
Polyphonic Midi Processing



S. Amari and H. Nagaoka. ***Methods of Information Geometry***. Oxford University Press, 2000

Information Geometry

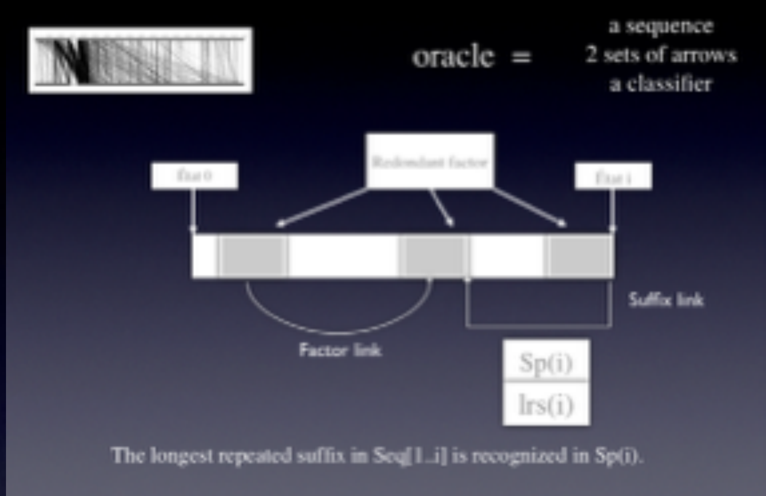
- **Machine Listening in an Information geometry** Framework : Distributing Sound Frames over **Riemannian Manifolds** with information metrics over parametric exponential probability spaces
- **Points** are probability distributions (approxed as **frequency domain descriptors**)
- **Distances** are **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

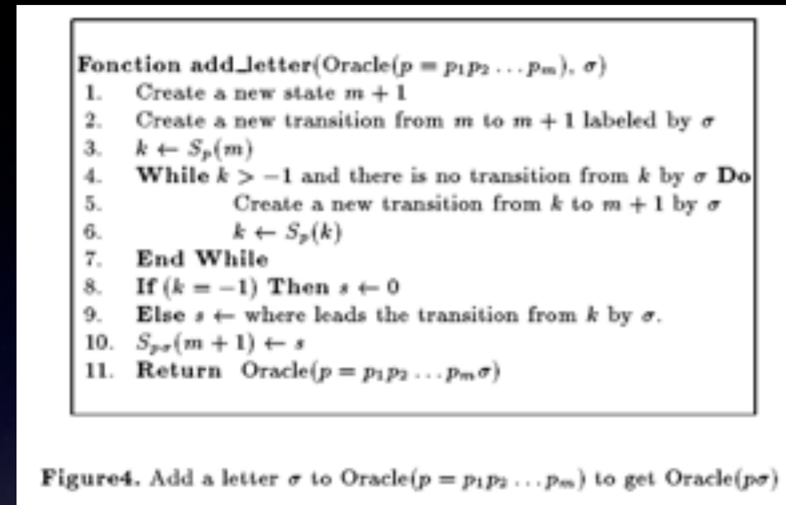


Cont A., Dubnov S., and Assayag G., *On the Information Geometry of Audio Streams with Applications to Similarity Computing*, **IEEE Transactions on Audio, Speech and Language Processing**, Vol. 19, no. 4, Pp. 837-846, May 2011.

Learn : Factor Oracle and Suffix Link Tree

Oracle structure





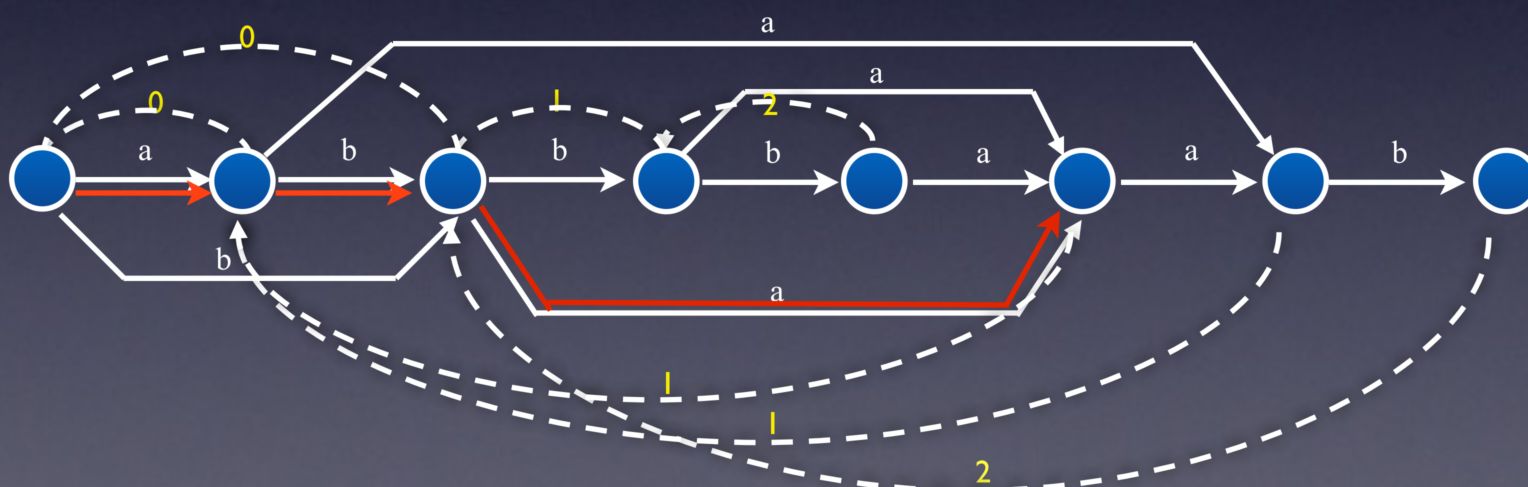
Fonction add_letter(Oracle($p = p_1 p_2 \ldots p_m$), σ)

1. Create a new state $m+1$
2. Create a new transition from m to $m+1$ labeled by σ
3. $k \leftarrow S_p(m)$
4. **While** $k > -1$ and there is no transition from k by σ **Do**
5. Create a new transition from k to $m+1$ by σ
6. $k \leftarrow S_p(k)$
7. **End While**
8. **If** ($k = -1$) **Then** $s \leftarrow 0$
9. **Else** $s \leftarrow$ where leads the transition from k by σ.
10. $S_{p\sigma}(m+1) \leftarrow s$
11. **Return** Oracle($p = p_1 p_2 \ldots p_m \sigma$)

Figure4. Add a letter σ to Oracle($p = p_1 p_2 \ldots p_m$) to get Oracle($p\sigma$)

abbbaab

Crochemore & al, 99

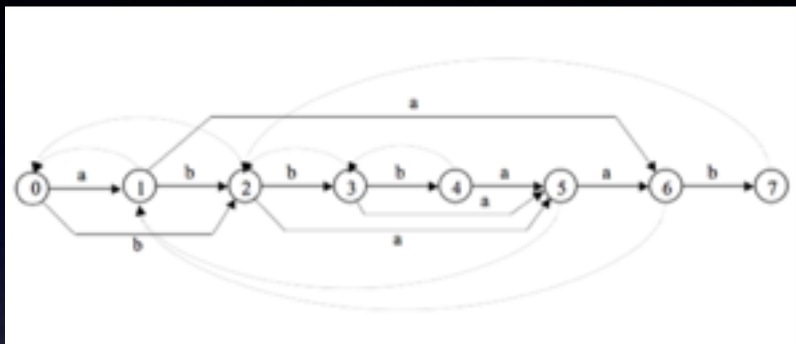


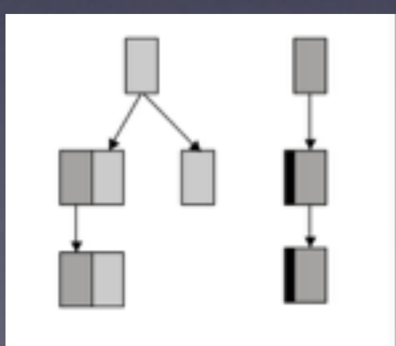
notice aba which is not a factor of the sequence

