Apprentissage de structures musicales en contexte d'improvisation

Ken Déguernel

UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Apprentissage de structures musicales en contexte d'improvisation

THÈSE

présentée et soutenue publiquement le 6 mars 2018

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention informatique)

par

Ken Déguernel

Composition du jury

Rapporteurs :	Elaine Chew	Professeur, Queen Mary University of London
	Maxime Crochemore	Professeur, King's College London
Examinateurs :	Florence Levé	Maître de Conférences, Université de Picardie
	Kamel Smaïli	Professeur, Université de Lorraine
Directeurs de thèse :	Emmanuel Vincent	Directeur de Recherche, Inria
	Gérard Assayag	Directeur de Recherche, Ircam

Institut National de Recherche en Informatique et en Automatique — UMR 7503 LORIA
Institut de Recherche et Coordination Acoustique/Musique — UMR 9912 STMS

Mis en page avec la classe thesul.

« conseillez-vous soigneusement »
- Érik Satie.

Remerciements

Merci à Emmanuel Vincent et Gérard Assayag.
Merci à Elaine Chew et Maxime Crochemore.
Merci à Florence Levé et Kamel Smaïli.
Merci à Darrell Conklin, Georges Bloch et Alain Dufaux.
Merci à Jérôme Nika et Karim Haddad.
Merci à Denis Jouvet et Hélène Zganic.
Merci à Louis Bourhis, Joël Gauvrit et Pascal Mabit.
Merci à Pierre Couprie, Rémi Fox et Bernard Lubat.
Merci à Damien, Léna, Axel, Adrien, Hugo, Léo et Lauréline.
Merci à Mathieu, Antoine, Diego, Aman et Baldwin.
Merci à Nathan, Sunit, Arie, Mathieu, Théo et Iñaki.
Merci à Guillaume, Angie, Thomas, Aleksi et Adrien.
Merci aux parents. Merci à Leslie.

À ma mère

Sommaire

1

Introduction

"Begin anywhere."

– John Cage

1.1 Cadre de la thèse

L'objectif de l'improvisation musicale automatique est de créer des agents musicaux numériques créatifs capables de s'introduire dans une scène musicale improvisée. Afin de pouvoir s'adapter à différentes scènes et différents styles musicaux, ces agents doivent être dotés de méthodes d'écoute artificielle, d'apprentissage et d'interaction. L'objectif de ces communications humain-agent numériques est de faire émerger de nouvelles dynamiques créatives d'expérience de jeu, esthétiquement satisfaisantes pour les musiciens, leur permettant de développer sur scène leur langage musical. La Figure 1.1 propose une représentation des différentes tâches constituant le fonctionnement d'un improvisateur numérique. Ce fonctionnement s'articule autour de trois tâches principales : l'écoute interactive, l'apprentissage de structures musicales et les dynamiques d'interaction dans l'improvisation. L'écoute interactive a pour objectif de constituer à partir d'un corpus ou en direct à partir du jeu des musiciens des données structurées pouvant être utilisées par le module d'apprentissage. Elle permet également de fournir des informations sur l'environnement, nécessaires au module d'interaction. Le module d'apprentissage a pour objectif de créer des modèles de mémoire capable de modéliser un style musical à partir d'un corpus ou à l'aide des données issues du module d'écoute et pouvant s'appuyer sur un scénario temporel connu a priori. Les modèles de mémoire sont alors utilisés pour la génération de musique dans le module d'interaction. Le module d'interaction utilise les modèles de mémoires ainsi que les informations de l'environnement musical (venant de musiciens ou d'autres agents numériques) afin de participer à la constitution du discours musical de la scène vivante de manière cohérente et créative.

Dans cette thèse, nous nous concentrons sur l'aspect apprentissage et sur la modélisation du style musical. Les travaux présentés font suite aux études sur l'utilisation de séquences avec des techniques issues de la théorie des langages formels pour la modélisation de styles musicaux. Nous nous baserons

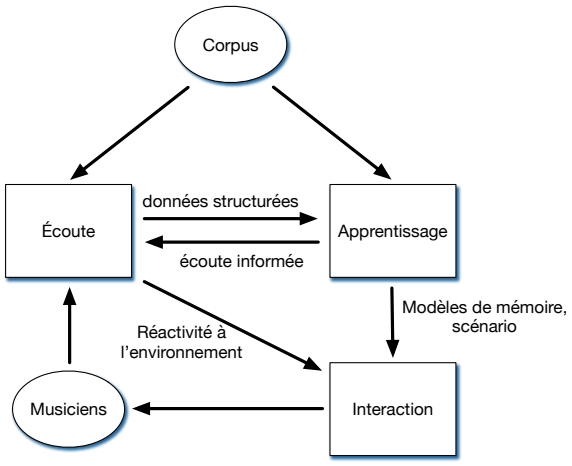

FIGURE 1.1 – Diagramme Écoute-Apprentissage-Interaction représentant la structure d'un improvisateur numérique.

sur les travaux sur l'improvisation automatique effectués à l'Ircam autour du logiciel *OMax* [Assayag et al., 2006b]. Le but de ce logiciel est de créer une mémoire musicale symbolique qui sera ensuite naviguée pour la génération ; cette mémoire pouvant être construite hors-ligne sur un corpus, ou en direct par l'écoute du jeu du musicien improvisant avec le système. Le modèle de mémoire utilisé est une structure d'automate issue de la théorie des langages formels et de la bio-informatique : l'oracle des facteurs [Allauzen et al., 1999; Crochemore et al., 2007b]. Cet automate permet une représentation linéaire de la mémoire et connecte chaque état de la mémoire à un état passé partageant un contexte similaire optimal, équivalent à un modèle de Markov d'ordre optimal, arbitraire à chaque état. Des heuristiques adaptées ont été créées pour la navigation de l'oracle [Assayag & Dubnov, 2004; Assayag & Bloch, 2007] afin d'étendre les capacités d'expression du modèle génératif. Une limite fondamentale de ce système est de pouvoir construire sa mémoire en ne suivant qu'une seule des dimensions d'une séquence musicale (par exemple, la hauteur) ; les autres dimensions (par exemple, l'harmonie, le rythme, le timbre, etc.) n'ayant pas d'impact ensuite dans le processus génératif de navigation de l'oracle et étant en pratique reproduites à l'identique. Cela est dû à l'explosion combinatoire des possibilités si l'on veut modéliser l'évolution sur plusieurs dimensions. Suite à ces recherches, les travaux autour de la thèse de Jérôme Nika [Nika, 2016] ont donné naissance au logiciel *ImproteK* dont l'objectif est d'introduire le concept de scénario temporel pour le guidage de l'improvisation automatique dans le cas de l'improvisation idiomatique. Cette méthode est appropriée en particulier pour les styles musicaux s'appuyant sur la présence d'une structure, comme, par exemple, une grille harmonique dans l'improvisation jazz ou rock. La génération est alors guidée par des mécanismes d'anticipation du scénario [Nika et al., 2017a] permettant, par exemple, à l'agent numérique de résoudre correctement des cadences. Une limitation de ce système est que le modèle de scénario utilisé n'est qu'une succession de symboles (par exemple, une suite d'accords) ne prenant pas en compte l'organisation structurelle à de plus hauts

niveaux (par exemple, un ensemble d'accords formant une fonction, ou l'organisation d'une grille harmonique en section, etc.). Les mécanismes d'anticipation et de parcours de la mémoire sont alors limités à une progression locale, alors qu'une même suite d'accords peut avoir différents rôles, selon sa position dans le scénario. La difficulté est que la forme varie d'un style musical à un autre, et peut montrer des irrégularités à l'intérieur même d'un morceau.

1.2 Contexte et motivations

Cette thèse se concentre sur l'intégration des structures multidimensionnelles et multi-niveaux de la musique dans l'apprentissage de l'agent numérique ; la structure représentant la teneur des connaissances d'un système [Korzybski, 1994]. Nous appelons ici dimensions musicales, les données musicales porteuses d'informations sémantiques comme la mélodie, l'harmonie, le rythme, etc., mais cela peut également s'étendre à des concepts plus abstraits comme le contour mélodique ou la densité et aussi à des descripteurs issues du traitement du signal musical comme les chromagrammes ou les centroïdes spectraux. Nous appelons structures multi-niveaux, une organisation du discours musical à plusieurs échelles temporelles, comme l'échelle du temps, de la mesure, de sections, etc. Cette organisation apparaît notamment dans un contexte d'improvisation idiomatique [Bailey, 1980] où l'expression du style est une des motivations principales de l'improvisateur et dépend souvent d'une organisation temporelle. C'est le cas par exemple du jazz, où les improvisations sur les standards ont lieu suivant des grilles harmoniques répétées *ad libitum*. Ochs [2000] décrit des stratégies d'organisations d'improvisations plus modernes mêlant l'idiomatique du jazz avec des procédés compositionnels, inspirés par la musique de Steve Lacy, Anthony Braxton, Cecil Taylor, etc. Le premier objectif est de représenter par des méthodes d'apprentissage automatique les relations et dépendances entre les différentes dimensions afin d'utiliser cette information lors du processus de génération. Le second objectif est, dans le cadre d'une improvisation guidée par un scénario temporel, de permettre à l'agent numérique d'apprendre et d'utiliser la connaissance de cette organisation à plusieurs échelles temporelles pour enrichir son expressivité.

Pour traiter ces questions, nous nous baserons sur des méthodes issues de la théorie des langages formels et de la modélisation probabiliste du langage que nous adapterons au traitement de séquences musicales. Les liens entre la linguistique et la musique ont été étudiés en détail par Lerdahl & Jackendoff [1983] qui se basent sur les travaux sur les grammaires génératives et transformationnelles de Noam Chomsky [Chomsky, 1965; Chomsky & Halle, 1968; Chomsky, 1975]. Ils expriment notamment la complexité du discours musical et ses aspects structurels multidimensionnels et multi-niveaux :

> « Les concepts fondamentaux de la structure musicale impliquent des facteurs tels que l'organisation du rythme et des hauteurs, une différenciation des dynamiques et des timbres, ainsi que des procédés motiviques et thématiques. Ces facteurs et leurs interactions forment des structures élaborées assez différentes, mais pas moins complexes que les structures linguistiques. »

Cette analyse de la musique avec de multiples dimensions liées par un réseau de connexions et de ses structures multi-niveaux se retrouve, de manière abstraite, chez David Shea [Shea, 2000] et de manière plus concrète dans des travaux d'analyse musicologique, comme par exemple sur la musique de John Coltrane [Schott, 2000] ou de Pierre Boulez [Campbell, 2014]. L'analyse de la pratique des musiques improvisées et en particulier du jazz a été étudiée et montre l'émergence de ces structures dans les improvisations, notamment dans la conception du temps [Siron, 2015] ou des interactions entre musiciens [Monson, 1996]. Les motivations musicologiques de cette thèse sont développées plus en détail dans le chapitre 2.

La modélisation probabiliste de la musique symbolique permet la modélisation simultanée de plusieurs dimensions musicales et de leurs dépendances. Cette méthode a été utilisée pour des tâches d'analyse automatique et d'extraction d'information musicale, inspirée par les méthodes de modélisation probabiliste des langages naturels avec l'utilisation, par exemple, de n-grammes. L'objectif est d'estimer la probabilité d'une séquence à partir d'un corpus d'apprentissage. Ces méthodes ont notamment été utilisées pour la modélisation de séquences d'accords [Raczyński & Vincent, 2014]. Des techniques de lissage [Chen & Goodman, 1998] peuvent être utilisées pour répondre au problème d'estimation des probabilités des séquences ayant un nombre faible, voire nul, d'occurrences dans le corpus d'apprentissage, mais pouvant apparaître dans le corpus de test. Une première limitation de ces méthodes est la taille du contexte qu'elles peuvent prendre en considération de manière efficace. En pratique, ces méthodes sont limitées au trigramme, équivalent à un modèle de Markov d'ordre $n = 3$. Conklin & Witten [1995]; Conklin [2013] utilisent des systèmes à points de vue multiples pour la prédiction et la classification de mélodies. L'ensemble des symboles sur chaque dimension forme un espace de possibilités obtenu par produit cartésien des dimensions ; l'estimation des séquences est alors effectuée sur cet espace. En pratique, l'explosion combinatoire d'un tel espace rend difficile cette estimation. L'interpolation de sous-modèles probabilistes, d'abord utilisé pour la modélisation du langage [Jelinek & Mercer, 1980; Klakow, 1998] peut être vu comme une approximation de cet espace et donc être une solution à ce problème. Ces méthodes ont été utilisées pour des tâches d'harmonisation automatique prenant en compte le contexte tonal, harmonique et mélodique [Raczyński et al., 2013a]. Une seconde limitation de ces méthodes est l'adaptation au style musical local. En effet, de telles méthodes créent un modèle moyen d'estimation du style basé sur l'ensemble du corpus d'apprentissage. Or la probabilité d'une séquence musicale, en particulier dans le contexte de l'improvisation, dépend beaucoup du contexte musical local qui est en train d'être joué. Nous voyons alors que la modélisation probabiliste permet de répondre à certaines limitations de la modélisation formelle et *vice versa*. D'un côté, les algorithmes développés autour du projet *OMax* permettent une modélisation forte du style local d'un musicien et la prise en compte d'un contexte local supérieur aux trigrammes. De l'autre côté, les méthodes d'interpolation de modèles probabilistes permettent une modélisation multidimensionnelle de la musique, tout en répondant aux problèmes dûs à des corpus restreints grâce à des méthodes de lissage.

1.3 Présentation des contributions

Combiner les modèles de séquences et l'approche probabiliste dans un contexte d'apprentissage continu d'improvisation est l'objectif principal de cette thèse. La combinaison de ces deux domaines nous permet de modéliser des principes cognitifs de communication entre mémoire à long terme et mémoire échoïque, d'interactions dans des dynamiques collectives et de compréhension d'organisation à plusieurs niveaux temporels. Nous voulons mettre en place des modèles efficaces permettant d'augmenter l'expressivité de l'agent numérique, en ayant en premier lieu une meilleure navigation de sa mémoire grâce à des connaissances multidimensionnelles et par la suite la capacité à générer de nouvelles séquences, combinant, tout en conservant la cohérence stylistique, des séquences de dimensions différentes observées dans le passé mais pas conjointement. Un apprentissage hors-ligne sur des corpus musicaux sera mis en place afin de représenter stylistiquement une forme de personnalité musicale de l'agent numérique pour pouvoir s'adapter à différents genres (jazz, rock, musique classique, etc.). L'agent numérique doit également être capable de s'adapter à une situation de jeu improvisée avec un style local et doit donc aussi répondre au problème de la quantité limitée de données pour la modélisation du style. Nous voulons également généraliser et effectuer un apprentissage automatique du concept de scénario proposé dans Nika [2016], permettant une modélisation multi-niveaux de la forme par un modèle hiérarchique et, par conséquent, un meilleur guidage de l'improvisation pour les mécanismes d'anticipation de scénario et de navigation de la mémoire.

Cette thèse présente trois contributions principales au domaine de l'informatique musicale avec l'introduction des aspects multidimensionnels et multiniveaux pour la génération musicale. De nouveaux paradigmes et modèles théoriques sont proposés inspirés par des procédés cognitifs de connaissances et d'interactions musicales.

La première contribution concerne l'introduction de la multidimensionnalité de la musique dans le cadre de l'improvisation automatique. Nous avons conçu un système capable de suivre la logique contextuelle d'une improvisation tout en enrichissant son discours musical à l'aide de connaissances multidimensionnelles d'une manière similaire à un musicien humain, formant un nouveau paradigme Intuition / Connaissance pour la génération musicale. D'une part, nous utilisons des modèles probabilistes permettant au système d'exploiter des données d'apprentissage multidimensionnelles et de prendre en compte les relations verticales entre les différentes dimensions considérées. Ces modèles sont alors interpolés et bénéficient de l'utilisation de méthodes de lissage et d'optimisation, faisant du modèle global une représentation inspirée du savoir musical qu'un musicien acquiert au cours de sa vie. D'autre part, nous utilisons un oracle des facteurs capable de représenter le contexte musical local et la logique de développement de motifs unidimensionnels d'un musicien. Cette méthode permet alors de générer des improvisations unidimensionnelles, mais prenant en compte des connaissances multidimensionnelles pour se guider.

La deuxième contribution fait suite à ces travaux. Nous avons voulu poursuivre cette recherche pour développer un nouveau système capable de générer des improvisations multidimensionnelles. Nous avons alors conçu un système

multi-agents modélisant l'interactivité entre plusieurs musiciens, ou entre plusieurs dimensions dans l'esprit d'un musicien. Les communications entre les agents sont réalisées à l'aide d'un graphe de *clusters*. Les connaissances de chaque agent sont représentées par des modèles probabilistes lissés, puis un algorithme de propagation de croyance est utilisé sur le graphe permettant un partage des connaissances et une estimation des probabilités marginales. De cette manière, chaque agent est capable de prendre une décision globale au regard de sa propre génération musicale et de ses connaissances internes et des connaissances externes venant des autres agents. Une fois de plus, le contexte local de chaque agent est représenté par un oracle des facteurs. Cette méthode permet alors de générer simultanément les différentes dimensions de l'improvisation de manière cohérente.

La troisième contribution concerne l'aspect multi-niveaux de l'improvisation pour créer des improvisations suivant une logique à court terme et à long terme. Nous nous sommes concentrés sur le cas de l'improvisation guidée par scénario dans un cas idiomatique. Nous avons modélisé la structure d'un scénario avec une grammaire syntagmatique et nous nous sommes intéressés à l'utilisation d'une telle grammaire pour diriger le processus génératif. Tout d'abord, cette généralisation du scénario permet au système de générer plusieurs variations d'une structure globale, permettant d'ajouter de la créativité dans sa réalisation tout en conservant son organisation hiérarchique multi-niveaux. Ensuite, cette information multi-niveaux est utilisée lors du processus de génération d'improvisations à l'aide de nouvelles heuristiques adaptées à cette information structurelle hiérarchique. L'improvisation générée suit la forme globale du scénario tout en respectant ses contraintes et est également capable de s'adapter à de nouvelles variations d'un scénario connu ce qui accroît ses possibilités créatives.

Les différents résultats obtenus autour de ces travaux ont été évalués par des musiciens et improvisateurs experts lors de sessions d'écoute afin de valider et parfaire les décisions scientifiques. Nous privilégions dans cette thèse une étude qualitative de nos générations plutôt que quantitative. Nous souhaitons privilégier une approche critique avec « un pied planté dans le processus de conception, l'autre pied planté dans la pratique réflexive de la critique » [Agre, 1997].

Ces travaux s'insèrent au cœur du projet ANR Dynamiques Créatives de l'Interaction Improvisée [1] dans la tâche "apprentissage interactif de structures musicales". Les différentes contributions décrites ci-dessus ont amené à la constitution de rapports et de maquettes logicielles pour ce projet. Dans le cadre de ce projet, j'ai notamment collaboré avec Jérôme Nika sur l'aspect scénario multi-niveaux ainsi que pour la rédaction d'un article haut niveau décrivant les différents travaux génératifs issus de DYCI2. J'ai également collaboré avec Diego Di Carlo et Antoine Liutkus pour un article de traitement du signal sur la réduction de repisse dans des enregistrements multi-canaux, sur lequel j'ai participé à la constitution et l'analyse d'un test perceptif. Bien que les modèles théoriques et les algorithmes présentés dans cette thèse se veuillent agnostiques et adaptables à n'importe quel genre musical, les applications pro-

1. `http://repmus.ircam.fr/dyci2`

posées en exemples se concentrent sur le cas du jazz, en vue de collaborations avec le Montreux Jazz Digital Project [2] dans le cadre du projet ANR DYCI2.

1.4 Publications

Mes travaux ont donné lieu à plusieurs publications listées ci-dessous. Les travaux des articles [Di Carlo et al., 2017] et [Di Carlo et al., 2018] n'apparaissent pas dans cette thèse, car le sujet de ces articles en est trop éloigné.
En sus de ces publications, les travaux exposés dans les chapitres 4 et 5 ont été présentés lors des *Journées des Jeunes Chercheurs en Acoustique, Audition et Signal Audio 2016* et ont reçu le prix *ex aequo* du meilleur poster par le jury. L'ensemble des travaux de cette thèse a également été présenté lors du *Workshop ImproTech Paris - Philly 2017.*

[Déguernel et al., 2018] Ken Déguernel, Emmanuel Vincent, & Gérard Assayag (2018). Probabilistic factor oracles for multidimensional improvisation. *Computer Music Journal*, 42(2).

[Déguernel et al., 2017b] Ken Déguernel, Jérôme Nika, Emmanuel Vincent, & Gérard Assayag (2017b). Generating equivalent chord progressions to enrich guided improvisation : application to rhythm changes. In *Proceedings of the 14th Sound and Music Computing Conference*, pages 399–406.

[Déguernel et al., 2016] Ken Déguernel, Emmanuel Vincent, & Gérard Assayag (2016). Using multidimensional sequences for improvisation in the OMax paradigm. In *Proceedings of the 13th Sound and Music Computing Conference*, pages 117–122.

[Di Carlo et al., 2018] Diego Di Carlo, Antoine Liutkus, & Ken Déguernel (2018). Interference reduction on full-length live recordings. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Proceesing.*

[Di Carlo et al., 2017] Diego Di Carlo, Ken Déguernel, & Antoine Liutkus (2017). Gaussian framework for interference reduction in live recordings. In *Proceedings of the Audio Engineering Society Conference on Semantic Audio.*

[Nika et al., 2017b] Jérôme Nika, Ken Déguernel, Axel Chemla-Romeu-Santos, Emmanuel Vincent, & Gérard Assayag (2017b). DYCI2 agents : merging the 'free', 'reactive' and 'scenario-based' music generation paradigms. In *Proceedings of the 43rd International Computer Music Conference.*

[Déguernel et al., 2017a] Ken Déguernel, Nathan Libermann, & Emmanuel Vincent (2017a). La musique comme une langue. In Commission française pour l'enseignement des mathématiques, editor, *Mathématiques et langages*, pages 18–19.

2. `http://metamedia.epfl.ch/`

1.5 Plan de la thèse

Dans le chapitre 2, nous parlerons des aspects musicologiques de cette thèse. Dans la partie 2.1, nous présenterons quelques questionnements musicologiques autour desquels nous énoncerons les différentes motivations qui ont poussé nos choix de recherche. Dans la partie 2.2, nous présenterons le corpus musical que nous avons constitué et qui est ensuite utilisé pour les différentes expériences présentées dans cette thèse.

Dans le chapitre 3, nous présenterons l'état de l'art sur la génération (composition et improvisation) automatique dans la partie 3.1 et sur l'apprentissage et la modélisation de structures multidimensionnelles et temporelles dans la partie 3.2.

Dans le chapitre 4, nous nous intéresserons à l'apprentissage multidimensionnel de la musique. Nous expliquerons dans la partie 4.1 comment construire un modèle de connaissance multidimensionnel par interpolation de sous-modèles probabilistes. Dans la partie 4.2 nous expliquerons comment utiliser ce modèle pour diriger la navigation d'un oracle des facteurs dans un paradigme Intuition / Connaissance. Ce paradigme est ensuite évalué par des musiciens experts dans la partie 4.3.

Dans le chapitre 5, nous parlerons de modèle d'interactivité entre dimensions ou entre musiciens pour la génération d'improvisations multidimensionnelles. Nous présenterons d'abord dans la partie 5.1 les outils théoriques utilisés et leur propriétés, issus de la théorie des modèles graphiques probabilistes. Nous expliquerons ensuite dans la partie 5.2 comment utiliser ces outils pour la création d'un modèle interactif générant plusieurs dimensions musicales de manière stylistiquement cohérente. Ces travaux sont ensuite évalués par des musiciens experts dans la partie 5.3.

Dans le chapitre 6, nous présenterons nos travaux sur l'aspect multi-niveaux de la musique pour l'improvisation guidée. Nous expliquerons d'abord dans la partie 6.1 comment utiliser des grammaires syntagmatiques pour la modélisation de structures multi-niveaux. Nous expliquerons ensuite dans la partie 6.2 comment effectuer un apprentissage automatique de cette structure multi-niveaux. Dans la partie 6.3, nous expliquerons comment exploiter cette structure dans le processus génératif de l'improvisation guidée. Ces travaux sont également évalués par des musiciens experts dans la partie 6.4.

Finalement, dans le chapitre 7, nous conclurons en résumant les contributions dans la partie 7.1 et présenterons les perspectives suite à cette thèse dans la partie 7.2.

2

Aspects musicologiques

"Information is not knowledge.
Knowledge is not wisdom.
Wisdom is not truth.
Truth is not beauty.
Beauty is not love.
Love is not music.
Music is the best."

– Frank Zappa

Dans ce chapitre, nous proposons en premier lieu quelques pistes de réflexions sur la nature et la pratique de l'improvisation, afin d'expliquer les différentes motivations musicologiques et les différentes influences qui ont guidé nos choix dans la conception de nos modèles informatiques pour l'improvisation automatique. Ces réflexions se basent principalement sur le livre *La partition intérieure* de Jacques Siron [Siron, 2015] et sur les travaux ethnographiques de John Bowers autour de l'improvisation [Bowers, 2002], auxquels s'ajoutent d'autres lectures personnelles [3]. Nous ne prétendons pas ici répondre à de quelconques recherches musicologiques mais simplement positionner nos travaux par rapport à différents questionnements musicologiques. Nous parlerons d'abord des spécificités de l'improvisation par rapport à d'autres pratiques musicales, puis de la complexité du discours musical inscrit dans un contexte culturel. Nous parlerons ensuite de l'organisation structurelle d'une improvisation et enfin de la notion d'interaction dans une scène improvisée. Dans un second lieu, nous présentons la base de données que nous avons constituée et qui a été utilisée pour les différents travaux effectués dans cette thèse.

3. Les différentes citations présentées dans ce chapitre (à l'exception de celles de Jacques Siron) ont été traduites de l'anglais vers le français par moi-même.

2.1 L'improvisation comme sujet de recherche musicologique

2.1.1 Improvisation et composition

Donner une définition de ce qu'est l'improvisation musicale et sa pratique est un sujet complexe. La définition la plus commune serait que l'improvisation est une composition spontanée. Cependant, dans ses *Elements of Improvisation* [Crispell, 2000], Marilyn Crispell écrit cette phrase entre parenthèses. Cette définition semble être une approximation de ce qu'est réellement l'improvisation, mais est incomplète et imprécise. Quelles sont les spécificités de l'improvisation qui demandent un apprentissage particulier et distinct des techniques de composition [Lewis, 2000a] ? Derek Bailey considère que l'attrait des musiques improvisées réside dans « l'accidentel, le fortuit et le moment ». D'un point de vue plus sociologique, cette vision implique que la pratique de l'improvisation en musique s'épanouirait dans ce que Garfinkel [1967] appelle la « merveilleuse contingence de la vie de tous les jours » [Bowers, 2002]. Pour sa part, Keith Jarrett décrit sa pratique de l'improvisation libre en solo comme un dialogue entre compositeur, improvisateur et interprète. Dans cette définition récursive se trouve une des problématiques principales de la définition de l'improvisation. Comment distinguer les pratiques de l'improvisation et de la composition ? Cette question semble simple au premier abord, mais placer une frontière entre ces deux pratiques est difficile.

L'improvisation a reçu de nombreuses critiques de la part de certains compositeurs considérant que la pratique de l'improvisation ne peut pas prétendre atteindre la complexité d'un travail de composition. Pierre Boulez en vient même à déclarer [Durant, 1989] :

> « Les instrumentistes ne possèdent pas l'invention — sinon ils seraient des compositeurs... La véritable invention demande une réflexion sur des problèmes qui n'ont encore en principe jamais été posés et la réflexion sur l'acte de création implique un obstacle à surmonter. Les instrumentistes ne sont pas des surhommes et leur réponse au phénomène d'invention est normalement de manipuler ce qui existe en mémoire. Ils se rappellent ce qui a déjà été joué afin de le manipuler et de le transformer. »

Cette vision purement moderniste (et polémique) de la musique marginalise complètement l'improvisation face à la composition, la composition demandant un travail trop important pour être réalisé en temps réel devant un public. La musique créée durant une improvisation pourrait paraître originale à l'oreille en apparence, mais pas en essence. « Ne pouvant pas être moderne, cela doit être primitif » résume Bowers sarcastiquement [Bowers, 2002].

Face à cette vision, Fred Frith répond :

> « Vous connaissez tous les clichés - l'improvisation est une "composition instantanée", une "composition spontanée", etc. ; bizarrement ce type d'expression n'existe pas dans l'autre sens. On n'entend pas les gens parler de la composition comme une "improvisation au ra-

> lenti" ! [...] Je pense que le processus de création implique une combinaison de pensée rationnelle, de choix intuitifs, d'une mémoire profonde et d'une volonté. Je pourrais appliquer tous ces mots à la fois à la composition et à l'improvisation. Le processus est différent et concerne différentes échelles temporelles, mais les autres aspects sont essentiellement similaires. »

Cette citation permet d'identifier des éléments nécessaires au processus créatif qu'ils soient appliqués à la composition ou à l'improvisation. Par exemple, les concepts d'intuition et de mémoire profonde sont présentés comme fondamentaux dans la pratique créative. La question de comment ces concepts s'expriment et sont influencés par l'activité créative effectuée se pose par conséquent. Si l'expression de ces concepts était identique, la différence entre composition et improvisation ne serait alors que temporelle.

David Rosenboom exprime une définition similaire et étend la conversation en considérant une synchronicité des activités de création et de prestation [Rosenboom, 1996] :

> « Mes définitions de composition et d'improvisation sont assez simples :
> Un compositeur est simplement un créateur de musique.
> L'improvisation est simplement une composition qui est entendue immédiatement, plutôt qu'entendue ultérieurement.
> Tout mélange de cela est parfaitement faisable. Les créateurs de musique pouvant être des interprètes, compositeurs, analystes, historiens, philosophes, écrivains, penseurs, réalisateurs, techniciens, programmeurs, designers et des auditeurs créatifs — et peut-être même le plus important, les auditeurs. »

On retrouve ici la complexité des liens entre improvisation et composition, peut-être même poussée à son extrême, tout le monde se retrouve alors au cœur du processus de composition et toutes les prestations musicales incluent de l'improvisation. Ceci explique l'apparition des auditeurs comme créateurs. Rosenboom explique que la musique étant une expérience partagée, elle ne peut prendre place sans participation active des auditeurs, ce qui les inclut dans le processus de composition. C'est ce qu'il définit par la suite comme *musique propositionnelle* dont l'objectif est de proposer des modèles d'une réalité musicale complète, se concentrant sur l'émergence dynamique de formes à travers l'évolution et la transformation [Rosenboom, 2000].

Les liens entre improvisation et composition ont également été mêlés en pratique pour la création d'*improvisations structurées*. Larry Ochs, inspiré par la musique d'Anthony Braxton et de Iannis Xenakis, intègre dans ses créations à la fois une part de composition et une part d'improvisation [Ochs, 2000]. Edgard Varèse a également pratiqué l'improvisation structurée dans le cadre d'ateliers avec des musiciens de jazz parmi lesquels Art Farmer et Charles Mingus. Les résultats de ces ateliers sont décrit dans [Johnson, 2012]. John Cage qui était présent a dit des pièces musicales issues de ces ateliers qu'elles « sonnaient comme du Varèse » [Cage, 1973]. À noter que certains extraits issus de ces ateliers ont par la suite été utilisés par Varèse pour son *Poème électronique*.

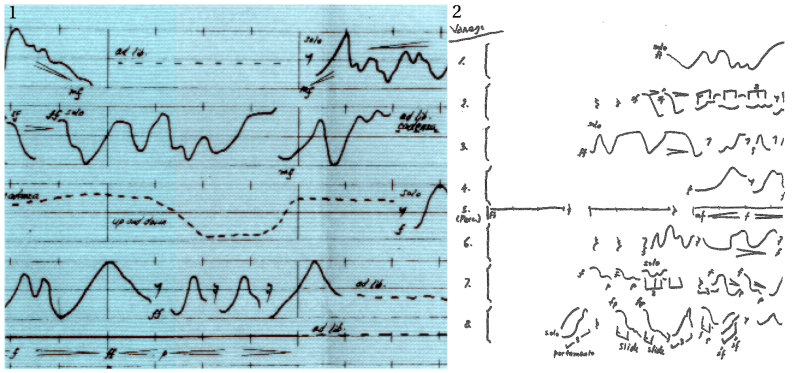

FIGURE 2.1 – 1. Extrait de la partition pour *Poème électronique* d'Edgard Varèse 1958 (source : `giorgiomagnanensi.com`). 2. Extrait de partition de Varèse utilisé pour une improvisation structurée 1957 [Nika, 2016].

La Figure 2.1 montre à gauche un extrait de la partition de *Poème électronique* et à droite la partition utilisée pour une improvisation structurée lors des ateliers. Cette partition est constituée d'une ligne par instrument. Chaque ligne est constituée d'informations sur le contour mélodique, sur le rythme et sur les dynamiques attendues. L'idée fondamentale de l'improvisation structurée est d'utiliser la composition comme outil structurel. D'une certaine manière, il s'agit de la création d'une idiomatique précise guidée par la volonté du compositeur dans lequel les improvisateurs peuvent s'exprimer. Un compositeur d'improvisation structurée doit alors trouver une balance entre sa volonté de contrôle et sa volonté de fournir des pistes pour l'improvisateur. D'un côté, il est clair qu'un système trop fermé retire toute liberté à l'improvisation et l'expression personnelle des musiciens, de l'autre si le système est trop ouvert de sorte que tout évènement musical est permis, l'improvisateur pourrait abandonner ce qui est écrit pour jouer librement.

Siron propose de caractériser les pratiques de l'improvisation et de la composition par l'opposition proposée par Chomsky entre compétence et performance dans le cadre de la linguistique générative [Chomsky, 1965], appliquée à la musique [Siron, 2015] :

> — « La compétence musicale est la connaissance de matériaux musicaux, ainsi que des règles musicales — les manières de façonner les matériaux, de les construire, de les assembler. Dans un sens général, on parle de modèle musical pour désigner les matériaux musicaux et les règles de leur utilisation.
> — La performance musicale est l'activité particulière qu'a un musicien à un moment donné ; c'est la mise en pratique d'un modèle musical, c'est une prise de parole individuelle. »

Dans ce cadre, la pratique de la composition se base sur des compétences sur un système musical. L'improvisation se base également sur des compétences sur un système musical, mais le cadre de ce système est plus lâche ; l'improvisation repose également sur la compétence de dépassement du système par l'extension

et la variation. L'improvisateur doit également avoir une compétence face à l'imprévu. Les prestations de l'improvisateur et du compositeur se distinguent par le fait que l'expression de l'improvisateur est directe alors que celle du compositeur est « hors du temps ».

