

# DYCI2

## Revue ANR à mi-parcours AAPG 2014

Mardi 10 janvier 2017

Gérard Assayag (Ircam, coord, WP3/4)

Emmanuel Vincent (Inria, coord WP2)

Sylvain Marchand (ULR, coord WP1)

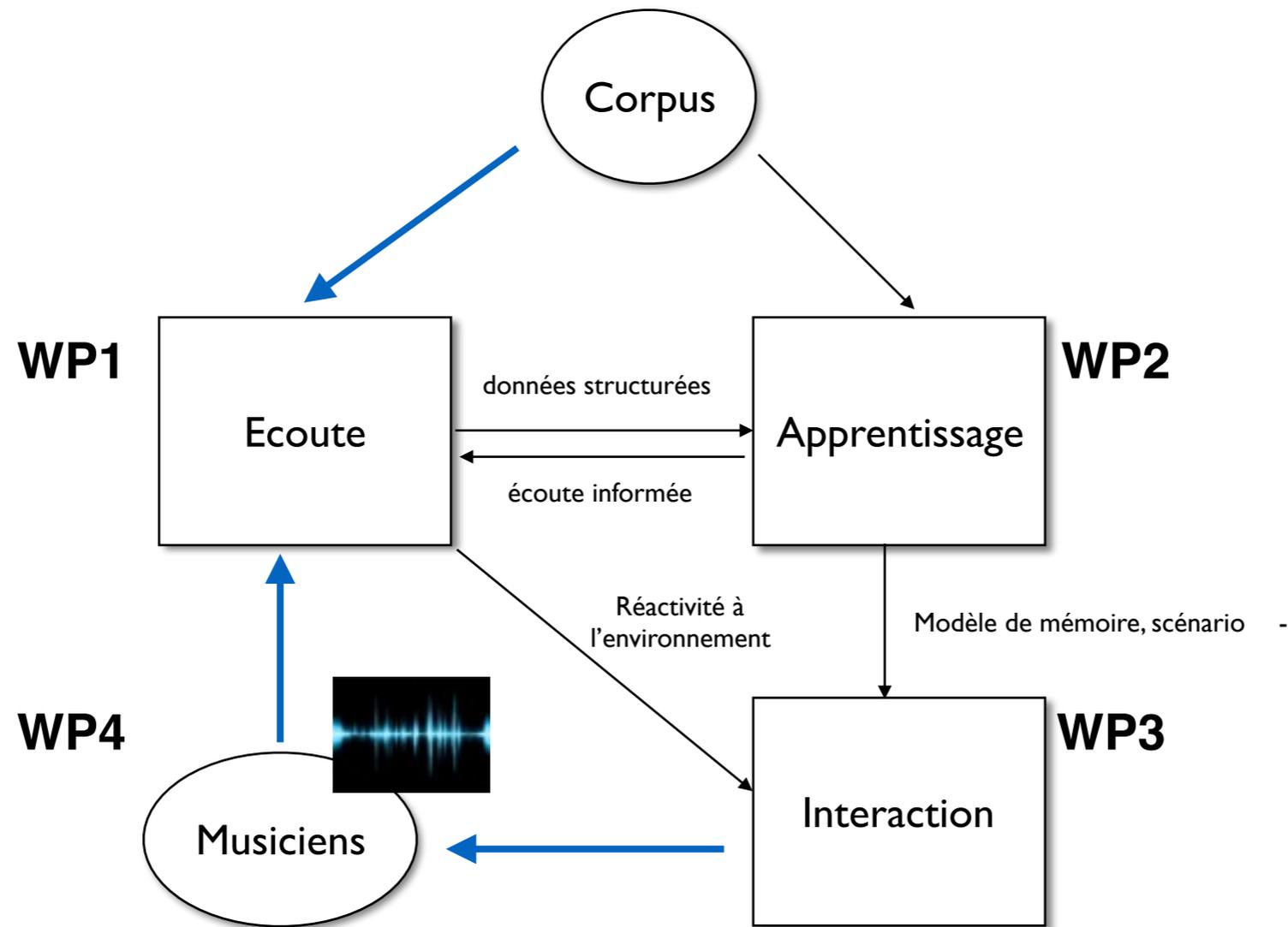
Antoine Liutkus (Inria, WP1)

Ken Deguernel, Nathan Libermann (doctorants WP2)

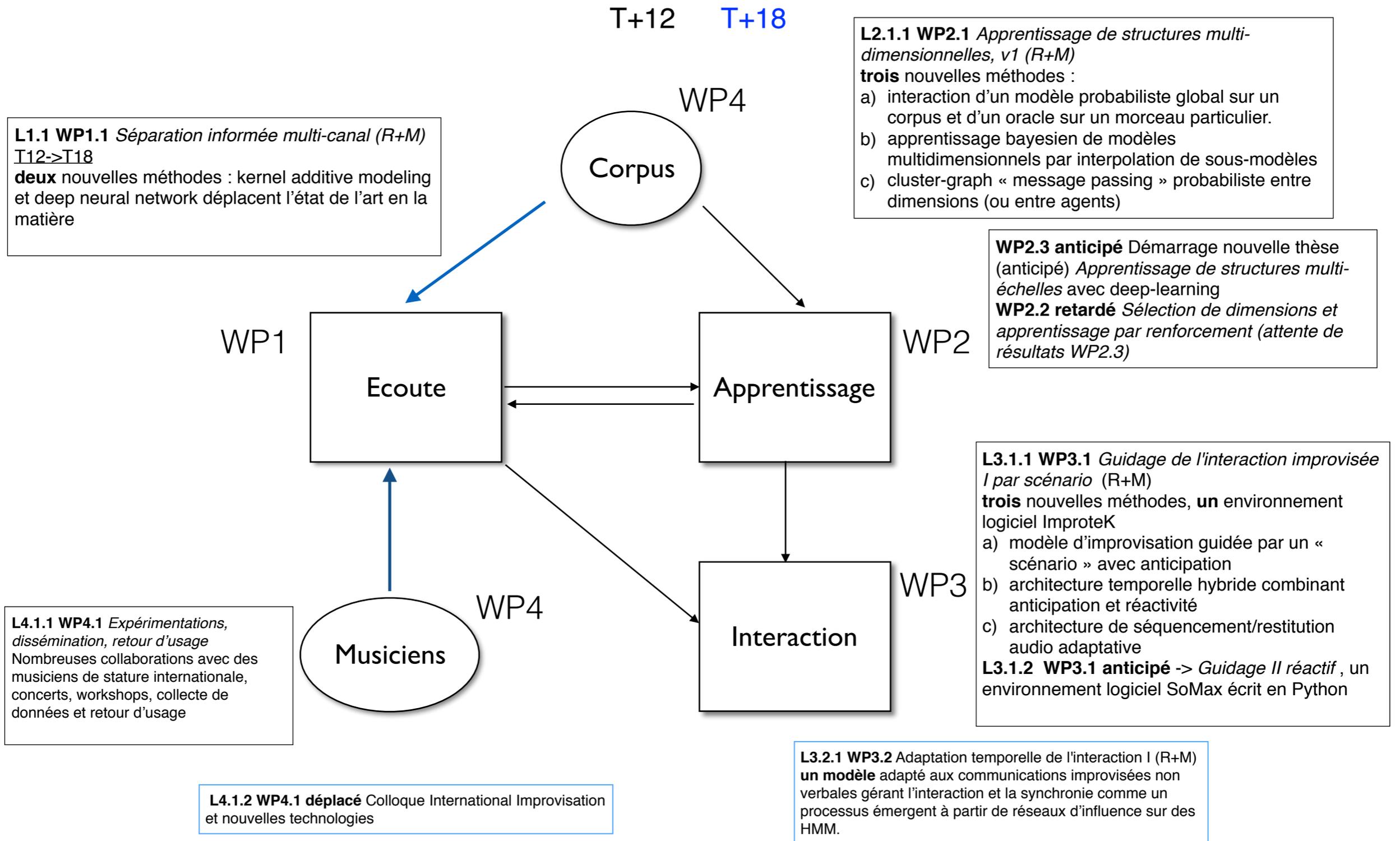
Jérôme Nika (postdoc WP3), Kevin Sanlaville (doctorant WP3), Axel Chemla (ing. WP3)

				Chronogramme														
				Mars15			Mars16			Mars17								
Partenaires				Année 1				Année 2				Année 3						
				Ir	In	U	3	6	9	12	15	18	21	24	27	30	33	36
<b>WP0</b>	<b>R</b>	<b>P</b>	<b>P</b>	Gestion du projet														
<b>WP1</b>	<b>P</b>	<b>P</b>	<b>R</b>	Ecoute informée créative														
1.1		✓	✓	Séparation informée multi-canal														
1.2	✓	✓	✓					Ecoute structurante										
1.3	✓	✓	✓									Décomposition / recomposition par imitation						
<b>WP2</b>	<b>P</b>	<b>R</b>		Apprentissage interactif de structures musicales														
2.1	✓	✓		Apprentissage de structures multi-dimensionnelles														
2.2	✓	✓						Sélection de dimensions et apprentissage par renforcement										
2.3	✓	✓										Apprentissage de structures multi-échelles						
<b>WP3</b>	<b>R</b>	<b>P</b>	<b>P</b>	Dynamiques d'interaction improvisée														
3.1	✓	✓		Guidage de l'interaction improvisée														
3.2	✓		✓					Adaptation temporelle de l'interaction										
3.3	✓	✓	✓					Mémoire, connaissance, contrôle des agents créatifs										
<b>WP4</b>	<b>R</b>	<b>P</b>	<b>P</b>	Intégration, expérimentation, validation et retour d'usage														
4.1	✓	✓	✓	Expérimentation, évaluation, dissémination														
4.2	✓	✓	✓												Intégration suite logicielle			

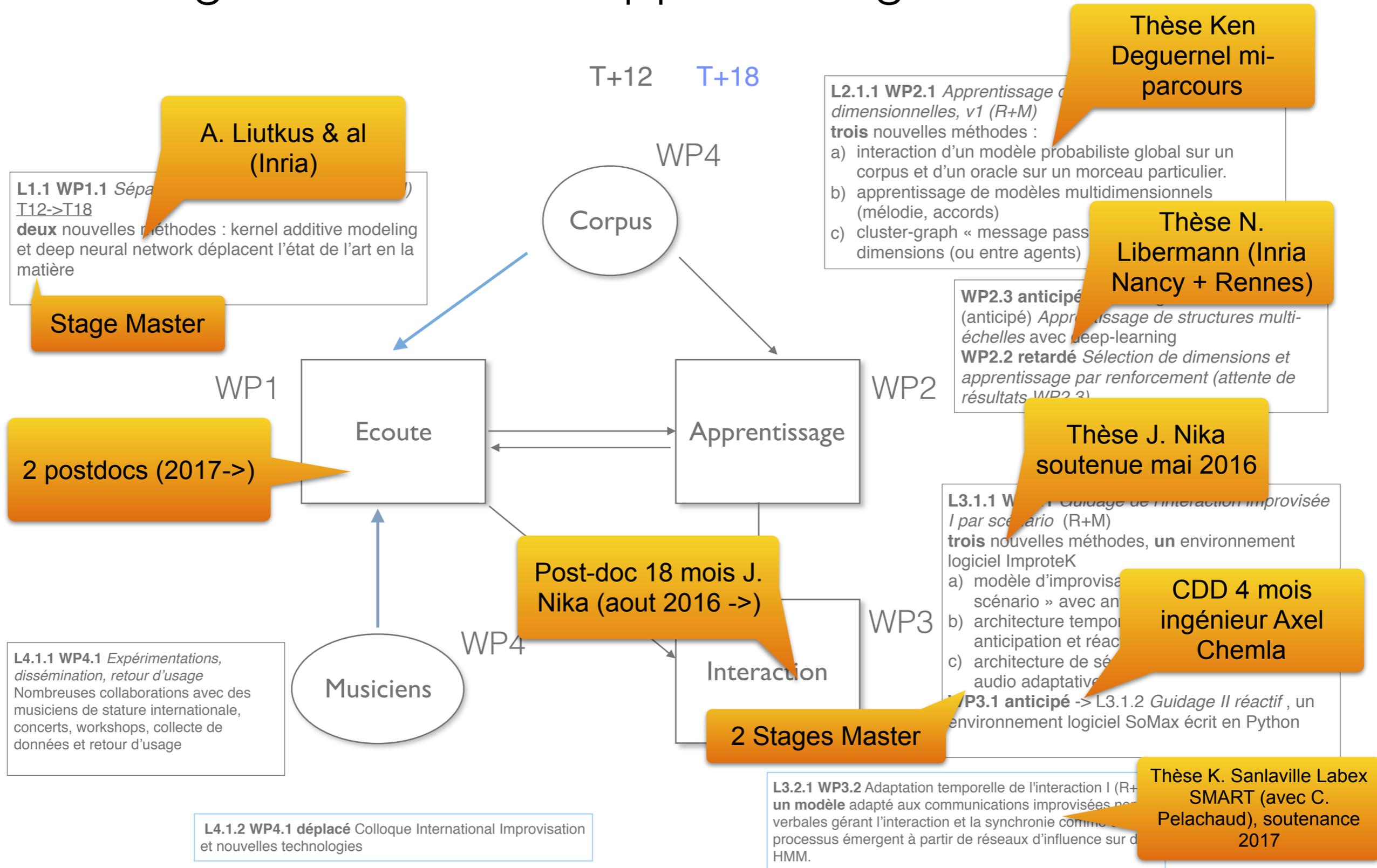
# Diagramme Ecoute Apprentissage Interaction