Listen + Learn : Factor Oracle and Suffix Link Tree

Signal / Symbolic articulation

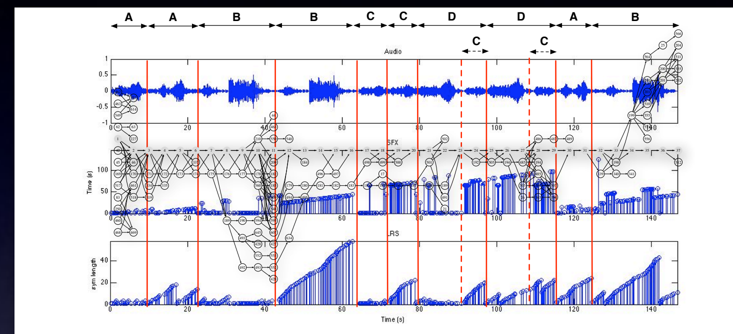
Factor Oracle + Suffix Link Trees

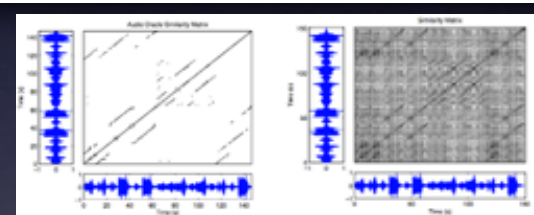


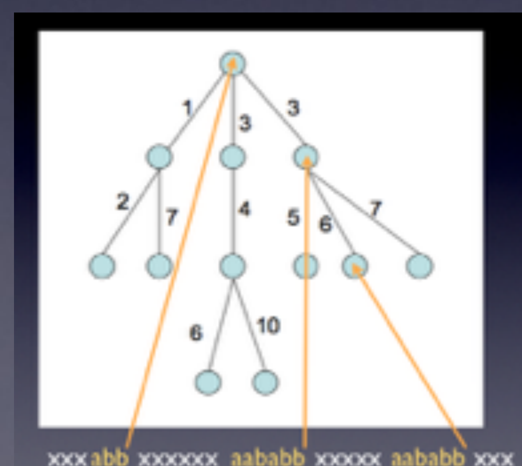


Suffix Link Trees form
a forest of trees fully
explaining
the algebraic partial
order of patterns
in learned sequence

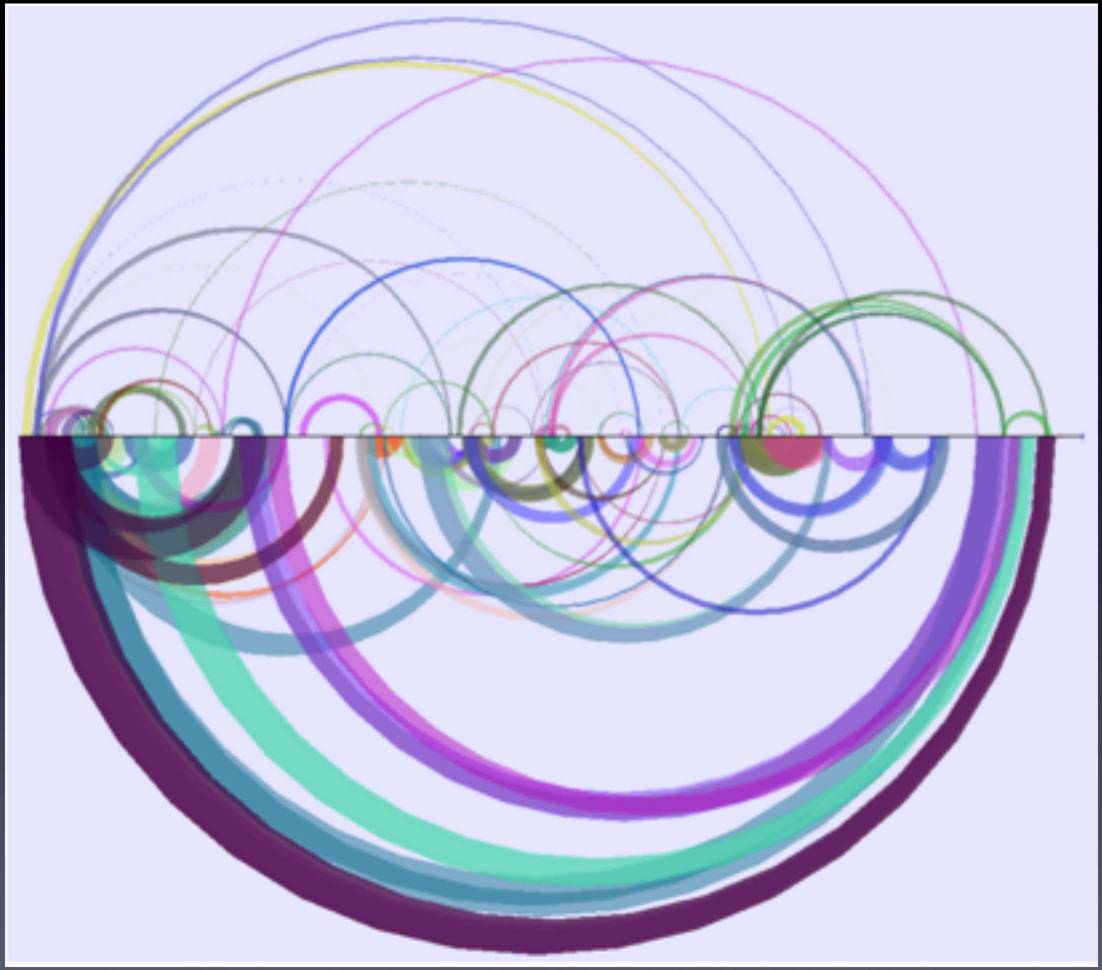
A. Cont, S. Dubnov Info-Geo + Similarity
on Beethoven's First Piano Sonata played by Gulda in 1958