Motivations : Premièrement, la caractéristique principale de l'improvisation qui ressort est le fait qu'il s'agit d'une expression artistique instantanée, qu'elle soit vue comme composition spontanée, ou comme part intégrale du processus d'interprétation. Les systèmes d'improvisation automatique proposés devront conserver des capacités de réponse en temps réel dans leur génération. Cela nous a encouragé à nous baser sur les travaux autour du paradigme OMax et de l'oracle des facteurs comme modèle génératif qui permet une construction de la mémoire linéaire en temps. Deuxièmement, un improvisateur doit posséder la compétence de gérer l'imprévu. Les modèles proposés devront inclure la possibilité d'une écoute active de l'environnement par rapport auquel ils pourront s'adapter. Les principes d'auto-critique face à l'aléatoire se manifestant par l'erreur (par exemple, la fausse note) ne sont pas étudiés dans cette thèse, mais nous ont encouragé à la direction d'un stage de master sur l'apprentissage par renforcement.

2.1.2 Complexité du discours musical

Un musicien, qu'il soit compositeur, improvisateur ou interprète, possède un ensemble de paramètres (que nous appelons dans cette thèse *dimensions*) sur lesquels il peut jouer : hauteur, rythme, dynamique, timbre, etc. Il est clair que la construction du discours musical repose sur une cohérence dans le contrôle de chacune de ses dimensions. David Shea écrit [Shea, 2000] :

> « La combinaison d'un élément avec un second, ou un autre, ne crée pas une synthèse ou une unité, mais plutôt une stratification ou un réseau de connexions créant un changement dans le temps et dans l'espace. L'ensemble, ou le 'un plus un' qui est deux, est toujours plus grand que la combinaison des éléments individuels. Une chose et son opposé ne sont pas deux, mais un réseau de connections que la pensée et le langage séparent et dont ils créent les différences, afin de pouvoir les distinguer les uns des autres. »

Le premier élément d'improvisation (qui est d'ailleurs également un élément de composition) proposé par Marilyn Crispell [Crispell, 2000] décrit justement cette combinaison d'éléments musicaux dans la construction du discours musical :

> « L'utilisation d'éléments rythmiques, mélodiques et harmoniques et de motifs (deux éléments ou plus unifiés pour former un tout logique) dans le développement d'une improvisation/composition »

Il est également important de comprendre que la pratique de l'improvisation varie selon les styles et les cultures. Improviser de la musique tonale, modale, atonale ou encore improviser librement sont des techniques différentes. Improviser des lignes mélodiques librement est différent d'improviser sur une

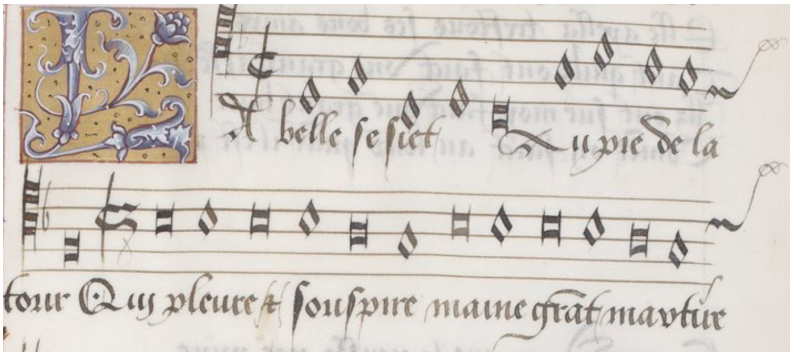

FIGURE 2.2 – Extrait du cantus firmus de *La belle se siet au pié de la tour*, chanson 89 issue du Manuscrit de Bayeux, recueil de chansons normandes du XV$^{\text{ème}}$ siècle.

base harmonique, ou d'improviser, par exemple, sur le timbre comme on peut le trouver dans certaines formes d'improvisation contemporaine [Coursil, 2008]. Nous entendons la musique à travers la connaissance d'un système musical basé sur un ensemble de dimensions musicales (relations harmoniques, cadences rythmiques, etc.), de la même façon que nous comprenons le langage car nous avons intégré une grammaire [Foucault, 1966].

Les formes d'improvisation sont nombreuses et leurs systèmes musicaux encore plus. Dès la Renaissance, la distinction entre la musique improvisée et la musique écrite existe. On parlait de la musique faite avec l'esprit (*con la mente*) et de musique faite avec la main (*con le mani*). Par exemple, la pratique du chant sur le livre (*cantus super librum*) se base sur différentes techniques d'improvisation sur une ligne mélodique donnée, appelée *cantus firmus* afin de réaliser du contrepoint : du gymel à deux voix jusqu'au contrepoint fleuri à cinq voix. La Figure 2.2 présente un exemple de *cantus firmus* issu du Manuscrit de Bayeux (XVème siècle). La ligne mélodique présentée est chantée et ornementée par une voix (la teneur) et d'autres lignes mélodiques sont improvisées autour de celle-ci. Janin [2014] présente les techniques et règles générales pour la pratique du chant sur le livre. Il faut également noter que les pratiques vont varier selon le lieu, le style et l'époque [Canguilhem, 2015; Fiorentino, 2013]. Dans la musique baroque et classique, l'improvisation était utilisée comme méthode de composition avec l'utilisation de *partimenti* : des scénarios avec basse et chiffrage sur lesquels les musiciens pouvaient improviser [Sanguinetti, 2012].

Le jazz est probablement le style de musique auquel l'improvisation est la plus associée. De nombreuses études se sont concentrées sur l'analyse musicale de ce style sur des regards mélodiques, harmoniques, rythmiques, etc. [Bailey, 1980; Levine, 1995]. Parmi ses spécificités, on peut, par exemple, parler des *blue notes* issues de la pratique du blues, l'utilisation de progressions harmoniques typiques comme l'*anatole*, des méthodes de substitution ou de coloration d'accords, etc. De plus, la pratique du jazz est souvent basée sur des standards, définissant un thème et une grille harmonique sur laquelle se basent les improvisations. La Figure 2.3 montre un extrait du standard *Giant Steps*

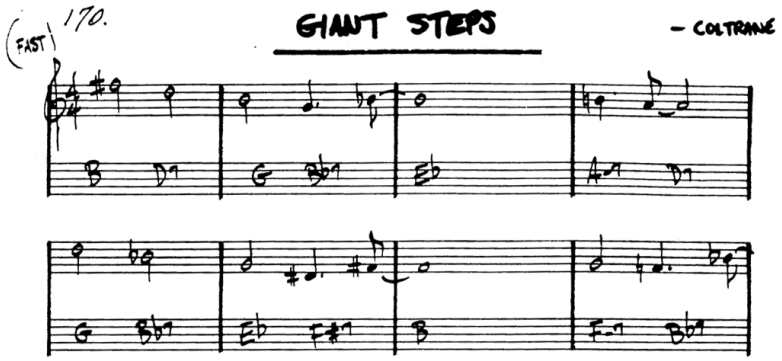

FIGURE 2.3 – Extrait du standard de jazz *Giant Steps* de John Coltrane. Partition issue du Real Book (5ème édition).

de John Coltrane, célèbre pour son organisation autour de *Coltrane Changes*, une progression d'accords basée sur un cycle de tierces majeures, courante dans la musique de Coltrane. Mais, encore une fois, la pratique de l'improvisation va dépendre du lieu, du style et de l'époque, les techniques varient entre le blues, le jazz tonal, le jazz modal, le *free jazz*, jusqu'à l'improvisation libre. La production musicale est « inculturée dans sa texture » : mélodies, harmonies, rythmes et autres éléments sont organisés comme une production culturelle.

Il ne s'agit ici que de certains exemples de la présence de l'improvisation dans différents styles musicaux issus de la culture occidentale et afro-américaine de la musique. L'improvisation est bien évidemment présente dans de nombreux autres contextes, comme par exemple l'improvisation collective modale dans la musique celtique, la musique arabe avec l'utilisation de maqams, la musique indienne avec l'utilisation de ragas, etc.

Motivations : Nous avons vu que la construction d'un discours musical dans le cadre de l'improvisation est complexe et s'organise autour d'un ensemble de dimensions musicales et de leurs combinaisons. Nous souhaitons que nos modèles génératifs soient capables de considérer l'ensemble de ces dimensions pour la création de nouvelles improvisations. Cela demande de modéliser des relations horizontales au sein d'une même dimension et des relations verticales entre plusieurs dimensions différentes. Ces modèles devront être des représentations efficaces, la prise en considération de toutes les combinaisons possibles étant d'une dimensionalité trop élevée. De plus, nous avons également vu que la pratique de l'improvisation est dépendante du style et de connaissances stylistiques intériorisées. Nous souhaitons que les modèles théoriques développés restent agnostiques pour pouvoir s'adapter aux différents styles. Nous ne voulons pas, par exemple, définir de règles strictes d'harmonie. Cependant, les experiences que nous présentons par la suite se basent sur un corpus du style *bebop*. Nous voulons également développer des modèles permettant des représentations influencées par l'intériorisation de connaissances stylistiques multidimensionnelles au sein d'un agent musical.

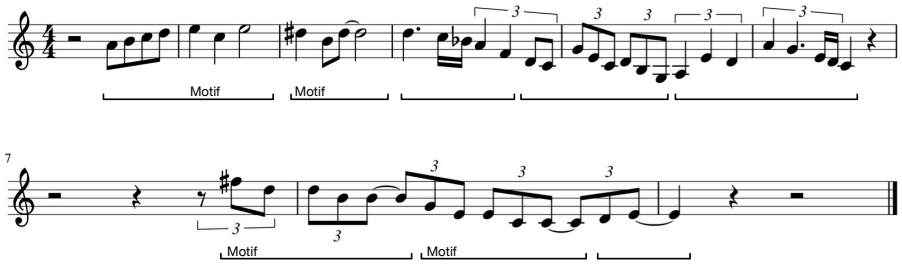

FIGURE 2.4 – Début de l'improvisation de Walt Fowler sur *The Torture Never Stops* de Frank Zappa, concert à Skedsmohallen 27/04/1988 (transcription K. Déguernel). Chaque portée représente une phrase mélodique. Les crochets représentent le découpage des phrases en gestes rythmiques.

2.1.3 Stuctures temporelles et niveaux de récit dans l'improvisation

Siron [2015] propose une organisation temporelle du discours musical mélodique en quatre *niveaux temporels* de récits emboîtés. Nous mettons ici en parallèle ces niveaux temporels avec les différents niveaux métriques.

- Le niveau de récit le plus bas est celui du *point sonore*. On considère ici une note, ou un son isolé. Ceci correspond au niveau métrique de la division du temps, au niveau du *tatum*.
- Le niveau suivant est celui du *geste rythmique*. On considère ici un motif mélodique possédant une dynamique propre. Ceci correspond généralement aux niveaux métriques variant du temps à la mesure.
- Le niveau suivant est celui de la *phrase mélodique*. On considère ici un ensemble d'événement formant une continuité poussée par une intention globale leur donnant un sens d'unité. Ceci correspond généralement aux niveaux métriques du groupe de mesures ou de la carrure.
- Le niveau de récit le plus haut est celui de *récit*. On considère ici un ensemble de phrases regroupées dans une « dramaturgie générale ». Ceci correspond au niveau métrique de la forme.

La Figure 2.4 montre un exemple d'organisation temporelle d'une improvisation. On illustre ici l'organisation du niveau du point sonore jusqu'au niveau de la phrase mélodique. Les points sonores sont les notes. Les gestes rythmiques sont identifiés par les changements de dynamique au sein des phrases et par l'utilisation de motifs développés.

Cette structuration du temps musical a également été observée dans les musiques traditionnelles d'Afrique Centrale dans Arom [1987] avec une organisation du flux rythmique basée sur une *pulsation*, décrite comme « une suite ininterrompue de points de repère », autour de laquelle se forme des *figures rythmiques* avec des organisation en cycle. Dans [Schott, 2000], John Schott effectue des analyses précises de la musique de John Coltrane dans lesquelles émergent également ces structures, y compris dans ses albums de *free jazz* comme *A Love Supreme*, *Ascension* ou *Meditations*. La Figure 2.5 présente un

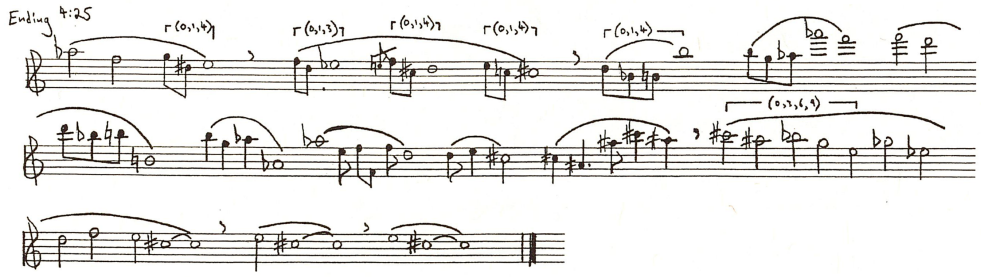

FIGURE 2.5 – Extrait de *Consequences* de John Coltrane de l'album *Meditation* (transcrit par John Scott) [Schott, 2000].

extrait de *Consequences* dans lequel on retrouve l'organisation de l'improvisation en plusieurs niveaux de récit.

Il est important de noter que différentes dynamiques de l'improvisation s'appliquent aux différents niveaux de récit formant des structures temporelles complexes et entrelacées, permettant une cohérence du discours musical. Par exemple, l'intonation va avoir un impact sur le point sonore, mais va ensuite permettre l'expression de tension et de détente dans le geste rythmique. De même l'organisation en tension/détente (sur les hauteurs, durées et intensités) des gestes rythmiques vont jouer un rôle important dans la constitution de phrase mélodique. Notons également que, bien que les niveaux de récit se basent sur les structures souvent strictes des niveaux métriques, un improvisateur se détache régulièrement des carrures pour se permettre plus de liberté.

Motivations : Nous voulons concevoir un modèle permettant la prise en considération des différents niveaux de récit dans la génération de l'improvisation, la construction de l'improvisation devant alors prendre en considération des dépendances à la fois à court et à long terme. Ceci nécessite une analyse des improvisations dans un corpus d'apprentissage. Nous nous consacrons au cas de l'improvisation idiomatique guidée décrite dans Nika [2016] basée sur un scénario temporel. Dans un premier temps, nous voulons utiliser des connaissances données sur les différents niveaux métriques du scénario pour approximer les différents niveaux de récit et utiliser cette information pour le guidage de l'improvisation. Dans un second temps, nous voulons réaliser une analyse automatique du corpus pour en extraire des scénarios et leurs différents niveaux métriques. L'objectif est de permettre au modèle génératif de considérer la forme globale d'un scénario et de construire des improvisations sur celui-ci sur plusieurs niveaux temporels, permettant une meilleure cohérence globale.

2.1.4 Improvisation et interaction

Une des choses distinguant l'improvisation musicale du langage naturel est le fait qu'il soit tout à fait possible de construire un discours à plusieurs avec une simultanéité des voix. L'interaction improvisée au sein d'un groupe est l'attrait principal déclaré par les musiciens des musiques improvisées, qu'il

s'agisse de la construction collective d'une même improvisation ou d'un échange entre soliste et accompagnateurs. Siron [2015] décrit l'évolution des relations entre l'individu et le groupe à travers l'histoire du jazz : le jazz New Orleans, proche de la fanfare où « les différentes voix se croisent dans une polyphonie où s'expriment ensemble tous les musiciens », puis le *bebop* où le thème est joué à l'unisson suivi d'une succession de solos sur une grille harmonique répétée, puis à partir des années 60 une plus grande expressivité de la section rythmique avec un jeu plus interactif, sans oublier le *free jazz* où « tous les improvisateurs contribuent à l'énergie collective », et où les rôles instrumentaux traditionnels sont remis en question.

L'interaction dans le cadre du jazz est régulièrement analysée par les musicologues comme une conversation. C'est une conversation où tout le monde parle en même temps, mais une conversation tout de même [Berliner, 1994]. Ralph Peterson Jr. dit [Monson, 1996] :

> « Vous voyez, ce qu'il se passe c'est que souvent, quand vous rentrez dans une conversation musicale, un membre du groupe va proposer une idée ou le début d'une idée et une autre personne va compléter cette idée ou son interprétation de cette même idée ; comment elle l'entend. Donc la conversation se fait par fragments et vient de différentes composantes, de différentes voix. »

Monson [1996] exprime en particulier comment la conversation s'organise localement comme une série de « prochains mouvements ». C'est par leur création du « prochain » que les musiciens vont exprimer leur compréhension de la conversation en cours [Bowers, 2002]. De nombreux exemples d'interactivité entre musiciens dans le jazz sont présentés dans [Monson, 1996] : de l'essence du *groove* et de la coordination rythmique entre musiciens (en particulier entre le bassiste et le batteur), de l'art de l'accompagnement de solos pour un pianiste, de l'organisation de questions/réponses entre musiciens, des échos mélodiques issus d'un solo répétés à la basse, etc. Les principes d'interaction se retrouvent également dans [Sudnow, 1978] qui décrit l'apprentissage du piano jazz. Ici, l'interaction ne se concentre pas sur le jeu entre musiciens, mais plutôt sur comment un pianiste construit son jeu comme un discours entre plusieurs voix jouées par le même instrumentiste.

Motivations : Nous voulons concevoir un modèle inspiré par l'interaction entre musiciens ou dans le cas décrit dans [Sudnow, 1978] l'interaction entre plusieurs voix. Une conception multi-agents semble appropriée pour pouvoir représenter une scène où plusieurs musiciens possèdent des connaissances différentes et agissent sur des instruments ou dimensions différents. Ceci permet également d'ouvrir ce modèle en intégrant un agent humain dans le modèle. La communication entre ces agents et une description des relations existant entre musiciens/voix sont nécessaires pour pouvoir représenter les principes conversationnels adéquats au système musical. Ceci permettrait au modèle génératif d'improviser sur plusieurs dimensions simultanément tout en assurant une cohérence globale et d'essayer de modéliser une dynamique de groupe.

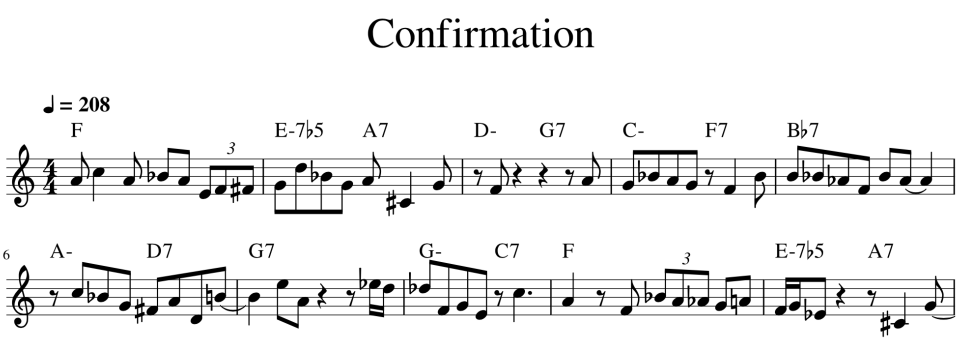

FIGURE 2.6 – Extrait du thème *Confirmation* de Charlie Parker issu de l'*Omnibook*.

2.2 Constitution d'une base de données

Les différentes motivations présentées dans ce chapitre nous ont poussé à constituer une base de données basée sur de la musique improvisée idiomatique, et contenant des informations musicales multi-dimensionnelles et multi-échelles. Cela nous permet d'effectuer des expériences sur les différents systèmes que nous avons développé. Cette base de données est basée sur l'*Omnibook* de Charlie Parker [Parker & Aebersold, 1978]. Elle est constituée de 50 thèmes et improvisations joués par Charlie Parker. L'avantage de cette base de donnée est qu'elle constitue un corpus d'apprentissage cohérent et spécifique au style *bebop* possédant des caractéristiques facilement identifiables par les musiciens. Ce corpus contient les informations mélodiques, rythmiques et harmoniques, permettant un traitement de données multidimensionnelles. Les informations harmoniques sont basées sur les accompagnements de piano et de contrebasse issus des enregistrements permettant de prendre en compte les différences entre chaque grille d'accords ainsi qu'une représentation simplifiée des échanges entre soliste et accompagnateur. L'ensemble des improvisations sont basées sur des grilles structurées, et le corpus contient des progressions d'accords célèbres du jazz comme le *blues* ou l'*anatole*. Les accords relevés sont divisés en cinq classes d'accords basées sur leur fondamental `X` : les accords majeurs (notés `X`), les accords mineurs (notés `X-`), les accords de dominante (notés $\texttt{X}^7$), les accords demi-diminués (notés $\texttt{X-}^7\flat5$) et les accords diminués (notés `Xo`). La Figure 2.6 montre un exemple issu de notre base de données. On peut y voir les informations sur les différentes dimensions.

L'*Omnibook* a été retranscrit à la main à l'aide de l'éditeur de partitions MuseScore. Les données ont ensuite été transformées en différents formats à partir de ce logiciel. Nous avons distribué cette base de données librement. Elle est disponible à partir du site du projet DYCI2 : `http://repmus.ircam.fr/dyci2/ressources`.

La base de données contient :
— un fichier PDF recueillant l'ensemble des 50 thèmes et improvisations,
— les fichiers MuseScore individuels pour chaque thème et improvisation,

— les fichiers Midi (sans information harmonique) individuels pour chaque thème et improvisation,
— les fichiers MusicXml individuels pour chaque thème et improvisation,
— un fichier de licence pour l'utilisation des données (licence de libre diffusion CC-BY-NC-SA 2.0 [4]),
— un script Python commenté permettant une traduction des fichiers MusicXml en données Python.

4. Creative Commons Attribution [BY] — Pas d'utilisation commerciale [NC] — Partage des conditions initiales à l'identique [SA] 2.0 `https://creativecommons.org/licenses/by-nc-sa/2.0/fr/`

3

État de l'art

"I'm very much an observer and a conduit of thoughts and ideas."

– Ian Anderson

Ce chapitre étudie les méthodes existantes d'improvisation automatique et d'apprentissage de structures musicales, afin de positionner les contributions de cette thèse, présentées dans les chapitres suivants dans leur contexte.

Dans un premier temps, nous présentons les méthodes d'improvisation automatique existantes. Dans la partie 3.1.1, nous résumons l'historique des recherches sur l'improvisation automatique à travers les travaux de modélisation du style. Dans la partie 3.1.2, nous présentons le paradigme d'improvisation automatique basé sur le système *OMax* et les différents projets gravitant autour de ce système. Dans la partie 3.1.3, nous décrivons la notion de guidage de l'improvisation, ses différents sens et les différentes méthodes de guidage. Dans la partie 3.1.4, nous présentons des systèmes interactifs d'improvisation automatique basés sur la communication entre plusieurs agents et leur adaptation dans des improvisations collectives. Enfin, la partie 3.1.5 propose une discussion résumant les limitations de ces systèmes.

Dans un deuxième temps, nous présentons les méthodes d'apprentissage automatique de structures musicales. Dans la partie 3.2.1, nous nous concentrons sur l'apprentissage de structures multidimensionnelles ; nous présentons des méthodes basées sur des modèles à points de vue multiples puis sur l'interpolation de sous-modèles ; nous étudions ensuite les méthodes de lissage et d'évaluation de tels modèles ; puis nous exposons quelques méthodes basées sur l'utilisation de réseaux de neurones. Dans la partie 3.2.2, nous nous intéressons à l'apprentissage de structures temporelles ; nous présentons d'abord des méthodes basées sur l'analyse musicale issues du domaine de la recherche d'information musicale ; puis nous présentons des modèles basés sur l'utilisation de grammaires formelles pour l'extraction de structures hiérarchiques. Finalement, la partie 3.2.3 résume les limitations de ces systèmes et les objectifs de nos travaux.

3.1 Improvisation automatique

3.1.1 Modéliser le style : composition, prédiction et improvisation automatique

La recherche sur l'improvisation automatique est issue des travaux sur la modélisation du style pour les systèmes de génération musicale. Le but de ces systèmes est de générer des séquences cohérentes en réutilisant des matériaux musicaux qu'ils ont appris. Il ne s'agit pas ici de construire un ensemble de règles ou une grammaire pour décrire des procédés compositionels [Lerdahl & Jackendoff, 1983; Chemillier, 2001] mais de construire une mémoire à partir d'un corpus hors-ligne ou en direct à partir du jeu d'un musicien puis d'utiliser une analyse de cette mémoire afin de trouver des chemins logiques dans ce corpus permettant de reformuler le matériau musical. Le processus génératif consiste alors à rejouer des séquences existantes dans la mémoire et à diverger du matériau originel en combinant des séquences qui n'étaient originellement pas les unes à la suite des autres dans la mémoire, afin de créer de nouvelles phrases musicales conservant la cohérence du style. L'arrangement des séquences est souvent basé sur des modèles de type markoviens afin de garantir la logique et la cohérence des séquences générées lors de la combinaison de différents éléments du matériau originel.

Les premières expériences de modélisation du style à partir d'un corpus correspondent aux premières tentatives de composition assistée par ordinateur dans les années 50 [Ames, 1989; Assayag, 1998]. En 1955, Fred et Carolyn Attneave utilisent des chaînes de Markov du premier ordre pour analyser et générer des mélodies de style Western décrites comme « convaincantes » [Ariza, 2005]. De même, Pinkerton [1956] utilise des chaînes de Markov du premier ordre pour générer des mélodies suite à l'analyse de 39 chansons pour enfants. La modélisation du style basée sur des méthodes probabilistes s'est poursuivie ensuite pour des tâches de prédiction et d'analyse avec les travaux de Conklin et al. Ces travaux utilisent des modèles de Markov d'ordre allant de 1 à 5 [Conklin & Cleary, 1988] puis des méthodes d'apprentissage automatique basées sur des n-grammes [Conklin & Witten, 1995; Conklin, 2003] pour prédire (à l'aide de mesures d'entropie) et pour générer des mélodies dans le style de la musique grégorienne ou dans le style de Bach. Les modèles de Markov sont également utilisé par Hall & Smith [1996] pour la génération de mélodies de blues. Par extension, des modèles de génération basés sur des réseaux de neurones ont été développés. Mozer [1994] utilise un réseau de neurones récurrent appris sur des mélodies de Bach et des mélodies traditionnelles européennes pour en extraire les régularités stylistiques et générer de nouvelles mélodies. Plus récemment, Bickerman et al. [2010] emploient des réseaux de neurones profonds afin de générer des mélodies de jazz. Le projet *WaveNet* [Van Den Oord et al., 2016] cherche à générer de l'audio en apprenant directement sur les formes d'ondes en utilisant un réseau de neurones convolutif. Cette méthode est également utilisée par *Google Brain* dans le projet *Audio DeepDream* [Ardila et al., 2016].

La première mention de l'utilisation de la modélisation du style pour l'improvisation est effectuée par Assayag et al. [1999]. Ils proposent de construire une mémoire musicale en utilisant un algorithme basé sur les travaux de Dub-

nov et al. [1998] (ensuite étendus par Dubnov et al. [2003]) pour la prédiction de mélodies. Cette méthode repose sur l'algorithme de Lempel-Ziv [Ziv & Lempel, 1978] permettant une prise en compte de contexte, similaire à une chaîne de Markov d'ordre variable. Une fois entraînée, la mémoire peut-être utilisée dans un système en temps-réel pour improviser une mélodie en contrepoint du jeu d'un musicien. Le *Continuator* [Pachet, 2002] utilise des méthodes similaires pour improviser automatiquement la suite d'une séquence musicale donnée. La séquence d'entrée est utilisée pour construire un arbre de préfixes qui est ensuite utilisé pour générer une continuation logique de la séquence. Ce système est alors capable d'apprendre et de générer des séquences musicales, en temps réel, sans connaissance *a priori* d'un style, à partir du jeu d'un musicien fourni en direct.

Un nouveau paradigme de modélisation du style a ensuite été créé avec le logiciel *OMax* [Assayag & Dubnov, 2004]. La modélisation du style s'effectue avec l'utilisation d'un automate fini déterministe appelé oracle des facteurs [Allauzen et al., 1999]. Cette structure s'inscrit au cœur des méthodes proposées dans cette thèse. Nous la présentons alors en détail dans la partie suivante. D'autres systèmes d'improvisation automatique plus récents sont ensuite présentés dans la partie 3.1.3.

3.1.2 Le paradigme *OMax* et l'oracle des facteurs

Dans cette partie, nous présentons tout d'abord l'oracle des facteurs, la structure servant de modèle de mémoire au système, puis nous présentons comment cette structure est utilisée dans *OMax*. Finalement, nous présentons différents développements et projets qui se sont créés autour de ce paradigme.

L'oracle des facteurs

L'oracle des facteurs est une structure issue de la théorie des langages formels et de la bioinformatique, créé par Crochemore et al. Il a initialement été conçu pour répondre au problème de construire un automate fini déterministe capable de reconnaitre le langage de l'ensemble des facteurs d'un mot [Crochemore & Rytter, 1994] pour des tâches de recherche de motifs. Étant donné un mot w, c'est-à-dire une séquence finie $w = w_1 w_2 ... w_m$ de lettres d'un alphabet Σ, le mot $x \in \Sigma^*$ est un facteur de w si et seulement si $w = uxv$, avec u et $v \in \Sigma^*$, où * est l'étoile de Kleene, c'est-à-dire, Σ^* est l'ensemble des séquences finies d'élements de Σ. D'autres structures sont utilisées pour cette tâche telles que les arbres des suffixes, les tableaux des suffixes [Crochemore et al., 2007a], les automates des facteurs ou encore les graphes compacts orientés acycliques de mots [Crochemore & Vérin, 1997] mais celles-ci souffrent globalement d'une complexité en mémoire importante et ne reconnaissent pas nécessairement l'ensemble des facteurs d'un mot. L'oracle des facteurs construit sur un mot w, permet de reconnaitre au moins l'ensemble des facteurs de w (il est parfois possible de reconnaitre des mots x qui ne sont pas facteurs de w et donc d'obtenir des faux positifs). Nous ne détaillons pas ici l'algorithme de construction de l'oracle des facteurs qui est présenté dans [Allauzen et al., 1999], nous nous intéressons ici plutôt à ses propriétés. Il s'agit d'un automate possédant exacte-

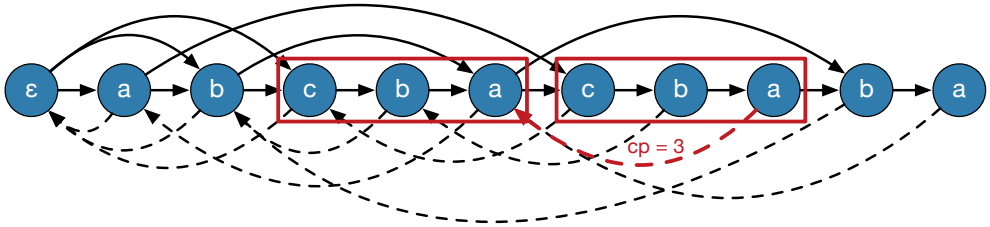

FIGURE 3.1 – Oracle des facteurs construit sur le mot $w = abcbacbaba$. Les transitions de l'automate sont représentées par les flèches continues. Les liens suffixiels, représentés par les flèches en pointillés, relient chaque état avec l'état passé le plus à gauche partageant le contexte commun le plus long.

ment $m+1$ états, un nombre linéaire de transitions et dont la construction est linéaire en temps et peut se faire de façon incrémentale. Lors de la construction de l'oracle des facteurs, des liens de construction appelés *liens suffixiels* sont utilisés pour relier chaque état w_i à l'état le plus à gauche où le plus long suffixe de $w_1...w_i$ est reconnu. La longueur du plus long suffixe reconnu pour chaque état peut être calculée pendant la construction de l'oracle en utilisant les méthodes proposées par Lefebvre & Lecroq [2000]. La Figure 3.1 montre un exemple d'oracle des facteurs construit sur le mot $w = abcbacbaba$, les flèches en trait continu représentent les transitions de l'oracle et les flèches en pointillés représentent les liens suffixiels. On peut remarquer, par exemple, que le lien suffixiel représenté en rouge relie des états partageant un suffixe commun encadré en rouge ; *cp* représente la taille de ce suffixe commun.

Modélisation et génération dans *OMax*

OMax est un logiciel d'improvisation automatique permettant d'effectuer des performances en temps réel [Assayag et al., 2006a] entre un musicien humain et une machine. Il se base sur un apprentissage continu du jeu du musicien en utilisant la construction incrémentale de l'oracle des facteurs. La séquence musicale d'entrée est découpée en éléments discrets, puis chaque élément est étiqueté suivant une dimension musicale. *OMax* propose, par exemple, un étiquetage des éléments par la hauteur, estimée à partir du signal audio par l'algorithme YIN [De Cheveigné & Kawahara, 2002], ou par le timbre, estimé par regroupement des coefficients cepstraux en échelle de mels (MFCC).

Dans *OMax*, l'oracle des facteurs n'est pas utilisé pour la recherche de motifs, mais comme modèle de mémoire exploitant les propriétés de cet automate. Premièrement, l'oracle des facteurs permet de conserver l'aspect séquentiel et l'organisation temporelle de la mémoire, contrairement, par exemple, aux chaînes de Markov qui regrouperaient les états partageant les mêmes étiquettes. En effet, un parcours horizontal de l'oracle des facteurs permet une retranscription exacte de la séquence d'origine. Deuxièmement, les liens suffixiels sont utilisés comme représentation des régularités dans la mémoire. Ils permettent de connecter des états de l'oracle partageant le même passé musical [Assayag & Dubnov, 2004]. Ils sont équivalents à un modèle de Markov d'ordre optimal, arbitraire à chaque état. La Figure 3.2 montre une visualisation de l'analyse

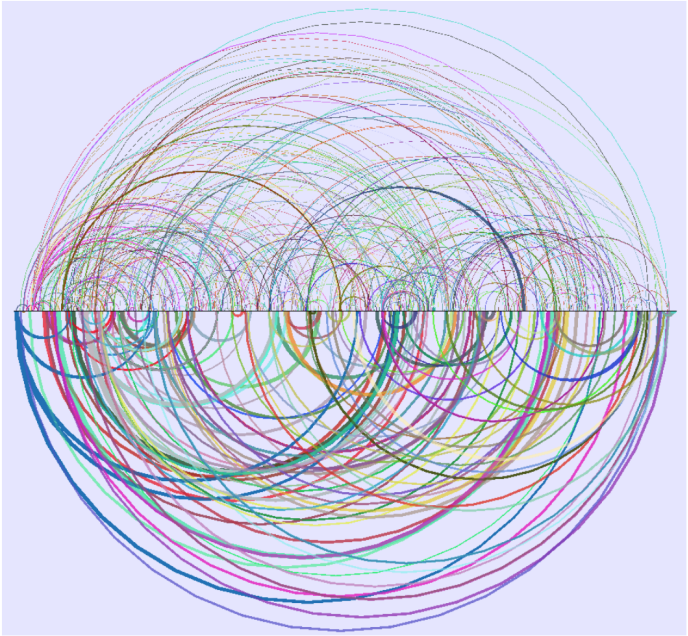

FIGURE 3.2 – Représentation de la mémoire dans *OMax* à l'aide des liens suffixes de l'oracle des facteurs. Deux oracles sont construits simultanément à partir d'un corpus musical : à partir de la séquence de hauteurs (en haut) ou à partir des MFCCs (en bas). L'axe horizontal représente l'axe temporel de la mémoire [Lévy et al., 2012].

de la mémoire fournie par les oracles des facteurs construits sur la hauteur (en haut) et sur les MFCCs (en bas) sur un exemple [Lévy et al., 2012]. Les arcs représentent les liens suffixiels des oracles.