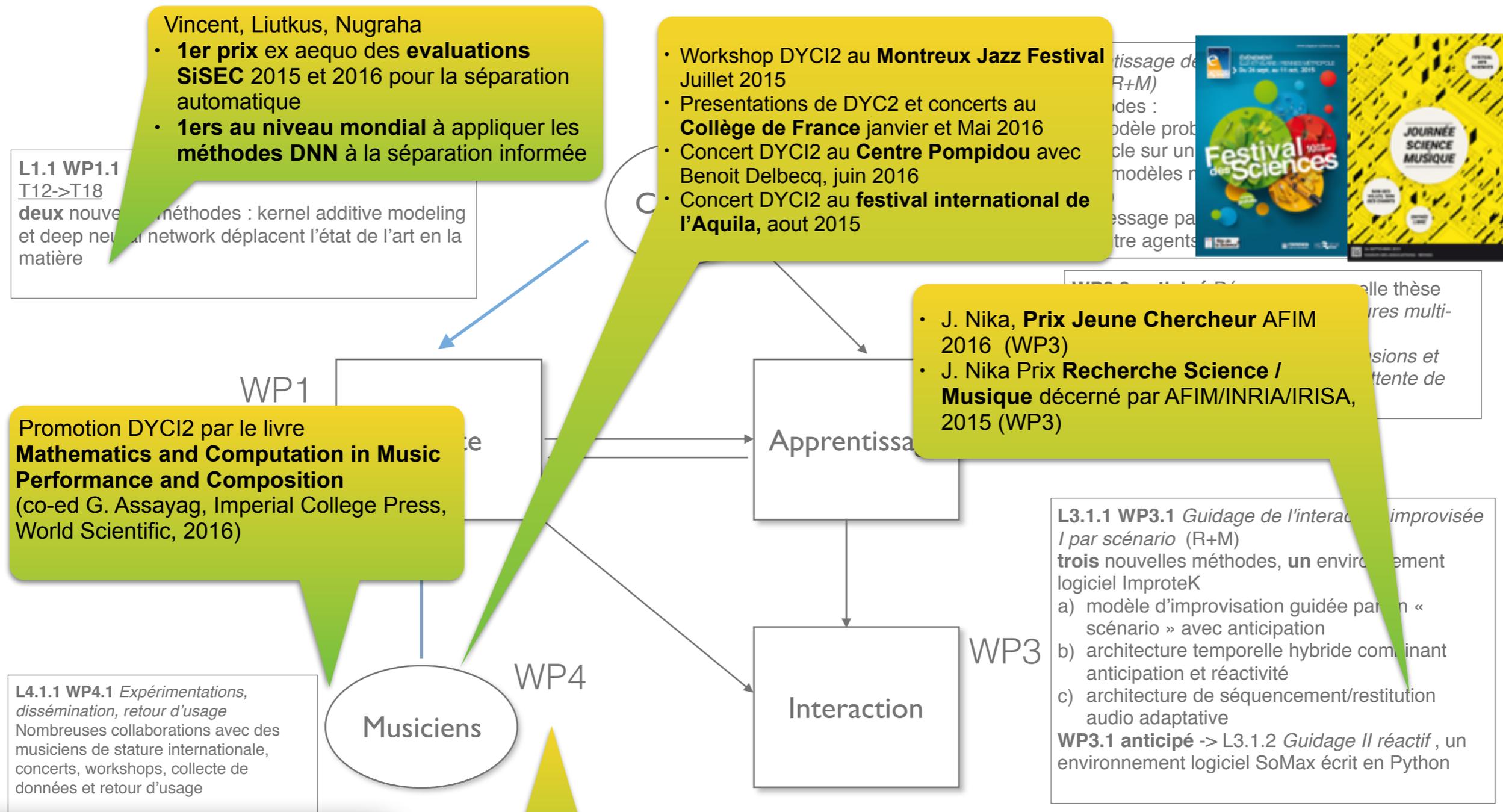
# Diagramme Ecoute Apprentissage Interaction



# Diagramme Ecoute Apprentissage Interaction



# Diagramme Ecoute Apprentissage Interaction

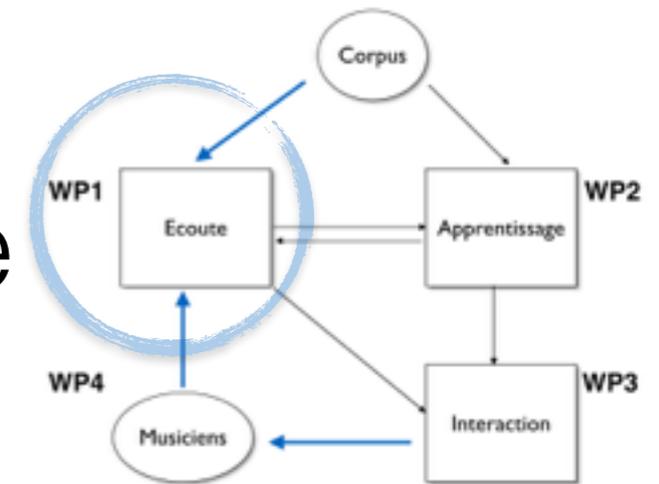


				Chronogramme														
				Mars15			Mars16			Mars17								
Partenaires				Année 1			Année 2			Année 3								
				Ir	In	U	3	6	9	12	15	18	21	24	27	30	33	36
<b>WP0</b>	<b>R</b>	<b>P</b>	<b>P</b>	Gestion du projet														
<b>WP1</b>	<b>P</b>	<b>P</b>	<b>R</b>	Ecoute informée créative														
1.1		✓	✓	Séparation informée multi-canal			→											
1.2	✓	✓	✓				→			→			→					
1.3	✓	✓	✓							→			→					
<b>WP2</b>	<b>P</b>	<b>R</b>		Apprentissage interactif de structures musicales														
2.1	✓	✓		Apprentissage de structures multi-dimensionnelles														
2.2	✓	✓					→			→			→					
2.3	✓	✓								←			Apprentissage de structures multi-échelles					
<b>WP3</b>	<b>R</b>	<b>P</b>	<b>P</b>	Dynamiques d'interaction improvisée														
3.1	✓	✓		Guidage de l'interaction improvisée			←			←			←					
3.2	✓		✓				→			→			→					
3.3	✓	✓	✓				→			→			→					
<b>WP4</b>	<b>R</b>	<b>P</b>	<b>P</b>	Intégration, expérimentation, validation et retour d'usage														
4.1	✓	✓	✓				→			→			→					
4.2	✓	✓	✓										Intégration suite logicielle					

# WP1

## Ecoute Informée Créative

Coord ULR, Sylvain Marchand  
Inria, E.Vincent, A. Liutkus



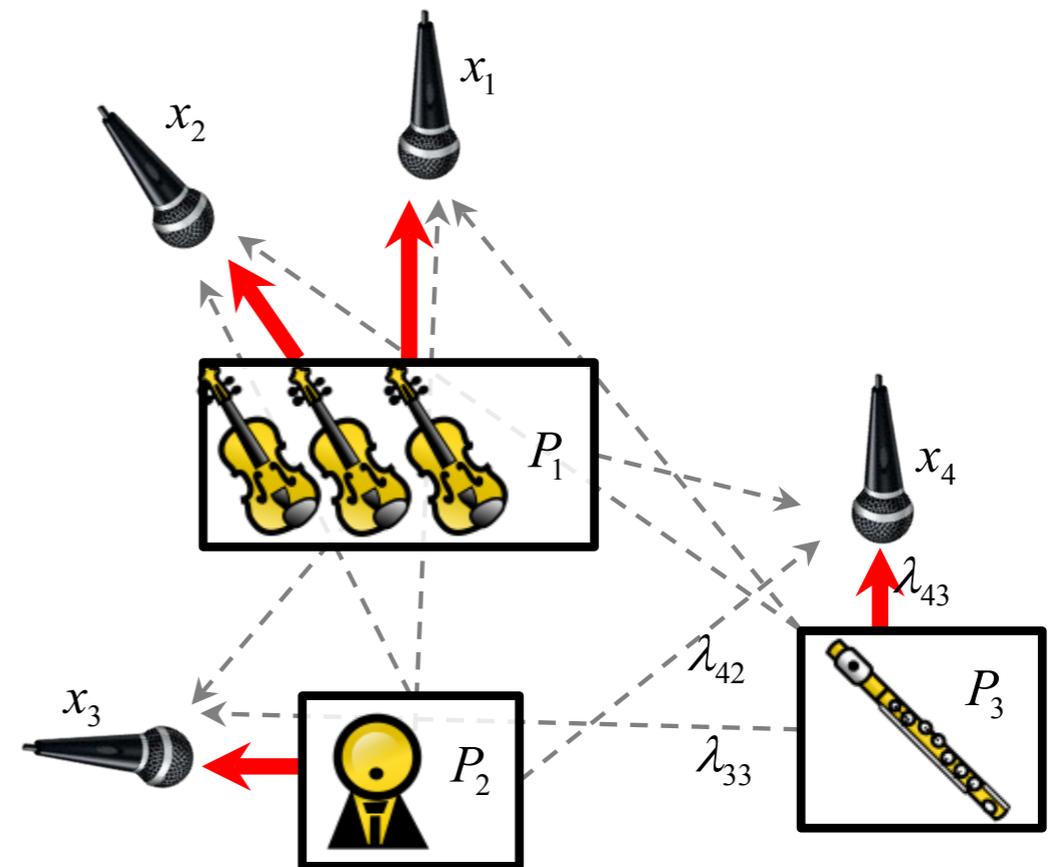
- Réorganisation : Sylvain Marchand (mutation de l'Université de Bretagne Occidentale — Brest vers l'Université de La Rochelle, transfert du projet effectif janvier 2017)
- ré-organisation à coût constant : 36 mois de doctorant transformés en 18+12 mois de post-doctorants (démarrage à T0+18), sous-tâches **WP1.2** et **WP1.3**
- travail jusqu'à T0+18 principalement effectué par le partenaire Inria (sous-tâche **WP1.1**)

# Écoute informée : écoute automatique + séparation de sources informée

- écoute automatique : (*Computational Auditory Scene Analysis*) difficulté de la polyphonie (plusieurs sources simultanées)
- séparation de sources informée : dans un problème inverse (ici séparer les sources), utiliser une **information a priori** pour obtenir une solution de meilleure qualité
- **Dimension verticale (WP1.1)** *Séparation informée multi-canal* : utiliser les différents canaux d'une prise de son multi-canal (problème : élimination des interférences)
- **Dimension horizontale (WP1.2)** *Écoute structurante* : prise en compte du temps : information du passé pour résoudre le problème à l'instant présent : structure musicale reconstruite en temps réel par une méthode inspirée de Shazam. Post-doc de 18 mois en cours de mise en place.
- **Application créative (WP1.3)** *Décomposition / recomposition par imitation* : boucle modèle, observation, et ré-injection d'information dans le modèle à partir de l'observation si elle se raccroche au modèle tout en apportant de la nouveauté : post-doc de 12 mois en cours de mise en place (en lien avec le post-doc précédent)

## L1.1 — Séparation de sources informée multicanal

En *live*, chaque micro enregistre un mélange de la source désirée et des autres sources (“repisse”).