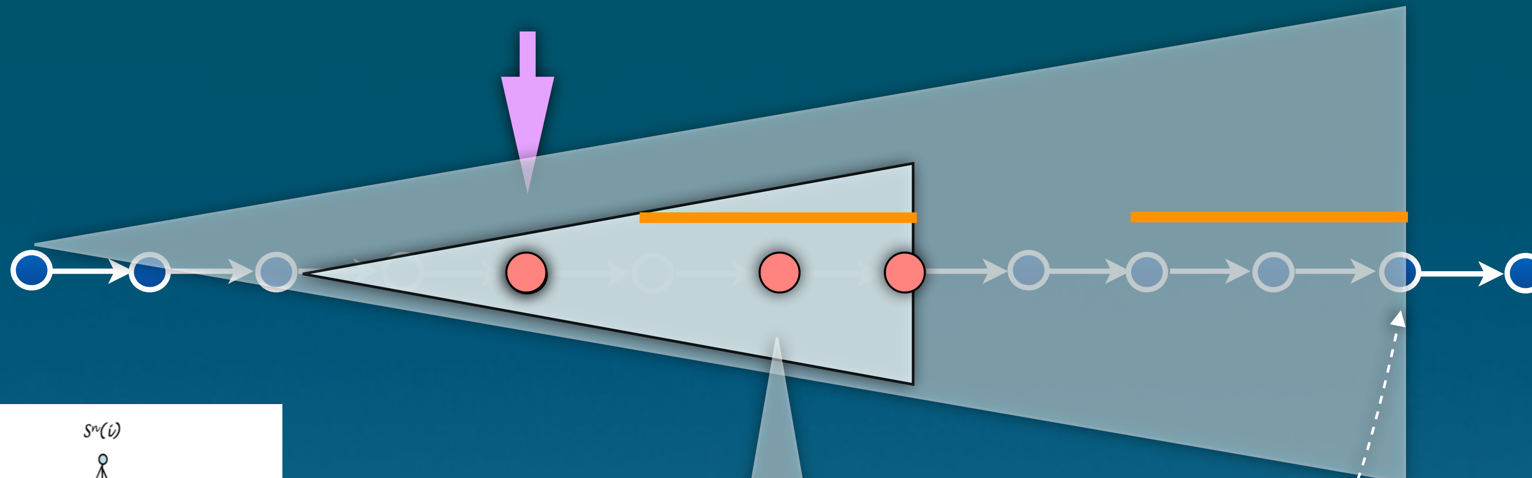


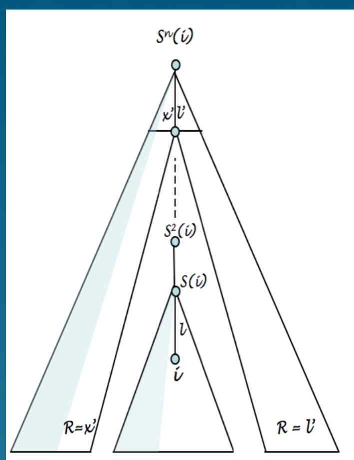


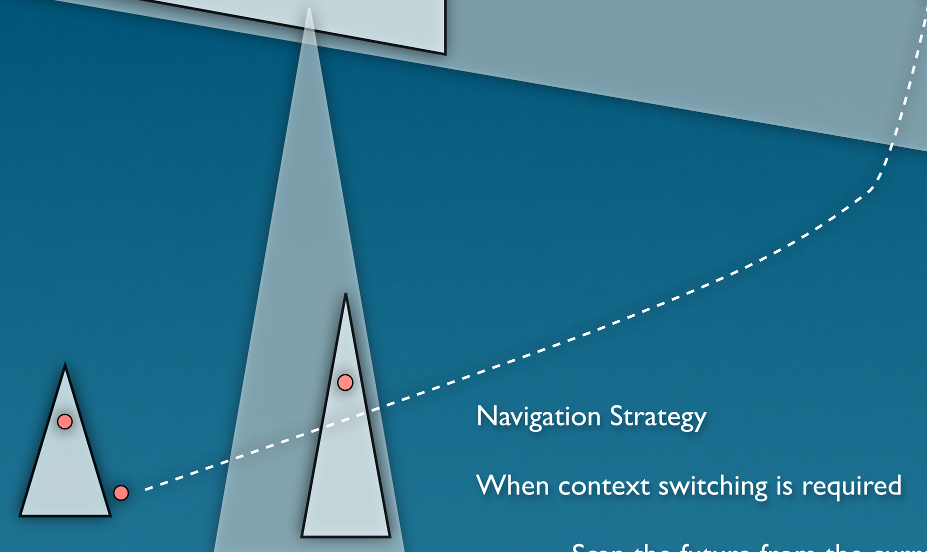
Oracle analysis of improvisation provide instant representation of musical structure i.e. how melodic, timbral, harmonic, rythmic recurrence and variation is organised



Generate : Navigation Heuristics in Suffix Link Trees







Choosen Solution

Navigation Strategy

When context switching is required

- Scan the future from the current point
- For each point explore tree, collect candidate solutions
- Sort by context length
- Apply probability distribution favoring context length and rhythmic quality
- Jump to selected candidate
- Put candidate into Taboo list for a while