Dans *OMax*, générer une improvisation s'effectue en parcourant cette mémoire musicale. L'idée principale est d'utiliser les liens suffixiels pour effectuer des parcours non-linéaires dans la mémoire, sautant d'un état à un autre partageant un passé musical similaire. Le postulat de base est que de tels parcours permettent la création de nouvelles phrases musicales, divergeant de la séquence originale, mais préservant une continuité logique du discours musical et conservant le style musical du corpus. Assayag & Bloch [2007] décrivent un ensemble d'heuristiques pour la navigation dans l'oracle des facteurs afin de générer des improvisations plus réalistes. On peut notamment citer l'utilisation d'un facteur de continuité pour éviter un discours musical trop décousu dû à des sauts dans la mémoire trop nombreux, ou l'utilisation d'une liste tabou pour éviter les boucles, etc. Les possibilités d'apprentissage et de génération temps-réel et simultanées d'*OMax* ont permis d'effectuer les premières expériences d'interaction musicale improvisée homme-machine. Les agents (humains et machine) étant directement et simultanément exposés à l'improvisation des autres agents, il émerge une nouvelle dynamique musicale analysée dans [Assayag et al., 2006b] par une boucle de rétroaction appelée *réinjection stylistique*.

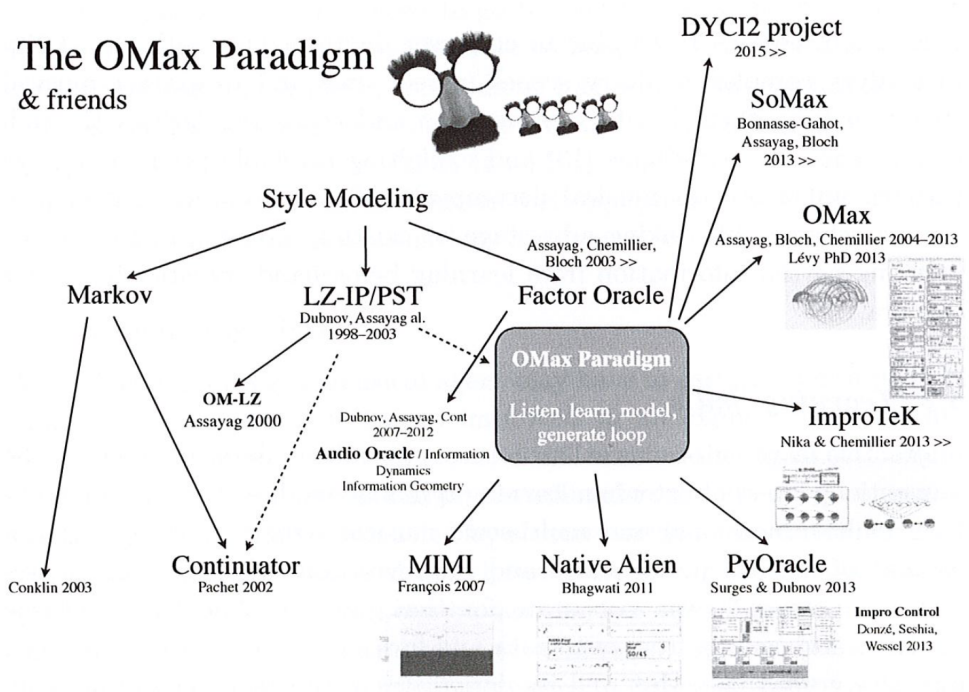

FIGURE 3.3 – La galaxie OMax représente l'ensemble des projets prennant place autour du paradigme OMax [Assayag, 2016].

La galaxie *OMax*

De nombreux projets et développements ont été créés autour d'*OMax* afin d'étendre les possibilités d'expressivité musicale ou de répondre à certains besoins issus de contextes d'improvisation particuliers. La Figure 3.3 représente une cartographie des différentes recherches ayant amené à la création d'*OMax* et de celles basées sur ce paradigme.

Une première piste de recherche consiste à ajouter des contrôles pour l'improvisateur humain ou pour un utilisateur actif durant la performance musicale. Par exemple, la visualisation de la modélisation de la mémoire dans *OMax* dont un exemple est montré dans la Figure 3.2, permet de localiser visuellement les points d'intérêts musicaux développés au cours de l'improvisation. Une interface a alors été développée pour permettre, par exemple, à un utilisateur de sélectionner des zones de la mémoire afin de diriger la navigation dans la mémoire ou de contraindre la génération à certaines sections. L'installation *Mimi4x* [François et al., 2010; Schankler et al., 2011] fournit une visualisation de quatre corpora musicaux sous forme de *piano roll* préparés à l'avance par un compositeur et sur lesquels des oracles des facteurs sont construits pour générer des improvisations. Une interface avec un contrôleur MIDI permet à l'utilisateur de construire une improvisation en choisissant les oracles qui doivent jouer, quand ils doivent jouer et en modifiant des paramètres de génération tels que le volume, la taille du facteur de continuité pour chaque oracle, etc. Maniatakos et al. [2010] proposent une architecture modulaire d'improvisation assistée par

ordinateur afin de formaliser à l'aide d'un graphe, le procédé de réinjection stylistique entre un improvisateur, une machine et un utilisateur actif donnant des instructions à la machine.

Une autre piste est celle de *PyOracle* [Surges & Dubnov, 2013] qui utilise un oracle des facteurs, appelé *Oracle Audio* dont l'alphabet est construit directement à partir de descripteurs audio définis par l'utilisateur [Dubnov et al., 2007]. La complexité et les répétitions dans le signal musical est mesurée pour sélectionner les descripteurs audio et évaluer l'*Oracle Audio*. Wang & Dubnov [2014] étendent ce modèle avec l'identification des grappes de trames audios lors de la construction de l'*Oracle Audio*, formant une nouvelle structure appelée *Variable Markov Oracle*. *PyOracle* est également utilisé en combinaison avec *CataRT* [Schwarz, 2007] par Einbond dans *CatOracle* [Einbond et al., 2016]. *CataRT* est un système permettant la construction d'une base de données de sons préenregistrés ou capturés en direct puis analysés à l'aide de descripteurs audio ; cette base de données est ensuite utilisée pour effectuer de la synthèse concaténative.

D'autres pistes de recherche, qui seront étudiées plus en détail dans les parties suivantes, incluent la modélisation de principes cognitifs de mémoire basés sur la rémanence auditive dans le logiciel d'accompagnement automatique SoMax [Bonasse-Gahot, 2014] (cf. *Guidage réactif*, partie 3.1.3) ; l'utilisation d'un scénario temporel connu *a priori* pour le guidage de l'improvisation dans un cadre idiomatique dans le logiciel ImproteK [Nika et al., 2017a] (cf. *Guidage structurel*, partie 3.1.3) ; ou encore l'utilisation de modèles d'adaptation temporelle des interactions dans une situation d'improvisation inspirés par la recherche sur la régulation de conversation pour des agents conversationnels [Sanlaville et al., 2015] (cf. partie 3.1.4).

3.1.3 Guidage de l'improvisation

La notion de guidage est apparue pour décrire deux stratégies distinctes de navigation d'une mémoire musicale. D'un côté, guider l'improvisation peut signifier un processus de réaction à un environnement effectué pas à pas. D'un autre côté, le guidage peut décrire l'utilisation de structures temporelles et de contraintes à long terme. Ces différentes notions *a priori* orthogonales ont pour objectif commun de permettre aux modèles génératifs une meilleure adaptation à un style voulu, que ce soit par une écoute active de l'environnement ou par des connaissances *a priori* sur celui-ci. Des travaux récents cherchent à combiner ses différentes méthodes de guidages Nika et al. [2017b].

Guidage réactif

La notion de *guidage réactif* se réfère à des processus où les prises de décision du modèle de génération se basent sur une écoute active de l'environnement en temps réel. L'objectif est de créer un système capable de s'adapter à des événements musicaux locaux et des changements de contexte musical au cours d'une improvisation afin d'obtenir une forte cohérence locale suivant le jeu de l'improvisateur humain de manière instantanée. Ces modèles sont particulièrement adaptés dans le cadre d'improvisations libres ou génératives.

Le système *Voyager* [Lewis, 2000b], développé par le compositeur et improvisateur George E. Lewis, analyse en direct le jeu d'un ou deux improvisateurs humains. Cette analyse est alors utilisée pour enclencher et guider des processus génératifs prédéfinis. Ce système est constitué d'un « orchestre improvisateur virtuel » avec 64 voix asynchrones, jouant en temps réel, possédant des comportements différents face aux jeux des improvisateurs (allant de la recherche d'une « union totale » à « l'indifférence absolue »). Les différentes voix peuvent avoir des comportements de groupe, ou s'activer simultanément sans synchronicité rythmique. *Voyager* a pour objectif de refléter les pratiques musicales et l'esthétique de la musique afro-américaine.

Le système *VirtualBand* [Moreira et al., 2013] propose d'extraire, en direct à partir du jeu d'un musicien, des descripteurs audio pour guider la génération d'agents virtuels. Les agents virtuels représentent les styles de différents musiciens organisés en une base de données audio constituée de tranches (mesures ou temps) stylistiquement représentatives. Les différentes réactions des agents sont modélisées par des liens entre agents pères (qui peuvent être un agent humain ou virtuel) et agents fils (agent virtuel) suivant un descripteur audio (chromagramme, centroïde spectral, énergie moyenne, etc.). L'agent fils réagit à l'information donnée par l'agent père, leur connexion spécifiant quelle tranche l'agent fils doit jouer à partir d'associations précédemment observées dans la mémoire.

Des systèmes d'accompagnement automatique se basent sur des principes similaires. Par exemple, le *Jambot* [Gifford & Brown, 2011] extrait des descripteurs audio du jeu de l'improvisateur pour fournir un accompagnement de percussions. La génération combine des comportements réactifs, basés sur l'imitation et des comportements proactifs, basés sur une analyse du contexte rythmique ; le passage d'un comportement à l'autre s'effectue par une mesure de confiance par rapport à la compréhension du contexte musical par le *Jambot*. De manière similaire, *Reflexive Looper* [Pachet et al., 2013] cherche à enrichir les possibilités des *loopers* (pédale d'effet permettant d'enregistrer une boucle et de la répéter *ad libitum*) en ajoutant un aspect réactif par rapport au jeu du musicien. *Reflexive Looper* effectue une classification de ce qui est en train d'être joué par le musicien suivant différents modes de jeu : ligne de basse, accords, mélodie. Le classifieur est entrainé de manière supervisée en amont sur un corpus à l'aide d'un séparateur à vaste marge (SVM) pour reconnaitre ces différents modes. Au lieu de rejouer à l'identique ce qui a été enregistré, l'accompagnement s'adapte au mode joué en temps réel par le musicien. Par exemple, si le musicien joue une mélodie, l'accompagnement automatique jouera les accords et la ligne de basse.

Le logiciel d'accompagnement automatique *SoMax* [Bonasse-Gahot, 2014] combine guidage réactif et modèle de mémoire en guidant sa génération à l'aide d'un *profil d'activité*. Il s'agit d'une fonction de score représentant la pertinence de chaque état dans la mémoire par rapport à l'environnement musical capturé en temps réel. Ce profil d'activité est calculé à partir de l'analyse du contexte sur différentes dimensions musicales. La Figure 3.4 illustre ce procédé. L'écoute active extrait des informations de contexte sur plusieurs dimensions musicales qui vont activer différentes zones de la mémoire. Le profil d'activité global est alors calculé en effectuant la somme pondérée de l'activité sur les

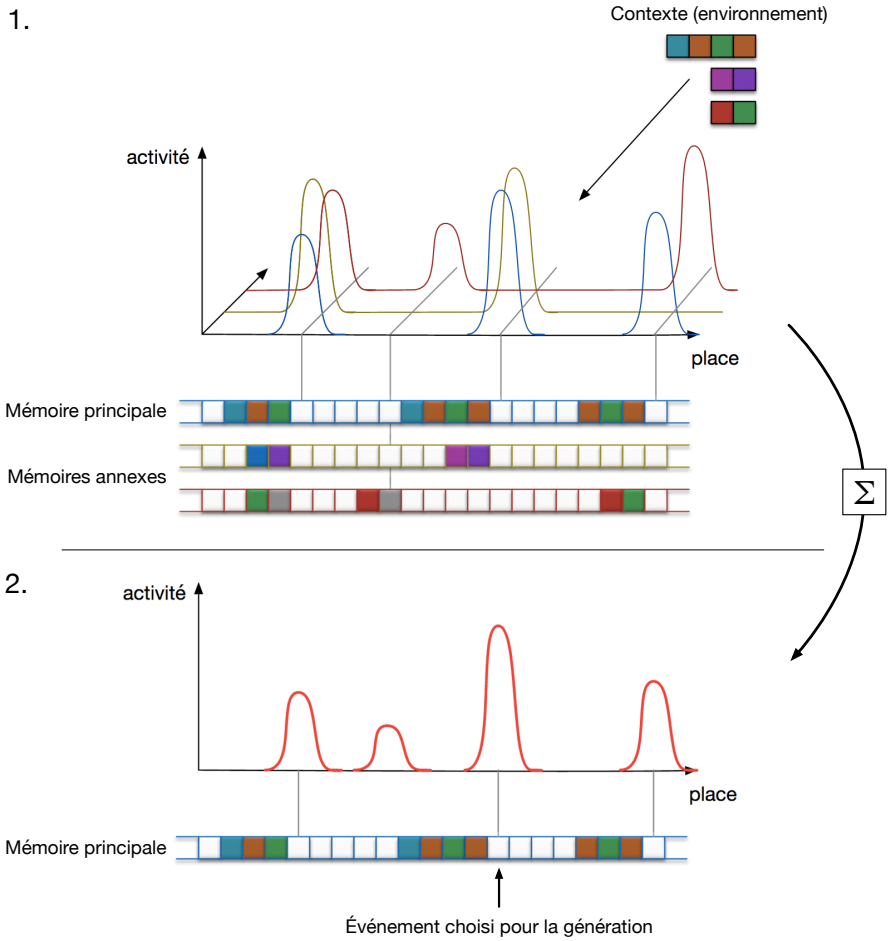

FIGURE 3.4 – Guidage réactif dans *SoMax* par écoute active du jeu d'un improvisateur humain. La mémoire du système est segmentée et représentée chronologiquement ici en plusieurs places. Cette mémoire est constituée d'une mémoire principale correspondant à ce que va jouer le logiciel et de mémoires annexes permettant de réagir à l'environnement sur plusieurs dimensions. 1. Le contexte active des zones de la mémoire selon différentes dimensions musicales. 2. Le profil d'activité global est calculé à partir des différentes dimensions. Les zones de la mémoire avec de fortes activités sont privilégiées lors de la génération [Chemla-Romeu-Santos, 2015].

différentes dimensions. Les poids de cette somme sont définis par l'utilisateur. Ce profil d'activité évolue dans le temps et se propage dans la mémoire en décroissant afin de modéliser le processus cognitif de rémanence auditive. De cette manière, l'accompagnement automatique va favoriser les événements de la mémoire partageant un contexte commun à l'improvisation en cours tout en suivant ses évolutions en privilégiant les zones de la mémoire à forte activité. Le processus de génération est également guidé par la logique interne de la mémoire musicale pour garantir une continuité cohérente de l'accompagnement. Cela est effectué en réinjectant la sortie du système comme événement

de l'environnement permettant au système d'avoir un processus d'auto-écoute.

Guidage structurel

La notion de *guidage structurel* se réfère à des processus dirigeant le modèle génératif par la connaissance *a priori* d'une structure temporelle ou séquentielle devant être suivie lors du processus de génération. L'objectif est d'être capable d'adapter la génération musicale à un ensemble de contraintes locales ou de spécifications temporelles à long terme représentatives d'un style musical, prédéfinies par l'utilisateur.

Certaines méthodes guident l'improvisation en privilégiant ou en forçant l'utilisation de certaines transitions à certains moments afin de respecter des règles prédéfinies pour respecter un style musical. Pachet & Roy [2011] utilisent des contraintes afin de guider la génération de séquences d'accords de Blues ou de mélodies utilisant des échelles de notes particulières à un style. Par exemple, les contraintes proposées pour la génération d'une grille de Blues sont : « commencer sur une tonalité donnée (par exemple, C^7) » — « terminer par un accord qui se résout sur la tonalité donnée précédemment (par exemple, G^7) » — jouer un quatrième degré de la tonalité initiale à la mesure 5 (par exemple, F^7) », etc. Cette méthode a également été utilisée pour forcer l'improvisation à respecter la métrique d'un morceau dans [Roy & Pachet, 2013]. De manière similaire, dans [Papadopoulos et al., 2014], une contrainte est formulée sur les séquences générées afin qu'elles ne comportent pas des fragments trop longs issus du matériau originel. Pour cela, ils construisent un automate d'ordre maximal éliminant les séquences trop longues pouvant être construites à partir d'une chaîne de Markov préalablement construite sur un corpus. Cet automate est alors utilisé pour générer des mélodies. Le projet *Flow Machine* [Ghedini et al., 2016] dirigé par François Pachet cherche à étendre ce principe à plus grande échelle, en appliquant un style (appris sur un corpus) à une structure choisie. Par exemple, *FlowComposer* [Papadopoulos et al., 2016] propose un outil d'aide à la composition, où l'utilisateur peut choisir un corpus stylistique et ensuite formuler des contraintes structurelles sur la mélodie ou l'harmonie ; le système génère alors une mélodie et une grille d'accords en fonction du style et des contraintes.

L'utilisation de spécifications formelles pour l'improvisation automatique a été proposée par Donzé et al. [2014] avec ce qu'ils appellent l'improvisation contrôlée. L'objectif est de générer une improvisation mélodique sur une grille d'accords donnée, en respectant des spécifications sur les hauteurs (par rapport aux accords) et les rythmes (par rapport à la métrique). La grille d'accords constitue ici une spécification temporelle à long terme. Pour cela, d'une part un oracle des facteurs basé sur les hauteurs et un oracle des facteurs basé sur le rythme sont construits à partir du jeu d'un musicien et d'une autre part des automates de contraintes rythmiques et harmoniques sont construits pour représenter les spécifications temporelles. La génération consiste alors à trouver un chemin qui respecte à la fois la modélisation du style fourni par les oracles des facteurs et les spécifications imposées par les automates. Ces travaux ont été étendus par Valle et al. [2016] qui utilisent des techniques d'exploration de données pour extraire des spécifications automatiquement à partir d'un

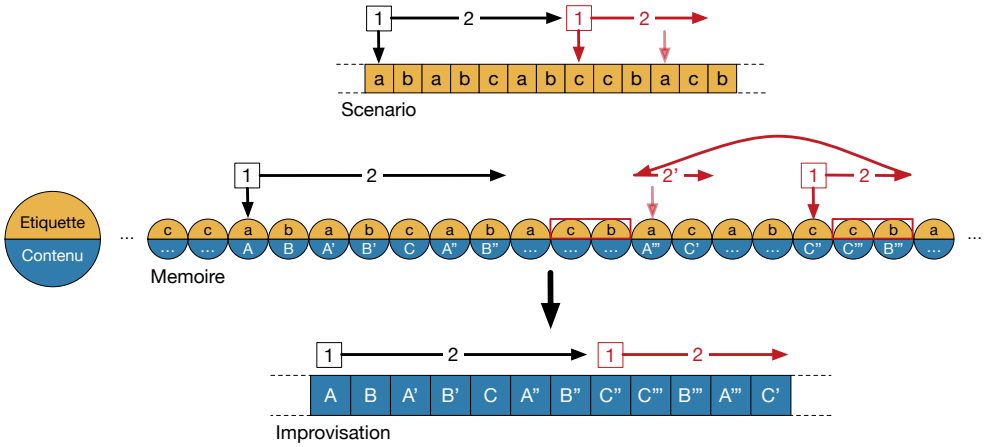

FIGURE 3.5 – Guidage structurel dans *Improtek* par l'utilisation d'un scénario temporel : exemple de deux phases de navigation (flèches noires et rouges). Chaque phase se déroule en deux étapes : une étape d'anticipation (notée 1) et une étape de navigation (notée 2) [Nika et al., 2017a].

corpus en utilisant des graphes de motifs appris à partir d'un ensemble de modèles de motifs. Cette méthode se base sur les travaux présentés dans [Li et al., 2010], ici appliqués à la musique. Une autre extension de l'improvisation contrôlée, pas encore appliquée à la musique, est présentée dans [Akkaya et al., 2016] avec l'utilisation de spécifications temporelles probabilistes représentées par des formules de logique du temps arborescent probabilistes (*Probabilistic Computation Tree Logic*) [Baier & Katoen, 2008].

Finalement, le projet *ImproteK* [Nika et al., 2017a] utilise les connaissances *a priori* d'une spécification temporelle, appelée *scénario*, afin d'introduire des principes d'anticipation comme guidage dans le processus de génération. Le scénario est représenté par une séquence de symboles, appelés *étiquettes* [Nika & Chemillier, 2014]. Selon le contexte musical, le scénario peut représenter une grille d'accords de jazz, l'activité de certains descripteurs audio, ou une structure formalisée à partir d'un alphabet pré-défini par le musicien (dans le cas de méta-compositions). La modélisation du style dans *ImproteK* se base sur *OMax* et utilise donc un oracle des facteurs. Cependant, contrairement à *OMax*, la mémoire n'est pas basée sur une séquence de hauteurs, mais sur une séquence de contenus musicaux organisées à partir des étiquettes de scénario. Le processus de génération dans *ImproteK* cherche à combiner une anticipation du scénario à venir et donc une prise en compte du futur du scénario, avec une navigation de la mémoire similaire à *OMax* permettant une cohérence mélodique vis-à-vis du style. La Figure 3.5 illustre le processus de génération en deux étapes dans *ImproteK*. Premièrement, l'étape d'anticipation cherche un événement dans la mémoire partageant un futur commun avec le scénario actuel. Cela est effectué en indexant les préfixes du suffixe du scénario actuel en utilisant les régularités du motif. De cette manière, en trouvant un tel préfixe, la continuité avec le futur du scénario est garantie. Deuxièmement, l'étape de navigation cherche des événements dans la mémoire partageant un contexte

commun avec ce qui a été joué afin d'effectuer un chemin non-linéaire dans la mémoire pour créer de nouvelles phrases musicales. Cela est effectué en utilisant les heuristiques développées dans *OMax* sur l'oracle des facteurs. Les travaux autour d'*ImproteK* ont donné lieu à des études de terrain et ce système a été utilisé lors de nombreux concerts avec des improvisateurs experts [Nika, 2016]. Il a également été utilisé comme cas d'étude pour le planificateur d'*OpenMusic* dans un système d'aide à la composition de processus musicaux [Nika et al., 2015; Bouche et al., 2017].

3.1.4 Systèmes interactifs

Rowe [1992] définit un système musical interactif comme un système capable, de manière similaire à un humain, « de changer de comportement en réponse à un signal musical. Une telle réactivité permet à ces systèmes de prendre part à des représentations en direct de musique écrite et improvisée ». Nous nous intéressons dans cette section aux applications pour les musiques improvisées (voir [Cont, 2008] pour un système interactif de suivi de partition pour la musique écrite). L'objectif est de profiter des capacités perceptives du système pour émuler des capacités cognitives d'adaptation et de participation à une situation d'improvisation collective. Ces systèmes s'adaptent bien aux systèmes multi-agents intelligents pour simuler des dynamiques de groupe.

Canonne & Garnier [2011] décrivent un modèle pour une situation d'improvisation collective libre avec plusieurs agents. Chaque agent possède des capacités d'intention et d'attention (avec une notion d'ennui) à plusieurs échelles temporelles définies par un système dynamique non-linéaire. Ces travaux montrent la possibilité d'émergence de dynamiques collectives dans une telle situation sans structure définie *a priori*. Différents comportements peuvent être alors obtenus (indirectement) pour chaque agent en jouant sur leurs capacités d'attention : un agent avec une grande capacité d'attention et qui ne s'ennuie pas facilement va jouer un rôle de *leader* et améliorer l'organisation globale de l'improvisation collective. Les capacités d'organisation collective et les dynamiques de groupe vont alors dépendre des comportements des différents agents. Également dans une situation d'improvisation libre collective, Kalonaris [2016] montre les capacités des modèles graphiques probabilistes [Koller & Friedman, 2009] pour représenter une telle situation de jeu. Kalonaris propose un système où tous les improvisateurs sont humains mais communiquent leurs intentions d'interactions avec un autre musicien à travers une interface sous forme de graphe où chaque nœud représente un musicien. Chaque musicien possède un graphe local propre et qui n'est pas connu des autres musiciens. L'ensemble des graphes individuels forme alors un réseau de Markov global. Les résultats montrent la validité du modèle pour la création artistique d'improvisations libres.

Sanlaville et al. [2015] développent un système d'interaction entre agents (numériques et humains) basé sur des mécanismes de régulation de la conversation [Ravenet et al., 2015]. Les agents possèdent différents comportements de prise de parole et d'écoute. Ces comportements sont modélisés par un automate à états finis représentant les différentes transitions possibles entre comportements. Les communications entre agents sont réprésentées par un modèle

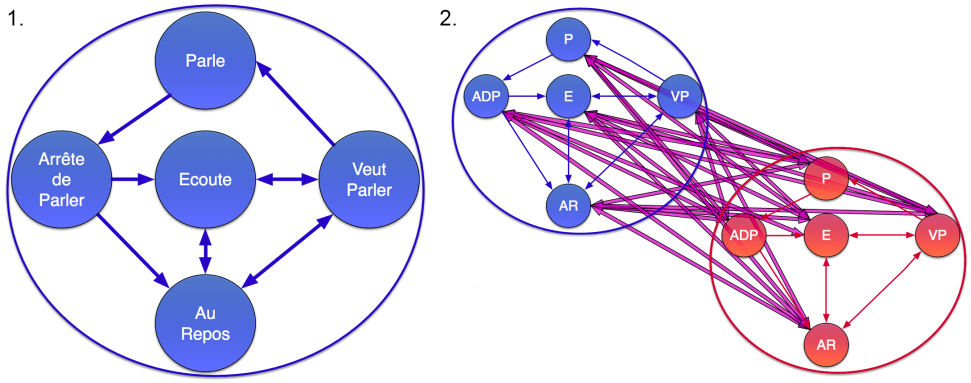

FIGURE 3.6 – 1. Automate représentant les différents comportements et les différentes transitions possibles entre comportements d'un agent conversationnel dans le cadre d'un système de régulation de conversation. 2. Modèle d'influence pour une conversation entre deux agents ; le comportement de chaque agent est influencé par le comportement de l'autre agent [Sanlaville et al., 2016].

d'influence sous la forme d'un modèle de Markov caché ; le changement de comportement de chaque agent est alors régulé par leur comportement présent, mais aussi par le comportement présent des autres agents conversationnels. Les différentes pondérations des transitions de ce modèle peuvent être apprises à partir de données réelles. La Figure 3.6 montre à gauche l'automate comportemental de chaque agent et à droite le modèle d'influence utilisé dans le cas d'une conversation entre deux agents. Un tel modèle permettrait de représenter les processus de prise de parole entre musiciens dans le cas d'une improvisation avec, par exemple, la notion de prise de solo et d'accompagnements [Sanlaville et al., 2016].

3.1.5 Discussion

Nous voyons à travers les différentes évolutions des systèmes d'improvisation automatique la volonté de capturer un style musical dans son ensemble, à travers les relations temporelles à différentes échelles et les relations entre différentes dimensions musicales.

Une limitation des premiers systèmes d'improvisation automatique est leur modèle de mémoire (n-grammes, oracle des facteurs, etc.) ne permettant pas de représenter efficacement l'information musicale multidimensionelle dans la mesure où il ne suit qu'une seule dimension. La notion de guidage propose quelques méthodes pour compenser ce problème. D'un côté, le guidage réactif permet de prendre en considération l'environnement musical présent lors de la génération et donc de considérer deux éléments distincts : l'environnement et la mémoire, les variables musicales issues de l'environnement pouvant être basées sur des dimensions musicales différentes que celle sur laquelle le modèle de mémoire est construit. Dans le cas de *SoMax* par exemple, les différents profils d'activité sont activés par l'écoute active de différentes dimensions en parallèle de l'activation de la mémoire principale. Cependant, il n'y a pas de réel appren-

tissage des liens entre ces différentes dimensions ou de détection des dimensions porteuses de sens (la pondération des profils d'activé est définie par l'utilisateur), la combinaison des dimensions n'étant basée que sur des co-occurences de leurs différents éléments. D'un autre côté, le guidage structurel permet également de prendre en considération une autre dimension musicale lors de la génération. Cela peut-être forcé par la définition de règles formelles à suivre lors de la génération, auquel cas le système perd son agnosticité ; ou dans le cas d'*ImproteK* le modèle de mémoire est construit en utilisant des étiquettes de scénario d'une dimension pour générer le contenu musical d'une autre dimension. Cependant, encore une fois, il n'y a pas de véritable apprentissage des liens entre les différentes dimensions, les contenus musicaux sont complètement subordonné aux étiquettes de scénario. En d'autres termes, le système ne pourrait donc pas improviser simultanément et à mesure les étiquettes de scénarios et les contenus associés, par exemple les étiquettes harmoniques d'une grille inédite et l'improvisation mélodique adaptée.

Une deuxième limitation est l'organisation des improvisations prenant en compte différents niveaux temporels. Comme nous l'avons vu dans la partie 2.1.3, une improvisation, en particulier dans un cadre idiomatique, se construit sur plusieurs niveaux de récits. Cette limitation a été partiellement résolue par le guidage structurel et notamment par le système *ImproteK* avec l'utilisation d'un scénario et l'anticipation. Cependant, dans ce système, les scénarios sont considérés simplement comme une séquence de symboles et ne présentent pas d'organisation hiérarchique. Les systèmes interactifs basés sur la régulation de conversation permettraient d'organiser la structure musicale à très haut niveau, mais ne permettent pas d'organiser hiérarchiquement le discours de chaque agent.

Dans la partie suivante, nous présentons des méthodes d'apprentissage automatique de structures multidimensionnelles et de structures temporelles pouvant constituer des pistes pour dépasser ces limitations.

3.2 Apprentissage et modélisation de structures multidimensionnelles et temporelles

3.2.1 Apprentissage de structures musicales multidimensionnelles

Les modèles probabilistes du langage tels que les n-grammes ont été utilisés dès les années 50 pour des tâches liées à la musique, par exemple par Pinkerton [1956] et Hiller & Isaacson [1959]. Ces tâches vont d'applications pratiques à des recherches théoriques. D'un point de vue applicatif, outre les tâches de composition assistée par ordinateur et d'improvisation automatique décrites plus en détail dans la partie 3.1, diverses tâches d'extraction d'information musicale (*Music Information Retrieval* ou MIR) ont été étudiées, comme par exemple l'identification de musique monophonique [Pickens, 2000] et polyphonique [Pickens et al., 2002], ou la classification automatique de suites d'accords [Pérez-Sancho et al., 2008]. D'un point de vue plus théorique, les n-grammes ont été utilisés pour l'analyse stylistique de la musique

[Ponsford et al., 1999; Dubnov et al., 2003], ou la modélisation cognitive de la perception de la musique [Ferrand et al., 2002; Eerola, 2004; Temperley, 2008; Pearce & Wiggins, 2012]. Cependant, dans l'ensemble des exemples cités précédemment, les n-grammes sont utilisés comme modèles pour représenter des mélodies comme une séquence de hauteurs, ou des progressions harmoniques comme une séquence d'accords. Dans cette partie, nous présentons plusieurs méthodes permettant de dépasser cette limitation. Dans un premier temps, nous présenterons les modèles à points de vue multiples proposés par Conklin et al. [Conklin & Witten, 1995]. Nous présenterons ensuite des méthodes d'interpolation de sous-modèles, plus proches des travaux présentés dans cette thèse. Nous étudierons des méthodes de lissage puis d'évaluation pour ces modèles. Finalement, nous présenterons des méthodes basées sur l'utilisation de réseaux de neurones.

Modèles à points de vue multiples

Conklin & Witten [1995] proposent l'utilisation de modèles à points de vue multiples (*multiple viewpoints systems*) pour la prédiction et la génération de musique multidimensionnelle. Ces modèles reposent sur des modèles de contexte qui sont définis par un corpus d'apprentissage formé de séquences sur un espace d'événements défini, un nombre d'occurrences dans le corpus de chaque séquence, et une méthode d'inférence pour calculer la probabilité d'un n-uplet. Par exemple, la probabilité conditionnelle $P(e|c)$ d'un événement e pour un contexte c donné est le nombre de fois où la séquence $c.e$ est observée dans le corpus, divisé par le nombre d'occurrence du contexte :

$$P(e|c) = \frac{\texttt{count}(c.e)}{\texttt{count}(c)} \quad . \tag{3.1}$$

Un modèle à points de vue multiples d'une mélodie est caractérisé par un ensemble d'attributs musicaux, appelés points de vue, d'une séquence mélodique, permettant, par exemple, de considérer la hauteur, le début et la durée de chaque note. Autour de ces points de vue de base peuvent être construits des points de vue dérivés. Par exemple, les points de vue `intervalle` et `contour mélodique` vont dériver du point de vue `hauteur`. Il est également possible de considérer des points de vue produits. Par exemple, le point de vue `hauteur` $\otimes$ `durée` correspond à l'ensemble des combinaisons possibles de hauteur et de durée. Ces modèles permettent aux n-grammes de tirer profit d'attributs musicaux arbitraires dérivés de la représentation basique du langage. Conklin & Witten [1995] proposent plusieurs modèles à points de vue multiples pour la prédiction et la génération de chorals de Bach. Pour cela un corpus de 100 chorals de Bach a été utilisé, divisé en deux sous-corpus : un corpus d'apprentissage de 95 chorals et un corpus de test contenant les cinq autres chorals. Pour chaque point de vue utilisé, la taille du contexte est fixé à deux ou trois éléments (c'est-à-dire trigramme ou quadrigramme). Tous les points de vues considérés sont évalués individuellement pour sélectionner le plus approprié.