But : extraire le signal de chaque source (piano, basse, voix...) ou ensemble de sources (percussions, accompagnement...).

- KAMIR : séparation informée par la position des micros  
→ conçu pour 16+ micros, temps de calcul < durée du signal
- dnnsep : séparation multicanal par *deep learning*  
→ conçu pour 1 à 8 micros, qualité de séparation maximale

# KAMIR (1)

Mouvements des sources  $\Rightarrow$  filtrage purement énergétique.

$$x_i(f, n) = \sum_{j=1}^J c_{ij}(f, n)$$

TFCT complexe du mélange au micro  $i$

image spatiale de la source  $j$

$$c_{ij}(f, n) \sim \mathcal{N}_c(0, v_j(f, n)\lambda_{ij}(f))$$

spectre de puissance commun à tous les micros

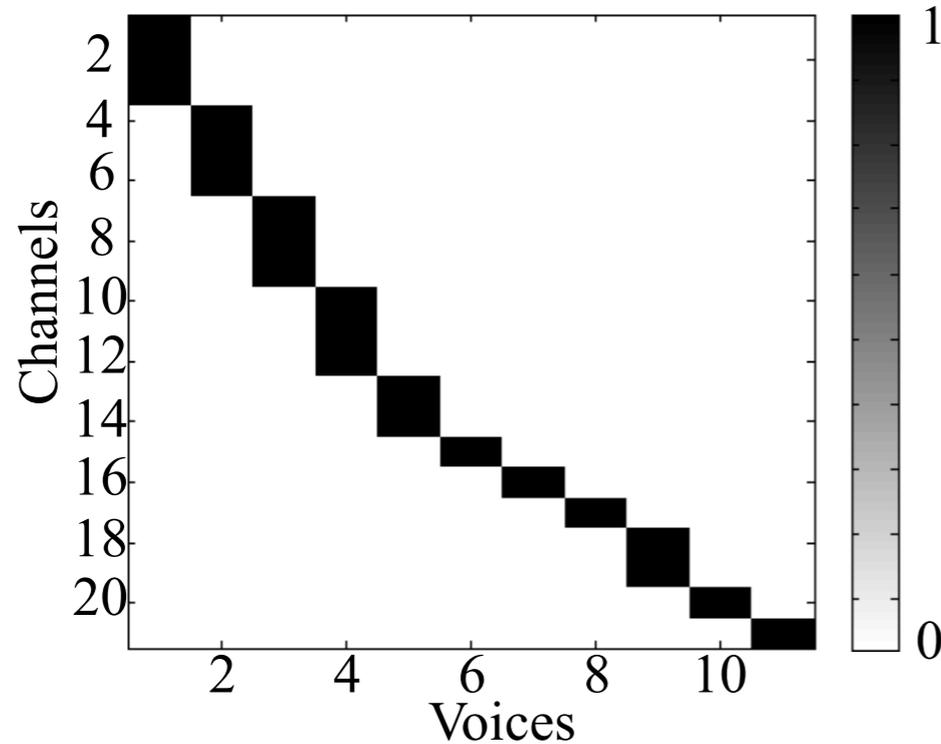
gain dépendant du micro (degré de "repisse")

Algorithme pseudo espérance-maximisation (EM) :

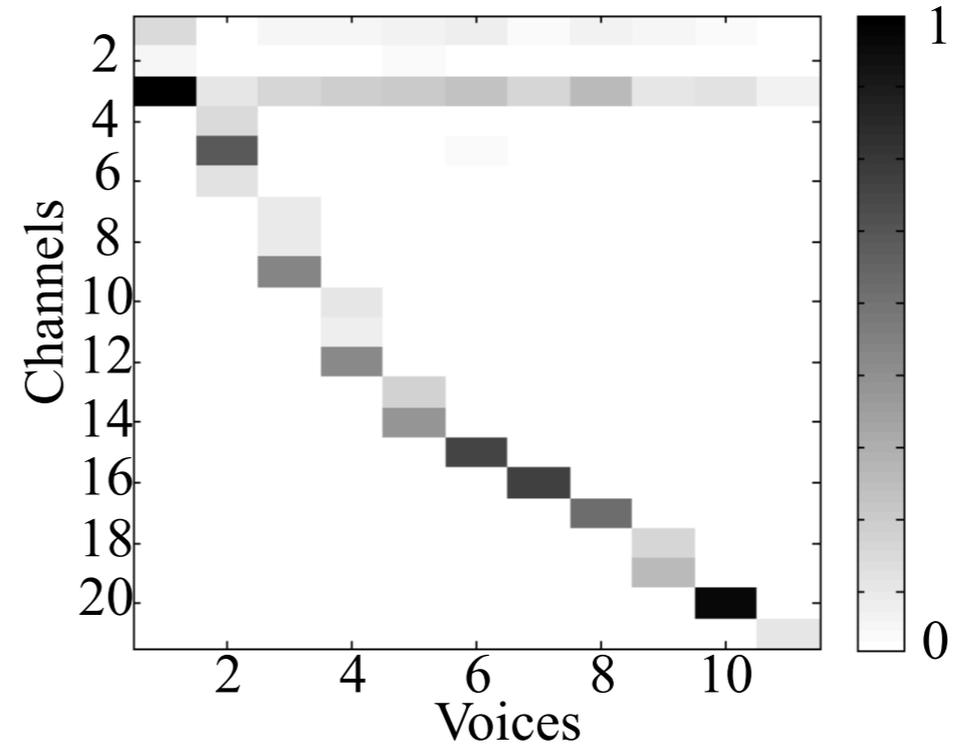
- initialisation :  $\lambda_{ij}(f) = 1$  si source prépondérante,  $\rho < 1$  sinon
- étape E : séparation de  $c_{ij}(f, n)$  par filtrage de Wiener
- étape M : mise à jour multiplicative de  $\lambda_{ij}(f)$  et  $v_j(f, n)$   
optionnel : filtrage médian de  $v_j$  par *kernel additive modeling*

# KAMIR (2)

Gains  $\lambda_{ij}(f, n)$  initiaux



Gains  $\lambda_{ij}(f, n)$  estimés



Micros



Sources estimées



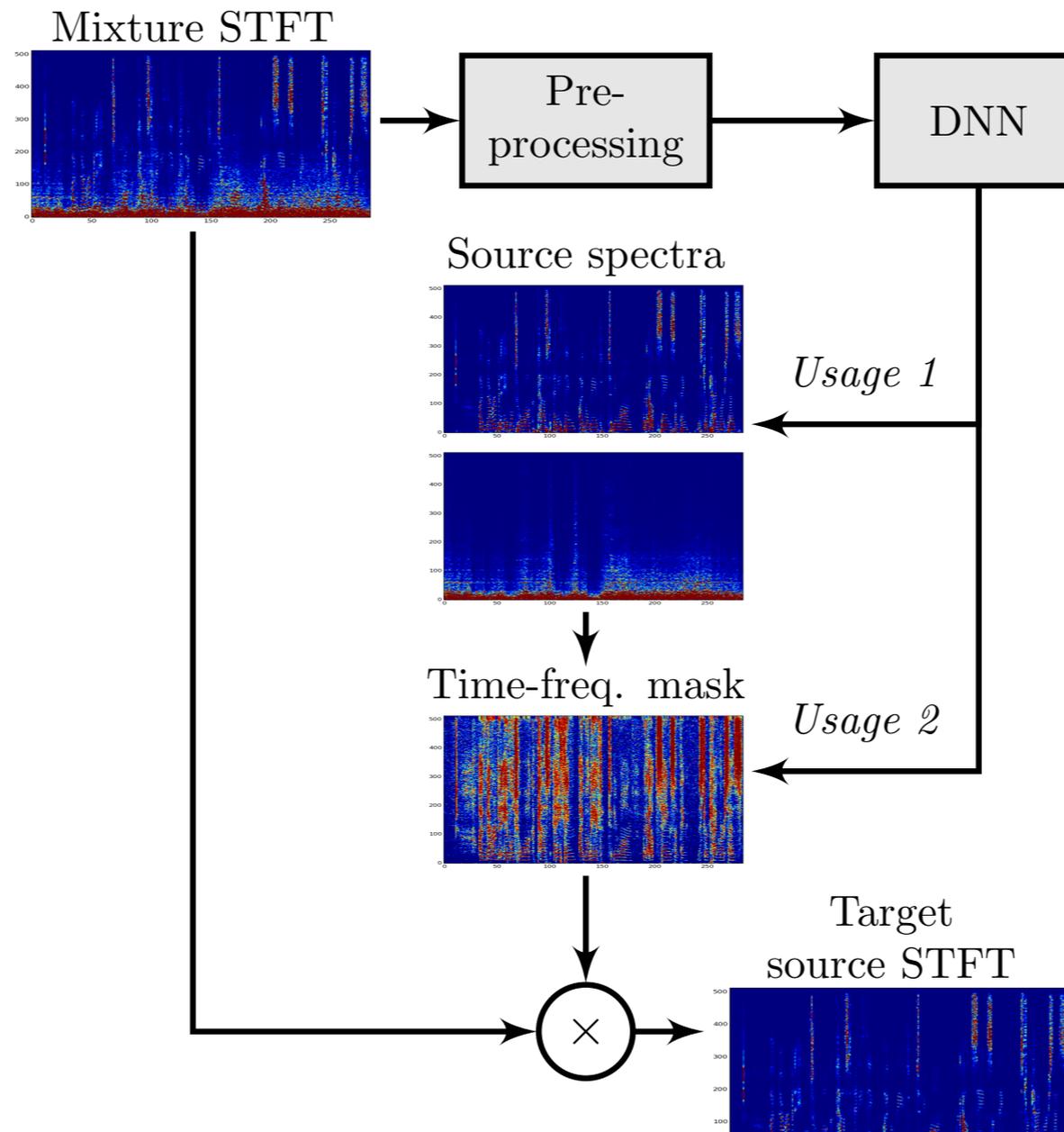
Démo

T. Prätzlich, R. Bittner, A. Liutkus, M. Müller, “Kernel additive modeling for interference reduction in multichannel music recordings”, ICASSP, 2015.

T. Prätzlich, R. Bittner, A. Liutkus, J. P. Bello, M. Müller, “Interference reduction for multitrack music recordings”, IEEE/ACM TASLP, submitted.

# dnnsep (1)

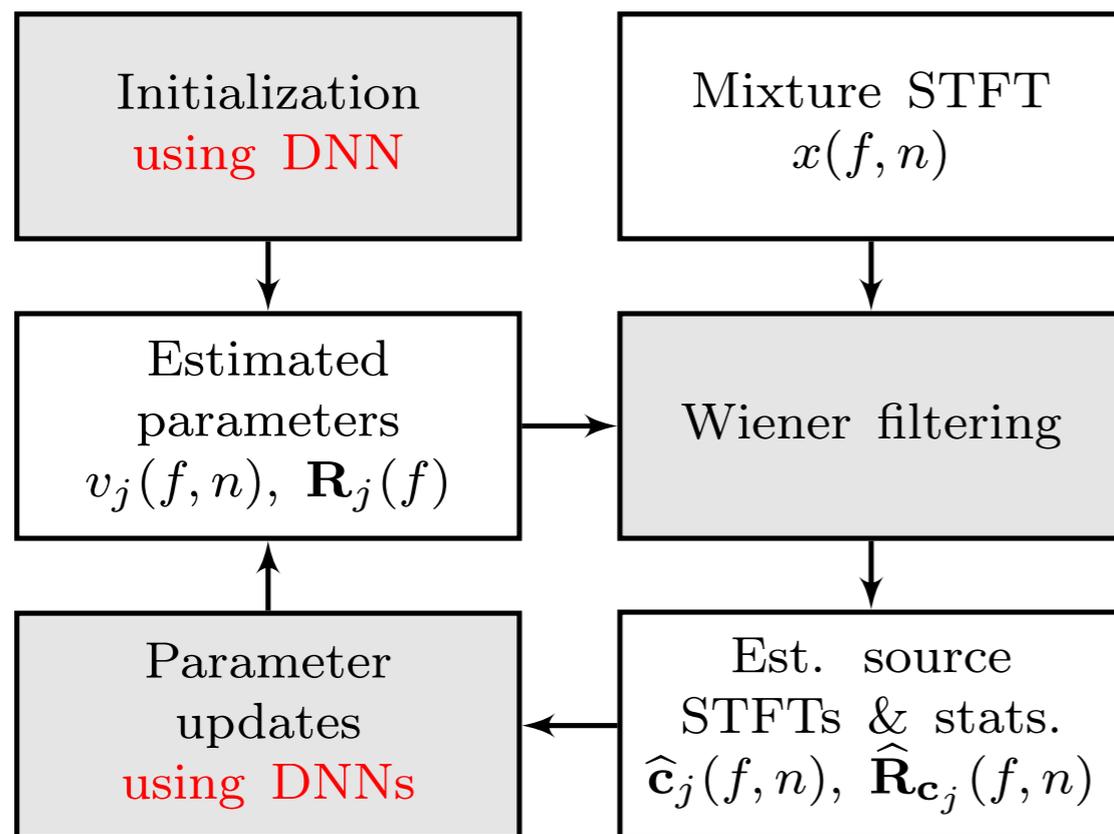
État de l'art : *deep learning* pour le rehaussement de la parole monocanal.