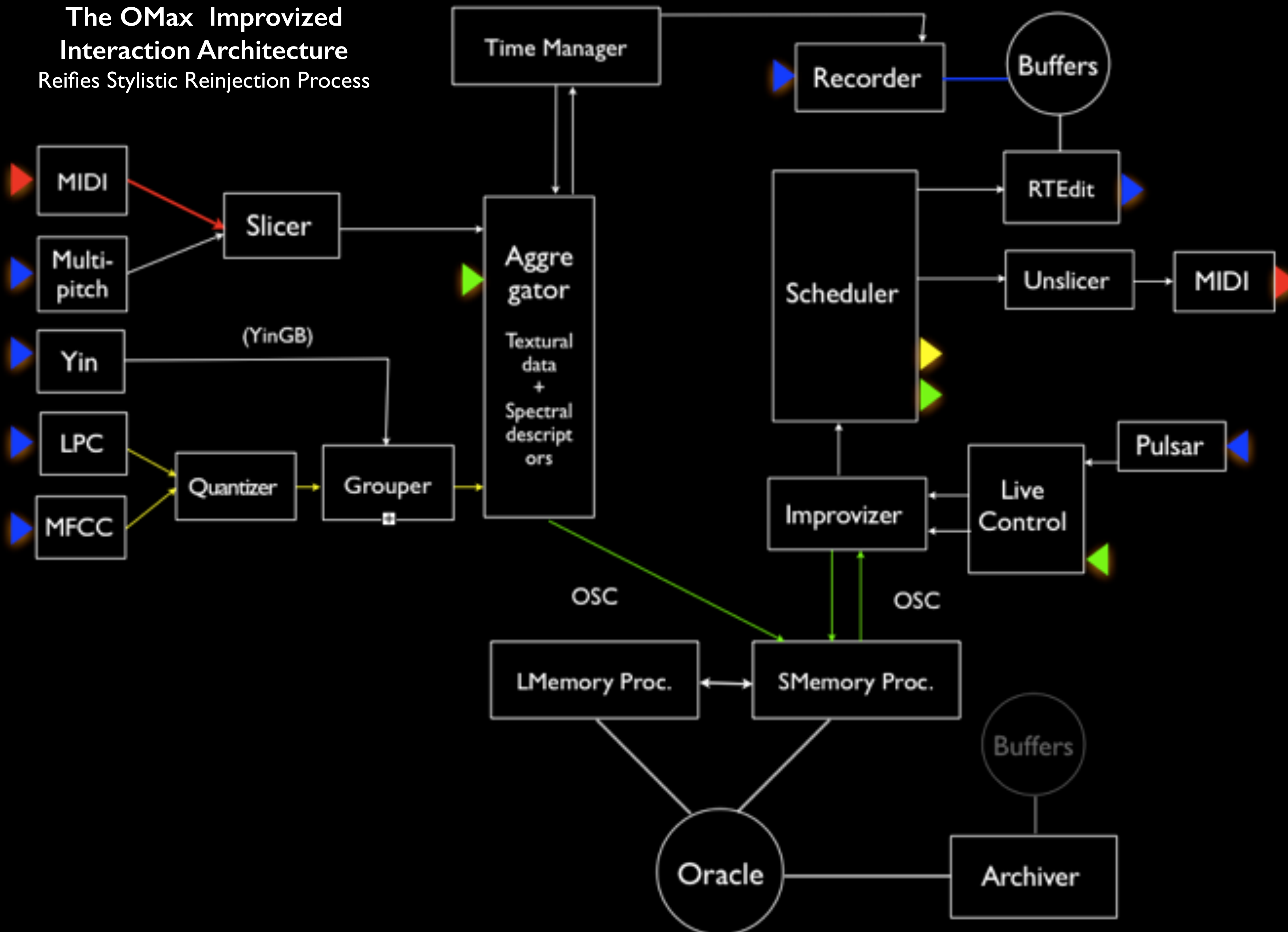
Captation
Listen

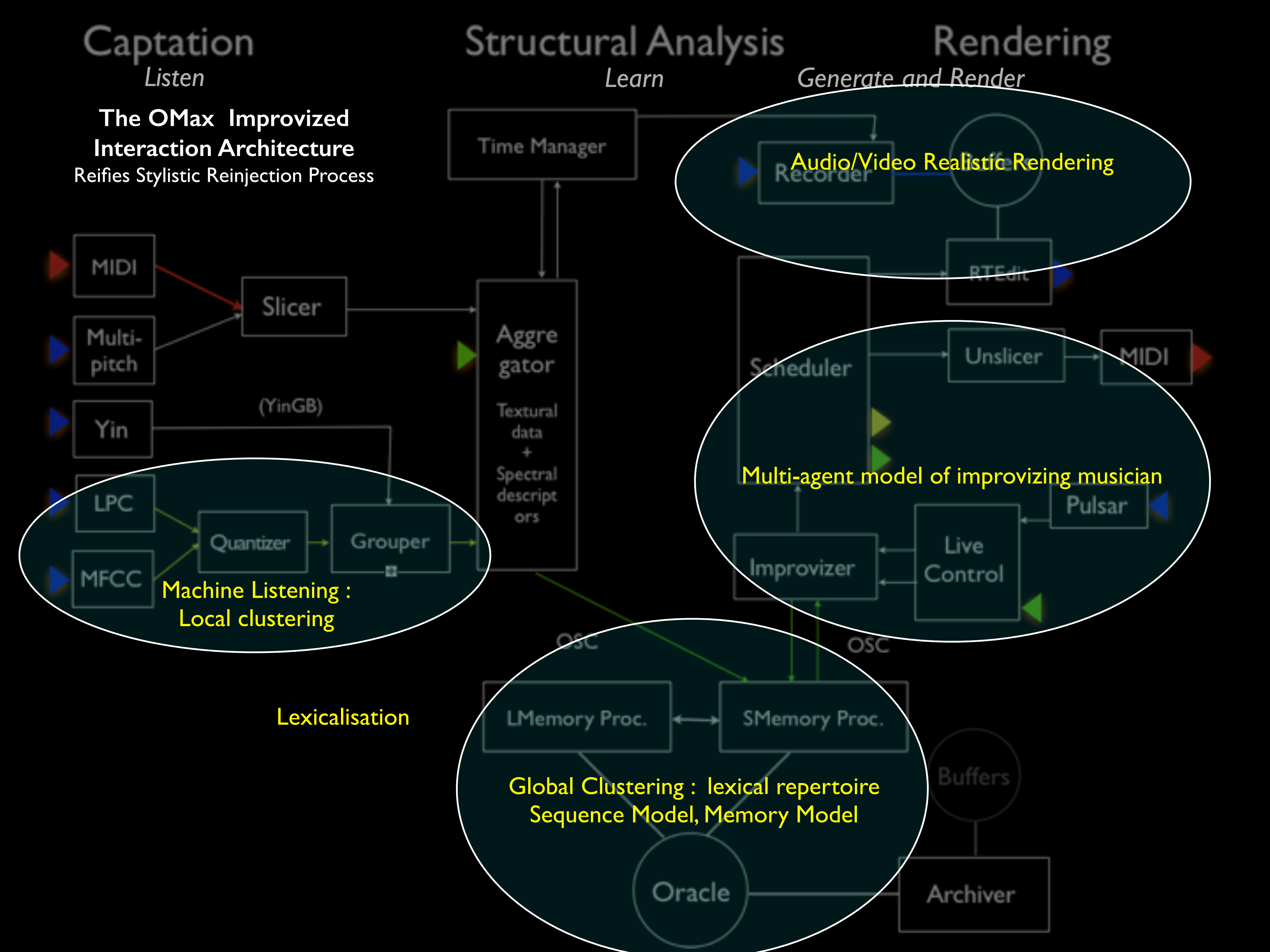
Structural Analysis
Learn

Rendering
Generate and Render

The OMax Improvized Interaction Architecture

Reifies Stylistic Reinjection Process




Captation
Listen
Structural Analysis
Learn
Rendering
Generate and Render
The OMax Improvized Interaction Architecture
Reifies Stylistic Reinjection Process
Time Manager
Audio/Video Realistic Rendering
Recorder
RTEdit
MIDI
Multi-pitch
Slicer
Aggregator
Textural data + Spectral descriptors
Yin
(YinGB)
LPC
Quantizer
Grouper
MFCC
Machine Listening : Local clustering
Scheduler
Unslicer
MIDI
Multi-agent model of improvizing musician
Pulsar
Improvizer
Live Control
OSC
OSC
Lexicalisation
LMemory Proc.
SMemory Proc.
Global Clustering : lexical repertoire
Sequence Model, Memory Model
Buffers
Oracle
Archiver

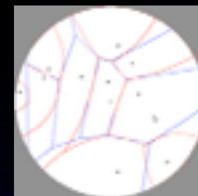
Fabrizio Cassol (Aka Moon) + OMax

-
-
-
-

Listening 2 : Musical Information Geometry

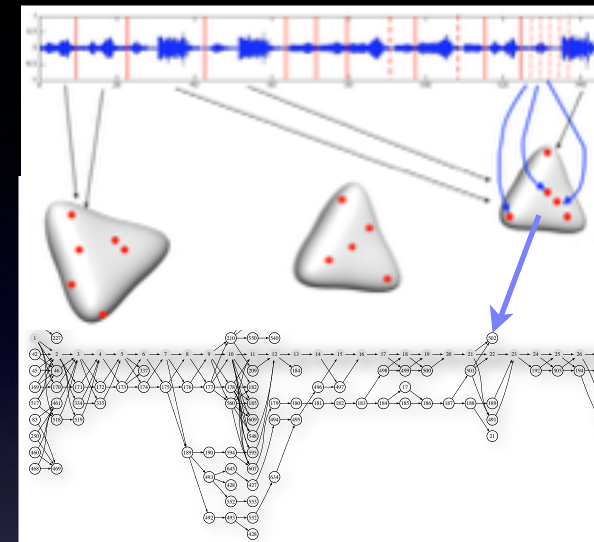
A framework for computational Info.Geo. applied to audio discovery

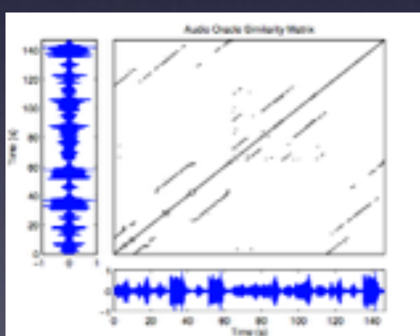
- From **Euclidean** Spaces to Riemann **Differential Manifolds**
- Statistical Inference
  - Differential Geometry
  - Information Theory
  - Signal Processing
- Novel Applications
  - Online Blind Change Detection / Segmentation
  - Online Audio Structure Discovery / Similarity Analysis

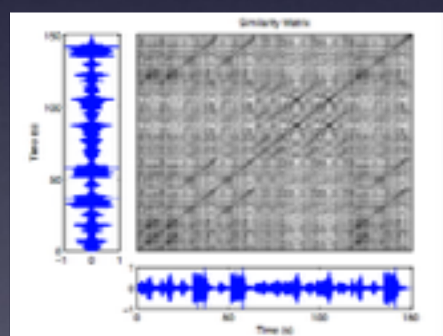


points
euclidean distances as in Voronoi Diagrams

Probability densities living on a differential manifold (Riemann surfaces)
Info-theoretical divergences (KL)
Geodesics







S. Amari and H. Nagaoka. *Methods of Information Geometry*. Oxford University Press, 2000

- **Machine Listening in an Information geometry** Framework : Distributing the Sound Flow over a Bregman Geometry
- **Riemannian Manifolds** with information metrics over parametric exponential probability spaces
- **Points** are parametric probability distributions (multinomial in our case, model perceptual descriptors associated to **incoming frequency domain signal frames**)
- **Distances** are **Bregman Divergences** : amount to relative entropy between perceptual descriptors
- Points are dynamically clustered into **Bregman Balls** (a generalization of Voronoi polyhedron)
- **Bregman Balls** become **symbols** in a formal language alphabet
- Enclosing Balls model stable and distinct musical units such as notes, chords etc.

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.

Listening : Musical Information Geometry

A framework for computational Info.Geo. applied to audio discovery

Validation on Beethoven's First Piano Sonata, third movement played by Gulda in 1958

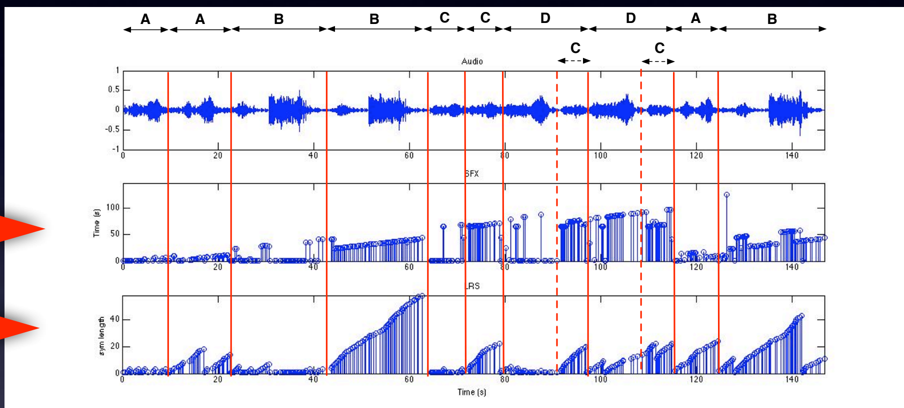
- *Using Constant-Q amplitude spectrum on Multinomial music information geometry*
- *150 seconds, > 9500 analysis frames, resulting to 440 states*

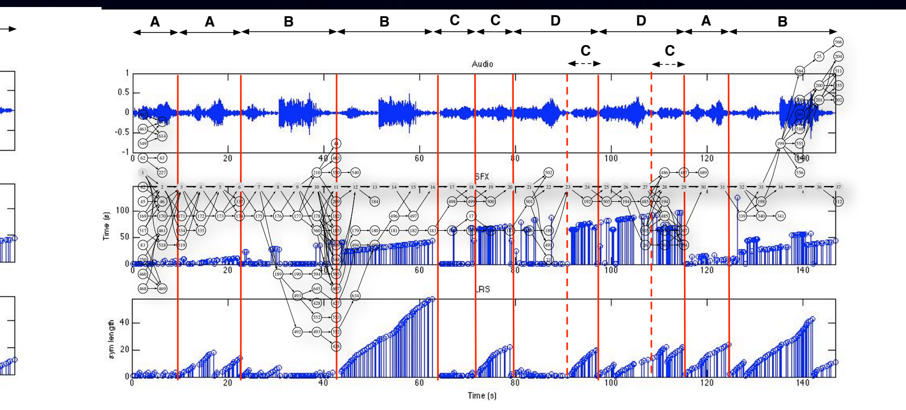
Sequence model imposed on the Information Geometry : Factor Oracle