Depuis, plusieurs développements autour des modèles à points du vue multiples ont été effectués et leur champ d'application s'est aggrandi. Conklin &

Anagnostopoulou [2001] utilisent des points de vue multiples en combinaison avec un arbre de suffixes pour la découverte de motifs dans les chorals de Bach. L'objectif est ici de chercher les plus longs motifs significatifs apparaissant suffisamment fréquemment dans le corpus et dans un nombre minimal fixé de chorals différents. Conklin [2002] étend ces travaux à la musique polyphonique en utilisant des points de vue verticaux et horizontaux afin de pouvoir prendre en considération les évènements simultanés de plusieurs voix. Par exemple, l'intervalle entre deux notes simultanées jouées par deux voix différentes peut-être considéré comme un point de vue vertical. Padilla & Conklin [2016] utilisent ces travaux pour une tâche de génération de musique dans le style de Palestrina (compositeur italien du XVIème siècle). Le corpus utilisé est constitué de 101 messes à deux voix composées par Palestrina. Des points de vue horizontaux et verticaux sont utilisés pour découvrir des motifs dans ce corpus. Encore une fois, les différents points de vue sont évalués individuellement pour selectionner le plus performant. Les motifs ainsi découverts sont alors utilisés par un algorithme glouton pour remplir, de gauche à droite, un patron de partition prédéfini en utilisant les motifs les plus spécifiques. Des méthodes de rétro-inspection sont ensuite utilisées dans les cas où les motifs proposés pour une certaine zone du patron ont des probabilités nulles. Conklin [2013] utilise des points de vue multiples pour la classification automatique de musique, appliquée à la reconnaissance de genre et de région pour la musique folklorique basque. Différents points de vue sont utilisés avec des méthodes bayésiennes de classification de texte [Peng et al., 2004] afin de calculer les probabilités *a posteriori* d'appartenance à une classe pour une séquence donnée.

Interpolation de sous-modèles

Une autre méthode pour modéliser des structures musicales multidimensionnelles est d'effectuer de l'interpolation de sous-modèles probabilistes. Nous présentons dans cette partie ses principes que nous utiliserons par la suite dans cette thèse. Les modèles présentés dans cette section sont également basés sur des modèles de contexte. Avec cette méthode, les relations verticales entre dimensions ne sont pas seulement considérées par les points de vues verticaux produits, par exemple : $\texttt{mélodie} \otimes \texttt{accord}$. Les modèles de contexte utilisés peuvent directement considérer une dimension comme contexte d'une autre, par exemple : $P(\texttt{mélodie}|\texttt{accord})$. Cela permet dans les calculs de probabilités de ne considérer que des symboles uni-dimensionnels. Par conséquent, l'alphabet des symboles est alors considérablement plus petit, ce qui réduit la combinatoire.

Raczyński et al. [2013a] proposent un modèle discriminatif [Jebara, 2004] pour une tâche d'harmonisation automatique. L'objectif est de prédire l'accord C_t à un instant t en fonction de l'ensemble des variables musicales présentes et passées. Ils veulent alors estimer

$$P(C_t|X_{1:t}) \ , \tag{3.2}$$

où $1:t$ représente l'ensemble des temps de 1 à t et X l'ensemble des autres variables musicales. Les différentes variables musicales contenues dans $X_{1:t}$

pouvant provenir de différentes dimensions, le modèle est capable de considérer plusieurs dimensions musicales et leurs liens pour la prédiction d'une dimension. Cependant, un tel modèle ne peut pas être utilisé en pratique du fait de sa trop haute dimensionalité ; l'ensemble des valeurs possibles de $X_{1:t}$ étant le produit cartésien de l'ensemble des valeurs possibles de chaque variable musicale X_t à chaque instant t ; la combinatoire devient trop importante (cf. [Whorley et al., 2013] pour une étude de la complexité en temps de la construction des espaces dans le cas des modèles à points de vue multiples dérivés et produits). Une approximation de ce modèle global est alors calculée en effectuant une interpolation de plusieurs sous-modèles P_i dont les conditions dépendent chacune d'un sous-ensemble de variables musicales $A_{i,t} \subset X_{1:t}$. Deux types d'interpolation sont possibles : l'interpolation linéaire proposée dans [Jelinek & Mercer, 1980] et l'interpolation log-linéaire proposée bien plus tard dans [Klakow, 1998]. L'interpolation de sous-modèles a notamment été utilisée dans la modélisation probabiliste du langage pour combiner des modèles de n-grammes de tailles différentes [Stolcke et al., 2011; Mikolov et al., 2011]. Ici, cette approche est généralisée pour la musique afin d'interpoler des modèles possédant des relations verticales et horizontales.

L'interpolation linéaire [Jelinek & Mercer, 1980] est définie par

$$P(C_t|X_{1:t}) = \sum_{i=1}^{I} \lambda_i P_i(C_t|A_{i,t}) \ , \tag{3.3}$$

où I est le nombre de sous-modèles interpolés et les λ_i sont les coefficients d'interpolation tels que

$$\lambda_i \geq 0 \quad \forall i \qquad \text{et} \qquad \sum_{i=1}^{I} \lambda_i = 1 \ . \tag{3.4}$$

L'interpolation log-linéaire [Klakow, 1998] est définie par

$$P(C_t|X_{1:t}) = Z^{-1} \prod_{i=1}^{I} P_i(C_t|A_{i,t})^{\gamma_i} \ , \tag{3.5}$$

où les γ_i sont les coefficients d'interpolation tels que

$$\gamma_i \geq 0 \quad \forall i \ , \tag{3.6}$$

et Z un facteur de normalisation défini par

$$Z = \sum_{C_t} \prod_{i=1}^{I} P_i(C_t|A_{i,t})^{\gamma_i} \ . \tag{3.7}$$

Ces formules se généralisent évidemment à n'importe quelle application, nous ne présentons ci-dessus que l'exemple présenté par Raczyński et al. [2013a]. Une méthode similaire aux travaux présentés dans [Raczyński et al., 2013a] a été utilisée pour la modélisation de séquences polyphoniques par Raczyński et al. [2013b], basée sur un réseau bayésien dynamique.

Lissage des modèles

Tous les modèles présentés dans cette partie, qu'il s'agisse des modèles à points de vue multiples ou des sous-modèles interpolés, demandent à être appris sur des corpus. Dans le cas de la modélisation probabiliste du langage et en particulier dans les applications musicales, il est courant que l'ensemble des éléments observés dans le corpus d'apprentissage n'inclut pas tous les éléments pouvant apparaitre dans le corpus de test, ou plus généralement que la distribution de certains éléments du corpus d'apprentissage ne correspondent pas à celle du corpus de test. Cela est dû à la taille trop restreinte des corpus d'apprentissage. Ce problème engendre du sur-apprentissage, c'est-à-dire une adaptation trop forte aux données d'apprentissage empêchant une généralisation correcte des modèles. Le terme de lissage vient du fait que ces techniques ont tendance à rendre les distributions de probabilité plus uniformes, en augmentant les éléments de faible probabilité et en baissant les éléments de forte probabilité. Chen & Goodman [1998] et Zhai & Lafferty [2004] présentent des études et comparaisons détaillées de différentes techniques de lissage. Nous présentons ici les concepts principaux de ces techniques.

La technique la plus simple de lissage est le *lissage additif* [Jeffreys, 1948]. Elle consiste à considérer que chaque élément possible apparait δ fois plus qu'il n'apparait réellement dans le corpus, avec généralement $0 < \delta \leq 1$.

$$P_{\texttt{add}}(e|c) = \frac{\delta + \texttt{count}(c.e)}{\sum_{e'} \delta + \texttt{count}(c.e')} \ . \tag{3.8}$$

L'estimation de Good-Turing [Good, 1953] est une extension de cette méthode qui consiste à considérer que chaque évènement apparaissant r fois dans le corpus d'apprentissage apparait en fait r^* fois avec

$$r^* = (r+1)\frac{n_{r+1}}{n_r} \ , \tag{3.9}$$

où n_r est le nombre d'événements apparaissant r fois dans le corpus d'apprentissage. Cette méthode de comptage est alors normalisée pour obtenir une probabilité. Si $c.e$ apparaît r fois, on considère :

$$P_{\texttt{GT}}(e|c) = \frac{r^*}{N} \qquad \text{avec } \ N = \sum_{r=0}^{\infty} n_r r^* = \sum_{r=1}^{\infty} n_r r \ . \tag{3.10}$$

En pratique, ces techniques sont peu utilisées seules, car elles fournissent des résultats peu convaincants [Gale & Church, 1994]. Les techniques de *lissage par repli* sont alors utilisées. Elles consistent à combiner le modèle que l'on souhaite lisser avec un modèle d'ordre inférieur. Cela permet de prendre en considération les fréquences d'apparition des événements d'ordres inférieurs et donc d'obtenir une meilleur estimation qu'avec un simple lissage additif. Pour cela, le lissage de Jelinek-Mercer utilise l'interpolation linéaire proposée dans [Jelinek & Mercer, 1980]. On a alors :

$$P_{\texttt{JM}}(e|c) = \lambda P(e|c) + (1-\lambda)P(e|c^*) \ , \tag{3.11}$$

avec c^* un sous-ensemble de c. Dans le cas des n-grammes, par exemple, un trigramme sera interpolé avec un bigramme, un bigramme avec un unigramme, etc. [Song & Croft, 1999]. Ces techniques peuvent s'utiliser récursivement. On peut alors remarquer que les techniques de lissage par repli sont une généralisation des techniques de lissage additif, dans le sens où, par récursion, le lissage par repli va au final utiliser une distribution uniforme de tous les éléments, comparable à un lissage additif avec $\delta = 1$. De nombreuses autres techniques de lissage par repli existent, notamment le lissage de Witten-Bell [Witten & Bell, 1991], le lissage de Katz [Katz, 1987], le lissage par *absolute discounting* [Ney et al., 1994], etc. (cf. [Chen & Goodman, 1998; Zhai & Lafferty, 2004] pour plus de détails).

Évaluation des modèles

Il existe différentes méthodes pour évaluer les modèles présentés dans cette partie. Ces méthodes peuvent être plus ou moins spécifiques par rapport aux applications précises de ces modèles, par exemple, des tests d'écoute peuvent être effectués pour les applications musicales. Une première méthode peut être d'utiliser la connaissance d'une vérité terrain. Par exemple, dans [Raczyński et al., 2013a], la prédiction de l'harmonie est évaluée sur un corpus de test, en effectuant des comparaisons entre l'accord prédit par les différents modèles proposés et l'accord réel dans le corpus. La précision (en pourcentage) des modèles peut alors être calculée.

Il est courant pour évaluer des modèles probabilistes d'utiliser des mesures d'*entropie*, afin de permettre une évaluation plus neutre et moins centrée sur l'aspect applicatif [Pearce, 2005]. Si on considère la fonction de masse $P(a \in \Sigma) = P(X = a)$ (équivalant à la densité de probabilité dans le cadre discret) d'une variable aléatoire X sur un alphabet Σ, l'entropie est définie par

$$H(P) = H(X) = -\sum_{a \in \Sigma} P(a) \log_2 P(a) \ . \tag{3.12}$$

Le théorème du codage de source (premier théorème de Shannon) montre que l'entropie estime le nombre moyen de bits par symbole nécessaire pour encoder le résultat d'un tirage de X [Shannon, 1948]. À noter également que, pour un alphabet donné, l'entropie a une borne supérieure $H_{\texttt{max}}$ qui est atteinte lorsque la probabilité d'apparition de chaque symbole est uniforme, c'est-à-dire quand $\forall a \in \Sigma,\ P(a) = \frac{1}{|\Sigma|}$, où $|\Sigma|$ est la taille de l'alphabet. Dans ce cas, on a

$$H_{\texttt{max}}(P) = \log_2 |\Sigma| \ . \tag{3.13}$$

L'entropie peut être interprétée comme le taux d'incertitude du système à choisir un symbole dans l'alphabet ; plus l'entropie est élevée, plus l'incertitude est grande.

L'*entropie croisée* représente la divergence entre le modèle et la distribution empirique d'un corpus [Chen, 2009]. Si on considère un modèle $P_{\texttt{M}}(e|c)$ et une

séquence $(e, c) = (e_1, c_1)...(e_T, c_T)$, l'entropie croisée $H_{\mathtt{M}}$ est définie par

$$H_{\mathtt{M}}(P_{\mathtt{M}}) = -\frac{1}{T}\log_2 P_{\mathtt{M}}(e|c) \tag{3.14}$$

$$= -\frac{1}{T}\sum_{t=1}^{T}\log_2 P_{\mathtt{M}}(e_t|c_t) \ . \tag{3.15}$$

Une autre mesure, équivalent à l'entropie croisée et couramment utilisée pour l'évaluation des modèles de langages, est la *perplexité*. La perplexité $PP_{\mathtt{M}}$ correspond à l'exponentielle en base 2 de l'entropie croisée :

$$PP_{\mathtt{M}}(P_{\mathtt{M}}) = 2^{H_{\mathtt{M}}(P_{\mathtt{M}})} \tag{3.16}$$

$$= 2^{-\frac{1}{T}\sum_{t=1}^{T}\log_2 P_{\mathtt{M}}(e_t|c_t)} \ . \tag{3.17}$$

L'entropie croisée et la perplexité sont utilisées pour de nombreuses applications issues de la modélisation probabiliste du langage comme la reconnaissance automatique de la parole, la traduction automatique, etc. [Haton et al., 2006] pour optimiser les modèles. Une telle optimisation des paramètres des modèles entraine une amélioration de leurs performances [Brown et al., 1992]. Dans les applications musicales, l'entropie croisée a également été utilisée pour optimiser les modèles par [Raczyński et al., 2013a] pour une tâche d'harmonisation automatique et par Conklin et al. pour sélectionner des modèles à points de vue multiples conçus pour différentes tâches [Conklin & Witten, 1995; Conklin, 2013].

Structures multidimensionnelles dans les réseaux de neurones

L'apprentissage de structures musicales multidimensionnelles a également été effectué avec l'utilisation de réseaux de neurones. Ceux-ci peuvent être utilisés pour représenter la distribution probabiliste des données d'un corpus. Bellgard & Tsang [1999] utilisent une machine de Boltzmann [Tsang & Bellgard, 1990] pour une tâche d'harmonisation automatique à quatre voix. Les notes jouées par les quatre voix à un instant t sont représentées par un vecteur d'activation. Ce vecteur représente l'ensemble de la tessiture ; si la $n^{\text{ème}}$ note du vecteur est jouée à l'instant t, alors la valeur dans cette position est 1, sinon cette valeur est 0. Cela permet de représenter les dépendances verticales entre les voix. Dans ces travaux, la génération n'est pas effectuée séquentiellement, de « gauche à droite », mais globalement en cherchant à minimiser l'énergie globale du système.

Plus récemment, l'émergence des réseaux de neurones profonds et leur popularité ont relancé la recherche sur les réseaux de neurones pour la génération musicale. Bickerman et al. [2010] génèrent des mélodies de jazz sur une suite d'accords donnée. Pour cela, ils utilisent un réseau de neurones profond basé sur une imbrication par couches de machines de Boltzmann restreintes. La Figure 3.7 illustre la structure d'un tel réseau ; chaque cadre représente une machine de Boltzmann restreinte. La profondeur de ce réseau permet au modèle d'apprendre des caractéristiques musicales plus complexes qu'une simple machine de Boltzmann. Pour l'apprentissage, des mélodies de quatre temps

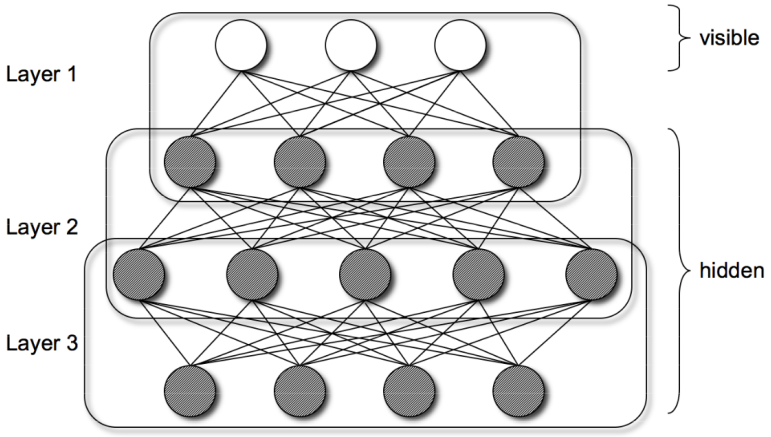

FIGURE 3.7 – Représentation d'un réseau de neurones profond avec 3 couches. Ce réseau est constitué par l'imbrication de machines de Boltzmann restreintes (représentées par les cadres). Les nœuds visibles sont activés par les vecteurs binaires d'entrées représentant la mélodie et les accords. Les nœuds cachés (*hidden*) apprennent des caractéristiques musicales à partir des données. Une fois l'apprentissage du réseau effectué, les nœuds visibles peuvent être considérés comme sorties du réseau [Bickerman et al., 2010].

sont utilisées ; chaque temps est divisé en douze tatums afin de pouvoir récupérer les doubles croches et les triolets ; pour chaque tatum, les notes jouées sont représentées par un vecteur de chroma et une octave, les activations sont faites de manière similaire que dans [Bellgard & Tsang, 1999]. Les accords sont représentés par un vecteur d'activation des notes présentes dans l'accord. Les vecteurs de mélodies et d'accords sont alors concaténés et fournis au réseau pour l'apprentissage. Un réseau de neurones similaire, utilisant des machines de Boltzmann restreintes conditionnelles, a été utilisé par Battenberg & Wessel [2012] pour classifier des motifs rythmiques de batterie, en utilisant les activations des différentes parties de l'instrument (grosse caisse, caisse claire, charleston...). Le projet MidiNet [Yang et al., 2017] utilise un réseau de neurones convolutif adversariel pour la composition de mélodie avec guidage structurel par une grille d'accord. Ce projet a été inspiré par le projet WaveNet [Van Den Oord et al., 2016] de Google DeepMind qui utilise également un réseau de neurones convolutif, mais pour la génération de musique à partir de données audio et d'une analyse des formes d'ondes. Schmidt & Kim [2013] utilisent des caractéristiques audio telles que des chromagrammes et des MFCC sur le spectre de sons percussifs pour extraire des informations mélodiques, harmoniques et rythmiques avec un réseau de neurones profonds basés sur des machines de Boltzmann restreintes.

3.2.2 Modélisation de structures musicales temporelles

Nous nous intéressons dans cette partie à un autre type de structure musicale : les structures temporelles multi-niveaux. Une modélisation de ces struc-

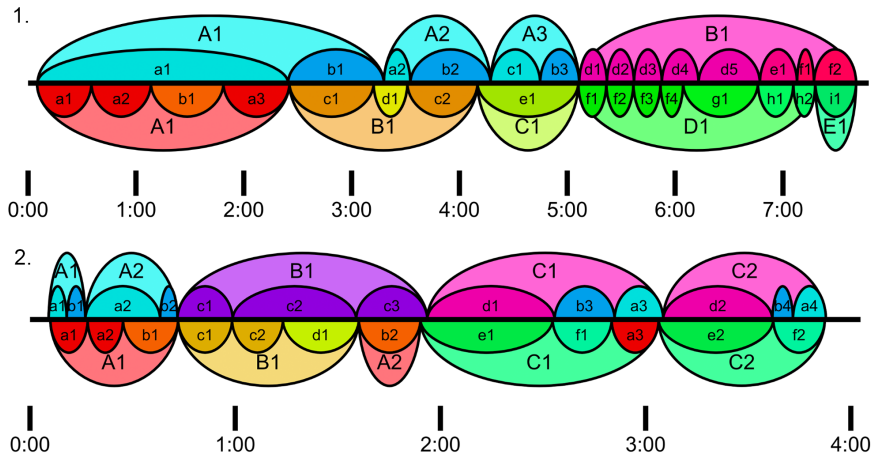

FIGURE 3.8 – Annotations de deux prestations du système *MiMi4x*. Chaque prestation (1. et 2.) est annotée par deux auditeurs ; les annotations du premier auditeur sont au dessus des axes, celles du second en dessous [Smith, 2014].

tures permettrait d'étendre les possibilités actuelles du guidage structurel de l'improvisation en permettant aux modèles génératifs d'organiser une improvisation sur plusieurs niveaux de récits.

La recherche de structures temporelles dans la musique est un problème complexe du domaine de la MIR. Les objectifs principaux sont d'extraire et d'analyser les formes musicales à différents niveaux d'organisation, typiquement, les quatre niveaux de récit présentés dans la partie 2.1.3. De manière similaire, inspiré des travaux de Snyder [2000], Bimbot et al. [2014] décrivent une organisation en trois niveaux : le *niveau acoustique* correspondant aux événements tels que les notes et les temps, le *niveau morpho-syntagmatique* correspondant aux motifs et aux mesures et le *niveau sectionnel* correspondant aux différentes parties d'un morceau et aux régularités à long terme. Des études de psycho-acoustique [Koelsch et al., 2013; Jackendoff & Lerdahl, 2006] ont montré que les musiciens, mais aussi les non-musiciens, en écoutant de la musique classique, appliquent des processus cognitifs capables de considérer les dépendances et relations musicales à long terme constituées d'une organisation hiérarchique de manière similaire à l'organisation des langues naturelles. L'analyse de la forme musicale est un sujet complexe, car la description de l'organisation structurelle d'un morceau de musique relève d'un certain niveau de subjectivité. Smith et al. [2013]; Smith [2014] s'intéressent aux différences d'annotations structurelles d'un même morceau de musique entre différents auditeurs. La Figure 3.8 montre les annotations d'organisation structurelle, correspondant aux niveaux morpho-syntagmatique et sectionnel, de deux auditeurs sur deux prestations différentes du système *MiMi4x* [Schankler et al., 2011]. On peut remarquer que les différences sont assez importantes. Ils expliquent principalement ces différences par deux phénomènes : premièrement, les auditeurs ne se concentrent pas nécessairement sur les mêmes caractéristiques ou aspects musicaux et deuxièmement, les analyses dépendent

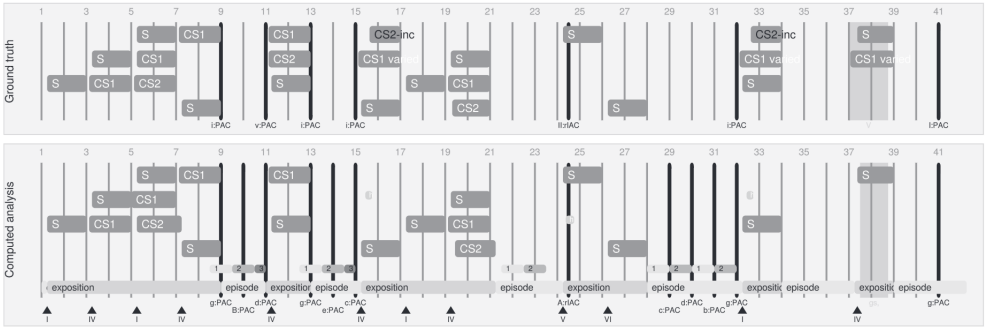

FIGURE 3.9 – Vérité terrain (en haut) et analyse automatique (en bas) [Giraud et al., 2015] de la Fugue No. 18 en Sol dièse mineur (BWV 863) de Bach avec le sujet (*S*) et contre-sujets (*CS1* et *CS2*). Dans la vérité terrain, *CS1-var* dénote une variation du premier contre-sujet et *CS2-inc* le deuxième contre-sujet incomplet.

des structures fondamentales trouvées au début de chaque prestation par les auditeurs (ces structures fondamentales se répercutant alors sur le reste de l'analyse).

Dans cette partie, nous présentons d'abord des travaux issus de l'extraction d'information musicale pour des tâches d'analyse, puis nous présentons des travaux se basant sur l'utilisation de grammaires formelles pour la représentation d'organisation syntactique de la musique.

Analyse musicale et MIR

Giraud et al. [2015] s'intéressent au cas particulier des fugues pour en extraire automatiquement leur forme, avec une détection des sujets et contre-sujets, des expositions et des épisodes, etc. Des connaissances *a priori* sur la forme Fugue sont utilisées pour identifier ses différents composants. Pour cela, ils utilisent une méthode d'extraction de motifs [Janssen et al., 2013] basée sur une adaptation de l'algorithme de Mongeau-Sankoff [Mongeau & Sankoff, 1990] utilisant la programmation dynamique. Cette méthode permet de déterminer le début et la fin des motifs principaux. Un modèle de Markov caché est ensuite utilisé pour définir la forme globale de la fugue (exposition, codetta, épisodes). La Figure 3.9 montre un exemple d'analyse de fugue avec l'identification et les apparitions du sujet et des contre-sujets au cours de la pièce. Les différentes expositions et épisodes sont ensuite identifiés. Des méthodes similaires sont utilisées pour la détection de forme Sonate dans [Bigo et al., 2017] et pour la reconnaissance de Thème et Variations dans [Giraud et al., 2014].

[Bimbot et al., 2016] présente le modèle *System & Contrast* pour la description de l'organisation structurelle d'un segment musical, appelé bloc, divisé en un ensemble de sous-blocs. Chaque bloc est défini d'une part par un *système porteur* représentant les relations entre les différents sous-blocs à l'aide d'un réseau matriciel et d'une autre part par un *contraste* sur le dernier sous-bloc décrivant les transformations s'écartant de la logique du système porteur. La Figure 3.10 montre un exemple d'application du modèle *System & Contrast*. La

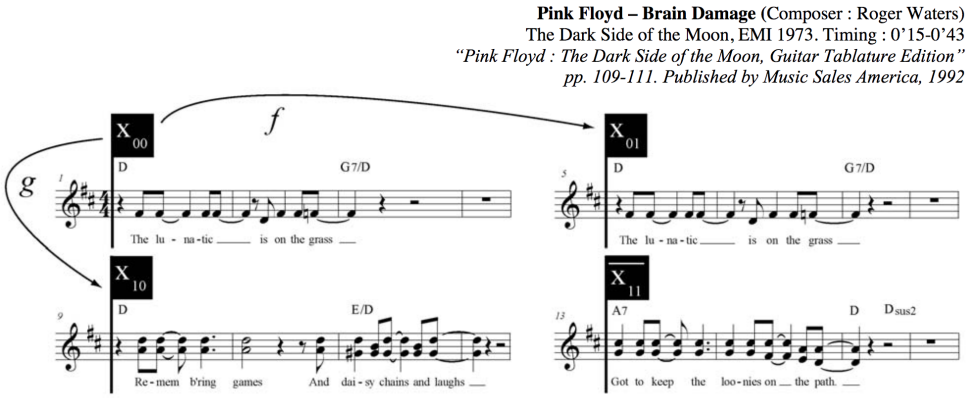

FIGURE 3.10 – Application du modèle *System & Contrast* sur le premier couplet de la chanson *Brain Damage* des Pink Floyd. Le couplet est décomposé en quatre séquences. Les fonctions f et g représentent respectivement les transformations entre la première séquence et la deuxième et troisième séquence. On constate que la dernière séquence n'est pas une application de $f \circ g$ sur la première séquence, formant ainsi un contraste [Bimbot et al., 2016].

séquence est ici divisée en quatre éléments $\{X_{00}, X_{01}, X_{10}, \overline{X}_{11}\}$ formant une structure *aabc*. La fonction f définit l'identité et représente la répétition du premier motif X_{00} en X_{01} et la fonction g décrit le changement de G^7/D vers E/D entre X_{00} et X_{10} ainsi que l'ajout d'une deuxième voix. Finalement, $\overline{X}_{11}$ forme un contraste dans le sens où il ne s'agit pas d'une répétition de X_{10} (application de f) ; $\overline{X}_{11}$ est un passage faisant sonner la fonction de dominante, il partage cependant certaines caractéristiques avec X_{10} comme l'utilisation d'une seconde voix.

Eck & Lapalme [2008] proposent d'utiliser un réseau de neurones récurrent pour apprendre des dépendences à long terme pour la génération de mélodies. Le réseau prend en entrée les accords et les notes jouées aux instants passés prédéfinis correspondant à la métrique du corpus (par exemple, t, $t-15$, $t-31$ et $t-63$ dans le cas d'un morceau de structure régulière multiple de 16) pour prédire la note jouée à l'instant $t+1$. L'objectif ici n'est pas d'extraire explicitement la structure hiérarchique de la musique, mais de trouver les influences à plusieurs échelles temporelles dans la construction d'une mélodie.

Grammaires formelles

Les grammaires formelles ont également été utilisées pour l'analyse musicale de structure temporelle. Les grammaires formelles se basent sur une séquence initiale appelée *axiome* et sur un ensemble de règles de ré-écriture permettant d'effectuer des dérivations d'une séquence de symboles en une autre [Chomsky, 1972]. Par exemple, la règle de ré-écriture : $aa \rightarrow ab$ permet de dériver la séquence *aabc* en *abbc* en modifiant l'occurrence de *aa*.

Lerdahl & Jackendoff [1983] décrivent un ensemble de règles (et approximations) pour la génération et l'analyse de structures métriques pour la musique

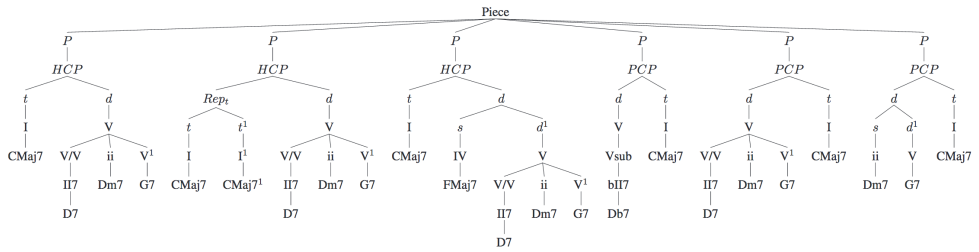

FIGURE 3.11 – Analyse syntaxique automatique du standard de jazz *Take the 'A' Train* à l'aide d'une grammaire structurelle de l'harmonie tonale [de Haas et al., 2009]. P représente une phrase musicale, PCP une cadence parfaite, HCP une demi-cadence et t, d, s représentent respectivement les fonctions de tonique, dominante et sous-dominante.

classique tonale ; ces règles peuvent être utilisées comme règles de ré-écriture dans une grammaire formelle ou comme procédé d'analyse pour effectuer des analyses par réduction représentant l'organisation hiérarchique d'une pièce de musique. Steedman [1984] décrit un ensemble de règles harmoniques pour la constitution d'une grammaire générative de grilles de blues en douze mesures. Ces règles ont été définies de manière empirique par l'analyse de grilles de blues modernes issues de [Coker, 1964]. Cette grammaire a ensuite été étendue par Steedman [1996] en montrant que les règles peuvent s'appliquer à d'autres types de jazz au-delà du blues à condition de modifier l'axiome de la grammaire. Cette grammaire a été implémentée dans le système *ImproteK* [Chemillier, 2004] permettant au système d'introduire des variations locales de la grille lors de la génération d'improvisations.

de Haas et al. [2009] utilisent une grammaire structurelle non contextuelle de l'harmonie tonale pour évaluer la similarité entre deux séquences d'accords. La grammaire utilisée est une extension de celle présentée dans [Rohrmeier, 2007]. L'avantage de cette grammaire est de fournir une description hiérarchique des éléments constitutifs d'une pièce de musique. Elle se base sur quatre niveaux d'analyse : le niveau de la phrase, le niveau fonctionnel (correspondant aux cadences), le niveau des degrés et le niveau de surface (correspondant aux accords effectifs). La Figure 3.11 présente l'analyse syntaxique d'un standard de jazz à l'aide de cette grammaire. Ce diagramme ne montre pas les règles grammaticales appliquées, mais permet d'observer la structure hiérarchique de l'analyse. La similarité entre deux séquences d'accords est alors effectuée en comparant les diagrammes issus de l'analyse syntaxique ; ils proposent plusieurs mesures de distance basées sur le plus grand arbre commun aux deux diagrammes.

Les différentes applications de grammaires présentées ici se basent sur une définition empirique des grammaires. Guichaoua [2017] propose une méthode d'induction automatique de grammaire pour la description de structure musicale sur l'harmonie d'un morceau de musique. Ils utilisent les méthodes d'induction de grammaire à dérivation unique minimal proposé par Gallé [2011] pour des séquences d'ADN. Cette méthode permet de détecter les structures

identiques mais n'est pas robuste aux légères variation. De plus, son application s'est limitée à l'analyse de morceaux de musique individuellement. Dans le cas des langages naturels, l'induction de grammaire se base très fortement sur un étiquetage morpho-syntaxique des différents composants d'une phrase [de Marcken, 2015; Klein & Manning, 2004]. Cependant, dans le cas de la musique, un tel étiquetage n'est pas possible, un même élément pouvant posséder différent rôles dans différents contextes (contrairement, par exemple, à un pronom qui reste un pronom quel que soit le contexte). Plus récemment, des méthodes d'induction de grammaire sans utiliser d'étiquetage morpho-syntaxique ont été développées pour essayer d'émuler l'apprentissage du langage chez les enfants. Börschinger et al. [2011]; Kwiatkowski et al. [2012] associent les phrases avec une représentation logique de leur sémantique. Cependant, cette représentation est propre aux langages naturels ; la notion de logique sémantique dans la musique est encore trop abstraite. Pate & Johnson [2016] effectuent de l'inférence de grammaires probabilistes non contextuelles directement à partir des mots sans utiliser d'information supplémentaire, à partir de la méthode d'inférence bayésienne variationelle de Kurihara & Sato [2004]. Cependant, l'apprentissage de tels modèles demande une quantité de données extrêmement importante, encore difficilement accessible dans le cas de la musique.

3.2.3 Discussion

Nous avons vu dans cette partie que les méthodes d'apprentissage de structures musicales multidimensionnelles permettent de prendre en considération les liens entre différentes dimensions, de considérer des représentations musicales plus complexes et d'évaluer les dimensions pertinentes à partir d'un corpus. Cependant, l'apprentissage de tels modèles prend du temps. Ces modèles ne peuvent donc pas être appris en direct avec le jeu d'un musicien. Une deuxième limite de ces modèles est qu'ils ne sont pas contraints par un cadre local et ne possèdent pas une représentation linéaire de la mémoire. Enfin, une dernière limite de ces systèmes, dans le cas de la génération de musique sur plusieurs dimensions, est la centralisation du processus de génération dans une seule entité, limitant la flexibilité des systèmes pour l'interaction entre différents agents, en particulier avec un agent humain.