## dnnsep (2)

Adaptation à la séparation d'enregistrements musicaux multicanal.

$$\mathbf{c}_j(f, n) \sim \mathcal{N}_c(\mathbf{0}, \underbrace{v_j(f, n)}_{\text{spectre de puissance}} \underbrace{\mathbf{R}_j(f)}_{\text{matrice de covariance spatiale}})$$

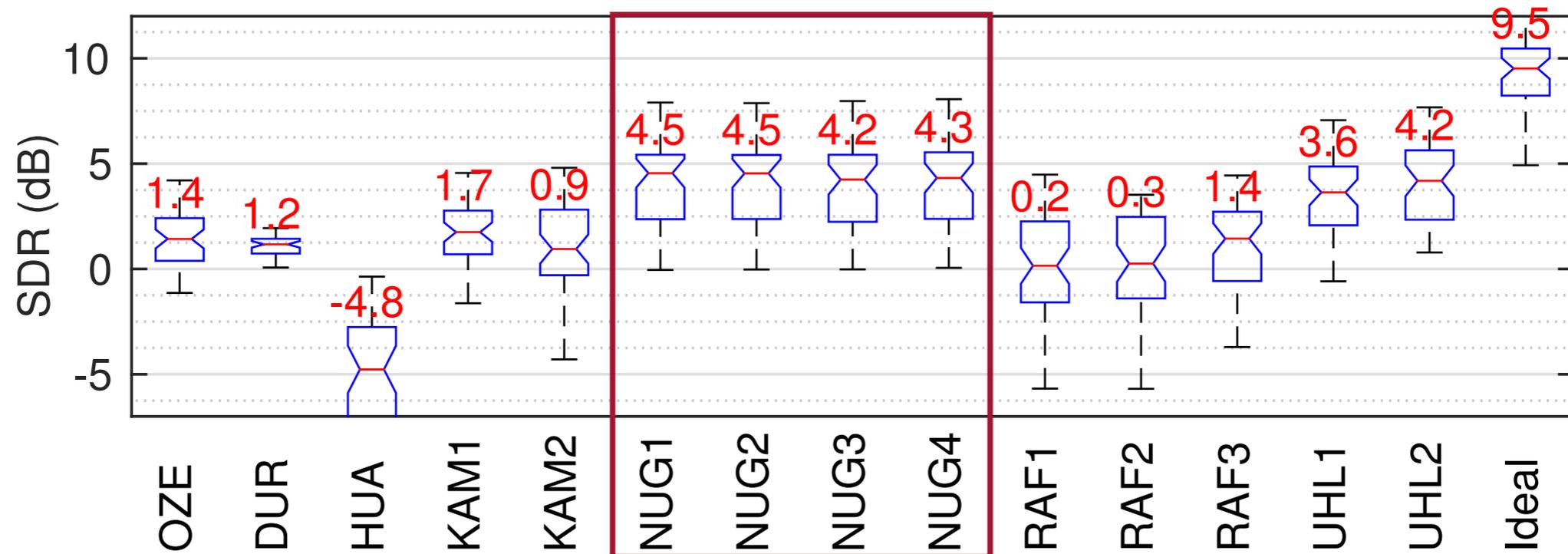


Algorithme EM si DNN appris par minimisation de la divergence d'Itakura-Saito et mise à jour "classique" de  $\mathbf{R}_j(f)$ .

Performance améliorée avec divergence de Kullback-Leibler et mise à jour pondérée  $\Rightarrow$  pseudo EM.

## dnnsep (3)

Classé 1er ex aequo des campagnes d'évaluation SiSEC 2015 et 2016 pour la séparation d'enregistrements musicaux.



## Démo

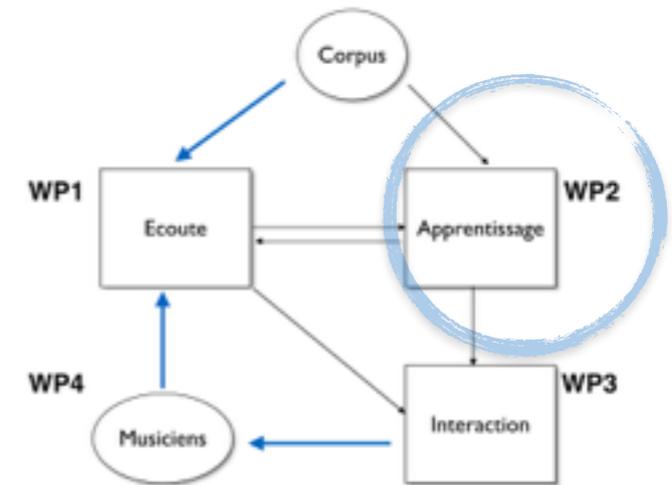
A. A. Nugraha, A. Liutkus, E. Vincent, "Multichannel audio source separation with deep neural networks", IEEE/ACM TASLP, 2016.

A. A. Nugraha, A. Liutkus, E. Vincent, "Multichannel music separation with deep neural networks", EUSIPCO, 2016.

# WP2

## Apprentissage interactif de structures musicales

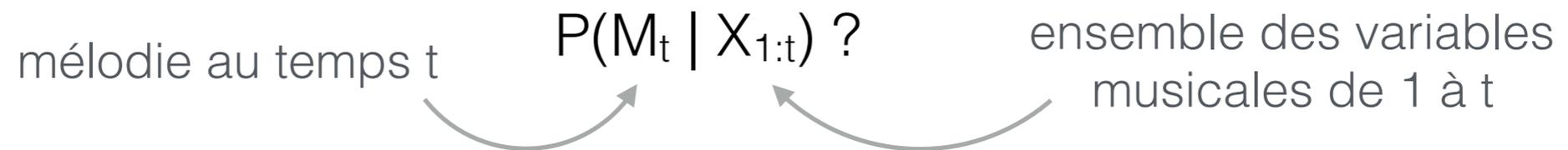
Coord Inria, E. Vincent  
Inria K. Deguernel  
Ircam G. Assayag, J. Nika



- WP2.1 Apprentissage de structures multi-dimensionnelles
- WP2.3 Apprentissage de structures multi-échelles

# Approche bayésienne avec interpolation de sous-modèles

Prédire la mélodie au temps  $t$  en prenant en compte l'ensemble de l'information musicale.



Approximation du modèle par interpolation de sous-modèles.

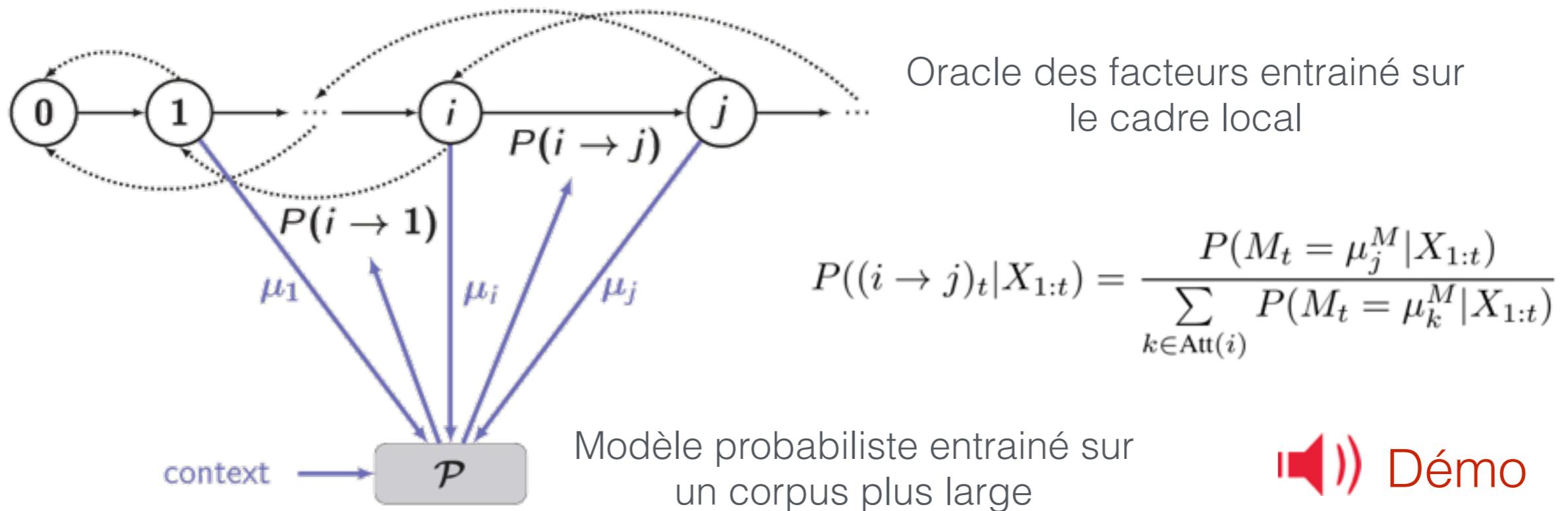
$$P(M_t | X_{1:t}) = \sum_i \lambda_i P_i(M_t | Z_{i,t}) \quad \text{avec } \sum_i \lambda_i = 1 \text{ et } \lambda_i \geq 0 \forall i$$

modèle global  $\leftarrow$  sous-modèles  $Z_{i,t} \subset X_{1:t}$

Exemples de sous-modèles :  $P(M_t | M_{t-1}, \dots, M_{t-n+1})$ ,  $P(C_t | C_{t-1})$ ,  
 $P(M_t | C_t)$ , ... accord au temps  $t$

Utilisation de méthodes de lissage pour éviter le sur-apprentissage et les probabilités nulles dues à des corpus d'entraînement trop petit.