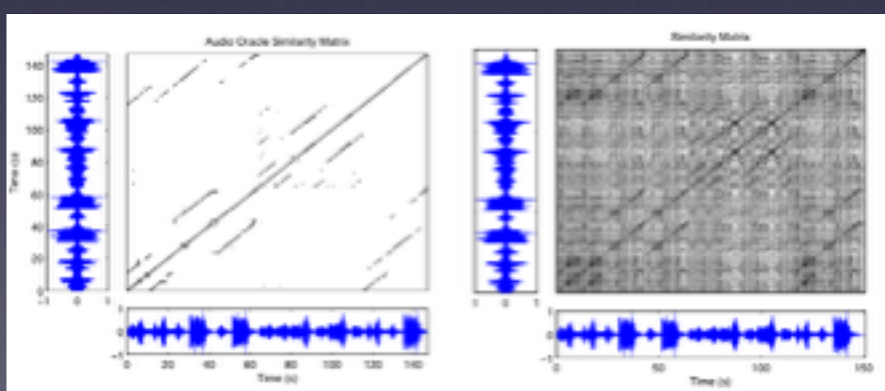
- *A Bregman Ball becomes a symbol in the musical alphabet associated to some state in the automaton*
- *Related states express all the motivic (stylistic structure) relations in the music*

Recall Structure

Recall Length

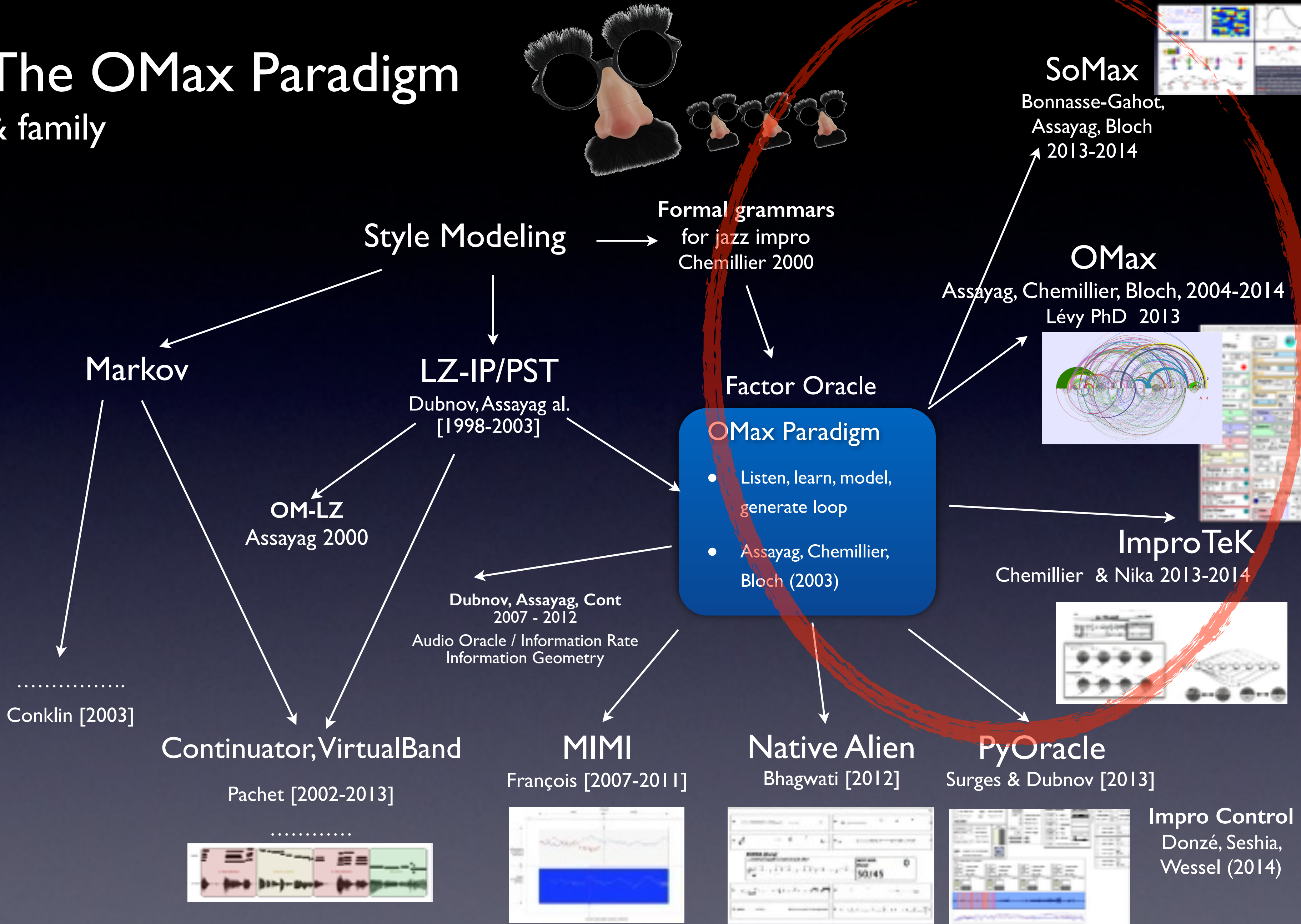






Equivalence to sparse signal correlation matrix

The OMax Paradigm
& family



OMax versatility

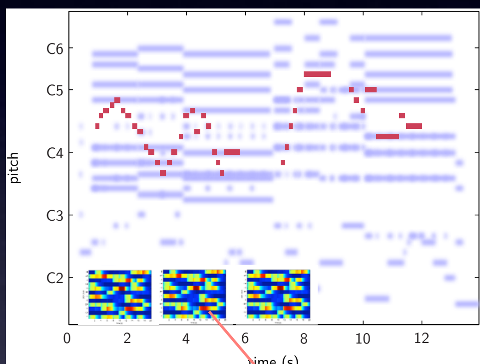
Rhythm and pulse derive from off-line learning of the multi-track (beat aligned) studio sessions

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

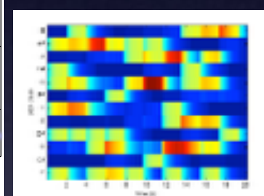
Going SoMax : listen carefully

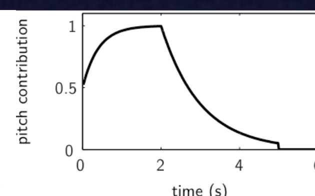
(ANR SOR2, Post-doc L. Bonnasse-Gahot)

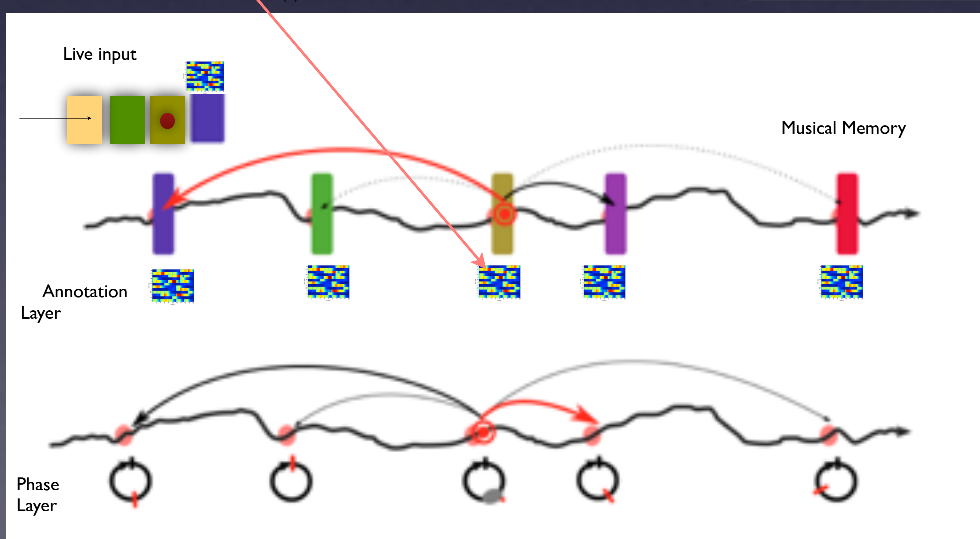
Listen to several channels (foreground, background, voices etc.), One for the main memory model, the others for automatic annotation
Create e.g. a solo memory model plus loose harmonic / textural annotation, or the other way round, or both.
At generation time, match annotation with features extracted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