L'apprentissage de structures temporelles a principalement été utilisé pour des tâches d'extraction d'information musicale. Les travaux de Eck et al. sur les réseaux de neurones et les travaux de Steedman & Chemillier sur les grammaires formelles pour la génération musicale considèrent une organisation structurelle de la musique, mais ne proposent pas d'organisation hiérarchique. L'utilisation de grammaire proposant une structure hiérarchique, n'a pas encore, à notre connaissance, été appliqué pour la génération musicale. De plus, les applications musicales basées sur les grammaires formelles se basent sur une création empirique des grammaires et n'utilisent pas de méthodes d'apprentissage automatique.

Dans les chapitres suivants nous proposons nos contributions pour répondre aux problématiques présentées dans la partie 3.1.5 et ici. Dans le Chapitre 4, nous présenterons un modèle combinant un apprentissage multidimensionnel de modèle probabiliste avec un oracle des facteurs pour la génération d'im-

provisations unidimensionnelles, pouvant être réactif, capable de prendre en compte les liens entre plusieurs dimensions et de s'adapter à un cadre local. Dans le Chapitre 5, nous décrirons un système capable de générer des improvisations multidimensionnelles à l'aide d'un système multi-agents sans centralisation du processus de génération. Enfin, dans le Chapitre 6, nous présenterons un système de guidage structurel utilisant une grammaire hiérarchique comme scénario pour le guidage.

4

Apprentissage multidimensionnel pour l'improvisation

"Each man is his own academy."

– Cecil Taylor

Ce chapitre présente la première contribution de cette thèse. Nous proposons une méthode pour générer des improvisations musicales unidimensionnelles nourries et guidées par des informations multidimensionnelles. L'objectif est de pouvoir effectuer un apprentissage sur des séquences multidimensionnelles afin de modéliser les relations entre différentes dimensions et utiliser cette information pour guider la navigation d'une mémoire musicale dans une situation d'improvisation. Cette méthode s'approche des travaux effectués autour de *SoMax* [Bonasse-Gahot, 2014] dans le sens où l'analyse de plusieurs dimensions peut permettre de prendre en compte des informations provenant à la fois d'une auto-écoute de l'improvisation générée et des informations de l'environnement musical issues d'une écoute active, formant ainsi un guidage réactif de l'improvisation. Cette méthode se distingue cependant par le fait que les différentes dimensions choisies ne sont pas séparées en différentes vues individuelles qui vont activer séparément des zones de la mémoire. Ici, nous voulons pouvoir représenter les liens entre les différentes dimensions afin de considérer des relations musicales plus complexes.

Nous présentons dans la partie 4.1 une utilisation des méthodes d'interpolation de sous-modèles appliquées à la prédiction de mélodies improvisées. Ces méthodes permettent une représentation des liens entre les différentes dimensions considérées. Nous évaluons le pouvoir de prédiction de ces méthodes de manière quantitative. Ensuite, dans la partie 4.2, nous présentons un paradigme de génération d'improvisations combinant des intuitions sur un contexte local et des connaissances multidimensionnelles. Ce paradigme est appliqué en combinant un oracle des facteurs et un modèle probabiliste multidimensionnel basé sur une interpolation de sous-modèles. Finalement, dans la partie 4.3 nous présentons une évaluation de notre méthode effectuée lors de sessions d'écoute avec des musiciens de jazz professionnels.

4.1 Apprentissage de connaissances multidimensionnelles

Dans cette partie, nous présentons une application des méthodes d'interpolation de sous-modèles probabilistes à la prédiction de mélodies improvisées. Nos travaux sont basés sur ceux de Raczyński et al. [2013a] présentés dans la partie 3.2. Dans la partie 4.1.1, nous rappelons le fonctionnement de la méthode en l'appliquant à la prédiction de mélodies improvisées, puis nous présentons les modèles et le corpus choisis. Dans la partie 4.1.2, nous évaluons notre modèle de manière quantitative.

4.1.1 Interpolation de modèles probabilistes pour la prédiction de mélodies improvisées

Rappel de la méthode

L'objectif est de créer un modèle probabiliste capable de prédire l'évolution mélodique d'une improvisation en utilisant des informations provenant de séquences multidimensionnelles. Par exemple, on imagine très bien que l'évolution mélodique d'une improvisation de style classique ou jazz va être intrinsèquement liée à une organisation harmonique. Nous voulons alors prédire la mélodie M_t jouée à l'instant t en considérant l'ensemble des variables musicales $X_{1:t}$ (provenant de plusieurs dimensions) sur l'ensemble des temps de 1 à t. Autrement dit, nous voulons estimer la probabilité $P(M_t|X_{1:t})$. Comme nous l'avons vu dans le chapitre précédent, l'estimation d'un tel modèle n'est pas faisable en pratique lorsqu'on utilise plusieurs dimensions sur plusieurs trames temporelles, car la combinatoire d'un tel modèle devient alors trop élevée [Whorley et al., 2013].

Nous allons alors approximer le modèle $P(M_t|X_{1:t})$ en effectuant une interpolation de sous-modèles P_i, chaque modèle étant basé sur un sous-ensemble différent de variables musicales $A_{i,t} \subset X_{1:t}$, plus facilement utilisable en pratique. L'interpolation des sous-modèles peut être effectuée de façon linéaire [Jelinek & Mercer, 1980], auquel cas nous avons

$$P(M_t|X_{1:t}) = \sum_{i=1}^{I} \lambda_i P_i(M_t|A_{i,t}) \ , \tag{4.1}$$

avec I est le nombre de sous-modèles interpolés et λ_i les coefficients d'interpolation tels que

$$\lambda_i \geq 0 \ \forall i \ \text{ et} \sum_{i=1}^{I} \lambda_i = 1 \ . \tag{4.2}$$

L'interpolation peut également être effectuée de façon log-linéaire [Klakow, 1998], auquel cas nous avons

$$P(M_t|X_{1:t}) = Z^{-1} \prod_{i=1}^{I} P_i(M_t|A_{i,t})^{\gamma_i} \ , \tag{4.3}$$

avec γ_i les coefficients d'interpolation tels que $\gamma_i \geq 0 \ \forall i$, et Z un facteur de normalisation défini par

$$Z = \sum_{C_t} \prod_{i=1}^{I} P_i(M_t|A_{i,t})^{\gamma_i} \ . \tag{4.4}$$

Les probabilités d'événements des modèles choisis sont estimées sur un corpus d'apprentissage à l'aide d'une fonction de comptage. Généralement, et en particulier dans le cas de l'improvisation dans un style donné, les corpus d'apprentissage sont de taille restreinte et il est donc fréquent que la distribution de certains éléments du corpus d'apprentissage ne corresponde pas à celle du corpus de test, voire que certains éléments du corpus de test n'apparaissent pas dans le corpus d'apprentissage. Les différents sous-modèles sont alors lissés pour éviter un sur-apprentissage (voir la partie 3.2.1 *Lissage des modèles* pour une présentation de différentes méthodes de lissage de modèles probabilistes). Cette méthode permet au système de considérer autant de sous-modèles que voulu (à condition qu'ils soient suffisamment simples). Les coefficients d'interpolation sont alors optimisés sur un corpus de validation, suivant une métrique choisie afin d'estimer au mieux le modèle global. Les sous-modèles porteurs du plus d'information vont recevoir un coefficient d'interpolation élevé et les sous-modèles peu porteurs de sens vont recevoir un coefficient d'interpolation proche de zéro (voir la partie 3.2.1 *Évaluation des modèles* pour une présentation de critères d'optimisation et d'évaluation des modèles).

Choix du corpus et des modèles

Pour valider l'utilisation de l'interpolation de sous-modèles pour la prédiction de mélodies improvisées, nous avons décidé d'effectuer un apprentissage sur les données issues du corpus de l'*Omnibook* présenté dans la partie 2.2. Nous avons utilisé des séquences multidimensionnelles issues de ce corpus contenant les informations mélodiques et harmoniques d'improvisations de Charlie Parker. Ceci permet d'avoir des données cohérentes et caractéristiques du style *bebop*. Nous avons divisé ce corpus en trois sous-corpus :

- un corpus d'apprentissage constitué de 40 thèmes et improvisations pour l'estimation des différentes probabilités des sous-modèles.
- un corpus de validation constitué de 5 thèmes et improvisations pour l'optimisation des coefficients d'interpolation et de lissage.
- un corpus de test constitué de 5 thèmes et improvisations pour l'évaluation de la capacité de prédiction du modèle.

Au vu des dimensions disponibles dans le corpus choisi, nous avons décidé d'utiliser deux sous-modèles très simples pour la prédiction de mélodies :

$$P_1(M_t|X_{1:t}) = P(M_t|M_{t-1}) \ , \tag{4.5}$$

$$P_2(M_t|X_{1:t}) = P(M_t|C_t) \ . \tag{4.6}$$

- P_1 est un bigramme sur la mélodie. Il est classique de considérer que la mélodie jouée à un instant t va être influencée par la mélodie qui la précède. Afin de ne pas être biaisé par l'apparition inégale des différentes tonalités dans le corpus ou par les problèmes de transposition,

nous allons considérer lors de l'apprentissage uniquement des relations relatives entre les mélodies. Ainsi, la probabilité de jouer un *fa* suivi d'un *sol* sera identique à la probabilité de jouer un *sol* suivi d'un *la*.

— P_2 représente les relations entre la mélodie et l'harmonie jouées à un même instant. Dans le style *bebop*, la mélodie est très fortement liée à l'harmonie, il nous a alors semblé important de représenter cette relation. De manière similaire au bigramme, pour éviter les problèmes de transposition, nous considérons des relations relatives entre la mélodie et l'harmonie. Ainsi, par exemple, la probabilité de jouer un *fa* sur un accord de F^{Δ} sera identique à la probabilité de jouer un *sol* sur un accord de G^{Δ}.

Apprentissage des modèles

Nous présentons ici comment nous avons effectué l'interpolation et l'apprentissage des sous-modèles P_1 et P_2 présentés dans la partie précédente pour estimer le modèle $P(M_t|X_{1:t})$. Ces modèles ont été interpolés linéairement ou log-linéairement en utilisant une combinaison de lissage additif et de lissage par repli avec $P(M_t)$ (ie. un unigramme) comme modèle d'ordre inférieur.

Pour l'interpolation linéaire [Jelinek & Mercer, 1980], on obtient le modèle suivant :

$$P(M_t|X_{1:t}) = \lambda_1 P(M_t|M_{t-1}) + \lambda_2 P(M_t|C_t) + \alpha P(M_t) + \beta U(M_t) \ , \quad (4.7)$$

où λ_1 et λ_2 sont les coefficients d'interpolation des deux sous-modèles, α et β sont respectivement les coefficients de lissage pour le lissage par repli et le lissage additif, avec

$$\lambda_1 + \lambda_2 + \alpha + \beta = 1 \ , \quad (4.8)$$

et U est la distribution uniforme.

Dans le cas de l'interpolation log-linéaire [Klakow, 1998], les sous-modèles doivent être lissés séparément, on obtient alors le modèle suivant :

$$P(M_t|X_{1:t}) = \frac{1}{Z} \prod_{i=1}^{2} (\delta_i P_i(M_t|A_{i,t}) + \zeta_i P(M_t) + \eta_i U(M_t))^{\gamma_i} \ , \quad (4.9)$$

où $\gamma_i \geq 0$ est le coefficient d'interpolation du modèle P_i, ζ_i et η_i sont respectivement les coefficients de lissage pour le lissage par repli et le lissage additif du modèle P_i, avec

$$\delta_i + \zeta_i + \eta_i = 1 \ \ \forall i, \quad (4.10)$$

et Z est la constante de normalisation définie par

$$Z = \sum_{M_t} \prod_{i=1}^{2} (\delta_i P_i(M_t|A_{i,t}) + \zeta_i P(M_t) + \eta_i U(M_t))^{\gamma_i} \ . \quad (4.11)$$

Les probabilités des sous-modèles P_1, P_2 et $P(M_t)$ sont estimées à partir du corpus d'apprentissage à l'aide d'une fonction de comptage :

$$P_1 = P(M_t|M_{t-1}) = \frac{\texttt{count}(M_{t-1}.M_t)}{\texttt{count}(M_{t-1})} \ , \tag{4.12}$$

$$P_2 = P(M_t|C_t) = \frac{\texttt{count}(M_t, C_t)}{\texttt{count}(C_t)} \text{ et} \tag{4.13}$$

$$P(M_t) = \frac{\texttt{count}(M_t)}{T} \tag{4.14}$$

Les coefficients d'interpolation et de lissage sont ensuite estimés sur le corpus de validation. Pour cela, nous les estimons de sorte à minimiser la mesure d'entropie croisée sur ce corpus :

$$H = -\frac{1}{T}\sum_{t=1}^{T} \log_2 P(M_t|X_{1:t}) \tag{4.15}$$

La convexité du problème n'étant pas certaine en particulier dans le cas de l'interpolation log-linéaire, les coefficients sont optimisés par la méthode du recuit simulé avec l'entropie croisée comme fonction d'énergie du système. Cette méthode est présentée dans l'Algorithme 1. Dans un cas convexe, une simple descente de gradient pourrait être utilisée.

Algorithme 1 Recuit simulé

$Temp \leftarrow Temp_0$
Initialisation aléatoire des coefficients d'interpolation et de lissage Θ
$H \leftarrow H(M, \Theta)$ ▷ Entropie croisée obtenue avec les coefficients Θ
Tant que $Temp > Temp_{\texttt{min}}$ **faire**
    $\Theta_{\texttt{nouveau}} \leftarrow \Theta + \Delta\Theta$ ▷ Légère variation des coefficients
    $\Delta H \leftarrow H(M, \Theta_{\texttt{nouveau}}) - H$
    Si $\Delta H < 0$ **ou** $\exp(-\Delta H/Temp) \geq U[0,1]$ **alors**
        $\Theta \leftarrow \Theta_{\texttt{nouveau}}$
        $H \leftarrow H(M, \Theta_{\texttt{nouveau}})$
    Fin Si
    $Temp \leftarrow 0.99 \times Temp$
Fin Tant que
retourner Θ, H

4.1.2 Évaluation des connaissances

Dans cette partie nous évaluons de façon quantitative les capacités de prédiction des modèles présentés dans la partie précédente. Pour chaque expérience, le modèle interpolé est comparé aux sous-modèles pris séparément. Après estimation des différentes probabilités des sous-modèles sur le corpus d'apprentissage et optimisation des coefficients des modèles sur le corpus de validation, nous utilisons ici le corpus de test pour obtenir les scores d'entropie croisée des modèles. Les tableaux présentés dans cette partie montrent les

Note à note	coefficients				Entropie croisée
	λ_1	λ_2	α	β	H
$P_1 + P_2$	**0,526**	**0,182**	**0,292**	**0**	**3,231**
P_1	0,801	0	0,199	0	3,285
P_2	0	0,756	0,244	0	3,300
unigramme seul	0	0	1	0	3,467

TABLEAU 4.1 – Entropie croisée (bits/note) pour la prédiction de mélodies improvisées note à note avec une interpolation linéaire. Les lignes du tableau présentent les résultats pour l'interpolation des sous-modèles de bigramme et mélodie/accord (P_1+P_2), puis pour le bigramme seul (P_1), puis pour le modèle mélodie/accord (P_2), puis pour l'unigramme seul.

coefficients obtenus après optimisation sur le corpus de validation et les scores d'entropie croisée obtenus sur le corpus de test.

Dans un premier temps, nous avons évalué la capacité de prédiction du modèle avec une interpolation linéaire dans le cas où la mélodie est considérée note à note. Dans ce cas, M_t correspond à la $t^{\text{ème}}$ note de la mélodie et C_t correspond à l'accord qui résonne lors de l'attaque de M_t (c'est-à-dire l'accord joué en même temps que la note s'il y en a un ou l'accord joué précédemment sinon). Nous obtenons alors les scores présentés dans le Tableau 4.1. Tout d'abord, nous pouvons constater que l'interpolation des deux sous-modèles permet d'obtenir un meilleur pouvoir de prédiction de mélodies improvisées au sens de l'entropie croisée que les sous-modèles pris séparément. Nous pouvons voir également que le bigramme sur la mélodie a une importance supérieure au sous-modèle relationnel entre mélodie et harmonie. Cela semble cohérent par rapport au style *bebop* du corpus. En effet, le *bebop* possède un langage très chromatique dont les phrases passent parfois par-dessus l'harmonie. Cependant, l'organisation des phrases est dirigée par cette harmonie, ce qui peut expliquer pourquoi une amélioration est obtenue dans le modèle interpolé. Nous pouvons aussi remarquer que le lissage additif n'est pas nécessaire dans cette situation ($\beta = 0$). Ceci peut être expliqué par le fait que l'ensemble des événements présents dans le corpus de test (que ce soit des couples de notes ou des couples note-accord) est également présent dans le corpus d'apprentissage.

Dans un second temps, nous avons testé comment l'interpolation s'adapte lorsque l'on considère la mélodie non pas note à note mais par trames temporelles. Nous avons fait varier la longueur de trame d'un demi-temps jusqu'à quatre temps. Dans ce cas, la mélodie M_t correspond à la liste des notes dont l'attaque a lieu dans la trame t, sans prendre en considération leur ordre temporel (c'est-à-dire que la séquence *fa sol la* jouée dans une même trame et la séquence *sol la fa* ont la même représentation). De même, C_t correspond à l'accord ou à la liste d'accords dont l'attaque a lieu dans la trame t, sans prendre en considération leur ordre temporel. Ici encore, nous utilisons une interpolation linéaire des modèles. Nous obtenons les scores présentés dans le Tableau 4.2. Encore une fois, nous pouvons constater que pour toutes les longueurs de trames considérées, l'interpolation des deux sous-modèles permet d'obtenir un

Trame = 1/2 temps	Coefficients				Entropie croisée
	λ_1	λ_2	α	β	$H(M)$
$P_1 + P_2$	**0.693**	**0.031**	**0.276**	**0**	**2.951**
P_1	0.714	0	0.286	0	2.956
P_2	0	0.623	0.377	0	3.224
unigramme seul	0	0	1	0	3.719
Trame = 1 temps	Coefficients				Entropie croisée
	λ_1	λ_2	α	β	$H(M)$
$P_1 + P_2$	**0.582**	**0.129**	**0.289**	**0**	**4.543**
P_1	0.672	0	0.328	0	4.572
P_2	0	0.639	0.361	0	4.881
unigramme seul	0	0	0.998	0.002	5.858
Trame = 2 temps	Coefficients				Entropie croisée
	λ_1	λ_2	α	β	$H(M)$
$P_1 + P_2$	**0.187**	**0.508**	**0.303**	**0.002**	**6.824**
P_1	0.392	0	0.602	0.006	7.382
P_2	0	0.671	0.327	0.002	6.937
unigramme seul	0	0	0.992	0.008	8.396
Trame = 4 temps	Coefficients				Entropie croisée
	λ_1	λ_2	α	β	$H(M)$
$P_1 + P_2$	**0.027**	**0.372**	**0.553**	**0.048**	**9.846**
P_1	0.082	0	0.762	0.156	10.500
P_2	0	0.390	0.557	0.053	9.868
unigramme seul	0	0	0.818	0.182	10.713

TABLEAU 4.2 – Entropie croisée (bits/trame) pour la prédiction de mélodies improvisées par trame avec interpolation linéaire. Les lignes du tableau présentent les résultats pour l'interpolation des sous-modèles de bigramme et mélodie/accord ($P_1 + P_2$), puis pour le bigramme seul (P_1), puis pour le modèle mélodie/accord (P_2), puis pour l'unigramme seul.

Note à note	coefficients		Entropie croisée
	γ_1	γ_2	$H(M)$
$P_1 + P_2$	**0,973**	**0,267**	**3,114**
P_1	1	0	3,285
P_2	0	1	3,300

TABLEAU 4.3 – Entropie croisée (bits/note) pour la prédiction de mélodies improvisées note à note avec une interpolation log-linéaire. Les lignes du tableau présentent les résultats pour l'interpolation des sous-modèles de bigramme et mélodie/accord (P_1+P_2), puis pour le bigramme seul (P_1), puis pour le modèle mélodie/accord (P_2).

meilleur pouvoir de prédiction de mélodies improvisées au sens de l'entropie croisée que les modèles pris séparément. Nous pouvons également remarquer que le choix de la longueur de trame a un impact important sur les coefficients des modèles. Le modèle de bigramme prend une importance plus grande pour des longueurs de trame courtes, alors que le modèle mettant en relation mélodie et harmonie prend le dessus pour des longueurs de trame plus élevées. Ceci peut s'expliquer par le fait que les accords changent en moyenne une fois toutes les mesures et donc que les trames de deux ou quatre temps caractérisent mieux l'évolution harmonique, alors que les trames d'un demi-temps ou d'un temps sont plus adaptées pour considérer une évolution mélodique. De plus, le lissage devient de plus en plus important lorsque la longueur de trame s'agrandit. Cela s'explique par le fait que les séquences considérées pour chaque trame deviennent de plus en plus longues et donc que la taille du vocabulaire s'agrandit en même temps que le nombre de trames considérées diminue. Il est important de noter que l'on ne peut pas comparer les scores d'entropie croisée pour des trames de longueurs différentes, car la taille de vocabulaire varie pour chaque taille de trame, ce qui a un impact dans le calcul de l'entropie croisée.

Finalement, nous avons voulu comparer la capacité de prédiction du système lorsque l'on utilise une interpolation log-linéaire dans le cas où la mélodie est considérée note à note. Nous obtenons les résultats présentés dans le Tableau 4.3. De manière similaire à l'interpolation linéaire, nous obtenons un meilleur pouvoir de prédiction de mélodies improvisées au sens de l'entropie croisée pour le modèle interpolé que pour les sous-modèles pris séparément. Encore une fois, les séquences étant prises note à note, le bigramme sur la mélodie a une importance supérieure au sous-modèle relationnel entre mélodie et harmonie. Cependant, si l'on compare les résultats obtenus entre l'interpolation linéaire et l'interpolation log-linéaire, nous pouvons remarquer que l'interpolation log-linéaire permet un meilleur pouvoir de prédiction au sens de l'entropie croisée. Bien qu'elle soit plus lourde en calcul du fait du plus grand nombre de coefficients et par la normalisation par Z, elle permet d'obtenir une amélioration relative du pouvoir de prédiction de 3,6% au sens de l'entropie croisée.

De manière générale, ces résultats sont encourageants. Bien que l'amélioration en terme d'entropie croisée est assez faible, l'interpolation de sous-modèles

permet d'obtenir un meilleur pouvoir de prédiction dans l'ensemble des cas étudiés, alors que nous utilisons seulement deux sous-modèles très simples. Les meilleurs résultats sont obtenus avec l'utilisation de l'interpolation log-linéaire. Nous utiliserons alors, par la suite, l'interpolation log-linéaire note-à-note pour les applications présentées dans les parties suivantes nécessitant un apprentissage de connaissances multidimensionnelles. Il est cependant difficile de définir si l'entropie croisée rend correctement compte de la qualité perçue dans le cas de la musique improvisée. L'entropie croisée mesure la capacité de prédiction et donc de reproduction d'un système, alors que l'improvisation n'est pas une pratique basée uniquement sur la reproduction, mais sur la variété dans l'expressivité du musicien. De plus, le modèle probabiliste forme un ensemble de connaissances globales qui ne prend pas en compte l'organisation locale d'une improvisation. Nous avons alors décidé de combiner ce modèle de connaissances avec un oracle des facteurs représentant une mémoire locale.

4.2 Diriger le parcours d'un oracle par des connaissances

4.2.1 Paradigme Intuition / Connaissance

Dans cette partie, nous présentons un paradigme d'improvisation automatique où la génération d'une improvisation s'effectue d'une manière inspirée de la pratique humaine. L'idée principale est de créer un agent possédant à la fois des connaissances globales apprises sur un large corpus et capable de s'adapter à un contexte local construit en ligne à partir du jeu d'un musicien ou hors-ligne sur un corpus de petite taille. De cette manière, l'agent peut enrichir son intuition sur un contexte local grâce à des connaissances passées. La formalisation de ce paradigme a été inspirée par un des éléments d'improvisation de Crispell [2000] (écrit pour Cecil Taylor et Anthony Braxton) :

> « Le développement d'un motif doit être fait d'une manière logique, organique et ordonnée (improvisation en tant que composition spontanée), pas cependant d'une manière préconçue, mais plutôt d'une manière basée sur l'intuition enrichie par des connaissances (de tout ce qui a été joué, écouté, des différents styles musicaux auxquels on a été exposé, etc., au cours d'une vie — y compris toutes les expériences personnelles) ; le résultat est un vocabulaire musical personnel. » [5]

Nous proposons ici une implémentation de ce paradigme combinant un modèle probabiliste obtenu par une interpolation de sous-modèles et un oracle des facteurs. Cette implémentation nous permet de bénéficier à la fois de l'apprentissage multidimensionnel des modèles probabilistes et des heuristiques développées par Assayag & Bloch [2007] pour le parcours d'un oracle des facteurs pour la génération d'une improvisation.

D'un côté, le modèle probabiliste multidimensionnel est créé pour représenter les connaissances et le style culturel d'un musicien que l'on souhaite émuler.

5. Traduit de l'anglais vers le français par moi-même.

Un ensemble de sous-modèles est sélectionné à partir des dimensions musicales que l'on souhaite prendre en considération. Ces sous-modèles sont interpolés et lissés pour créer le modèle probabiliste global (voir la partie précédente). Ce modèle probabiliste est appris hors-ligne, avant la génération, sur un corpus relativement important de séquences multidimensionnelles représentatives du style musical souhaité. Le modèle probabiliste fournit des connaissances plus larges de la musique que l'oracle des facteurs, grâce à son apprentissage sur un corpus plus important et permet au système de considérer des informations multidimensionnelles. Cela permet de considérer pendant la génération des structures de plus haut-niveau, comme par exemple des relations harmoniques, au lieu de simples séquences logiques.

D'un autre côté, un oracle des facteurs est construit, en ligne à partir du jeu d'un musicien, ou hors-ligne sur un corpus de petite taille (généralement, un morceau), de manière similaire à *OMax*. Les transitions de l'oracle des facteurs sont construites en suivant la dimension que l'on souhaite générer. Cela représente le contexte local de l'improvisation. L'oracle des facteurs impose un développement séquentiel logique et organique des motifs qui sont générés et permet au système de prendre en compte un contexte plus long que le modèle probabiliste. Cela est garanti par les liens suffixiels connectant chaque état de l'oracle avec l'état précédent partageant le contexte commun le plus long et par les heuristiques développées pour *OMax* assurant l'utilisation de liens suffixes reliant des états avec une longueur de contexte supérieure à un minimum.

La combinaison de ces deux éléments permet alors de générer des improvisations unidimensionnelles suivant la dimension sur laquelle les transitions de l'oracle des facteurs sont construites, cependant, ces improvisations sont guidées par les connaissances multidimensionnelles du modèle probabiliste. De plus, la modélisation de plusieurs dimensions par le modèle probabiliste permet au système de pouvoir prendre en considération des informations de l'environnement musical lors de la génération. Ainsi, ce système est adapté pour le guidage réactif de l'improvisation.

Grâce à ce paradigme, il est également possible de considérer la création d'un musicien virtuel hybride où les connaissances stylistiques apprises par le modèle probabiliste ne correspondent pas au style présent dans le contexte local, en faisant varier les corpus d'apprentissage. Ainsi, sur un même contexte local (c'est-à-dire même oracle des facteurs), deux agents ne possédant pas le même modèle probabiliste, que ce soit au niveau du corpus utilisé ou des sous-modèles considérés, ne guideront pas l'improvisation de la même manière et donc entraîneront des improvisations différentes.

4.2.2 Guider l'oracle des facteurs à partir de ses connaissances

Dans cette partie, nous rentrons dans les détails du guidage de l'improvisation par un modèle probabiliste dans un oracle des facteurs (voir la partie 3.1.2 pour une présentation de l'oracle des facteurs).

Soit un modèle probabiliste P capable d'estimer la probabilité d'une dimension à partir d'un ensemble de variables musicales issues de dimensions différentes, appris sur un large corpus de séquences multidimensionnelles (voir la partie 4.1 pour la construction du modèle probabiliste). Soit $\mu = \mu_1 ... \mu_m$ une

séquence musicale multidimensionnelle jouée en direct par un musicien ou lue dans un corpus de petite taille. Chaque élément μ_i de la séquence est constitué d'un ensemble de dimensions : $\mu_i = \{\mu_i^1, \mu_i^2, ..., \mu_i^l\}$. L'oracle des facteurs est créé en ne suivant qu'une de ces dimensions (celle sur laquelle on souhaite improviser) avec l'Algorithme 2 [Allauzen et al., 1999; Lefebvre & Lecroq, 2000]. L'oracle est constitué d'états allant de 0 à m, l'état i possédant le contenu musical μ_i. Notons $S(i)$ l'état cible du lien suffixiel sortant de l'état i. Notons également $S^{-1}(i)$ l'ensemble des états dont le lien suffixiel a pour cible l'état i.

Algorithme 2 Construction d'un oracle des facteurs

Entrée : $\mu = \mu_1...\mu_m$ ▷ Dans cet algorithme, pour les étiquettes, nous ne considérons que la dimension choisie.
Sortie : Oracle des facteurs de μ
1: Créer l'état 0
2: $S(0) \leftarrow -1$
3: **Pour** i allant de 1 à m **faire**
4: Créer l'état i
5: Créer une transition de l'état $i-1$ à l'état i étiquetée par μ_i
6: $k \leftarrow S(i-1)$
7: **Tant que** $k > -1$ **et** qu'il n'existe pas de transition partant de k étiquetée par μ_i **faire**
8: $k \leftarrow S(k)$
9: **Fin Tant que**
10: **Si** $k = -1$ **alors**
11: $s \leftarrow 0$
12: **Sinon**
13: $s \leftarrow$ la cible de la transition partant de k étiquetée par μ_i
14: **Fin Si**
15: $S(i) \leftarrow s$
16: **Fin Pour**

L'improvisation générée à partir de l'oracle sera alors une nouvelle séquence de contenu musical $\mu_{t_1}\mu_{t_2}...\mu_{t_T}$ jouée aux instants allant de 1 à T. Posons également $\nu_1\nu_2...\nu_T$ les événements de l'environnement au cours de la génération. Ces éléments sont également constitués d'un ensemble de dimensions. Le choix des dimensions considérées pour les contenus musicaux et pour les événements de l'environnement dépendent des dimensions qui ont un impact dans P.

Considérons que le système a improvisé sur les instants de 1 à $T-1$ et qu'à l'instant $T-1$, l'état courant dans l'oracle est l'état $t_{T-1} = i$. Nous souhaitons alors générer sur l'instant T. La première étape de la navigation consiste à déterminer, à partir de l'état courant i, l'ensemble des états atteignables $\mathrm{Att}(i)$. Cette étape est effectuée en suivant les heuristiques présentées dans Assayag & Bloch [2007]. Tout d'abord, on construit un ensemble d'états constitué de l'état $i+1$ et de la liste des états suivant ceux connectés à i par un réseau de liens suffixiels. Ce réseau d'états connectés est constitué de :

1. $S(i)$, l'état cible du lien suffixe de i,

2. $S^{-1}(i)$, l'ensemble des états cibles des liens suffixes inverses de i,
3. $S^{-1}(S(i))$, l'ensemble des états cibles des liens suffixes inverses de $S(i)$
4. les états reliés par les liens suffixiels inverses sortant des états obtenus précédemment,
5. les états obtenus par une itération de l'étape 4, jusqu'à obtention de l'état 0 de l'oracle des facteurs.

Des contraintes sont alors appliquées à cet ensemble d'états pour déterminer les états atteignables. Les contraintes appliquées sont :

— la mise en place d'un facteur de continuité : ce facteur permet d'éviter des sauts trop fréquents dans le parcours de l'oracle des facteurs. Ainsi, si le facteur de continuité à l'instant courant est inférieur à un seuil fixé, seul l'état $i+1$ suivant l'état courant i sera atteignable.

— une longueur minimale de contexte commun : les liens suffixiels reliant deux états partageant un contexte commun d'une longueur inférieure à un seuil fixé ne sont pas pris en considération. Cela permet de mieux garantir la cohérence musicale lorsque des sauts sont effectués dans le parcours de l'oracle des facteurs.

— la mise en place d'une liste tabou : cette liste permet d'éviter les boucles lors de la navigation de l'oracle des facteurs et donc d'éviter trop de répétitions dans la génération de l'improvisation. La liste tabou correspond aux M dernières cibles de sauts dans la navigation ainsi que leur voisinage. La taille M de la liste est fixée arbitrairement.

L'étape de détermination des états atteignables est résumée dans l'Algorithme 3.

Algorithme 3 États atteignables à partir de l'état i

Entrée : État i, facteur de continuité minimum $cf_{\texttt{min}}$, facteur de continuité courant cf, contexte commun minimum $cc_{\texttt{min}}$, liste tabou $Tabou$
Sortie : Liste des états atteignables Att(i)

Si $cf < cf_{\texttt{min}}$ **alors**
    Ajouter l'état $i+1$ à Att(i)
Sinon
    Ajouter l'état $i+1$ et la liste des états suivants ceux connectés à i par un réseau de liens suffixiels à Att(i)
    Pour chaque état j dans Att(i) **faire**
        Si la taille du contexte commun entre $i+1$ et $j < cc_{\texttt{min}}$ **alors**
            Retirer l'état j de Att(i)
        Fin Si
        Si j est dans $Tabou$ **alors**
            Retirer l'état j de Att(i)
        Fin Si
    Fin Pour
Fin Si
retourner Att(i)

Une fois Att(i) obtenu, nous souhaitons obtenir les probabilités de transitions $P(t_T = j | t_{T-1} = i, \mu_{t_1} ... \mu_{t_{T-1}}, \nu_1 ... \nu_T)$ pour chaque état $j \in$ Att(i),

où i est l'état joué à l'instant $T-1$ et j l'état joué à l'instant T. Pour cela, l'ensemble du contenu musical μ_i de l'état i, l'ensemble du contenu musical μ_j pour chaque état atteignable, l'ensemble du contenu musical de l'improvisation $\mu_{t_1}\mu_{t_2}...\mu_{t_{T-1}}$ ainsi que les événements issus de l'environnement $\nu_1\nu_2...\nu_T$ sont envoyés au modèle probabiliste, afin d'obtenir un score pour chaque transition possible. La longueur du passé de l'improvisation générée et des événements de l'environnement envoyé au modèle probabiliste dépend du nombre de trames passées ayant un impact dans $\mathcal{P}$. Finalement, ces scores sont normalisés pour obtenir les probabilités de transition vers chaque état j. Enfin, pour la génération, une transition est choisie au hasard en suivant ces probabilités. Cela permet de privilégier les transitions à forte probabilité tout en conservant un indéterminisme dans la génération de l'improvisation. La Figure 4.1 illustre le processus global de la génération de l'instant T à partir de l'instant $T-1$.