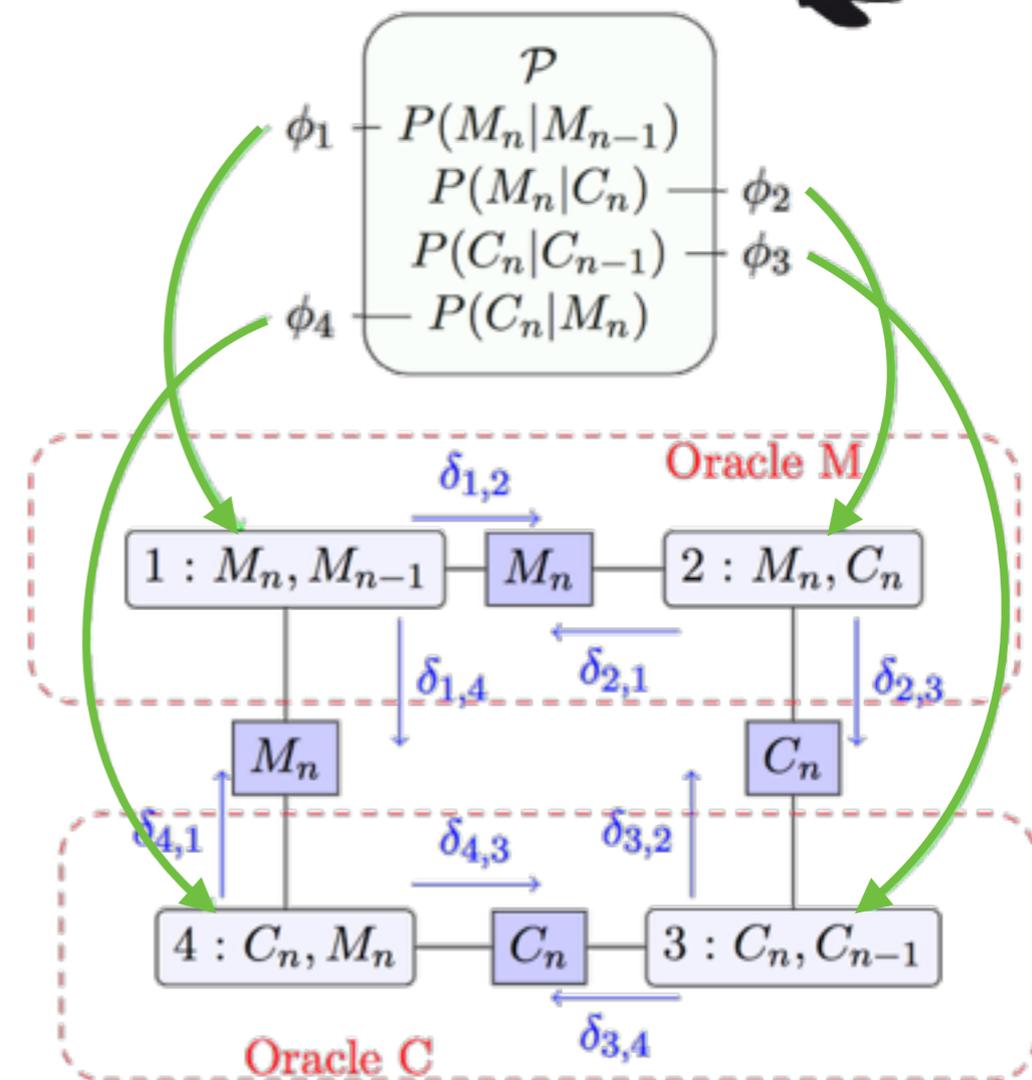
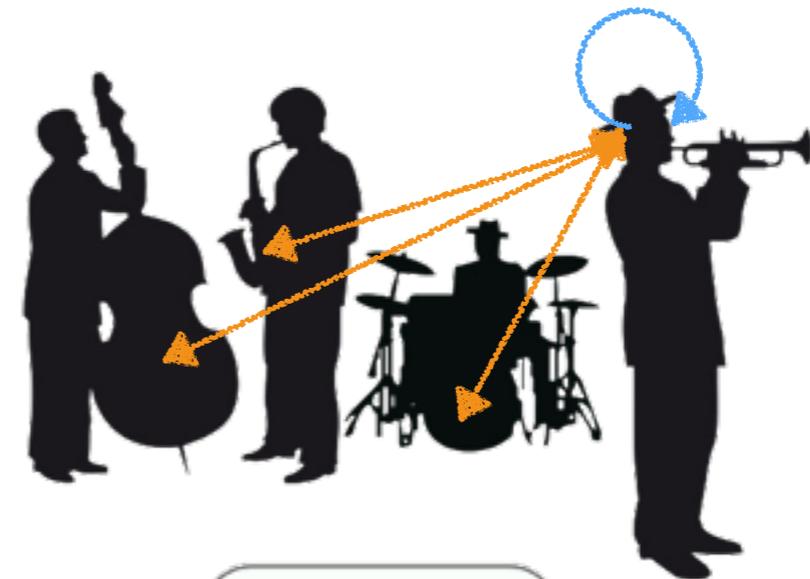
# Oracle des facteurs et modèles probabilistes



- Le module probabiliste représente les connaissances du système.
- L'oracle des facteurs représente le cadre local de l'improvisation jouée.
- L'oracle des facteurs a deux impacts dans ce modèle :
  - il agit comme une contrainte sur le module probabiliste,
  - il permet au système de prendre en compte un contexte plus long.
- La navigation dans l'oracle est guidée par des connaissances multidimensionnelles.

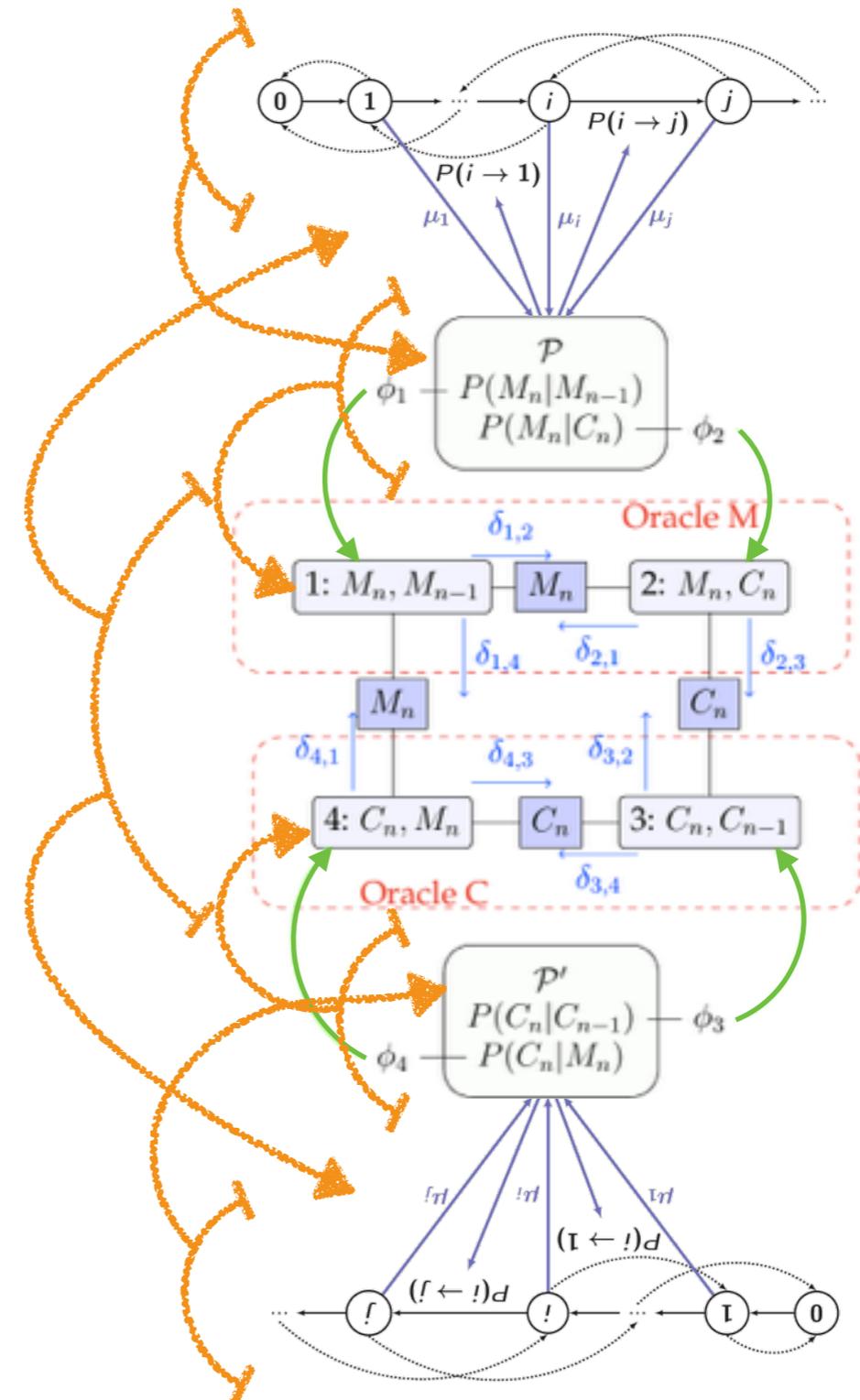
# Communication entre oracles par graphe de clusters

- *But : Générer des improvisations multidimensionnelles*
- Plusieurs oracles sont créés pour représenter les dimensions.
- Les sous-modèles probabilistes servent de facteurs dans le graphe de cluster.
- Chaque oracle est représenté dans le graphe par un sous-ensemble de clusters.
- Les oracles communiquent via le graphe de cluster par passage de messages : *Algorithme de propagation de croyance*.



# Une étape de la génération dans le cas mélodie + accords

1. Les oracles fournissent leurs états atteignables.
2. Les croyances initiales de chaque cluster sont calculées.
3. Passage de messages par l'algorithme de propagation de croyances.
4. Estimation de  $P(M_n)$  et  $P(C_n)$ .
5. Le graphe de cluster donne les probabilités de transition pour chaque oracle.
6. Génération aléatoire suivant les probabilités de transition.



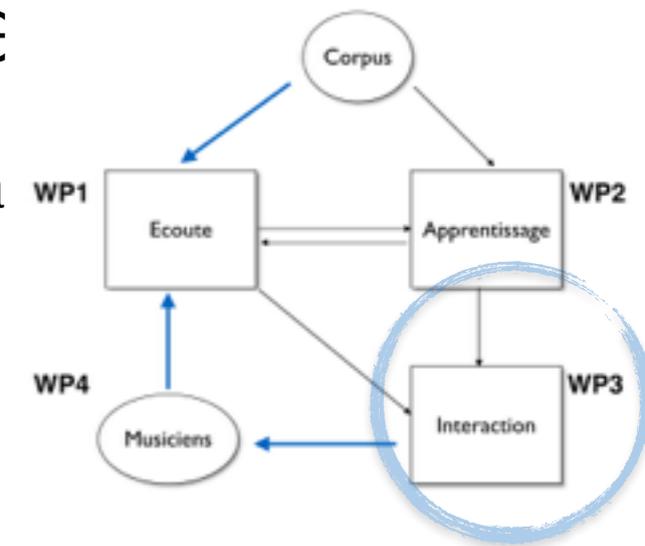
# WP3

## Dynamiques d'interaction improvisé

Coord. Ircam, G. Assayag

Ircam J. Nika, K. Sanlaville, F. Bevilacqua, A. Chemla

Inria K. Deguernel



- WP3.1 Guidage de l'interaction improvisée (I par scénario, II par écoute réactive)
- WP3.2 Adaptation temporelle de l'interaction
- WP3.3 Mémoire, connaissance, contrôle des agents créatifs

# Improvisation musicale homme-machine

*Conception d'agents créatifs capables d'improviser sur scène  
avec des musiciens humains*

**Perception** : *écouter* pour apprendre et/ou réagir.

**Mémoire** : mémoire *long-terme* et mémoire *instantanée*.

**Initiatives / stratégies** : dépendant du contexte musical et des connaissances a priori associées.

Improvisation libre ? Pulsée ? Jazz, blues, ..., « idiomatique » ?