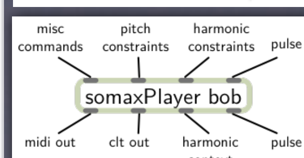


Extract the chroma Harmony of any musical stream using echoic memory and leaky integration (Toiviainen & Krumhans, Durational accent, Parncutt)









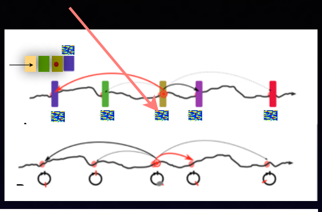
2 artificial agents playing from a musical memory they have learned may act as both live input and musical memory one for another

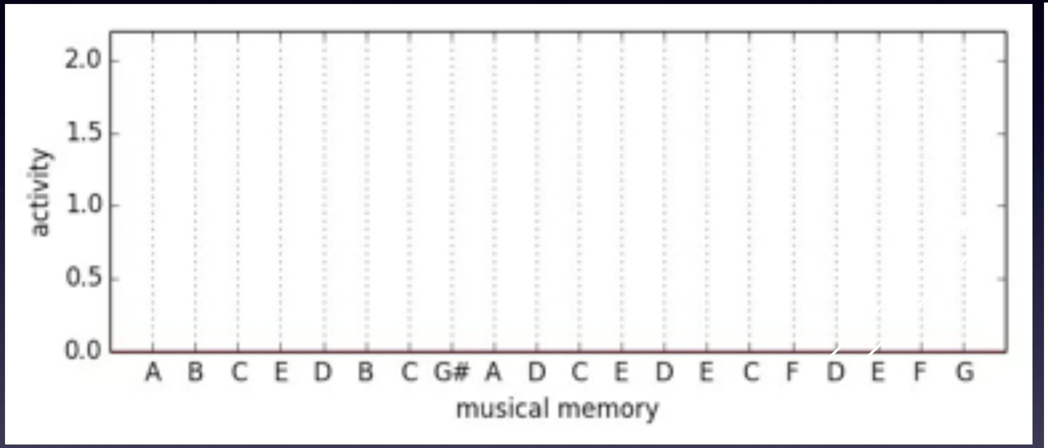
exemple : one agent learns the right hand of a piano piece, with, as annotatin, the chroma harmony of the left hand.
Another agent can play a stream whose chroma harmony is extracted in real time. The 2 agents interact on behalf of the common chroma harmony, i.e. agent 1 will jump to follow the harmony induced by agent 2.

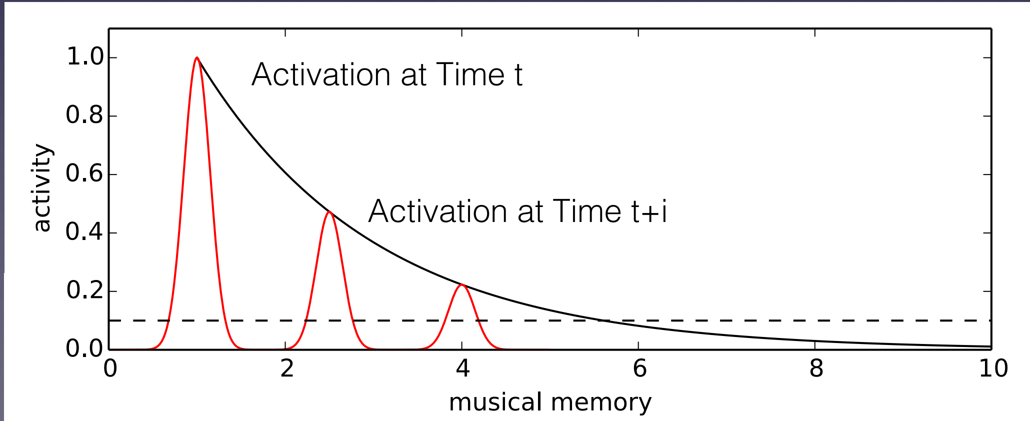
SoMax

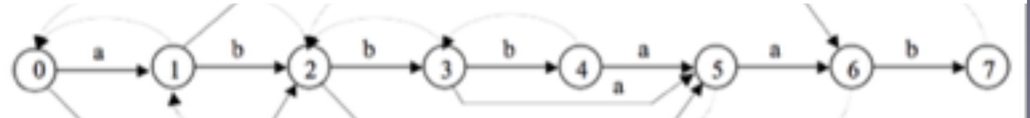
Memory Activation Scheme

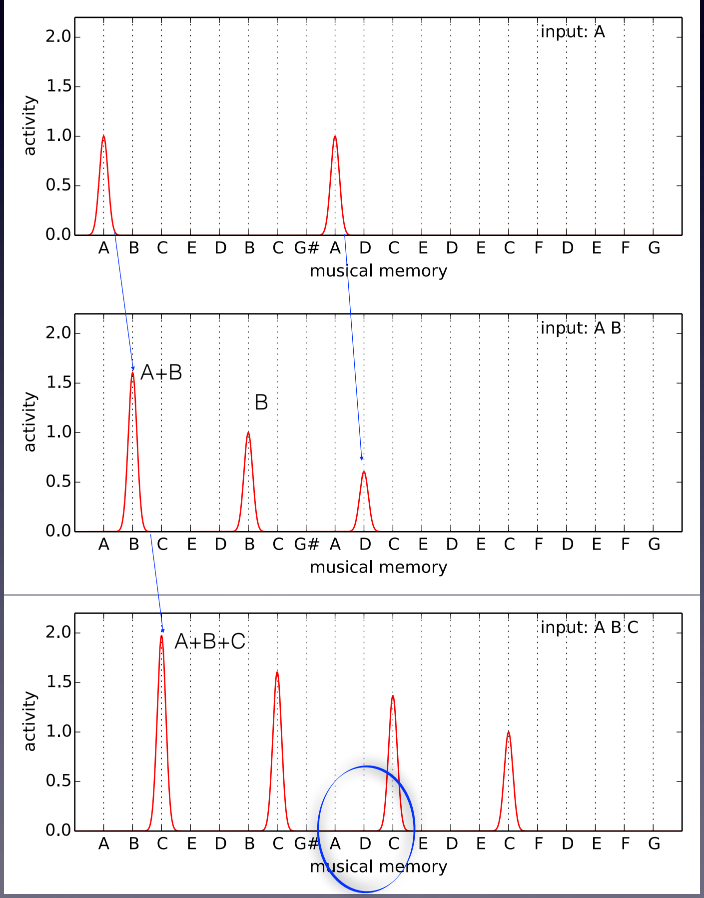
Addressing the **cartographical blindness**, the **evidence accumulation**, and the **cognitive persistence** questions







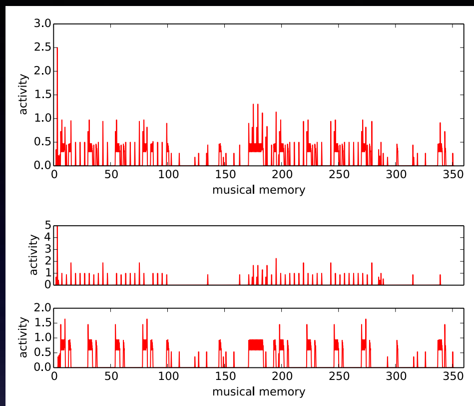


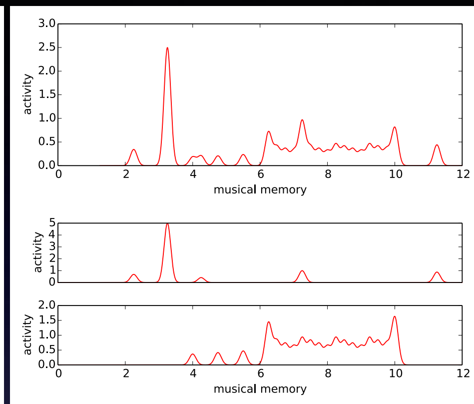


Fuzzy pattern
escape from the purely markovian sequence logic

SoMax

Summing up activation profiles of parallel annotation views including self listening