Par exemple, supposons que l'on souhaite générer la mélodie M_{t_T} de manière réactive à une harmonie C_T de l'environnement. Notre modèle probabiliste est une interpolation des modèles

$$P_1(M_{t_T}|\mu_{t_1}...\mu_{t_{T-1}},\nu_1...\nu_T) = P(M_{t_T}|M_{t_{T-1}}) \text{ et} \tag{4.16}$$

$$P_2(M_{t_T}|\mu_{t_1}...\mu_{t_{T-1}},\nu_1...\nu_T) = P(M_{t_T}|C_T) \ . \tag{4.17}$$

L'oracle des facteurs est construit selon la dimension mélodique. Supposons que l'improvisation jusqu'à l'instant $T-1$ nous a amené à l'état i de l'oracle des facteurs et a généré la séquence mélodique $\mu^M_{t_1}...\mu^M_{t_{T-1}}$. On détermine à partir des propriétés de l'oracle des facteurs les états atteignables $\mathrm{Att}(i)$. À l'instant T, l'harmonie jouée dans l'environnement est ν^C_T. Pour chaque état j de $\mathrm{Att}(i)$, le score $P(M_{t_T} = \mu^M_j|\mu_{t_1}...\mu_i,\nu_1...\nu_T)$ pour la transition $i \to j$ correspond à l'interpolation de

$$P_1(M_{t_T} = \mu^M_j|M_{t_{T-1}} = \mu^M_i) \tag{4.18}$$

et de

$$P_2(M_{t_T} = \mu^M_j|C_T = \nu^C_T) \tag{4.19}$$

Les scores sont normalisés pour obtenir les probabilités de transition, on a alors :

$$P(t_T = j|t_{T-1} = i,\mu_{t_1}...\mu_{t_{T-1}},\nu_1...\nu_T) = \frac{P(M_{t_T} = \mu^M_j|\mu_{t_1}...\mu_i,\nu_1...\nu_T)}{\sum_{k\in\mathrm{Att}(i)} P(M_{t_T} = \mu^M_k|\mu_{t_1}...\mu_i,\nu_1...\nu_T)} \ . \tag{4.20}$$

Finalement, chaque agent prend une décision aléatoire en suivant les probabilités de transitions.

Cette méthode de navigation permet donc bien de suivre les informations contextuelles fournies par l'oracle sur la dimension improvisée, tout en pouvant prendre en considération un ensemble d'autres dimensions provenant soit du jeu de l'improvisation sur l'ensemble des dimensions, soit d'événements issus de l'environnement. De plus, contrairement à *OMax* qui infère uniquement sur le passé de la mémoire dans l'oracle des facteurs, cette méthode permet également

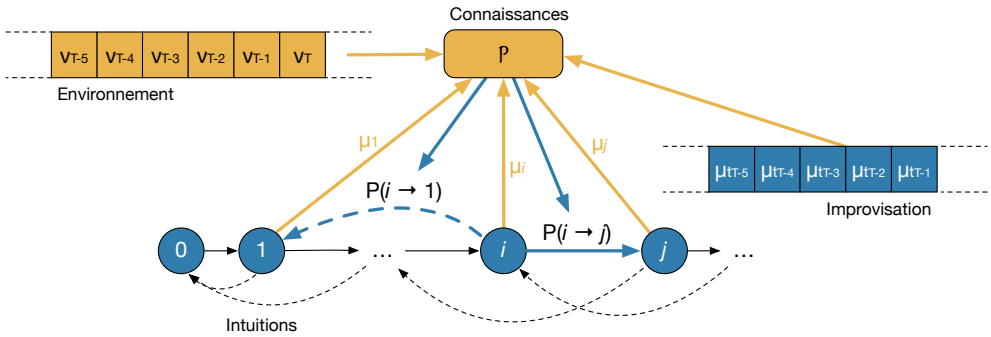

FIGURE 4.1 – Utilisation d'un modèle probabiliste multidimensionnel $\mathcal{P}$ avec un oracle des facteurs. Dans cet exemple, à partir de l'état i, supposons que l'oracle fournit les états atteignables 1 et j. En utilisant les informations des événements de l'environnement, de l'improvisation passée, du contenu musical de l'état actuel et du contenu musical des états atteignables, le modèle probabiliste calcule les probabilités de transitions pour chaque état atteignable. La transition effectuée est choisie au hasard en suivant ces probabilités.

d'inférer sur le passé de l'improvisation générée ($\mu_{t_1}...\mu_{t_{T-1}}$). De cette manière, les intuitions du parcours de l'oracle des facteurs sont enrichies par les connaissances du modèle probabiliste. Ainsi, nous pouvons générer une improvisation unidimensionnelle prenant en compte des informations multidimensionnelles et pouvant profiter d'un guidage réactif.

4.3 Évaluation et retours des musiciens

Pour évaluer notre modèle, nous avons effectué deux expériences. Pour chaque expérience, nous avons pratiqué un test d'écoute avec des musiciens professionnels afin d'obtenir une évaluation qualitative des générations. Les improvisations générées se basent sur le style de l'*Omnibook* de Charlie Parker (voir partie 2.2). Nous avons fait appel à trois musiciens et improvisateurs de jazz professionnels :

— Pascal Mabit, saxophoniste et professeur de jazz, diplômé du Conservatoire National Supérieur de Musique et de Danse de Paris,
— Louis Bourhis, contrebassiste de jazz, diplômé de la Haute École de Musique de Lausanne,
— Joël Gauvrit, pianiste et professeur de jazz et de musique classique, diplômé du Conservatoire National Supérieur de Musique et de Danse de Lyon. [6]

Les tests d'écoute avec les musiciens ont été faits séparément. Étant des jazzmen professionnels, ces musiciens sont très familiers avec la musique de Charlie Parker et sont donc qualifiés pour fournir des retours pertinents sur les improvisations générées.

6. Des biographies plus détaillées des musiciens sont fournies en Annexe B.

4.3.1 Sur le guidage de l'improvisation avec un modèle probabiliste

Choix des paramètres

Tout d'abord, nous avons voulu vérifier si l'utilisation d'un modèle probabiliste permet de guider de manière efficace la navigation dans un oracle des facteurs. Nous nous intéressons en particulier dans cette expérience au guidage d'une mélodie par une harmonie issue de l'environnement. Pour cela nous avons généré des improvisations libres en utilisant l'*Omnibook* de Charlie Parker (voir partie 2.2) comme corpus. Ce corpus a été divisé en trois sous-corpus : un corpus d'apprentissage constitué de 40 thèmes et improvisations, un corpus de validation constitué de 5 thèmes et improvisations et un corpus de test constitué de 5 thèmes et improvisations. L'apprentissage du modèle probabiliste a été fait en utilisant une interpolation log-linéaire de deux sous-modèles :

— $P_1(M_{t_T}|\mu_{t_1}...\mu_{t_{T-1}}, \nu_1...\nu_T) = P(M_{t_T}|M_{t_{T-1}})$: un bigramme sur la mélodie,

— $P_2(M_{t_T}|\mu_{t_1}...\mu_{t_{T-1}}, \nu_1...\nu_T) = P(M_{t_T}|C_T)$: un modèle représentant les liens entre la mélodie et l'harmonie.

Les probabilités des sous-modèles ont été estimées sur le corpus d'apprentissage. Les sous-modèles ont ensuite été lissés par un mélange de lissage par repli et de lissage additif, puis les coefficients d'interpolation et de lissage ont été optimisés sur le corpus de validation (voir partie 4.1.1). Les improvisations ont été générées en suivant deux méthodes :

1. un oracle des facteurs seul construit sur la mélodie d'un morceau du corpus de test,
2. un oracle des facteurs construit sur la mélodie du même morceau combiné avec le modèle probabiliste.

La grille d'accords du morceau selectionné est alors fournie au système comme séquence d'événements de l'environnement et est donc suivie lors de l'utilisation du modèle probabiliste dans la méthode 2. Cependant, pour ne pas biaiser l'écoute des musiciens avec la méthode 1, pour laquelle la mélodie jouée n'est pas informée par la grille, les accords ne sont pas joués, seule la mélodie est entendue pour les deux méthodes. Des exemples de génération pour cette expérience sont disponibles sur : `http://repmus.ircam.fr/dyci2/demos/probabilistic_fo`.

Résultats et analyse

Chaque musicien a écouté une douzaine d'improvisation, plus ou moins, selon son vouloir, en utilisant les deux méthodes présentées précédemment, à partir de deux morceaux issus du corpus de test : *Anthropology* (un anatole) et *Donna Lee* (une grille plus complexe avec un langage mélodique très chromatique).

Dès les premiers exemples, les trois musiciens ont noté une différence claire entre les deux méthodes par rapport à l'organisation des phrases dans l'improvisation. D'un côté, lors de l'écoute d'exemples générés à partir de la première méthode, les trois musiciens soulignent un manque de cohérence et d'unité de

l'improvisation. Les différents éléments de l'improvisation rappellent effectivement le style de Charlie Parker, mais l'arrangement de ces différents éléments manque de sens. Mabit dit :

> « Pour l'instant c'est un peu l'idée que je me ferais de ce qu'on ferait d'un solo de Charlie Parker si on le mettait dans un mixeur, on appuie sur le bouton et ça redistribue, ça éclate un peu la chose. Mais de temps en temps, on peut reconnaître des bribes. A priori, ce n'est pas l'idée que j'aurais d'une improvisation dans le style de Charlie Parker, j'ai plutôt l'impression que ça ressemblerait plus à une pièce de musique contemporaine dans laquelle on aurait intégré, par un moyen ou un autre, des éléments de ces choses-là, un peu à la manière d'un *zapping*. Quelque chose d'un peu aléatoire. Il n'y a pas forcément une notion ou une sensation d'unité. »

Bourhis utilise une image similaire en faisant remarquer que les improvisations générées par cette méthode forment un « *patchwork* » d'éléments caractéristiques du jeu de Charlie Parker et du morceau joué, mais qui ne respecte pas la construction du morceau. Il remarque cela particulièrement sur *Anthropology* où l'improvisation joue une triade de D, caractéristique du passage sur le pont, à des moments complètement inappropriés qui font « sortir complètement du défilement » du morceau. De plus, Mabit remarque que, par conséquent, l'harmonie n'est pas claire, ou arrangée de manière aléatoire dans la conduite mélodique de l'improvisation (à l'exception des passages où l'improvisation fait de longues citations du thème) :

> « La notion d'évolution harmonique, elle prend son sens dans une continuité. Une succession d'accords a un sens pour nous parce qu'elle est basée sur un système de tension/résolution. Pour l'instant, le logiciel ne prend pas tout à fait cela en compte, ou alors il juxtapose ça de manière complètement aléatoire. On n'entend pas vraiment d'harmonie. On entend du note pour note, ou une suite de phrases. Mais même dans les phrases, il n'y a pas forcément toujours de notion harmonique, mis à part quand il place une citation du thème, alors là on entend les polarités. »

À l'inverse, en écoutant des exemples générés à partir de la deuxième méthode, Mabit était capable de savoir sur quel accord l'improvisation était jouée en se basant uniquement sur la mélodie. En particulier, il a trouvé qu'il y avait une succession claire des centres tonaux. Malgré cela, l'improvisation préserve le style global de Charlie Parker grâce au contexte local fourni par l'oracle des facteurs. Sur *Donna Lee*, Mabit a également noté que cette méthode n'est pas non plus trop contrainte par l'harmonie et arrive à exprimer l'ensemble des tonalités :

> « Déjà, effectivement, on entend plus une succession d'accords, en tout cas, des centres tonaux. Là tout le début était complètement en Bb majeur, avec effectivement des tournures de phrases, des ornementations, des choses comme ça qui font vraiment langage *bebop*

> et Charlie Parker. [...] on entend vraiment le passage au quatrième degré, il fait sonner le quatrième degré mineur et voilà il revient. »

Bourhis fait exactement la même analyse sur la clarté des développements harmoniques sur *Donna Lee*. Il fait remarquer particulièrement que cette clarté permet un suivi beaucoup plus aisé de l'improvisation et permettrait donc d'accompagner le musicien de façon plus confortable :

> « On est beaucoup plus à la maison quand on entend ça, si je peux me permettre cette expression. Par exemple, on entend clairement la modulation au quatrième degré ou le relatif mineur sur les deux endroits qui sont les plus caractéristiques. Il les fait dans l'ordre. On voit bien qu'il suit quelque chose. Ce qui fait que beaucoup plus facilement, on comprend. »

Gauvrit, s'intéressant plus à l'organisation mélodique des phrases et des phrases entre elles, remarque également une amélioration dans le suivi, avec l'utilisation de la deuxième méthode. Les phrases lui semblent moins décousues et plus structurées, avec une notion de développement de motifs et de « suivi dans les idées ». Bourhis dit que cette organisation donne au système une plus grande « crédibilité ». Gauvrit précise :

> « J'ai le sentiment que les matériaux sont plus développés, qu'il y a plus d'unité dans ce qui est fait, qu'il reprend des éléments qu'il vient de faire, qu'il les reprend en renversant, en miroir. Les phrases sont longues, c'est plus conjoint. [...] C'est plus cohérent. [...] On a l'impression qu'il développe son idée quoi. Qu'il y a une idée qui est développée. Pas sur l'ensemble, mais du coup, quelque chose qui ressemble assez à la réalité parce que tu as une idée qui en amène une autre, qui en amène une autre, qui en amène une autre... »

Cependant, en plus de quelques erreurs harmoniques, certains moments des improvisations générées restent parfois flous, en particulier sur le pont d'*Anthropology* à cause d'un manque de compréhension de la forme globale de la progression d'accords. Plus généralement, les improvisations ont du sens d'un point de vue harmonique à une échelle locale, mais manquent de construction et de logique par rapport à leur position dans la grille. Cette remarque était attendue dans le sens où ce problème existe dans tous les systèmes d'improvisation basés sur *OMax* et notre méthode ne cherche pas à résoudre ce problème. Mabit dit :

> « Une fois qu'il aura compris le principe de forme, ça sera encore mieux, parce que là, pour l'instant, j'ai l'impression qu'il prend les accords les uns après les autres. Mais il ne sait pas plus, il n'a pas de notion de fonctions, tout ça. Ce qu'il fait, ça fonctionne avec les accords mais ce n'est quand même pas toujours très sensé. »

Bourhis déplore également un manque d'anticipation harmonique du système. La génération n'est pas dirigée par l'accord suivant et par conséquent n'effectue pas de conduites vers l'harmonie à venir. Bourhis à l'impression que l'improvisateur « sait ce qu'il fait, mais pas où il va ». Encore une fois, cette remarque

était attendue, car le système possède un guidage purement réactif sans notion d'anticipation. L'introduction de principes d'anticipation similaires à ceux d'*ImproTek* [Nika et al., 2017a] ou des travaux de Nika et al. [2017b] pourrait potentiellement permettre de résoudre ce problème. Bourhis dit :

> « Il y a certains trucs où tu as l'impression qu'il est surpris. [...] Il ne dirige pas ses phrases, cela expliquerait pourquoi il fait des résolutions, il arrive sur un accord qu'il n'avait pas anticipé et du coup il fait des chromatismes pour repartir autre part. C'est un peu l'impression qui est donnée. »

Pour conclure, Bourhis compare les deux méthodes en faisant le parallèle avec des élèves de classe de jazz en disant :

> « T'en as qui fait un peu n'importe quoi ou qui a entendu un peu ce morceau et qui du coup fait un peu ce dont il peut se souvenir là où son oreille veut l'emmener, alors que l'autre aurait beaucoup plus travaillé. C'est ça l'impression que j'ai. C'est assez marrant à voir. »

Cette première expérience montre des résultats encourageants. L'impact du modèle probabiliste sur le guidage de l'oracle des facteurs est entendu clairement par les jazzmen professionnels et les improvisations générées avec ce modèle sont préférées par rapport aux improvisations générées sans ce modèle. Certaines limitations communes aux deux méthodes ont été relevées, en particulier sur le manque de considération de la forme globale et d'anticipation.

4.3.2 Sur le choix du corpus

Choix des paramètres

Nous avons effectué une deuxième expérience afin de valider si le paradigme Intuition / Connaissance permet effectivement de créer des musiciens virtuels hybrides. Nous souhaitons déterminer si une différence pouvait être entendue lors de l'utilisation de corpus différents pour l'apprentissage du modèle probabiliste. Encore une fois, nous nous intéressons en particulier au guidage d'une mélodie par une harmonie issue de l'environnement. Pour cela nous avons utilisé deux corpus différents : le corpus de l'*Omnibook* divisé en trois sous-corpus identiquement à l'expérience précédente ; et un corpus de musique classique basé sur Wikifonia, contenant les informations mélodiques et harmoniques que nous avons divisé en deux sous-corpus, un corpus d'apprentissage constitué de 850 morceaux et un corpus de validation constitué de 75 morceaux. Pour chaque corpus, un modèle probabiliste différent est créé en faisant une interpolation log-linéaire des deux sous-modèles présentés dans l'expérience précédente. Encore une fois, les probabilités des sous-modèles ont été estimées sur le corpus d'apprentissage. Les sous-modèles ont ensuite été lissés par le même mélange de lissage par repli et de lissage additif, puis les coefficients d'interpolation et de lissage ont été optimisés sur le corpus de validation. Les improvisations ont été générées en suivant deux méthodes :

1. un oracle des facteurs construit sur la mélodie d'un morceau issu du corpus de test de l'*Omnibook* avec un modèle probabiliste appris sur le corpus de l'*Omnibook*,
2. un oracle des facteurs construit sur la mélodie du même morceau issu du corpus de test de l'*Omnibook* avec un modèle probabiliste appris sur le corpus de musique classique.

Comme dans la première expérience, la grille d'accords du morceau sélectionné est fournie au système comme séquence d'événements de l'environnement pour permettre un guidage réactif du système. Cependant, dans cette expérience la progression d'accords est jouée en arrière-plan de l'improvisation pour faire ressortir les relations entre la mélodie et l'harmonie. Les improvisations générées ne possèdent pas d'informations rythmiques (seules des croches et des demi-pauses sont jouées) afin d'éviter des décalages rythmiques. Des exemples de génération pour cette expérience sont disponibles sur : `http://repmus.ircam.fr/dyci2/demos/corpus_choice`.

Résultats et analyse

Chaque musicien a écouté une douzaine d'improvisations, plus ou moins, selon son vouloir, sur les mêmes morceaux que dans la première expérience : *Anthropology* et *Donna Lee*.

En premier lieu, Mabit a remarqué que l'idée générale du style de Charlie Parker est toujours conservée, même lors de l'utilisation du corpus de musique classique. Cela peut être expliqué par une dominance du contexte local fourni par l'oracle des facteurs. Les éléments de l'improvisation provenant de Charlie Parker sont forts et définissent le centre d'intérêt principal. Cependant, après plusieurs écoutes, Mabit trouve que, lorsque le corpus de musique classique est utilisé, les improvisations semblent viser plus les notes des accords que lorsque le corpus de l'*Omnibook* est utilisé. Les improvisations semblent plus prudentes et par conséquent plus logiques d'un point de vue harmonique. Mabit exprime alors une préférence pour les improvisations générées avec le corpus de musique classique :

> « Le plus crédible à mon avis, c'est celui avec le corpus de musique classique parce que, tout simplement, à l'oreille ça marche mieux. Ça marche mieux, parce qu'il y a aussi une plus grande prise en compte des espaces harmoniques, il y a une plus grande prise en compte de ce qui se passe sur chaque accord. [...] Du coup à l'oreille ce qu'il se passe, c'est qu'on a quand même l'impression d'avoir quelqu'un qui joue l'harmonie, mais qui prend un peu de libertés. »

Gauvrit souligne la différence entre les deux corpus de manière similaire, en faisant remarquer que, lorsque le corpus classique est utilisé, les improvisations générées donnent des choses « moins altérées ». Cependant, en terme de cohérence par rapport au style de Charlie Parker, les improvisations générées avec le corpus de l'*Omnibook* sont plus crédibles. Gauvrit exprime alors une préférence pour les improvisations générées avec le corpus de l'*Omnibook*, jugeant les improvisations générées avec le corpus de musique classique moins surprenantes et du coup moins intéressantes :

> « Effectivement, il doit y avoir moins d'incohérences harmoniques pures avec le corpus classique qu'avec le corpus *Omnibook*. [...] Il y a pas mal de trucs qui sonnent *out*, un peu tordu, dans le deuxième, mais en même temps, c'est ça qui fait que ça sonne crade un peu, mais c'est ça qui fait aussi un peu plus jazz. [...] C'est moins surprenant celui qui a appris le classique et plus juste en même temps. Plus dans les clous, plus académique. »

Gauvrit fait également remarquer qu'avec le corpus de musique classique, les improvisations évoquent plutôt un style antérieur au *bebop*, plus proche des chansons de l'origine du jazz, estompant alors le style de Charlie Parker :

> « Dans celui qui a appris le classique ça fait des choses qui sont hors style, je trouve. [...] Régulièrement ça me fait penser à des choses qui sont effectivement très très peu altérées qui font penser à Gershwin dans le texte. Tu vois ? Plutôt *Broadway*. »

Bourhis précise l'idée de surprise générée par les deux modèles exprimée par Gauvrit, en remarquant que les improvisations générées avec le corpus de l'*Omnibook* donnent les résultats les plus crédibles stylistiquement mais également les résultats les plus étranges :

> « C'est comme si la version classique était sur une ligne assez régulière dans les surprises qu'elle engendre chez nous, aussi bien en mauvais qu'en bien. J'ai l'impression que quand c'est l'*Omnibook* ça renforce tout ça. C'est-à-dire qu'il y a des moments surprenants dans le bon sens dans certaines phrases et dans le mauvais sens avec des moments encore plus forts. »

Bourhis résume les propos de Mabit et Gauvrit en expliquant que, d'un côté, les improvisations générées avec le corpus de musique classique sont plus strictes d'un point de vue harmonique et, de l'autre côté, les improvisations générées avec le corpus de l'*Omnibook* sont plus représentatives du style de Charlie Parker grâce à une meilleure logique mélodique à l'intérieur des phrases :

> « Il y a clairement un truc qui diffère entre les deux. Dans les versions *Omnibook*, il se permet plus de [...] faire des passerelles entre des accords par rapport à des plans plus extensifs de Charlie Parker. C'est-à-dire qu'il ferait des *patterns* à l'intérieur d'un motif qui déborderaient un peu plus sur l'harmonie. Il serait moins dans le moule, mais il aurait un peu plus pour but de définir ce qu'est le langage de Parker. [...] Dans la version classique, c'est mieux organisé dans le détail, la façon dont il souligne l'harmonie est plus claire, plus exploitée et mieux agencée entre les accords. Après ce qui est différent, c'est si tu veux faire le plus proche de Charlie Parker et plus proche d'une improvisation jazz, ou quelque chose de plus acceptable harmoniquement. C'est complètement différent. »

Les résultats de cette expérience sont également encourageants. Les trois musiciens sont capables de faire une distinction entre les improvisations générées avec des corpus différents. La préférence entre les générations faites avec

les deux corpus varie, cela dépendant de l'esthétique recherchée et des goûts personnels des musiciens. Cependant, les différentes remarques effectuées sur les générations faites avec les deux corpus semblent correspondre à ce à quoi on peut s'attendre étant donné le contenu des corpus. L'impact du corpus d'apprentissage et donc des connaissances du système a un impact actif sur la navigation dans l'intuition du système. Ceci valide la possibilité de créer des musiciens hybrides grâce au paradigme Intuition / Connaissance. Mabit et Gauvrit ont tout deux fait remarquer qu'il serait intéressant de voir si en réalisant un corpus de connaissance à partir de travaux de musicologie définissant les véritables influences stylistiques de Charlie Parker (par exemple, Buster Smith, Lester Young, Stravinsky...) il serait possible de reconstituer un avatar plus réaliste de Charlie Parker.

4.3.3 Résumé des résultats

Les retours des musiciens obtenus lors des sessions d'écoute sont très satisfaisants. Malgré l'usage de seulement deux sous-modèles probabilistes très simples pour constituer les connaissances du système dans les expériences effectuées, le guidage de l'improvisation est déjà clair et permet une nette amélioration de l'organisation des improvisations vis-à-vis des dimensions considérées. De plus, l'utilisation d'un corpus d'apprentissage pour établir les connaissances du système a un véritable impact sur le guidage des improvisations. Il est donc possible grâce à cette modélisation de créer des musiciens hybrides permettant d'effectuer des variations stylistiques tout en conservant un contexte local commun.

Ces résultats nous ont encouragé à étendre ces travaux pour pouvoir générer des improvisations multidimensionnelles en faisant interagir plusieurs agents basés sur ce système.

5

Interactivité entre dimensions/musiciens

"I'm seeking to have an art that is engaged as a way for saying Hurray for unity."

– Anthony Braxton

Dans ce chapitre, nous continuons les travaux précédents sur les aspects multidimensionnels de la musique dans un contexte d'improvisation. Nous proposons une méthode pour générer des improvisations musicales multidimensionnelles. Contrairement à Valle et al. [2016], nous ne voulons pas utiliser des symboles multidimensionnels pour la génération afin d'éviter les risques de sur-apprentissage dus à l'explosion combinatoire de la taille de l'alphabet lors de la combinaison de plusieurs dimensions. Kalonaris [2016] a montré la validité des modèles graphiques probabilistes pour représenter une situation d'improvisation libre entre plusieurs musiciens. Nous proposons alors un système multi-agents inspiré par les interactions entre musiciens dans une scène collective ou entre plusieurs dimensions dans l'esprit d'un musicien et où la communication entre les agents est réalisée à l'aide d'un graphe de *clusters*. Chaque dimension est représentée par un agent, qui suit le système Intuition / Connaissance présenté dans le chapitre précédent. Afin d'effectuer un partage des connaissances entre les différents agents, nous utilisons un algorithme de propagation de croyance. Ainsi, chaque agent est capable de prendre une décision, nourrie de ses connaissances internes et des connaissances externes venant des autres agents, pour la génération de la dimension qu'il dirige. Ce système multi-agents permet alors de générer simultanément les différentes dimensions de l'improvisation, tout en conservant une cohérence sur les relations entre dimensions.

Nous présentons dans la partie 5.1 les outils théoriques sur les modèles graphiques probabilistes que nous utilisons dans notre système. Nous présentons d'abord les graphes de *clusters* et leurs propriétés, puis l'algorithme de propagation de croyance. Ensuite, dans la partie 5.2, nous présentons comment utiliser ces outils pour la génération d'improvisations multidimensionnelles à l'aide du paradigme Intuition / Connaissance. Finalement, dans la partie 5.3, nous évaluons cette méthode lors de sessions d'écoute avec des musiciens de jazz professionnels.

5.1 Représentation graphique des interactions

Dans cette partie, nous présentons les outils théoriques que nous utilisons pour modéliser les interactions entre dimensions. Dans la partie 5.1.1, nous définissons la notion de graphe de *clusters* ainsi que les propriétés qu'un tel graphe doit respecter. Puis, dans la partie 5.1.2, nous présentons l'algorithme de propagation de croyance sur un graphe de *clusters* et ses propriétés en termes de convergence et de conservation de l'information. L'ensemble des outils présentés dans cette partie est issu de Koller & Friedman [2009].

5.1.1 Graphes de *clusters*

Définitions

Considérons X un ensemble de variables aléatoires . On définit un facteur ϕ comme une fonction de $\text{Val}(X)$ dans $\mathbb{R}$. On peut noter que la notion de facteur inclut à la fois les probabilités jointes et les probabilités conditionnelles. Dans notre cas, ils correspondent aux sous-modèles probabilistes que nous utilisons, par exemple, un bigramme $P(M_t|M_{t-1})$. L'ensemble des variables aléatoires utilisées par un facteur est appelé la portée du facteur (en anglais, *scope*) et est notée $\text{Scope}[\phi]$.

Un graphe de *clusters* $\mathcal{G}$ pour un ensemble de facteurs $\Phi = \{\phi_1, ..., \phi_K\}$ sur un ensemble de variables aléatoires X est un graphe non orienté pour lequel :

- à chaque nœud est associé un sous-ensemble de variables aléatoires $\mathcal{C}_f \subset X$, appelé *cluster*,
- à chaque arête entre deux *clusters* $\mathcal{C}_f$ et $\mathcal{C}_g$ est associée un *sep-ensemble* non vide $\mathcal{S}_{f,g} \subseteq \mathcal{C}_f \cap \mathcal{C}_g$. Le sep-ensemble $\mathcal{S}_{f,g}$ correspond aux variables sur lesquelles les *clusters* $\mathcal{C}_f$ et $\mathcal{C}_g$ vont pouvoir communiquer.

Chaque facteur $\phi_k \in \Phi$ est attribué à un *cluster* $\mathcal{C}_{\alpha(k)}$ tel que $\text{Scope}[\phi_k] \subseteq \mathcal{C}_{\alpha(k)}$. La croyance initiale du *cluster* $\mathcal{C}_f$ est définie par :

$$\psi_f(\mathcal{C}_f) = \prod_{k;\alpha(k)=f} \phi_k \ . \tag{5.1}$$

La Figure 5.1 montre un exemple d'un ensemble de facteurs $\Phi = \{\phi_1, ..., \phi_7\}$ sur un ensemble de variables aléatoires $X = \{A, B, C, D, E, F\}$, attribué sur un graphe de *clusters*. On peut remarquer que pour ce graphe de *clusters*, l'attribution des facteurs pourrait être différente. Par exemple, le facteur ϕ_2 aurait pu être attribué au *cluster* $\mathcal{C}_1$ au lieu du *cluster* $\mathcal{C}_2$, ces deux *clusters* partageant les variables B et C formant la portée de ϕ_2. Dans ce cas, les croyances initiales de $\mathcal{C}_1$ et $\mathcal{C}_2$ auraient été, respectivement, $\psi_1 = \phi_1.\phi_2$ et $\psi_2 = \phi_3$.

Propriétés

Afin d'être valide, par rapport à un ensemble de facteurs Φ, un graphe de *clusters* doit satisfaire les deux propriétés suivantes.

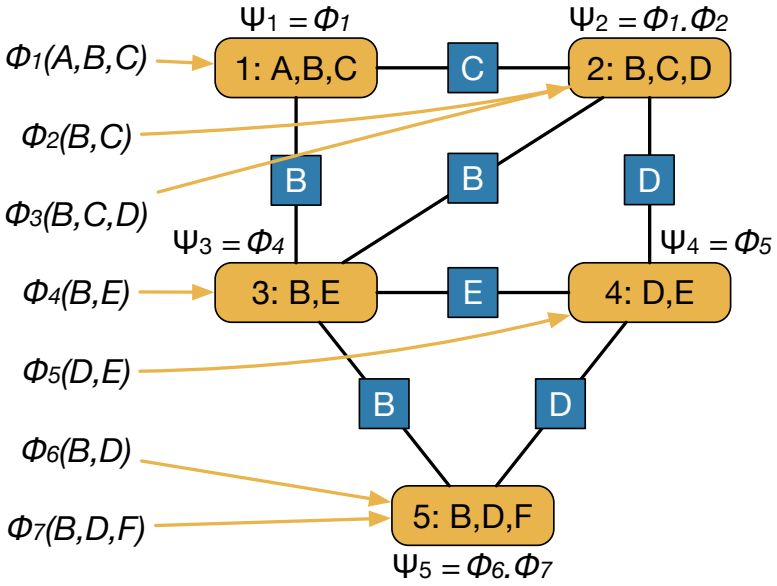

FIGURE 5.1 – Exemple d'une attribution de facteurs sur un graphe de *clusters*.

Propriété 1 : *Pour chaque facteur* $\phi_k \in \Phi$*, il existe un cluster* $\mathcal{C}_f$ *tel que* $Scope[\phi_k] \subseteq \mathcal{C}_f$.

Cela permet de garantir que tous les facteurs de Φ peuvent être attribués à un *cluster* et plus généralement de s'assurer que l'ensemble de l'information que l'on souhaite prendre en compte soit inclus dans le graphe de *clusters*.

Propriété 2 : *Pour chaque paire de clusters* $(\mathcal{C}_f, \mathcal{C}_g)$ *et pour chaque variable aléatoire* $A \in \mathcal{C}_f \cap \mathcal{C}_g$*, il existe un unique chemin entre* $\mathcal{C}_f$ *et* $\mathcal{C}_g$ *sur lequel tous les clusters et sep-ensembles incluent la variable* A.

Une propriété équivalente consiste à vérifier si pour chaque variable aléatoire A, l'ensemble des *clusters* et des sep-ensembles incluant A forme un arbre. Cette propriété a deux conséquences. Premièrement, l'existence de ce chemin garantit que l'ensemble de l'information sur une variable aléatoire A va pouvoir voyager sur l'ensemble des *clusters* incluant A. Deuxièmement, l'unicité de ce chemin évite que l'information sur une variable tourne en rond et génère des *fausses rumeurs*.

Notons que l'exemple donné en Figure 5.1 respecte ces deux propriétés. Premièrement, chaque facteur a pu être attribué à un *cluster* correspondant à leur portée. Deuxièmement, les ensembles de *clusters* et des sep-ensembles incluant chacune des variables forment des arbres (par exemple, pour la variable B, les arêtes $\{(1,3);(2,3);(3,5)\}$ et les nœuds $\{1,2,3,5\}$ forment un arbre). La Figure 5.2 montre des exemples de graphes de *clusters* non valides, car ne respectant pas la Propriété 2. Le graphe de *clusters* de gauche ne respecte pas l'existence du chemin pour la variable aléatoire B ; le *cluster* 2 est isolé. L'information sur B ne pourra pas voyager entre le *cluster* 2 et les autres

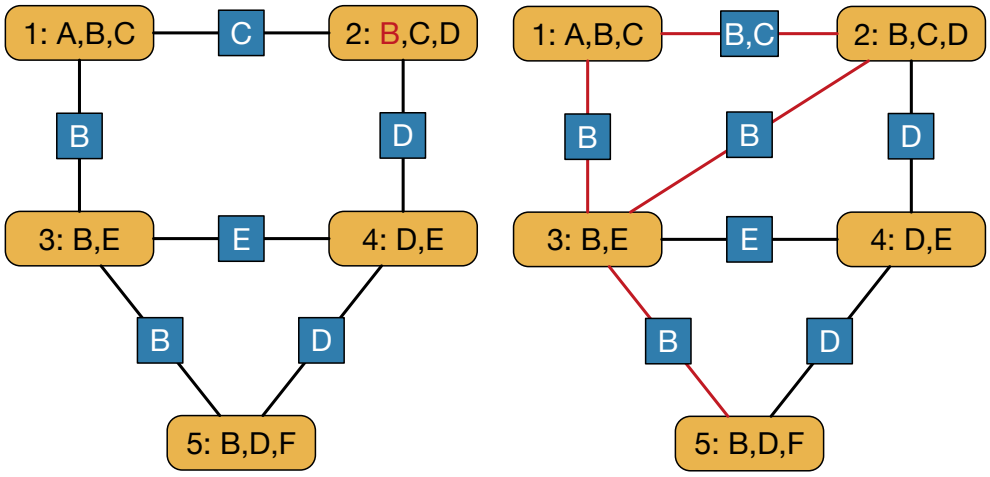

FIGURE 5.2 – Exemples de graphe de *clusters* ne respectant pas la Propriété 2.

clusters, pouvant amener à des divergences sur cette variable. Le graphe de *clusters* de droite ne respecte pas l'unicité du chemin pour la variable aléatoire B en raison de l'existence d'une boucle entre les *clusters* 1, 2 et 3. Cela peut entrainer la propagation de fausses rumeurs, le *cluster* 1 communiquant des informations sur B au *cluster* 2, puis le *cluster* 2 transmettant ces informations au *cluster* 3, qui retransmet ces informations au *cluster* 1, amplifiant alors ses croyances initiales. L'inférence d'information sur cette variable devient alors complètement biaisée.