# Adaptation au contexte musical

Contexte musical

Purement réactif

Planification long-terme

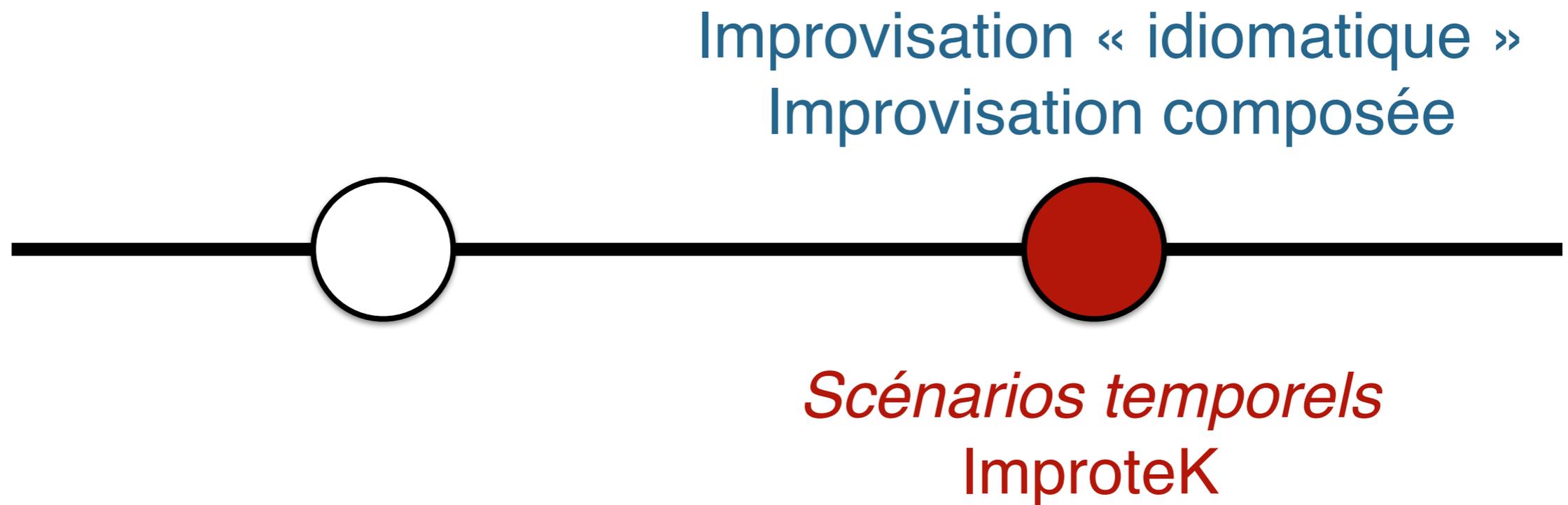


Approche

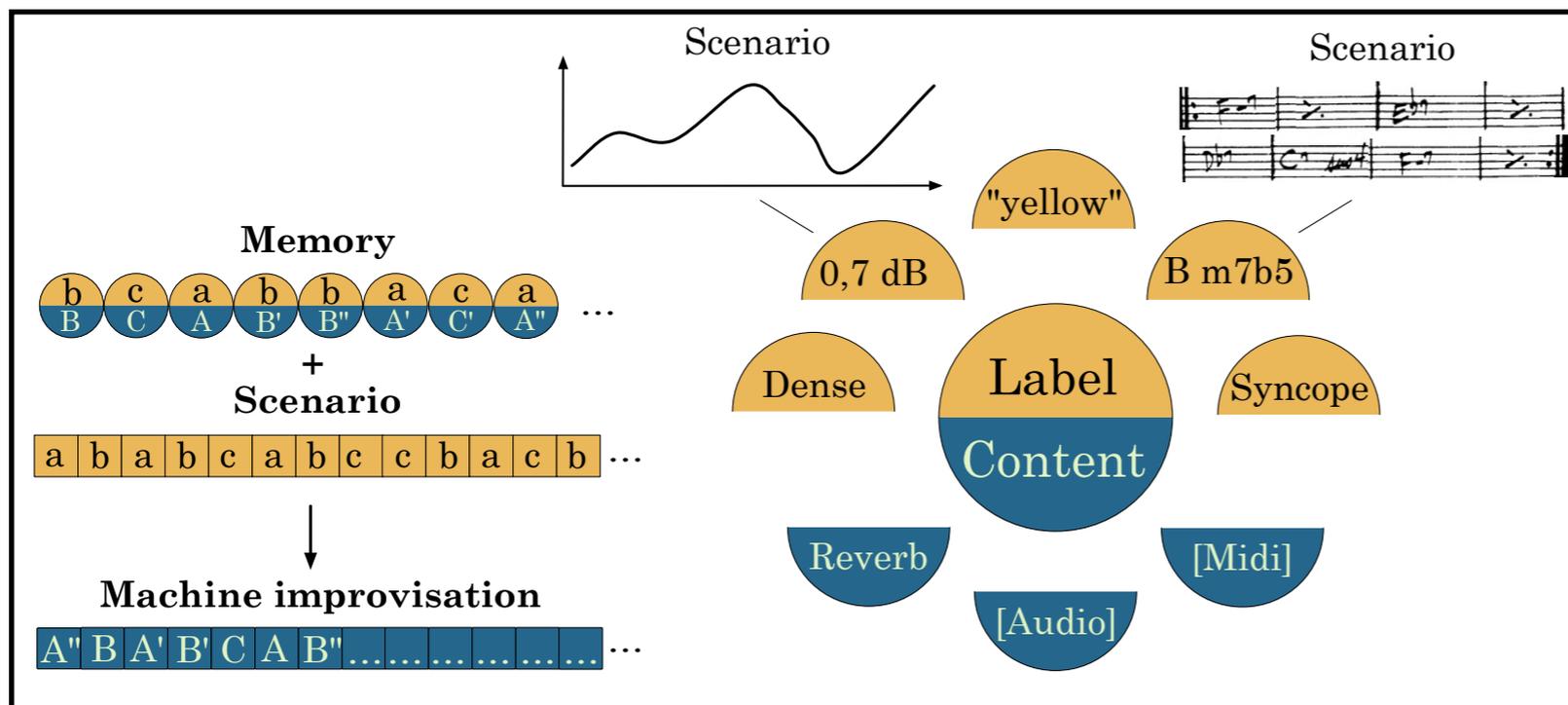
# Adaptation au contexte musical (1)

Purement réactif

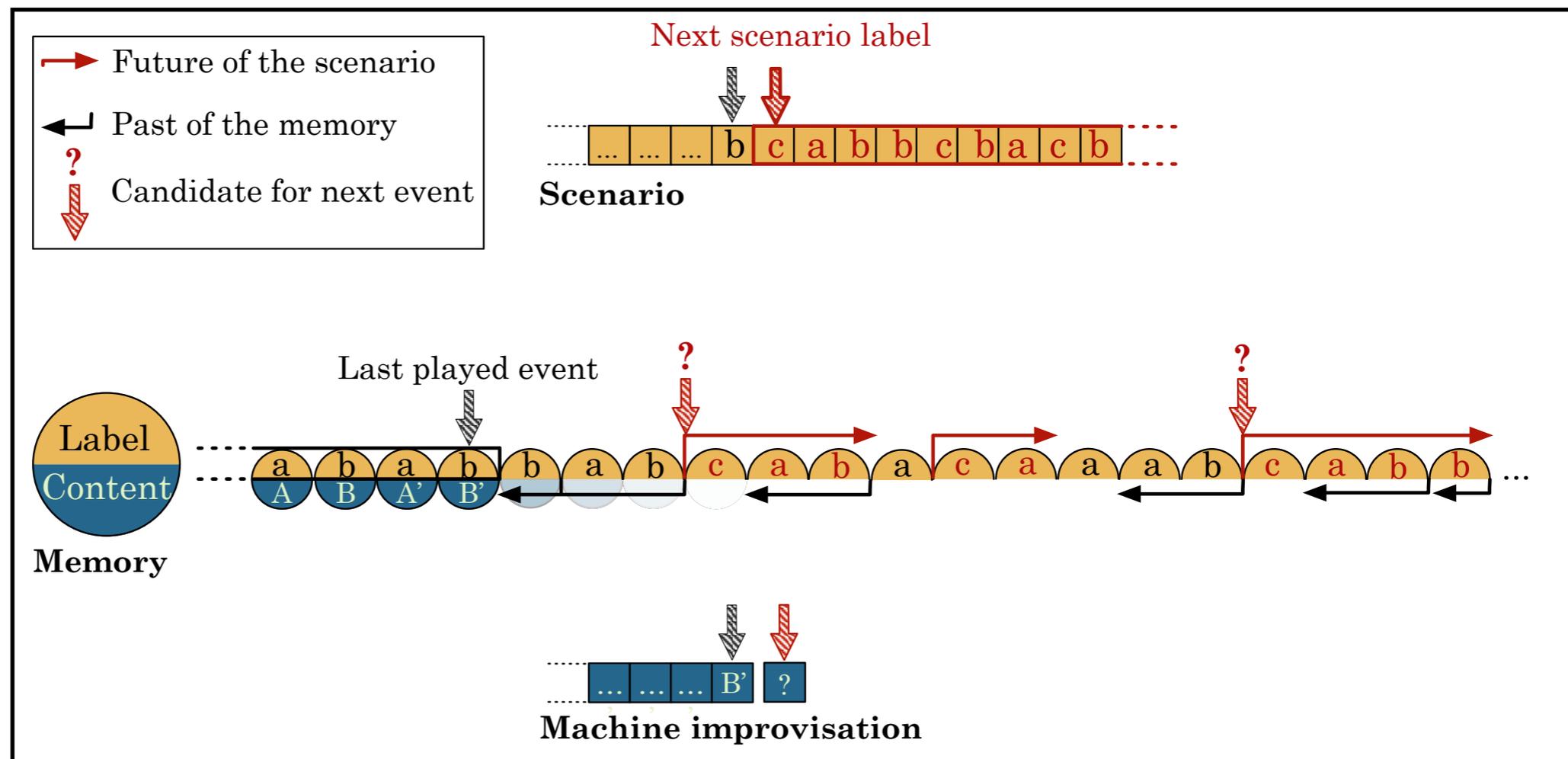
Planification long-terme



# Adaptation au contexte musical (1) : Scénario



WP 3.1.1  
ImproteK



# Adaptation au contexte musical (2)

Purement réactif

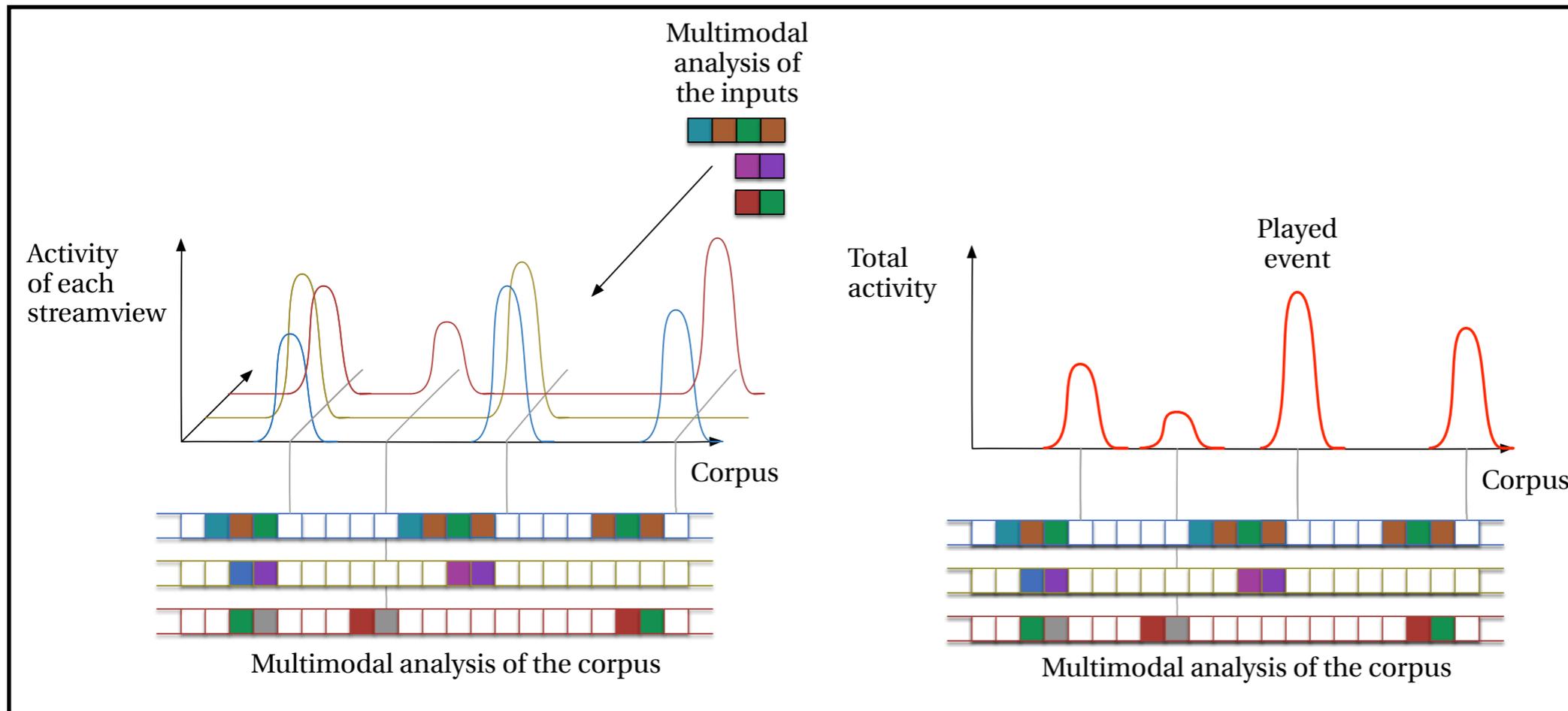
Synchronisation musical flottante  
Accompagnement