Beat/Phase Profile

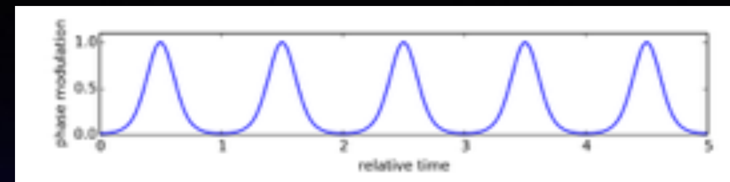
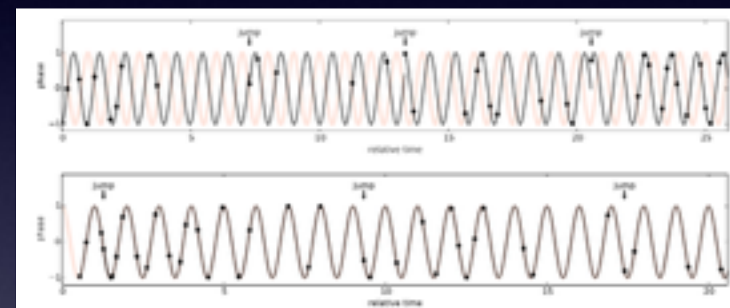
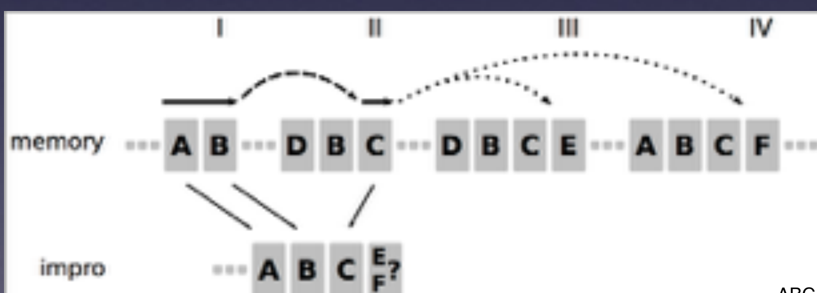
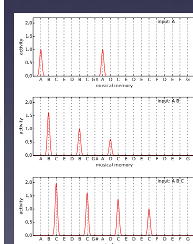


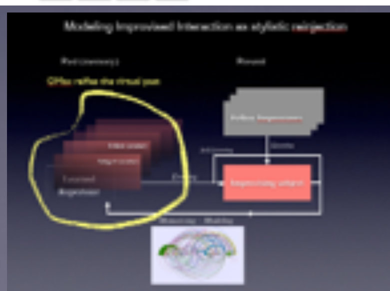
Figure 14: Phase modulation function, $\exp(\eta[\cos(2\pi(\xi - \xi_{target})) - 1])$. In this illustration, $\xi_{target} = 0.5$ and $\eta = 2.0$.