5.1.2 Propagation de croyance sur un graphe de *clusters*

Présentation de l'algorithme

L'algorithme de propagation de croyance est basé sur le passage de messages probabilistes entre *clusters*. Le message envoyé par le *cluster* $\mathcal{C}_f$ au *cluster* $\mathcal{C}_g$ sur les variables du sep-ensemble $\mathcal{S}_{f,g}$ est noté $\delta_{f\to g}(\mathcal{S}_{f,g})$ et est défini par :

$$\delta_{f\to g}(\mathcal{S}_{f,g}) = \sum_{\mathcal{C}_f \setminus \mathcal{S}_{f,g}} \psi_f \prod_{h\in(N_f\setminus\{g\})} \delta_{h\to f} \ , \tag{5.2}$$

où N_f est le voisinage de $\mathcal{C}_f$.

Par exemple, dans le graphe de *clusters* de la Figure 5.1, les messages passés entre les *clusters* 1 et 3 sont :

$$\delta_{1\to 3}(B) = \sum_{A,C} \psi_1(A,B,C)\delta_{2\to 1}(C) \ , \tag{5.3}$$

$$\delta_{3\to 1}(B) = \sum_{E} \psi_3(B,E)\delta_{2\to 3}(B)\delta_{4\to 3}(E)\delta_{5\to 3}(B) \ . \tag{5.4}$$

On peut remarquer que le message $\delta_{f\to g}(\mathcal{S}_{f,g})$ ne dépend pas du message $\delta_{g\to f}(\mathcal{S}_{f,g})$. Cela permet d'éviter la répétition de l'information reçue d'un *cluster* vers le même *cluster* qui entraînerait, comme dans les cas où l'unicité de la Propriété 2 n'est pas respectée, la propagation de fausses rumeurs.

Les étapes de l'algorithme de propagation de croyance sont définies dans l'Algorithme 4.

Algorithme 4 Propagation de croyance

Entrée : Ensemble de facteurs Φ, graphe de *clusters* respectant les Propriétés 1 et 2.

1: Attribuer chaque facteur $\phi_k \in \Phi$ à un *cluster* $\mathcal{C}_{\alpha(k)}$.

2: Définir les croyances initiales :

$$\psi_f(\mathcal{C}_f) = \prod_{k:\alpha(k)=f} \phi_k \ . \tag{5.5}$$

3: Initialiser tous les messages à 1.

4: Répéter la mise à jour de tous les messages :

$$\delta_{f\to g}(\mathcal{S}_{f,g}) = \sum_{\mathcal{C}_f \setminus \mathcal{S}_{f,g}} \psi_f \prod_{h\in(N_f\setminus\{g\})} \delta_{h\to f} \ . \tag{5.6}$$

5: Calculer les croyances finales :

$$\beta_f(\mathcal{C}_f) = \psi_f \prod_{h\in N_f} \delta_{h\to f} \ . \tag{5.7}$$

Pour chaque *cluster*, la croyance finale est un nouveau facteur, basé sur les croyances initiales mises à jour par l'inférence de l'information provenant des autres *clusters*. La croyance finale $\beta_f(\mathcal{C}_f)$ est une approximation de la probabilité marginale $P(\mathcal{C}_f)$ appelée *pseudo-marginale*.

Propriétés de l'algorithme

Pour un graphe de *clusters* quelconque, la convergence de l'algorithme de propagation de croyance n'est théoriquement pas garantie. De plus, dans un graphe de *clusters*, la convergence est une propriété locale, c'est-à-dire que certains messages peuvent converger très rapidement tandis que d'autres peuvent ne jamais converger. De manière générale, plus le graphe de *clusters* est complexe, plus les chances d'une convergence totale sont faibles. Cependant, empiriquement, certaines méthodes ont été développées pour améliorer les chances de convergence de l'algorithme. Premièrement, l'ordre dans lequel les messages sont mis à jour dans l'étape 4 de l'Algorithme 3 a une influence sur la vitesse et les chances de convergence de l'algorithme. Deux variantes principales de mise à jour des messages ont été étudiées : la propagation de croyance synchrone où l'ensemble des messages sont mis à jour en parallèle et la propagation de croyance asynchrone où les messages sont mis à jour l'un après l'autre. Expérimentalement, la propagation de croyance asynchrone fournit de meilleurs résultats à la fois en terme de vitesse de convergence et de probabilité de convergence. Deuxièmement, il est possible de lisser les messages afin d'amortir les

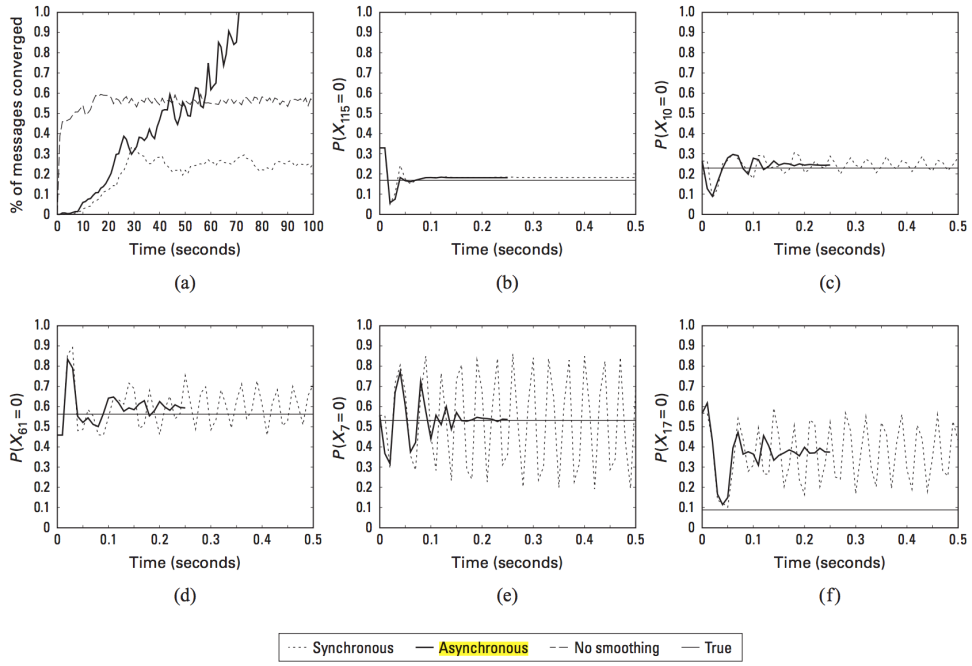

FIGURE 5.3 – Étude de la convergence de plusieurs variantes de l'algorithme de propagation de croyance sur un réseau d'Ising de taille 11×11. (source : Koller & Friedman [2009]).

oscillations dans les messages. Dans ce cas, la mise à jour du message devient :

$$\delta_{f\to g}(\mathcal{S}_{f,g}) \leftarrow \lambda\Big(\sum_{\mathcal{C}_f \setminus \mathcal{S}_{f,g}} \psi_f \prod_{h\in(N_f\setminus\{g\})} \delta_{h\to f}\Big) + (1-\lambda)\delta_{f\to g}^{\text{precedent}}(\mathcal{S}_{f,g}) \ . \quad (5.8)$$

Bien que le lissage puisse ralentir la convergence de l'algorithme, expérimentalement, il agrandit fortement ses chances de convergence.

La Figure 5.3 montre un cas d'étude de l'algorithme de propagation de croyance sur un réseau d'Ising [Ruozi, 2012] de taille 11×11 (avec 121 variables aléatoires). La Figure 5.3 (a) montre le pourcentage de messages ayant convergé en fonction du temps pour la propagation de croyance synchrone lissée (en pointillés), la propagation de croyance asynchrone lissée (en trait continu) et la propagation de croyance asynchrone non lissée (en trait discontinu). Tout d'abord, on peut voir que la version asynchrone de l'algorithme donne un résultat nettement supérieur à la version synchrone. On peut également remarquer que le lissage permet à l'algorithme de converger complètement. La Figure 5.3 (b) montre une pseudo-marginale qui converge rapidement pour les deux versions, synchrone et asynchrone. Les Figures 5.3 (c – e) montrent des pseudo-marginales où la version asynchrone converge, mais pas la version synchrone. La Figure 5.3 (f) montre une pseudo-marginale où les deux versions ne tendent pas vers la bonne valeur. Bien que la convergence théorique de cet algorithme ne soit pas garantie, il fournit de bons résultats en pratique, à l'exception des cas très complexes avec plus d'un millier de variables aléatoires, ce qui n'est pas le cas dans nos applications.

Il est également possible de montrer que la propagation de croyance sur un graphe de *clusters* s'effectue sans perte d'information sur la distribution initiale. En effet, si l'on considère l'ensemble des facteurs Φ sur l'ensemble des variables aléatoires X, la distribution définie par Φ est

$$\tilde{P}_\Phi(X) = \prod_{\phi \in \Phi} \phi \ . \tag{5.9}$$

À chaque étape de l'algorithme, on obtient les croyances β_f sur les *clusters* et $\rho_{f,g}$ sur les sep-ensembles suivantes :

$$\beta_f(\mathcal{C}_f) = \psi_f \prod_{h \in N_f} \delta_{h \to f} \ \text{ et } \ \rho_{f,g}(\mathcal{S}_{f,g}) = \delta_{f \to g}\delta_{g \to f} \ . \tag{5.10}$$

On a donc, à chaque étape de l'algorithme, le résultat suivant :

$$\tilde{P}_\Phi(X) = \frac{\prod\limits_f \beta_f(\mathcal{C}_f)}{\prod\limits_{(f,g)} \rho_{f,g}(\mathcal{S}_{f,g})} \ . \tag{5.11}$$

En effet,

$$\frac{\prod\limits_f \beta_f(\mathcal{C}_f)}{\prod\limits_{(f,g)} \rho_{f,g}(\mathcal{S}_{f,g})} = \frac{\prod\limits_f \psi_f \prod\limits_{h \in N_f} \delta_{h \to f}}{\prod\limits_{(f,g)} \delta_{f \to g}\delta_{g \to f}} \tag{5.12}$$

$$= \prod_f \psi_f = \prod_{\phi \in \Phi} \phi = \tilde{P}_\Phi(X) \tag{5.13}$$

car chaque message $\delta_{f \to g}$ apparait exactement une fois au numérateur et au dénominateur dans (5.12). Ce résultat montre que l'algorithme de propagation de croyance sur un graphe de *clusters* ne dilue pas l'information et effectue seulement une *reparamétrisation* de la distribution initiale et donc que l'ensemble de l'information est conservé [Koller & Friedman, 2009].

5.2 Interactions multidimensionnelles

Notre objectif est de faire communiquer plusieurs agents (dimensions ou musiciens) à travers un graphe de *clusters* en utilisant l'algorithme de propagation de croyance. Les agents sont constitués chacun d'un oracle des facteurs et d'un ensemble de sous-modèles probabilistes. La navigation de chaque oracle des facteurs est alors guidée par les croyances finales de graphe de *clusters* sur leur dimension respective. Ces croyances finales sont obtenues à la fois à partir des sous-modèles probabilistes définissant les croyances initiales du graphe de *clusters* et par le passage de messages de l'algorithme de propagation de croyance. De cette manière, chaque agent va influencer la navigation des autres agents. La navigation simultanée sur tous les oracles permet alors de générer des improvisations multidimensionnelles cohérentes et respectant les relations entre dimensions. Pour cela, nous présentons tout d'abord dans la partie 5.2.1 comment construire un graphe de *clusters* adapté aux dimensions et aux sous-modèles considérés. Puis, nous présentons dans la partie 5.2.2 le processus complet de navigation et de communication des oracles des facteurs.

5.2.1 Construction du graphe de clusters

Dans cette partie, nous présentons la première contribution de ce chapitre. Nous proposons une méthode de construction de graphe de *clusters* pour effectuer une interaction entre dimensions garantissant le respect des propriétés nécessaires présentées dans la partie 5.1.1. L'oracle des facteurs correspondant à chaque agent n'intervient pas dans la construction du graphe de *clusters*. Celui-ci est construit hors ligne, uniquement par rapport aux sous-modèles probabilistes considérés par les différents agents.

Considérons un cas général avec L agents chacun sur une dimension de X^1 à X^L. Chaque agent possède un ensemble de sous-modèles probabilistes $\Phi^l = \{\phi_1^l, ..., \phi_{K_l}^l\}$. Ces sous-modèles correspondent aux facteurs utilisés pour le graphe. Posons que ϕ_1^l est un sous-modèle horizontal de type n-gramme :

$$\phi_1^l = P(X_t^l | X_{t-1}^l, ..., X_{t-n+1}^l) \ . \tag{5.14}$$

On suppose que pour chaque dimension X^l, le n-gramme ϕ_1^l est le sous-modèle possédant la plus grande étendue temporelle sur cette dimension parmi tous les facteurs de $\Phi = \Phi^1 ... \Phi^L$. On suppose également que pour chaque facteur, l'ensemble de l'information considérée provient d'un sous-ensemble des L dimensions, c'est-à-dire qu'il n'y a pas d'information extérieure aux dimensions considérées. La portée de chaque facteur ϕ_k^l est l'ensemble des variables aléatoires apparaissant dans le sous-modèle. Par exemple, la portée d'un bigramme sur la mélodie est $\text{Scope}[P(M_t|M_{t-1})] = \{M_t, M_{t-1}\}$.

Premièrement, nous construisons pour chaque facteur ϕ_k^l un *cluster* $\mathcal{C}_k^l$ tel que $\mathcal{C}_k^l = \text{Scope}[\phi_k^l]$. De cette manière, nous nous assurons que chaque facteur peut être attribué à un *cluster* (voir partie 5.1.1, Propriété 1). La croyance initiale du *cluster* $\mathcal{C}_k^l$ est alors $\psi_k^l = \phi_k^l$.

Deuxièmement, afin de s'assurer que l'ensemble des *clusters* et des sep-ensembles pour chaque variable aléatoire forme un arbre dans le graphe de *clusters* (voir partie 5.1.1, Propriété 2), nous allons construire les arêtes du graphe en reliant chaque *cluster* $\mathcal{C}_1^l$, construit pour le facteur horizontal ϕ_1^l, aux autres *clusters* $\mathcal{C}_k^{l'}$ tels que

$$\mathcal{C}_1^l \cap \mathcal{C}_k^{l'} \neq \varnothing \ , \tag{5.15}$$

l' pouvant être égal à l. Le sep-ensemble entre ces *clusters* est donc

$$S_k^{l,l'} = \mathcal{C}_1^l \cap \mathcal{C}_k^{l'} \ . \tag{5.16}$$

Comme nous avons supposé qu'il n'y a pas d'information extérieure aux L dimensions considérées et que l'étendue temporelle sur la dimension X^l du facteur ϕ_1^l est supérieure aux autres modèles, par cette construction, on garantit l'existence du chemin entre tous les *clusters* partageant des variables aléatoires communes, ceux-ci étant tous relié au même $\mathcal{C}_1^l$. De plus, comme aucune autre arête n'est créée, on obtient pour chaque variable aléatoire X_k^l un arbre de profondeur 2 dont la racine est le *cluster* $\mathcal{C}_1^l$. Cette construction permet alors bien de respecter la Propriété 2.

La Figure 5.4 montre un exemple de graphe de *clusters* construit par cette méthode pour un cas à trois dimensions. Ici, nous avons pris un cas extrême

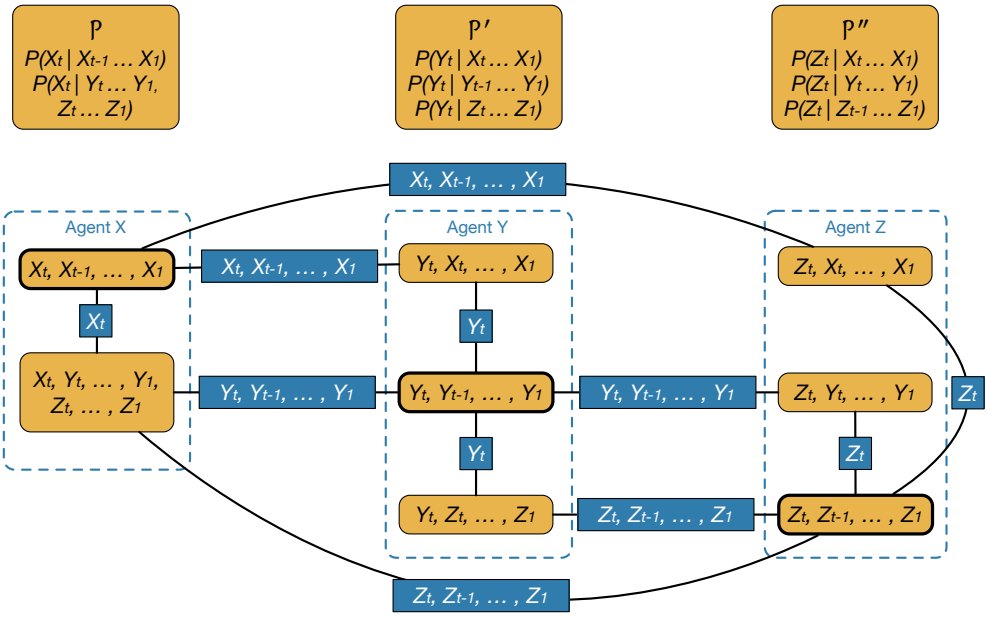

FIGURE 5.4 – Exemple de construction de graphe de *clusters* garantissant le respect des Propriétés 1 et 2 (voir partie 5.1.1) dans le cas avec trois dimensions.

où chaque agent prend en considération l'ensemble de l'information jouée sur les trois dimensions. Le haut de la figure montre les sous-modèles considérés par chaque agent. Les *clusters* encadrés en gras correspondent aux *clusters* $\mathcal{C}_1^l$. On peut vérifier que par cette construction, les deux propriétés du graphe de *clusters* sont bien respectées.

Si l'on souhaite prendre en considération des informations externes aux dimensions considérées comme, par exemple, des événements de l'environnement, il est toujours possible de construire un graphe de cette manière, en ignorant d'abord cette information pour la construction des arêtes, puis, *a posteriori*, ajouter les arêtes manquantes pour relier les différents *clusters* utilisant cette information, en faisant attention à respecter la Propriété 2. Pour un ensemble de facteurs donné, des graphes de *clusters* différents de celui proposé ici pourraient être construits. Cependant, cette méthode garantit le respect des deux propriétés pour tout ensemble de facteurs et, de plus, permet (en fonction des facteurs considérés) une interactivité équilibrée entre les différentes dimensions, dans le sens où il n'y a pas d'agent central faisant le lien entre plusieurs agents.

De manière plus concrète, considérons l'exemple où nous souhaitons générer une improvisation à la fois sur les dimensions mélodiques et harmoniques. Cela pourrait représenter, par exemple, le jeu d'un pianiste en improvisation libre jouant à la fois une mélodie à la main droite et des accords à la main gauche, ou les interactions entre un pianiste et un saxophoniste dans un contexte d'improvisation libre. Prenons un exemple similaire à celui étudié dans le chapitre précédent, où chaque agent possède un oracle des facteurs construit sur un contexte local et deux sous-modèles probabilistes appris sur un corpus :

— un bigramme sur leur dimension propre : $P(M_t|M_{t-1})$ pour l'agent mélodique et $P(C_t|C_{t-1})$ pour l'agent harmonique,

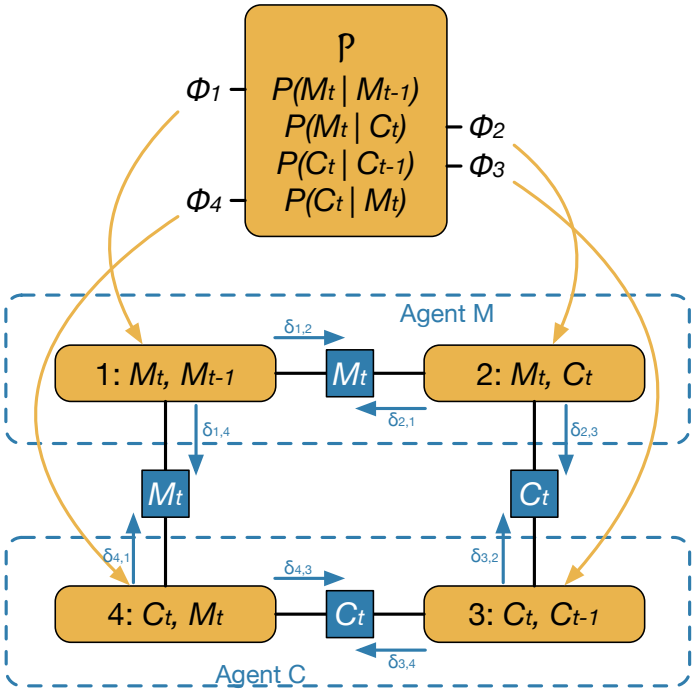

FIGURE 5.5 – Exemple de graphe de *clusters* pour une interaction entre un agent mélodique et un agent harmonique. Les *clusters* 1 et 2 correspondent aux sous-modèles probabilistes de l'agent mélodique et les *clusters* 3 et 4 correspondent aux sous-modèles probabilistes de l'agent harmonique.

— un modèle représentant les liens entre la mélodie et l'harmonie : $P(M_t|C_t)$ pour l'agent mélodique et $P(C_t|M_t)$ pour l'agent harmonique.

On construit un *cluster* pour chacun de ces sous-modèles, puis les sep-ensembles sont construits afin de relier les *clusters* construits sur les bigrammes avec les autres *clusters* partageant des variables aléatoires communes. La Figure 5.5 montre le graphe de *clusters* obtenu. Encore une fois, il est facile de vérifier que ce graphe de *clusters* respecte les deux propriétés de construction.

5.2.2 Propagation de croyance entre oracles des facteurs sur un graphe de *clusters*

Dans cette partie, nous expliquons le processus de génération d'improvisation multidimensionnelle et comment s'effectue le guidage de la navigation dans plusieurs oracles communiquant via un graphe de *clusters* grâce à des connaissances multidimensionnelles fournies par des sous-modèles probabilistes. Nous considérons ici que les différents éléments sont déjà construits (voir Lefebvre & Lecroq [2000]; Assayag et al. [2006a] pour la construction des oracles des facteurs, la partie 4.1.1 pour l'apprentissage des sous-modèles probabilistes et la partie 5.2.1 pour la construction du graphe de *clusters*).

Soit un ensemble d'agents musicaux chacun constitué :

— d'un ensemble de sous-modèles probabilistes $\Phi^l = \{\phi^l_1, ..., \phi^l_{K_l}\}$ appris sur un large corpus de séquences multidimensionnelles puis lissés,

— d'un oracle des facteurs construit sur une séquence $\mu^l = \mu^l_1 ... \mu^l_m$ jouée en direct par un musicien ou lue dans un corpus de petite taille.

Un graphe de *clusters* a été construit à partir de l'ensemble des sous-modèles probabilistes $\Phi = \bigcup_l \Phi^l$. Chaque sous-modèle probabiliste $\phi_k \in \Phi$ a été attribué à un *cluster* $\mathcal{C}_{\alpha(k)}$. L'improvisation multidimensionnelle générée à partir des oracles sera alors la superposition des nouvelles séquences $\mu^l_{t_1} \mu^l_{t_2} ... \mu^l_{t_T}$ jouées aux instants allant de 1 à T.

Considérons que le système a improvisé sur les instants de 1 à $T-1$. Posons $t^l_{T-1} = i^l$ l'état courant dans l'oracle des facteurs de l'agent l à l'instant $T-1$. Nous souhaitons alors définir les états t^l_T pour chaque oracle à l'instant T. La première étape consiste à déterminer pour chaque oracle l'ensemble des états atteignables $\mathrm{Att}(i^l)$. Pour cela on applique l'Algorithme 2, présenté dans la partie 4.2, basé sur les heuristiques présentées par Assayag & Bloch [2007]. Les probabilités des différents sous-modèles sont alors normalisées par rapport aux états atteignables. La croyance initiale de chaque *cluster* $\mathcal{C}_f$ est alors déterminée :

$$\psi_f(\mathcal{C}_f) = \prod_{k:\alpha(k)=f} \phi_k \ . \tag{5.17}$$

Les messages $\delta_{f,g}$ entre les *clusters* $\mathcal{C}_f$ et $\mathcal{C}_g$ reliés par un sep-ensemble $\mathcal{S}_{f,g}$ sont ensuite initialisés à 1, puis mis à jour un nombre arbitraire de fois, en fonction de la vitesse de convergence du graphe de *clusters*, de manière asynchrone avec lissage en utilisant l'équation (5.8). Les croyances finales $\beta_i(\mathcal{C}_i)$ sont alors déterminées :

$$\beta_f(\mathcal{C}_f) = \psi_f \prod_{h \in N_f} \delta_{h \to f} \tag{5.18}$$

où N_f est le voisinage de $\mathcal{C}_f$. Les pseudo-marginales $P(X^l_{t_T})$ pour chaque dimension peuvent alors être estimées à partir de n'importe quel *cluster* contenant la variable aléatoire $X^l_{t_T}$. En théorie, si l'algorithme de propagation de croyance a convergé et a effectué une inférence exacte de l'information, on a pour tout cluster $\mathcal{C}_f$ contenant la variable $X^l_{t_T}$:

$$\sum_{\mathcal{C}_f \setminus X^l_{t_T}} \beta_f(\mathcal{C}_f) = P(X^l_{t_T}) \ . \tag{5.19}$$

Lorsque toutes ces sommes sont égales, on dit que le graphe de *clusters* est calibré. Par exemple, à partir d'un bigramme, on aurait :

$$P(X^l_{t_T}) = \sum_{X^l_{t_{T-1}}} \beta(X^l_{t_T}, X^l_{t_{T-1}}). \tag{5.20}$$

Ces pseudo-marginales définissent alors des scores pour chaque transition vers les états atteignables. Ces scores sont normalisés pour obtenir les probabilités de transition dans les oracles des facteurs. Pour l'oracle des facteurs de l'agent l :

$$P(t^l_T = j^l) = \frac{P(X^l_{t_T} = \mu^l_j)}{\sum_{k \in \mathrm{Att}(i^l)} P(X^l_{t_T} = \mu^l_k)} \tag{5.21}$$

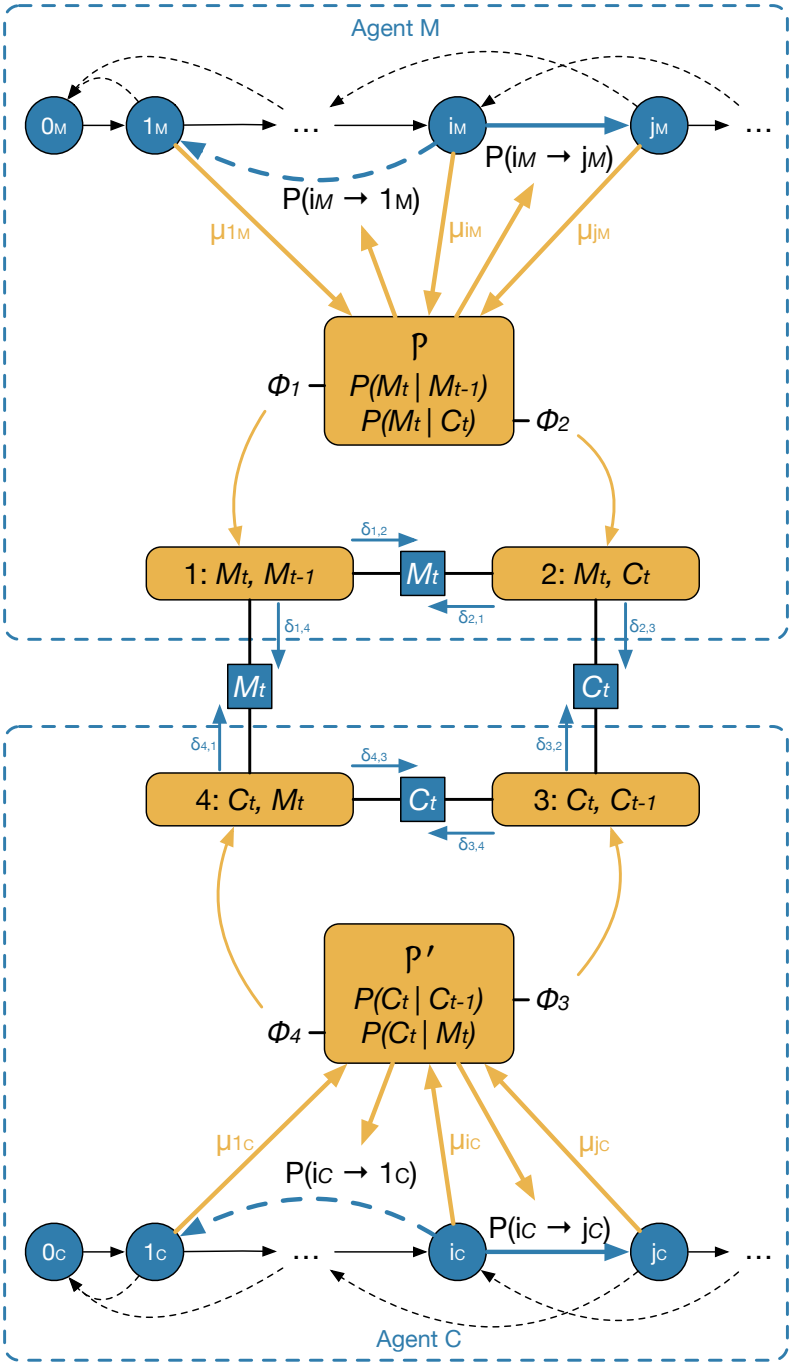

FIGURE 5.6 – Communication entre un agent mélodique et un agent harmonique à travers un graphe de *clusters*. Chaque agent possède un oracle des facteurs sur sa dimension propre et des connaissances multidimensionnelles sous forme de sous-modèles probabilistes. Les agents communiquent leurs possibilités de navigation dans leur oracle respectif, puis ils prennent une décision globale de navigation à partir de leur propre connaissance et des connaissances de l'autre agent.

Finalement, chaque agent prend une décision aléatoire en suivant les probabilités de transition.

La Figure 5.6 montre une application pour l'exemple de la partie précédente pour la génération d'une improvisation mélodique et harmonique avec deux agents possédant chacun un bigramme sur leur dimension propre et un modèle représentant les liens entre la mélodie et l'harmonie. Le graphe de *clusters* est représenté au milieu de la figure. Dans cet exemple, à l'instant $T-1$, l'agent M est dans l'état i^M et a pour états atteignables les états 1^M et j^M. De même, à l'instant $T-1$, l'agent C est dans l'état i^C et a pour états atteignables les états 1^C et j^C. À partir de ces données, l'algorithme de propagation de croyance s'effectue sur le graphe de *clusters*. Les pseudo-marginales $P(M_{t_T})$ et $P(C_{t_T})$ sont alors calculées. Par exemple,

$$P(M_{t_T}) \simeq \sum_{M_{t_{T-1}}} \beta(M_{t_T}, M_{t_{T-1}}) \ . \tag{5.22}$$

Les probabilités de transition dans l'oracle des facteurs de chaque agent sont alors déterminées par normalisation des scores obtenus à partir de la pseudo-marginale correspondante.

Cette méthode de génération permet alors bien de conserver une génération cohérente sur chaque dimension grâce à l'utilisation d'oracles des facteurs, tout en guidant cette génération à partir de connaissances multidimensionnelles internes de chaque agent et des connaissances externes provenant des autres agents, les pseudo-marginales étant calculées à partir de l'ensemble de l'information fournie au graphe de *clusters*. De cette manière, les agents prennent une décision cohérente vis-à-vis de leur propre génération et de la génération des autres agents. De plus, un des avantages de cette méthode est qu'il est possible de considérer une situation où les sous-modèles et les oracles de chaque agent sont appris sur des corpus différents émulant des styles de différents musiciens (y compris les sous-modèles faisant le lien entre dimensions), créant ainsi une individualité propre à chaque agent. Ce modèle est alors plus flexible qu'un système avec des connaissances centralisées et des probabilités jointes.

5.3 Évaluation et retours des musiciens

5.3.1 Choix des paramètres

Pour évaluer ce modèle d'interaction entre dimensions par propagation de croyance sur un graphe de *clusters*, comme pour le chapitre précédent, nous avons réalisé des tests d'écoute avec des musiciens professionnels afin d'obtenir une évaluation qualitative des générations. Une fois encore, les improvisations générées sont basées sur le style de l'*Omnibook* de Charlie Parker (voir partie 2.2). Nous avons alors fait appel aux mêmes trois musiciens et improvisateurs de jazz professionnels :

— Pascal Mabit, saxophoniste et professeur de jazz, diplômé du Conservatoire National Supérieur de Musique et de Danse de Paris,
— Louis Bourhis, contrebassiste de jazz, diplômé de la Haute École de Musique de Lausanne,

— Joël Gauvrit, pianiste et professeur de jazz et de musique classique, diplômé du Conservatoire National Supérieur de Musique et de Danse de Lyon. [7]

Les tests d'écoute ont été faits séparément.

Nous nous sommes intéressés au cas d'une improvisation avec un agent mélodique et un agent harmonique dans une situation d'improvisation libre. Chaque agent possède deux sous-modèles probabilistes. L'agent mélodique possède les sous-modèles suivants :

— $\phi_1 = P(M_{t_T}|M_{t_{T-1}})$: un bigramme sur la mélodie,
— $\phi_2 = P(M_{t_T}|C_{t_T})$: un modèle déterminant quelle mélodie jouer sur quel accord.

L'agent harmonique possède les sous-modèles suivants :

— $\phi_3 = P(C_{t_T}|C_{t_{T-1}})$: un bigramme sur les accords,
— $\phi_4 = P(C_{t_T}|M_{t_T})$: un modèle déterminant quel accord jouer sur quelle mélodie.