Planification long-terme

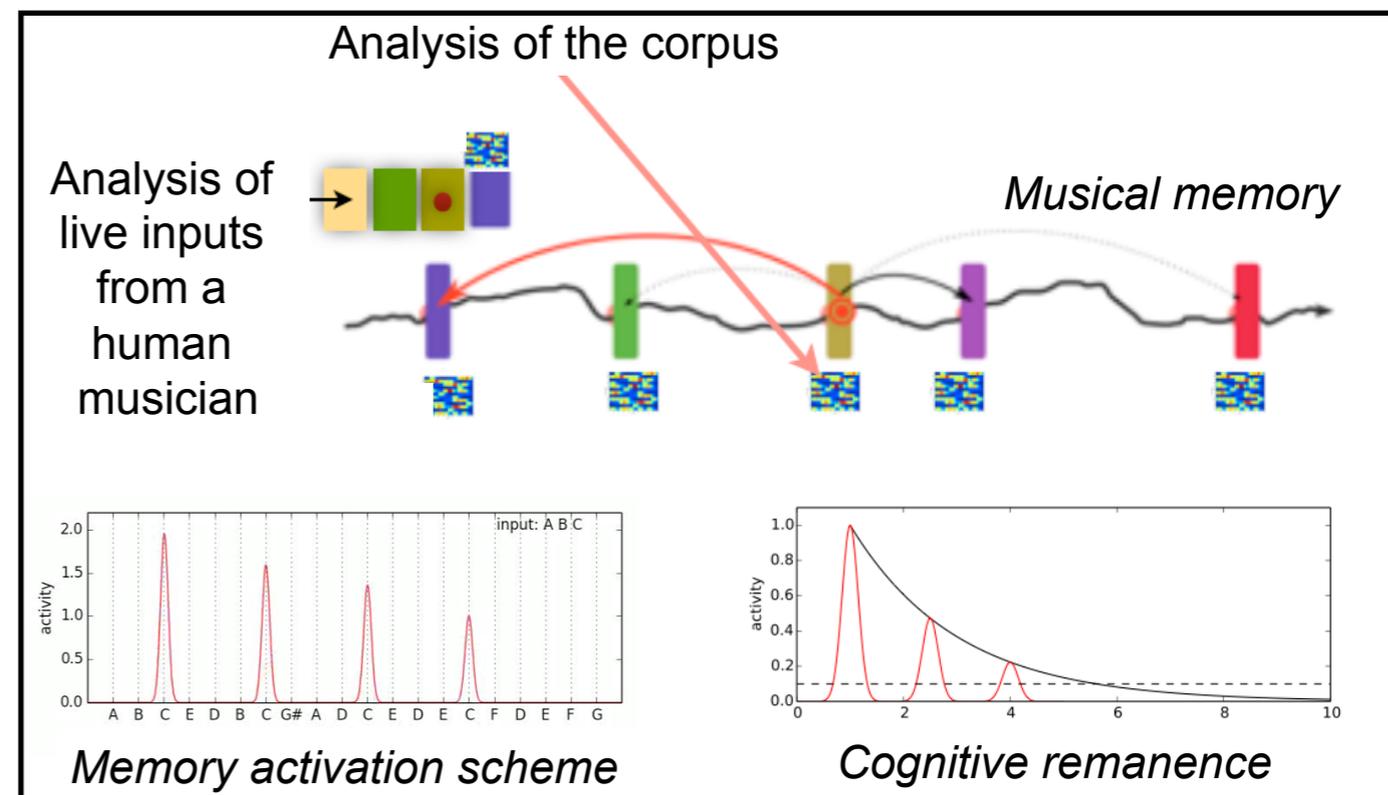


*Écoute réactive*  
Somax

# Adaptation au contexte musical (2) : *écoute réactive*



WP 3.1.2  
Somax



# Conformity / Scenario - Digression / Memory



Scenario: chord chart - Memory: live audio  
ImproteK music improvisation software, with Rémi Fox

## *Rent party*

Rémi Fox (saxophone), Jérôme Nika (ImproteK),  
Repetitions for a performance at Montreux Jazz Festival 2015.



**Scenario:** simple chord progression

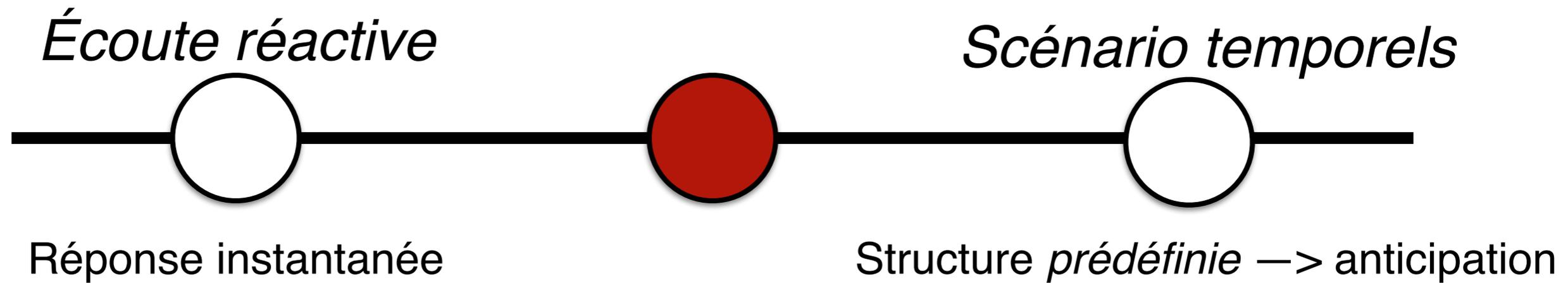
II: Cm7 Bb7 | AbMaj7 Bb7 :II (*Rent party*, Booker T. Jones)



**Memory:** live audio segmented and labeled online —> same chord progression

| 1 voice

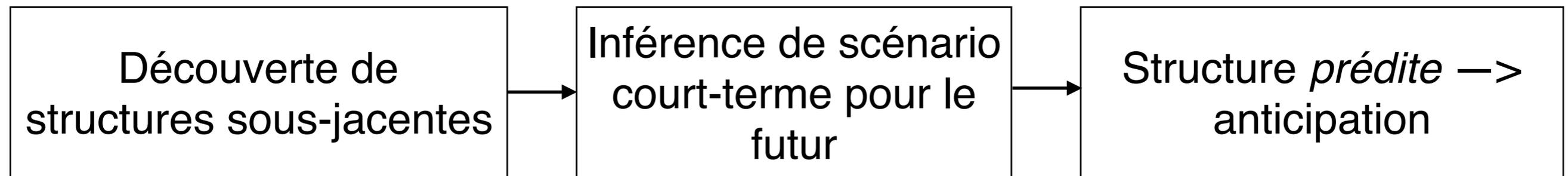
# Adaptation au contexte musical : perspectives



Fusion des paradigmes « écoute réactive » et « scénario ».  
Improvisation structurée temporellement sur une dimension musicale et réactive à l'environnement sur une autre dimension.

Nika, Bouche, Bresson, Chemillier, Assayag, *Guided improvisation as dynamic calls to offline models*. Sound and Music Computing 2015.  
Chemla—Romeu-Santos, *Guidages de l'improvisation*. Master's thesis ATIAM (UPMC), 2015.  
Bouche, Nika, Chechile, Bresson, *Computer-aided composition of musical processes*. Journal of New Music Research, 2016.

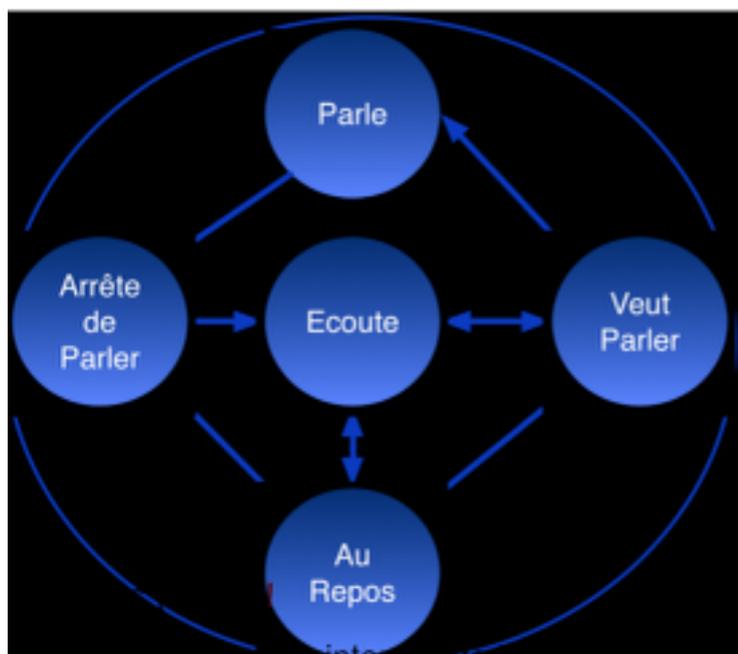
Perception —> Expectation —> Prédiction —> Anticipation



Bazin, *Deep learning for music structure analysis and prediction*. Master's thesis ATIAM (UPMC) - ENS Cachan, 2016.

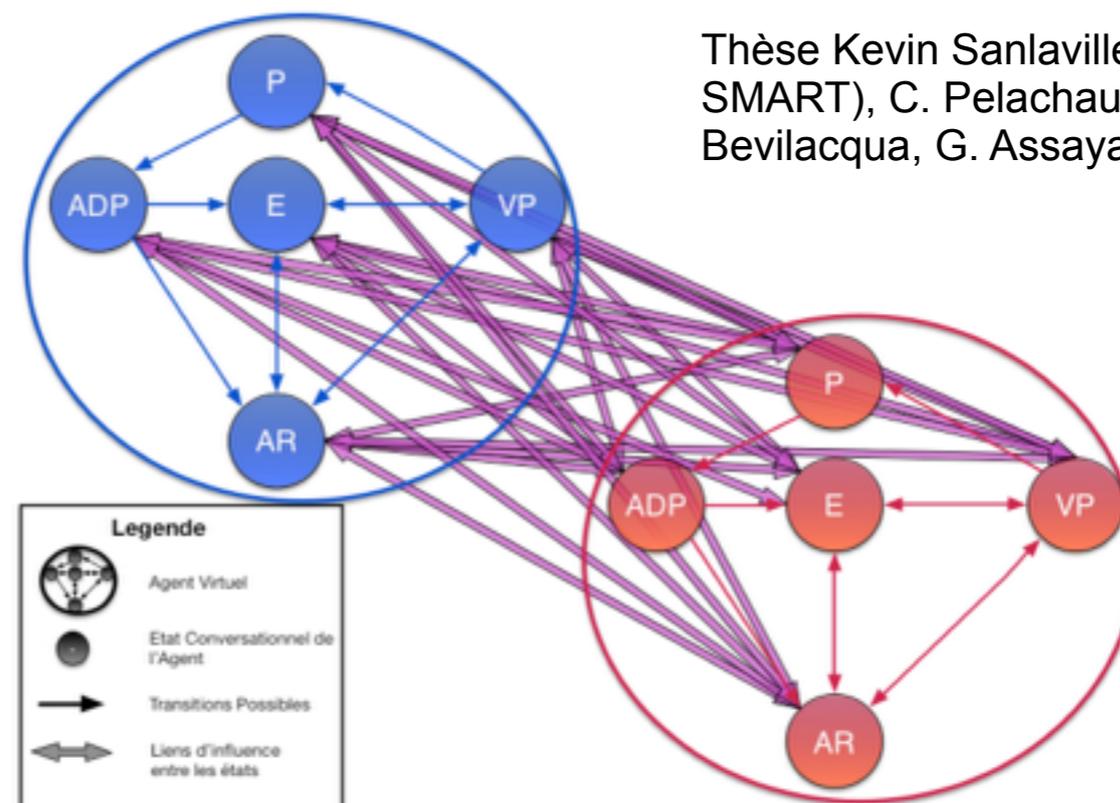
# Adaptation temporelle de l'interaction

Modéliser les comportements internes des agents de sorte que les règles conversationnelles globales (synchronie, turn taking) émergent d'elles même.