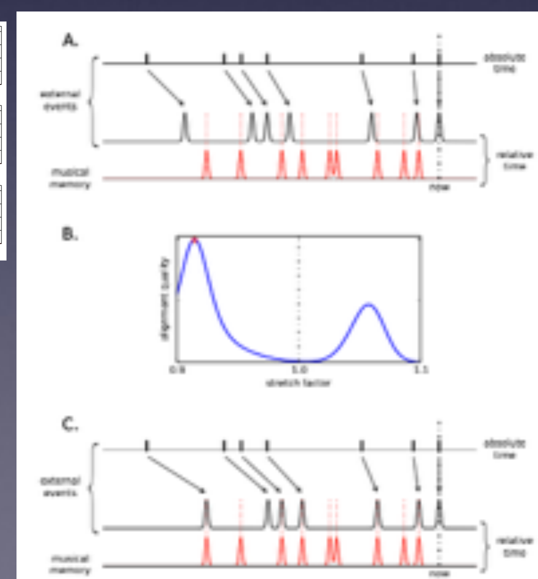


Self Listening : listening to the memory, or to the generation



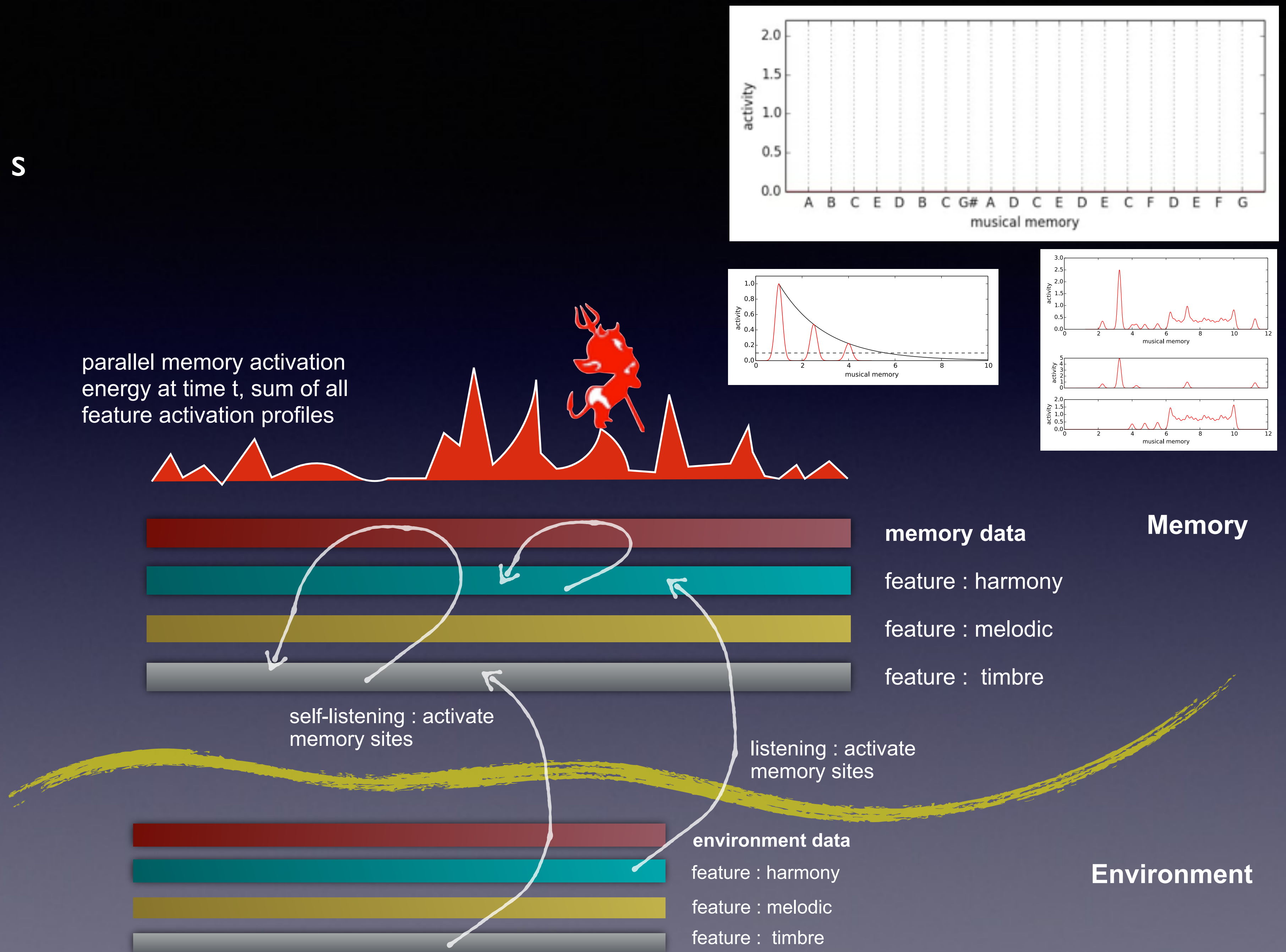






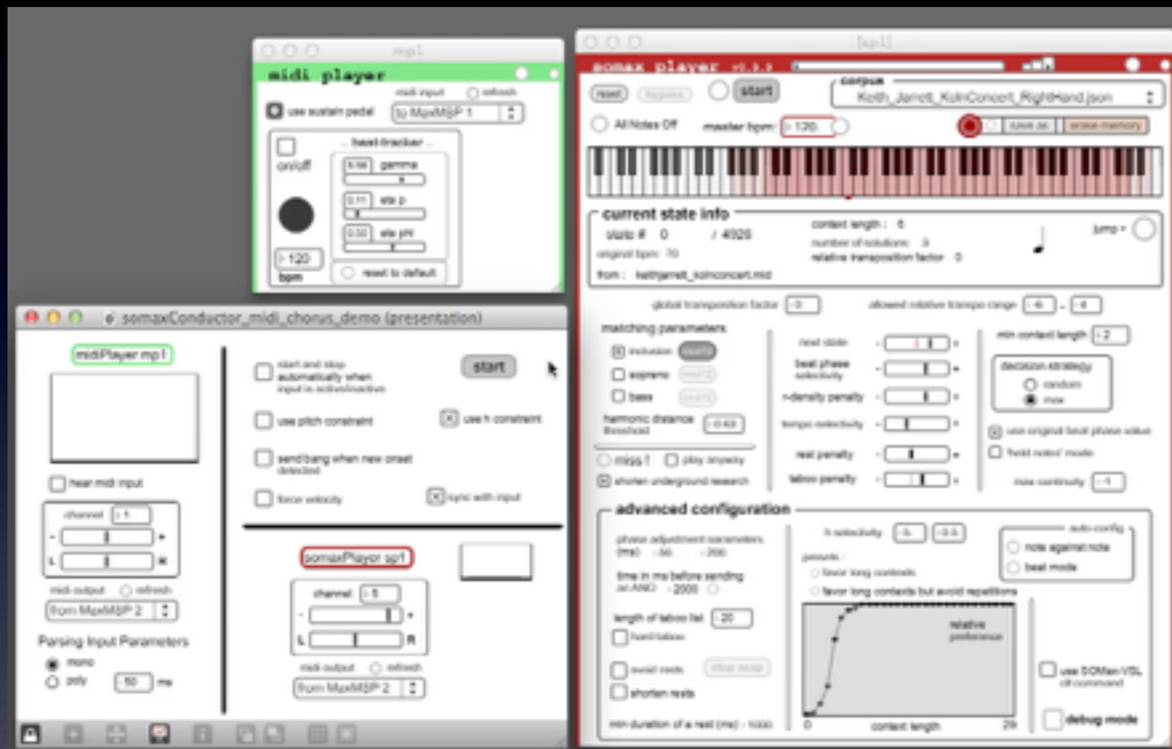
Flexible Time

S

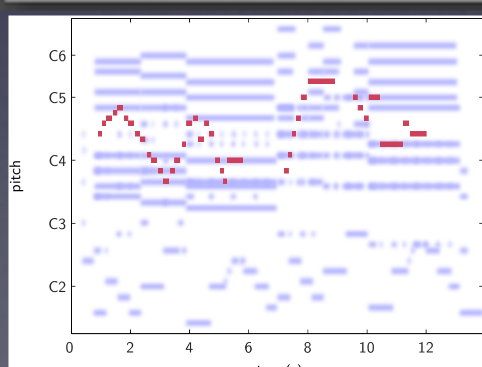


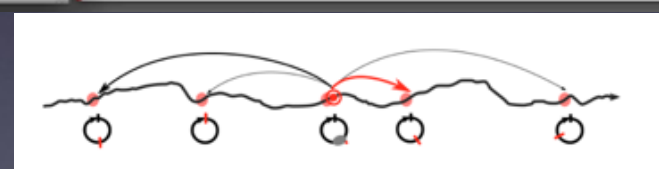
2.0
1.5
1.0
0.5
0.0
activity
A B C E D B C G# A D C E D E C F D E F G
musical memory
activity
musical memory
activity
musical memory
parallel memory activation energy at time t, sum of all feature activation profiles
memory data
Memory
feature : harmony
feature : melodic
feature : timbre
self-listening : activate memory sites
listening : activate memory sites
environment data
feature : harmony
feature : melodic
feature : timbre
Environment

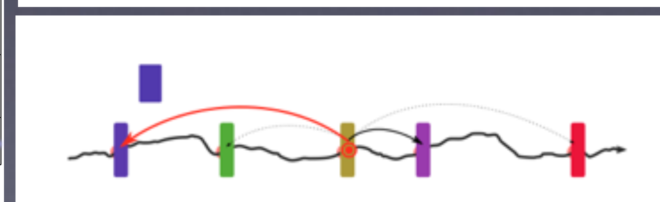
Going SoMax : Autumn in Köln



s



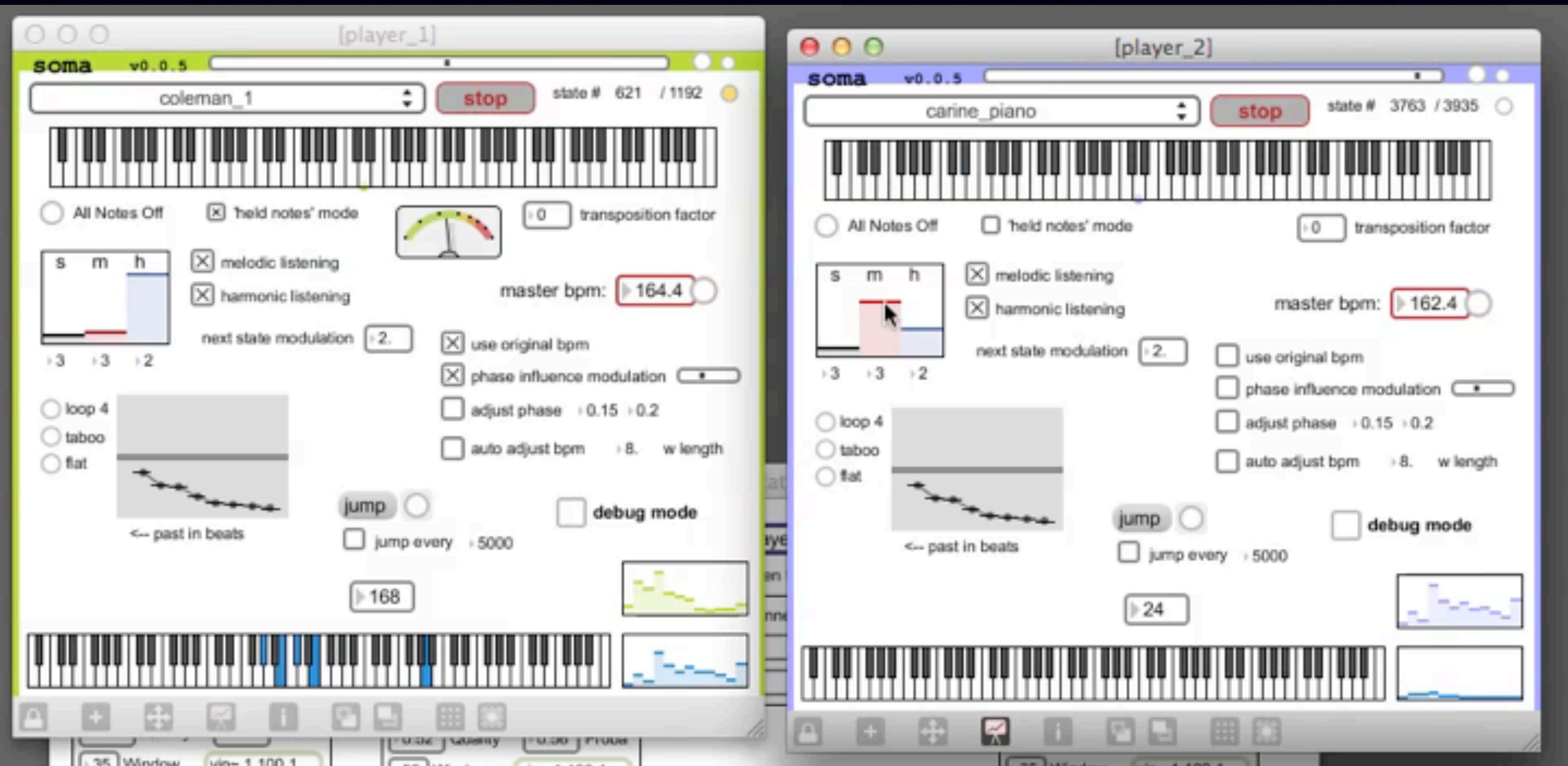




2 corpus-based Somax agents co-improvise freely on behalf of chroma harmony.

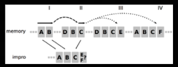
1. An audio recording of Steve Coleman
2. A Midi recording of Carine Bonnefoy Piano standards

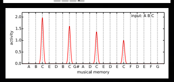
(1) mostly harmonic listening, (2) mostly melodic listening

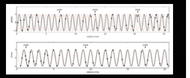


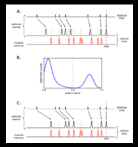
In the mood of Time Remembered (Rémi Fox + Bill Evans)

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



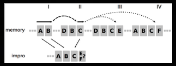


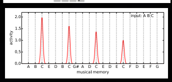


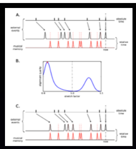


Schoenberg revisited

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.







Remi’s Clones

Rémi’s 1st impro is learned, then 2 clones are launched (01:39) which listen to Remi melodically/ harmonically. Rémi can drive the 2 agents to unison or heterophony by evoking initial material. Clones also alternate strong self listening and strong sensitivity to environment.