Modèles de Markov Cachés (HMM) !

Etats du HMM = Etats conversationnels internes de l'agent  
Observations du HMMs = signaux verbaux et non verbaux



Thèse Kevin Sanlaville (Labex SMART), C. Pelachaud, F. Bevilacqua, G. Assayag

Modèles d'influence [Dong 2007]

Plusieurs HMMs

Les transitions ne dépendent plus seulement de l'état du modèle mais des états des autres modèles (les autres agents) selon une pondération appelée « influence ».

Expérimentation : Base d'apprentissage quatre agents GRETA dans un environnement créé dans UNITY, résultat de la simulation : stabilité des agents, répartitions similaires des temps de parole (monopole, cacophonie, turn taking).

# WP4

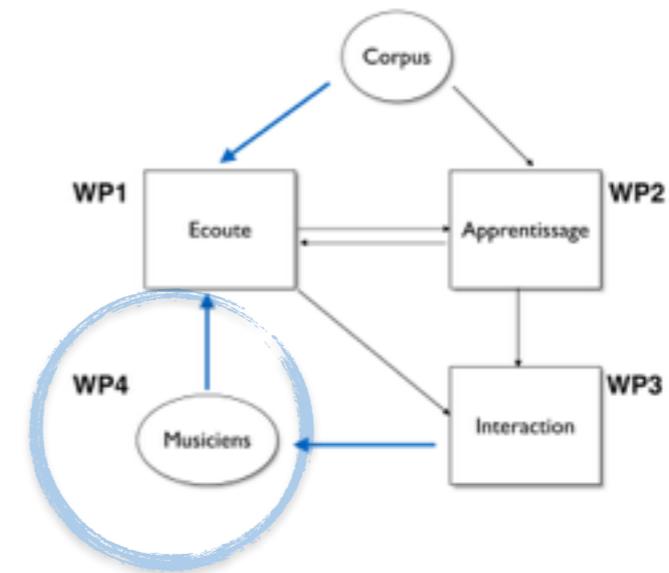
Expérimentation, dissémination, retour d'usage, intégration

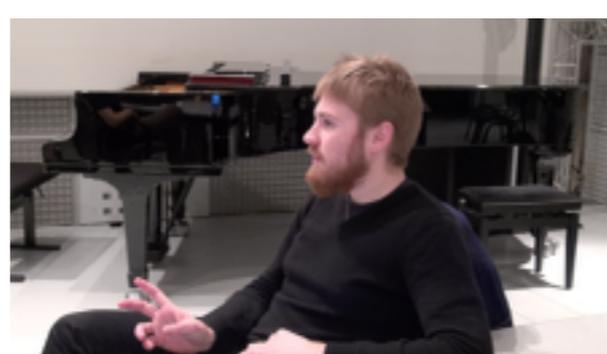
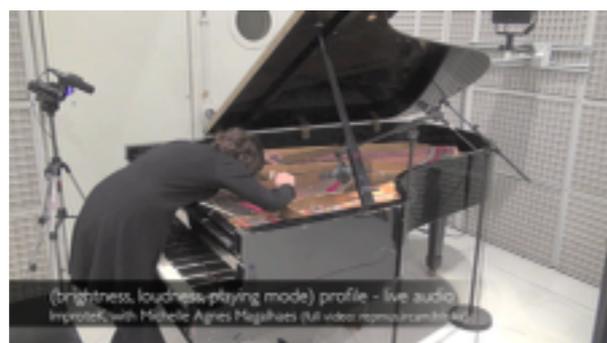
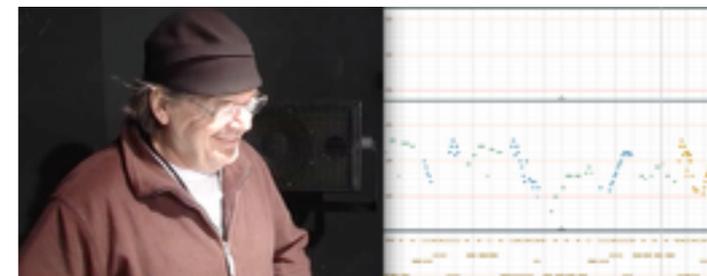
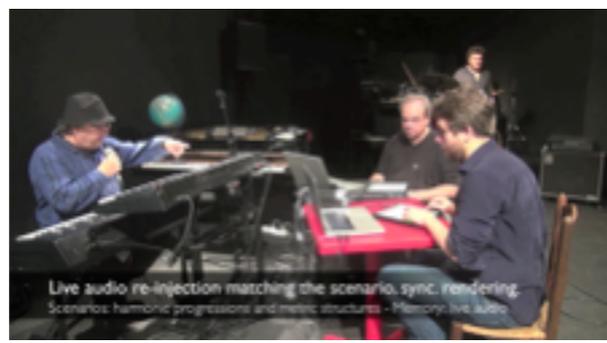
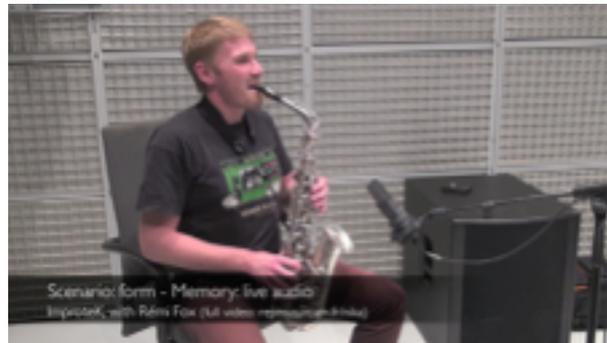
Ircam coord G. assayag

Ircam J. Nika

Inria E. Vincent, K. Deguernel, A. Liutkus

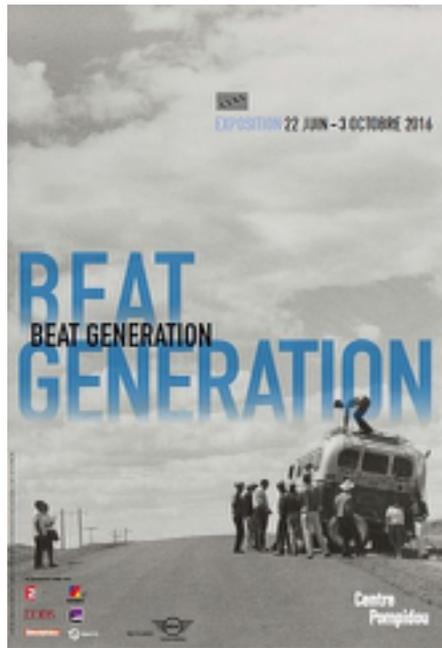
ULR S. Marchand





Vidéos de sessions d'expérimentation et entretiens disponibles sur le site du projet (~ 40 vidéos)

# 15 concerts depuis 2015



« Secret Heroes », **Festival Manifeste** de l'Ircam, Grande Salle, Centre Pompidou, 22 juin 2016, avec Benoît Delbecq et al



Projet « Jazz augmenté » avec la compagnie Lubat : **Tournée Madagascar-Mayotte**, mai 2016



« Edith Piaf, Elizabeth Schwarzkopf, Billie Holiday, tre donne del 1915 », H. Sellin, G. Bloch, **Festival internazionale Pietre che cantano**, L'Aquila, Italie, 21 août 2015



« Edith Piaf, Elizabeth Schwarzkopf, Billie Holiday, tre donne del 1915 », H. Sellin, G. Bloch, **Conservatoire Rameau, GDR Esthétique Arts & Sciences**, Paris, 6 février 2016



Workshop et concert « Musician and machine », **Montreux Jazz festival**, juillet 2015



**Festival d'Uzeste**, Un concert par jour du 16 au 23 août 2015 avec les musiciens du festival

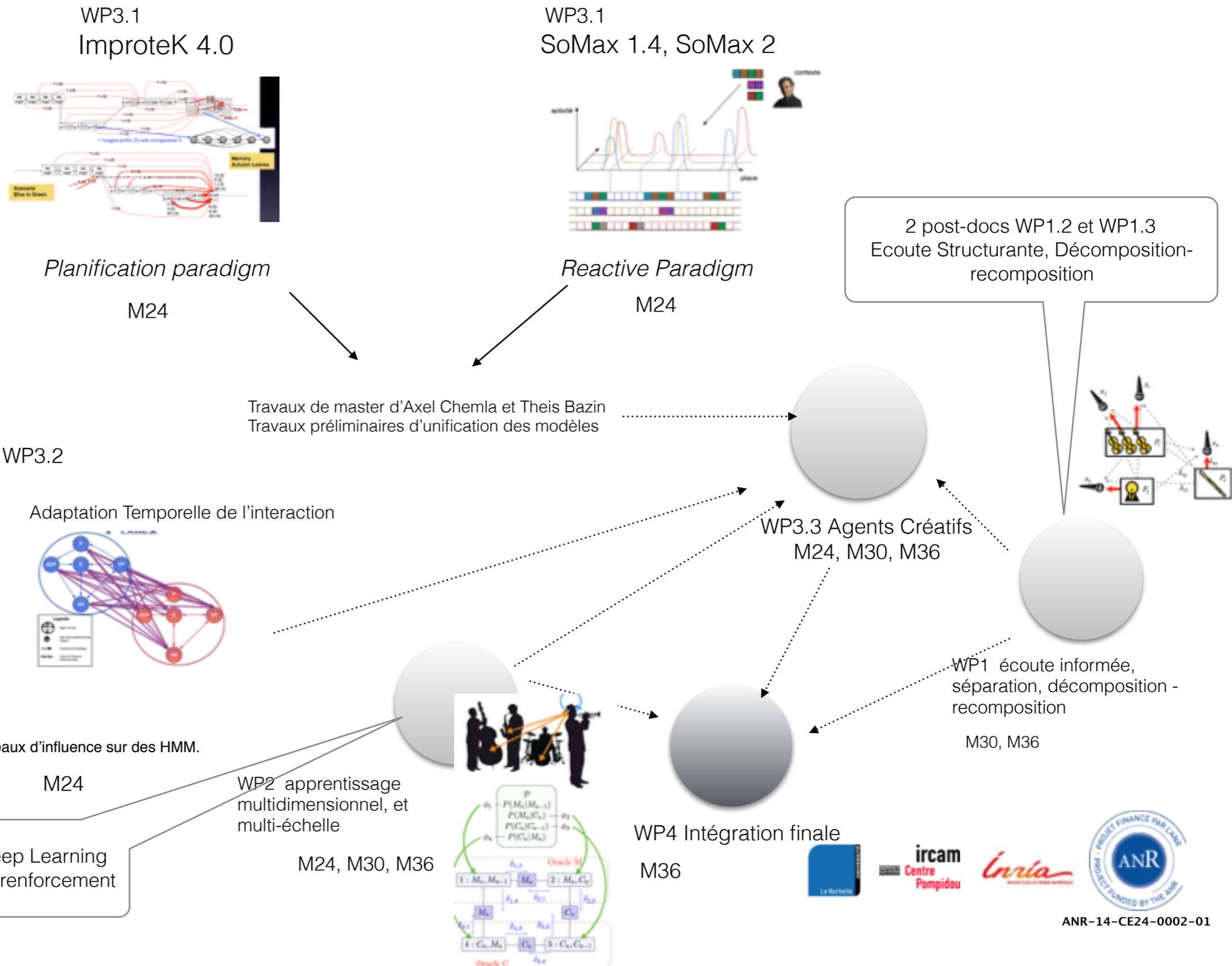


« Ateliers Innatendus » école itinérante France-Belgique organisée par l'IRI et le philosophe **Bernard Stiegler**. Ateliers **Mons Capitale de la culture 2015**, concert lors du **Festival de Jazz de Tournai 2015**



Conférence et Concert en co-improvisation avec les systèmes Omax et ImproTeK, Bernard Lubat, **Collège de France**, Paris, 27 mai 2016

# DYCI2 : Dynamique à partir de T18



MERCI