Le corpus de l'*Omnibook* a été divisé en trois sous-corpus : un corpus d'apprentissage constitué de 40 thèmes et improvisations, un corpus de validation constitué de 5 thèmes et improvisations et un corpus de test constitué de 5 thèmes et improvisations. Les probabilités des sous-modèles ont été estimées sur le corpus d'apprentissage. Ces sous-modèles ont été lissés par un mélange de lissage par repli et de lissage additif. Les coefficients de lissage ont été optimisés sur le corpus de validation. L'agent mélodique et l'agent harmonique possèdent également chacun un oracle des facteurs construit respectivement note-à-note et accord-à-accord sur des thèmes et improvisations issus du corpus de test. Dans cette expérience, les deux oracles des facteurs sont construits sur les mêmes morceaux. Les deux agents communiquent alors via le graphe de *clusters* présenté précédemment dans la Figure 5.6. À chaque étape, tous les messages sont mis à jour 10 fois, de manière asynchrone et lissée.

Pour éviter les décalages rythmiques, l'improvisation mélodique ne possède pas d'information rythmique (seules des croches et des demi-pauses sont jouées). L'improvisation harmonique joue les accords avec les durées présentes dans le morceau sur lequel l'oracle des facteurs harmonique est construit (généralement, deux temps ou une mesure). Un accord joué est alors fixé pour sa durée, les messages envoyés par l'agent harmonique font état d'une probabilité égale à 1 pour cet accord pendant toute sa durée. La superposition de la génération des deux agents est alors jouée.

Des exemples de génération d'improvisations multidimensionnelles mélodiques et harmoniques sont disponibles sur : `http://repmus.ircam.fr/dyci2/demos/cluster_graph`. Les Figures 5.8 et 5.7 montrent également des relevés d'improvisations multidimensionnelles générées avec les oracles des facteurs construits respectivement sur *Anthropology* et *Donna Lee*.

5.3.2 Résultats et analyse

Chaque musicien a écouté une douzaine d'improvisations, plus ou moins, selon son vouloir, générées à partir du contexte local de deux morceaux issus

7. Des biographies plus détaillées des musiciens sont fournies en Annexe B.

du corpus de test : *Donna Lee* et *Anthropology*.

Dès les premières écoutes, Mabit remarque une certaine forme de réalisme des générations et que les résultats obtenus pourraient représenter un cas d'étude de situations réelles :

> « C'est marrant. Ça donne vraiment l'impression... Ça donnerait vraiment ça en vrai si on disait qu'on se met avec un pianiste qui connait un peu cette musique là, on lui dit "vas-y, t'improvises un truc"... En fait, ça fait un peu *trip* du CNSM des classes d'improvisation expérimentale [...], des cours d'improvisation où on se dit voilà, on a travaillé un mois, deux mois sur *Donna Lee*, seulement *Donna Lee*, on a bossé à fond. Maintenant, on joue *Donna Lee*, on sait quels sont les accords et maintenant on les ré-agence. »

Les trois musiciens ont fait remarquer que le réalisme des générations provenait principalement des progressions harmoniques générées. Gauvrit dit « qu'au niveau harmonique, ça marche toujours ; l'enchainement des accords fonctionne. ». Les successions d'accords générées sont crédibles par rapport aux standards de jazz et il serait facile d'improviser dessus. Mabit précise cette pensée sur les deux morceaux, tout d'abord sur les improvisations générées à partir de *Donna Lee* :

> « L'harmonie avait complètement du sens. Ce n'était pas forcément *Donna Lee*, mais en tout cas, il y avait des progressions qui étaient complètement logiques et ça pourrait être n'importe quelle chanson de jazz. Ca faisait un peu standard de base avec premier degré, arrivée au quatrième. Un truc, que du connu. Tu m'aurais généré n'importe quelle grille avec ces éléments-là, j'aurais improvisé d'oreille sans aucun souci. »

Puis sur les improvisations générées à partir d'*Anthropology* il précise que même si la forme globale de l'anatole en 32 mesures n'est pas respectée par l'harmonie, ses différents éléments apparaissent de manière réaliste :

> « Ça ressemblait vraiment à un anatole au niveau de l'harmonie. [...] Ça n'a pas fait de *AABA*, mais le défilé faisait vraiment une première partie en B♭ avec une modulation vers le quatrième degré, le pont et retour en B♭ avec un petit bout de quatrième degré, tout ça... C'était presque plaisant en fait, parce qu'on pouvait entendre presque l'anatole, mais ça n'en était pas un, mais quand même, ça ressemblait à quelque chose. »

Ceci peut-être remarqué sur l'exemple montré dans la Figure 5.7. La progression harmonique consiste ici principalement en une série de cellules-anatoles (`I I II- V`^{7} ; `I VI`^{7} `II- V`^{7} ou `III- VI`^{7} `II- V`^{7}) avec un passage vers le quatrième degré aux mesures 7 et 8. Cela donne une véritable sensation d'être sur le *A* d'un anatole, mais sans respect de la forme globale car la modulation au quatrième degré serait attendue aux mesures 5 et 6. Bourhis exprime une idée similaire en précisant que cette cohérence harmonique des générations provient du fait que les blocs harmoniques tels que les cellules-anatoles sont généralement exprimés dans leur intégralité. Ceci est sans doute dû à l'utilisation de l'oracle des facteurs harmonique, garantissant une logique locale de la suite d'accords :

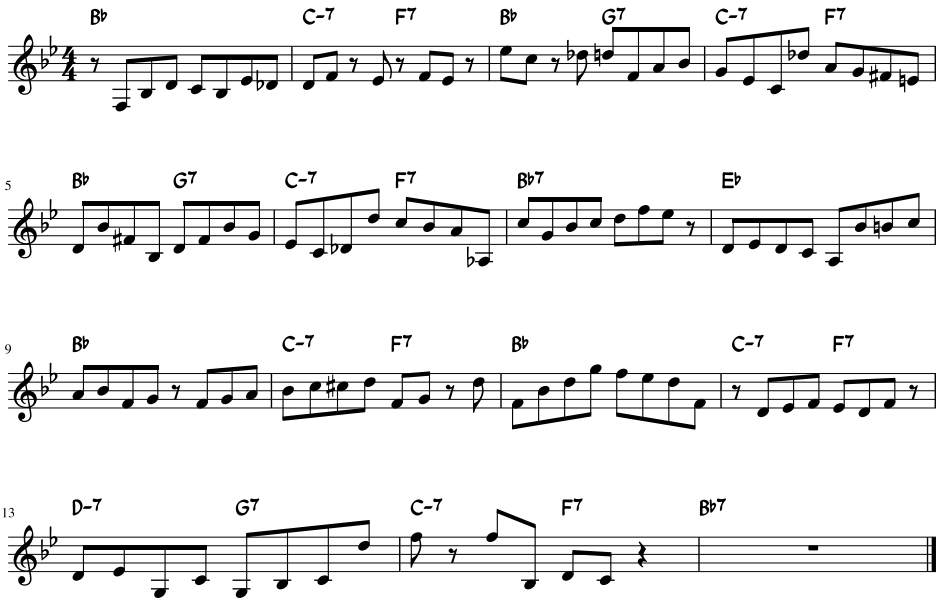

FIGURE 5.7 – Exemple d'improvisation multidimensionnelle avec interaction entre un agent mélodique est un agent harmonique par le graphe de *clusters* présentés dans la Figure 5.6. Les oracles des facteurs mélodique et harmonique sont construits sur *Anthropology* de Charlie Parker. Les connaissances multidimensionnelles des deux agents sont apprises sur l'*Omnibook*. La mélodie et les accords sont improvisées simultanément par le système.

> « Ce qui est marrant, par exemple, ce qui est bien, ce qui renforce la crédibilité des suites d'accords, c'est qu'il coupe très rarement les `II- V`^{7} `I` ou des trucs comme ça. [...] J'ai l'impression que vous lui avez donné une certaine connaissance de l'harmonie, une priorité des suites et des cadences. »

Les progressions harmoniques fonctionnant, les critiques se sont alors plutôt dirigées sur l'organisation mélodique de l'improvisation. D'après les trois musiciens, la qualité des improvisations mélodiques générées est similaire à celles générées avec le modèle probabiliste du chapitre précédent. Bourhis dit :

> « Là on est, en fait, au niveau du vocabulaire assez similaire aux versions de *Donna Lee* d'avant où il avait travaillé la grille sauf qu'on n'est pas guidé, parce que la grille elle a un sens aussi à l'intérieur d'elle-même. À la mélodie on entend des progressions harmoniques qui sont claires, qui sont connues et qui sont justement attendues par nous quand on écoute. [...] Le fait qu'il y ait une sous-dominante mineure qui va sur [la tonique], c'est attendu et ça fait que ça guide de façon indirecte (puisque lui aussi quand même il joue sur des accords), et que donc, même si c'est parfois un peu maladroit à l'intérieur, on a quand même des résolutions qui sont là. »

Gauvrit précise cette idée en expliquant que comme, dans la première expérience du chapitre précédent, les constructions mélodiques sont assez cohérentes dans le sens où on a bien « une idée qui en amène une autre, qui en amène une autre, etc. » . Cependant, dans la construction globale de l'improvisation, des améliorations pourraient être faites sur une cohérence dans l'utilisation des notes de couleur :

> « Et les phrases par rapport à la version non entraînée du départ, elles sont toujours plus cohérentes. Tu as rarement des trucs où tu te dis, il n'y a aucun lien entre les choses mélodiquement. [...] Je reviens toujours à la même chose, c'est ce truc de cohérence stylistique. Ca passe de très peu altéré à très altéré de façon très soudaine. [...] Les couleurs modales parfois c'est hyper coloré, puis après, on se retrouve dans quelque chose de complètement diatonique de façon très soudaine. »

Cela se remarque, par exemple, dans la Figure 5.7 sur *Anthropology*. On peut voir que l'improvisation est très diatonique jusqu'au milieu de la mesure 4, puis devient très altérée sur les mesures 4 à 6, puis retourne sur un jeu très diatonique par la suite. Cependant, on peut voir que dans tous les cas les notes jouées sur les premiers temps des accords sont cohérentes harmoniquement par rapport au style (utilisation des toniques, tierces, quintes, septièmes, etc.). Dans une improvisation de jazz, on s'attendrait typiquement à une progression graduelle de la complexité harmonique de la mélodie. Cette critique était assez attendue, dans le sens où notre système ne prend pas en compte une organisation globale de l'improvisation, mais cherche uniquement à répondre à des attentes locales. Mabit suggère également que les passages altérés pouraient être dus à la taille trop réduite du contexte local, ne donnant pas suffisamment de matériel à l'improvisateur :

> « La mélodie... peut-être que comme le corpus est un peu réduit du coup, il essaie absolument de prendre dans son corpus et du coup ça donne parfois des choses complètement étranges. »

De plus, on retrouve également les critiques communes à l'ensemble des systèmes issus du paradigme *OMax*, comme par exemple la présence de sauts d'octave dans la mélodie dus à la représentation des notes utilisée pour éviter un vocabulaire trop important :

> « Il y a toujours un truc inévitable qui nous fait sortir de la probabilité que ça existe. Par exemple, il y a des sauts monstrueux. [...] La cohérence, elle se fait parfois un peu... pas virer, mais des choses vont prendre le dessus sur la cohérence globale. »

Cela se remarque, par exemple, dans la Figure 5.8, sur *Donna Lee*, entre la mesure 11 et la mesure 12.

À propos des interactions entre les deux agents, les trois musiciens ont fait remarquer que les liens entre la mélodie et l'harmonie ont du sens et sont réalistes. Les agents semblent collaborer avec une volonté commune pour le développement de l'improvisation. Mabit dit :

> « Ce qui est intéressant, c'est qu'il y a quand même une notion de suivi. [...] Les directions sont les mêmes pour les deux voix et du coup, il n'y a pas trop de problème de ce point de vue là. »

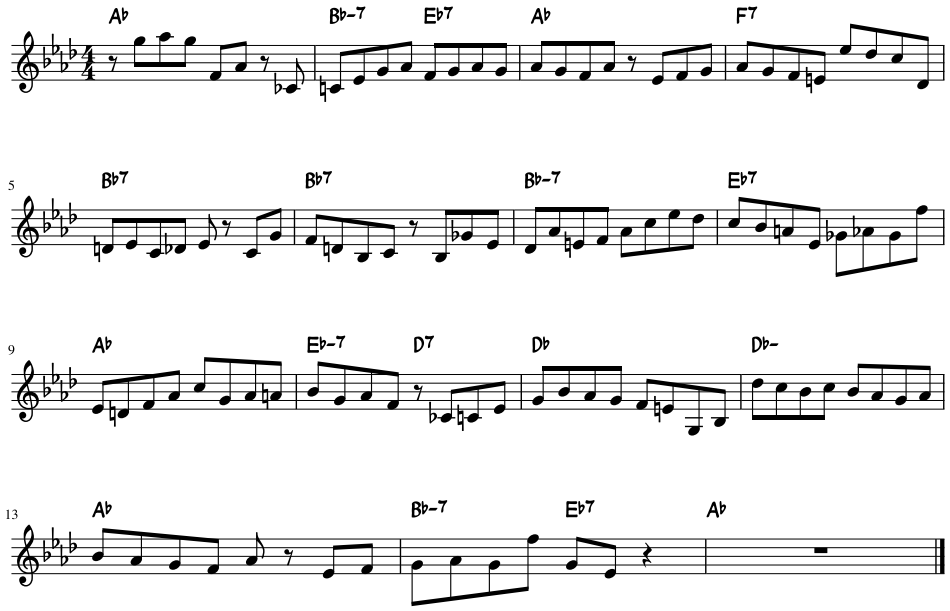

FIGURE 5.8 – Exemple d'improvisation multidimensionnelle avec interaction entre un agent mélodique et un agent harmonique par le graphe de *clusters* présentés dans la Figure 5.6. Les oracles des facteurs mélodique et harmonique sont construits sur *Donna Lee* de Charlie Parker. Les connaissances multidimensionnelles des deux agents sont apprises sur l'*Omnibook*. La mélodie et les accords sont improvisées simultanément par le système.

D'un autre côté, Bourhis déplore un peu un manque d'anticipation et d'organisation de l'improvisation orientée vers le futur. En effet, les sous-modèles utilisés permettent aux agents de communiquer uniquement sur le passé et sur le présent de l'improvisation. Ainsi, les principes d'anticipation harmonique tels que les conduites de voix ne sont pas possibles. Ceci peut alors s'entendre dans les exemples générés, en particulier sur les longues plages harmoniquement statiques :

> « Mais c'est pareil, s'il n'y a pas assez de vision du long terme, il ne peut pas préparer des notes, c'est-à-dire que quand tu as des modulations ça se prépare, tu vois ? Ça fait qu'il ne peut rien préparer et du coup moins adoucir pour l'oreille, c'est un peu bourrin. Mais ça marche pas mal quand même sur une mesure. Du coup, vu qu'il y a plus de possibilités, il y a plus de pièges. Quand c'est deux temps, il a une certaine sécurité parce que ça change vite et du coup on passe à autre chose. Alors que plus la période est longue et plus la marge d'erreur devient importante, du moins par rapport à ce qu'on attend, à ce qu'on connait de Charlie Parker. »

Gauvrit analyse ce phénomène similairement. Cependant, il pousse cette analyse plus loin, en remarquant que bien qu'il n'y ait pas de conduite mélodique vers l'accord suivant, une certaine cohérence existe du fait que l'agent harmo-

nique va aussi s'adapter pour justifier le jeu mélodique passé. Cette méthode donne par conséquent des résultats corrects, mais semble « à rebours » de la façon dont un musicien prévoirait son improvisation :

> « Dans les phrases mélodiques, la plupart du temps il va vers l'accord suivant, mais la conduite, elle vient tard. C'est une perception que j'ai. [...] C'est étonnant du coup qu'ils soient quand même capables de faire des mouvements conjoints vers une note de l'accord suivant, parce que c'est beaucoup le cas. [...] Bah oui, en fait... il choisit juste la note qui est juste à côté de là où il en est arrivé [...] Ce n'est pas une question de « je vais jusque là », c'est « maintenant que j'ai fait ça... » Ça justifie à chaque fois *a posteriori* ce qu'on a fait avant. »

De plus, Mabit trouve que la cohésion entre les deux agents est parfois trop forte, rendant l'agent mélodique subordonné à l'agent harmonique. La notion de réactivité entre les agents est du coup effacée par le fait que les agents prennent constamment des décisions communes ce qui limite les possibilités d'exploration musicale et par conséquent entraine des improvisations pouvant sonner un peu ternes. Encore une fois, ce phénomène peut-être amplifié par la taille très restreinte du contexte local utilisé (un seul thème et improvisation pour la construction de l'oracle des facteurs) :

> « En fait ce qu'il se passe c'est que comme les deux voix savent à peu près, enfin savent exactement ce qu'il se passe l'une avec l'autre à tous les instants, c'est le moment où ils en savent trop. C'est le moment où on en sait trop et du coup ça gène un peu le jeu. Nous quand on joue, ou même le quartet de Wayne Shorter, le groupe où les gars ils se connaissent depuis dix ans, ils jouent ensemble tout le temps et c'est incroyable ça communique tout ça. Personne ne peut savoir ce que les autres vont faire tout le temps à chaque instant. Et c'est le principe de l'improvisation qu'il y a une notion de réactivité. [...] Là où un vrai humain pourrait avoir envie de faire des choix parfois complètement irrationnels pour un ordinateur. C'est peut-être ça aussi qui va être dur à intégrer dans l'improvisation, c'est l'irrationalité de l'esprit humain. »

5.3.3 Résumé des résultats

Dans l'ensemble, les retours des musiciens sont très satisfaisants. La génération des deux agents est cohérente à la fois dans leur individualité, mais aussi dans le jeu collectif. La communication des agents grâce au graphe de *clusters* permet de représenter efficacement des situations d'interaction, les deux agents semblant développer l'improvisation avec une volonté commune et cela malgré l'utilisation d'un graphe de *clusters* très simple basé sur un nombre de sous-modèles probabilistes faible. Gauvrit, enthousiaste, dit en plaisantant :

> « C'est chouette. Ouais bravo ! *(rires)*. S'il swinguait un peu plus, je vais au concert, je paie ma place. »

Cependant, certaines limitations reviennent entre les systèmes développés dans ce chapitre et dans le chapitre précédent à propos du manque d'anticipa-

tion du futur et d'organisation globale de l'improvisation. Nous nous sommes alors intéressés par la suite à la modélisation de structures musicales à plusieurs niveaux temporels dans le cas d'improvisations avec un guidage structurel.

6

Improvisation sur un scénario à plusieurs niveaux temporels

"Improvised music involves a lot of intuition and I like developing intuition."

– Fred Frith

Dans ce chapitre, nous quittons l'aspect multidimensionnel de l'improvisation pour nous concentrer sur son organisation à plusieurs niveaux temporels. Par exemple, dans le cas d'un morceau de jazz, la grille d'accords est souvent structurée à plusieurs échelles temporelles différentes. Certains groupes d'accords (petite échelle) peuvent former des fonctions tonales ou modales (moyenne échelle) et ces fonctions peuvent être organisées en différentes sections (grande échelle). Nous nous intéressons ici au cas de l'improvisation guidée par un guidage structurel avec l'utilisation d'un scénario temporel (voir partie 3.1.3). Notre objectif est premièrement de pouvoir représenter l'organisation temporelle hiérarchique d'un scénario par un *scénario multi-niveaux*[8]. Pour cela, nous proposons d'utiliser une grammaire syntagmatique [Chomsky, 1972] permettant de générer plusieurs variations d'une même structure globale et de représenter l'organisation hiérarchique à plusieurs niveaux temporels. Nous voulons ensuite utiliser cette information multi-niveaux lors du processus de génération de l'improvisation à l'aide de nouvelles heuristiques adaptées à cette information. Ces travaux se veulent comme une extension des travaux proposés dans *ImproteK* par Nika [2016]; Nika et al. [2017a].

Nous présentons dans la partie 6.1 comment une grammaire basée sur une analyse linguistique en constituants peut représenter la structure hiérarchique d'un scénario temporel pour constituer des scénarios multi-niveaux. Nous construisons, puis évaluons, une telle grammaire avec des musiciens professionnels pour un cas particulier de scénario multi-niveaux : une grille d'accords de jazz appelée anatole. Ensuite, dans la partie 6.2, nous proposons une méthode d'apprentissage automatique d'une telle grammaire basée sur des méthodes probabilistes de sélection de séquences de mots partageant une information mutuelle [Zitouni et al., 2000]. Dans la partie 6.3, nous proposons

8. Par la suite, nous choisissons le terme « niveau » plutôt qu'« échelle » pour décrire l'organisation temporelle afin d'éviter la confusion avec la notion d'échelle de notes.

une nouvelle heuristique pour générer une improvisation sur un scénario multi-niveaux. Cette heuristique est basée sur les principes d'anticipation de scénario proposés par Nika et al. [2017a]. Finalement, dans la partie 6.4, nous présentons une évaluation de notre méthode de génération effectuée lors de sessions d'écoute avec des musiciens de jazz professionnels.

6.1 Utiliser une grammaire pour modéliser une structure multi-niveaux

Dans cette partie, nous présentons comment utiliser une grammaire pour représenter l'organisation hiérarchique d'une grille d'accords en plusieurs niveaux temporels. Dans la partie 6.1.1, nous présentons le type de grammaire que nous allons utiliser : les grammaires syntagmatiques. Puis, dans la partie 6.1.2, nous construisons, avec un musicien professionnel, une grammaire syntagmatique pour représenter l'aspect multi-niveau d'une grille de jazz très populaire : l'anatole. Cette grammaire est ensuite évaluée par des musiciens professionnels dans la partie 6.1.3.

6.1.1 Grammaire syntagmatique

Les définitions présentées dans cette partie sont adaptées de Chomsky [1956, 1972] et de Hopcroft & Ullman [1979].

Soit un ensemble de symboles X, on note X^* l'ensemble des séquences finies d'éléments de X. Une grammaire $G = (N, \Sigma, R, s)$ est définie par :

— un ensemble N de symboles non terminaux,
— un ensemble Σ de symboles terminaux, disjoint de N,
— un élément particulier $s \in N$ appelé axiome,
— un ensemble fini de règles de réécriture

$$R \subset (N \cup \Sigma)^* N (N \cup \Sigma)^* \rightarrow (N \cup \Sigma)^* \ . \tag{6.1}$$

Une règle de réécriture, généralement notée $u \rightarrow v$ peut être interprétée comme une instruction signifiant « remplacer u par v ». Le langage $L(G)$ engendré par une grammaire G est l'ensemble des séquences de symboles terminaux pouvant être créé à partir de l'axiome s en utilisant les règles de réécriture de R :

$$L(G) = \{w \in \Sigma^* | s \xrightarrow{*} w\} \ , \tag{6.2}$$

où $\xrightarrow{*}$ représente l'utilisation d'un ensemble fini de règles de réécriture de R pour réécrire s en w.

Une *grammaire syntagmatique* est un type particulier de grammaire, présentée dans Chomsky [1956]. Ces grammaires se basent sur une description linguistique d'un langage au niveau syntaxique par une analyse en constituants, c'est-à-dire par une décomposition des fonctions linguistiques au sein d'une structure hiérarchique. Chomsky [1972] propose un exemple de grammaire syntagmatique pour la construction de phrases simples en anglais, présenté ici dans la Grammaire 1.

Dans cet exemple, les symboles non terminaux sont écrits en italique et les symboles terminaux en police normale. La règle 1 peut être lue comme

Grammaire 1 Exemple de grammaire syntagmatique pour la construction de phrases en anglais

1: $Phrase \rightarrow GN + GV$
2: $GN \rightarrow Article + Nom$
3: $GV \rightarrow Verbe + GN$
4: $Article \rightarrow$ a, the...
5: $Nom \rightarrow$ man, ball...
6: $Verbe \rightarrow$ hit, took...

« réécrire $Phrase$ par $GN + GV$ », c'est-à-dire qu'une phrase est constituée d'un *groupe nominal* suivi par un *groupe verbal*. Une interprétation similaire peut être effectuée pour les règles 2 et 3. La règle 4 spécifie que le symbole non terminal $Article$ peut être réécrit par l'un des symboles terminaux proposés. Une interprétation similaire peut être effectuée pour les règles 5 et 6.

Une *dérivation* est la séquence de règles de réécriture utilisée pour obtenir une certaine phrase. Par exemple, pour obtenir la phrase « the man hit a ball », la dérivation présentée dans le Tableau 6.1 a été utilisée. Les numéros à droite de chaque ligne indiquent la règle de réécriture utilisée pour construire cette ligne à partir de la ligne précédente.

$Phrase$	
$GN + GV$	1
$Article + Nom + GV$	2
$Article + Nom + Verbe + GN$	3
the $+Nom + Verbe + GN$	4
the + man $+Verbe + GN$	5
the + man + hit $+GN$	6
the + man + hit $+Article + Nom$	2
the + man + hit + a $+Nom$	4
the + man + hit + a + ball	5

Tableau 6.1 – Dérivation pour obtenir la phrase « the man hit a ball » à partir de la Grammaire 1.

Une dérivation peut être également représentée sous forme de diagramme. La Figure 6.1 montre le diagramme pour la dérivation utilisée pour la phrase « the man hit a ball ». Ce diagramme est moins riche en informations que la dérivation, car l'ordre d'application des règles de réécriture n'apparaît pas. Le diagramme ne retient de la dérivation que ce qui est essentiel pour déterminer la structure syntagmatique. Il permet alors de voir clairement la structure syntaxique hiérarchique de la phrase et une visualisation de l'analyse en constituants.

6.1.2 Une grammaire syntagmatique pour les anatoles

Afin de montrer l'utilisation d'une grammaire syntagmatique pour créer un scénario multi-niveaux, nous avons décidé de créer une telle grammaire pour

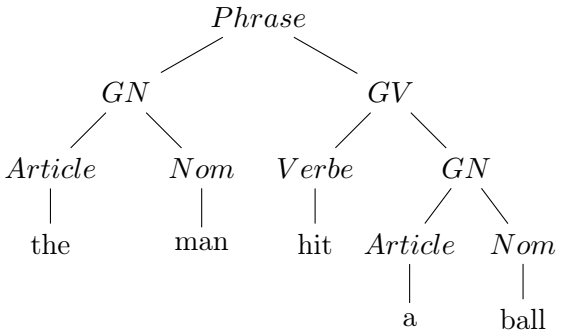

FIGURE 6.1 – Diagramme de la dérivation de la phrase « the man hit a ball ».

A	I VI-	II- V^7	I VI-	II- V^7
	I I^7	IV IV-	I VI-	II- V^7
A	I VI-	II- V^7	I VI-	II- V^7
	I I^7	IV IV-	I V^7	I
B	III^7	III^7	VI^7	VI^7
	II^7	II^7	V^7	V^7
A	I VI-	II- V^7	I VI-	II- V^7
	I I^7	IV IV-	I V^7	I

FIGURE 6.2 – Grille d'accords du thème *I Got Rhythm* de George Gershwin. Les accords sont marqué en degrés par rapport à la tonalité initiale du morceau.

une progression d'accords très utilisée dans le jazz traditionnel et le *bebop* : l'*anatole*. Nous décrivons en premier lieu la structure d'un anatole du point de vue de l'analyse musicale (une description plus précise peut-être trouvée dans [Siron, 2015]), puis nous formalisons une grammaire syntagmatique appropriée représentant cette analyse hiérarchique de la progression d'accords.

Analyse de l'anatole

L'anatole (en anglais *rhythm changes*) est une progression d'accords de 32 mesures, basée sur le thème *I Got Rhythm* de George Gershwin. Le terme *rhythm changes* est en fait un diminutif de « *Chord changes of I Got Rhythm* » (en français « progression d'accords de *I Got Rhythm* »). Le terme français « anatole » a été inventé par Jean-Claude Fohrenbach. La Figure 6.2 montre la progression d'accord originale de *I Got Rhythm*.

La structure globale de l'anatole consiste en une forme $AABA$, où la section B appelée le pont rentre en contraste avec les sections A.

— D'un côté, les sections A sont des sections de huit mesures avec des accords changeant rapidement avec généralement deux accords par mesure

et restant proches de la tonalité initiale. Ces sections sont constituées de :

- une série de deux *cellules-anatoles* sur les mesures 1&2 puis 3&4. La cellule-anatole est une fonction tonale de deux mesures basée sur un fragment du cycle des quintes. Sa forme la plus caractéristique est `I VI- II- V`, mais elle est régulièrement modifiée par les substitutions classiques (par exemple, `I I II-` $\mathtt{V}^7$, ou `III- VI`^{7} `II-` $\mathtt{V}^7$, etc.).
- une *cellule-christophe* sur les mesures 5&6 faisant sonner le IVème degré (sous-dominante). Le terme « christophe » a également été inventé par Jean-Claude Fohrenbach d'après le thème *Christopher Columbus*. Sa forme caractéristique est `I` $\mathtt{I}^7$ `IV IV-`, mais elle est également régulièrement modifiée par les substitutions classiques. Par exemple, `V- I`^{7} `IV` $\sharp$`IVo`, ou `I` $\mathtt{I}^7$ $\mathtt{IV}^7$ $\flat$$\mathtt{VII}^7$, etc.
- une dernière cellule-anatole sur les mesures 7&8. À l'exception du premier *A*, cette cellule-anatole peut être remplacée par une *cadence-boucle* ayant pour but soit de conclure dans l'accord de tonique soit d'anticiper la fonction d'accord suivante. Une cellule-anatole est alors un cas particulier de cadence-boucle. Par exemple, une cadence-boucle peut être `II-` $\mathtt{V}^7$ `I I`.

— D'un autre côté, la section *B* est une section de huit mesures avec une progression plus lente où chaque accord est généralement joué sur deux mesures, basé sur des accords de dominante suivant le cycle des quintes ($\mathtt{III}^7$ $\mathtt{VI}^7$ $\mathtt{II}^7$ $\mathtt{V}^7$) donnant une sensation de changement tonal. Les improvisateurs ont tendance souligner ces changements en insistant sur les notes guides (c'est-à-dire la tierce et la septième) de ces accords de dominante. Encore une fois, des substitutions peuvent être effectuées, par exemple, les deux mesures de $\mathtt{V}^7$ peuvent être remplacées par une mesure de `II-` suivie d'une mesure de $\mathtt{V}^7$ (ou de $\flat\mathtt{II}^7$).

L'anatole est un cas d'étude intéressant pour notre application, car il existe une grande quantité de variations autour de cette progression d'accords. C'est d'ailleurs l'attrait principal pour les musiciens qui peuvent modifier la progression à la volée et effectuer différentes substitutions au cours de l'improvisation, tant que la forme globale et les différentes fonctions sont présentes. La progression d'accords jouée peut alors être différente à chaque itération de la grille. Parmi les anatoles les plus célèbres, on peut citer *Anthropology*, *Dexterity*, *Oleo*, *Rhythm-A-Ning*, *The Eternal Triangle*...

Utiliser une grammaire syntagmatique pour analyser et générer des anatoles semble alors approprié. En considérant les accords, les fonctions tonales et les sections comme constituants, nous pouvons créer une grammaire syntagmatique représentant la structure hiérarchique des anatoles et où les accords sont les symboles terminaux.

Construction de la grammaire

Afin de créer une grammaire syntagmatique pour l'anatole, nous avons analysé avec Pascal Mabit, musicien professionnel et professeur de jazz, l'ensemble des anatoles issus du corpus de l'*Omnibook* (voir partie 2.2). Ce sous-corpus contient 26 variations différentes d'anatoles.

Nous définissons d'abord les notations utilisées pour les différents constituants décrits dans la partie précédente et jugés nécessaires pour la création de la grammaire :

— On note τ une cellule-anatole. On note également $\tau_{\texttt{I}}$ le sous-ensemble des cellules-anatoles commençant par le degré `I` ($\tau_{\texttt{I}} \subset \tau$).
— On note σ une cellule-christophe.
— On note ω une cadence-boucle (à noter que $\tau \subset \omega$).
— On note $\delta_{\texttt{X}}$ les fonctions d'accords de dominante sur le degré `X` trouvées dans la section B.

La grammaire syntagmatique pour les anatoles que nous avons créée à partir de l'analyse du sous-corpus de l'*Omnibook* est présentée dans la Grammaire 2.

Grammaire 2 Grammaire syntagmatique pour les anatoles

1: $Anatole \rightarrow A_1 + A_2 + B + A$
2: $A_1 \rightarrow \tau_{\texttt{I}} + \tau + \sigma + \tau$
3: $A_2 \rightarrow \tau_{\texttt{I}} + \tau + \sigma + \omega$
4: $A \rightarrow A_1,\ A_2$
5: $B \rightarrow \delta_{\texttt{III}} + \delta_{\texttt{VI}} + \delta_{\texttt{II}} + \delta_{\texttt{V}}$
6: $\tau_{\texttt{I}}, \tau, \sigma, \omega, \delta_{\texttt{III}}, \delta_{\texttt{VI}}, \delta_{\texttt{II}}, \delta_{\texttt{V}}$ sont appris sur le corpus.

— La règle 1 décrit la structure en $AABA$ de l'anatole. Ces sections de huit mesures sont les plus grands constituants après la progression d'accords complète.
— Les règles 2 et 3 décrivent la composition d'une section A en quatre fonctions tonales de deux mesures. Une section A commence toujours avec un élément de $\tau_{\texttt{I}}$ pour affirmer la tonalité au début de la section avec un accord de degré `I`. Ensuite, un autre τ est joué, suivi par le passage plagal σ. La différence entre un A_1 et un A_2 se situe sur les deux dernières mesures. D'un côté, A_1 se termine par une cellule-anatole τ, de l'autre côté A_2 se termine par une cadence-boucle ω.
— La règle 4 indique que la dernière section A peut soit être un A_1 ou un A_2.
— La règle 5 décrit la composition d'une section B en quatre fonctions tonales de deux mesures. Chaque δ décrit un déplacement tonal sur le cycle des quintes en suivant les degrés `III VI II` et `V`.
— La règle 6 spécifie que le passage des symboles non terminaux (ici fonctions tonales) aux symboles terminaux (ici accords) est appris sur un corpus d'apprentissage. Chaque fonction tonale est composée d'une séquence de quatre accords d'une durée de deux temps (avec la possibilité de répétition). Apprendre les symboles terminaux sur un corpus permet de considérer un ensemble large de possibilités pour chaque fonction et d'effectuer une forme de modélisation du style du corpus pour la génération de scénario multi-niveaux.

La Figure 6.3 montre le diagramme d'une dérivation de cette grammaire pour l'un des anatoles de l'*Omnibook* sur le thème *Celerity*.

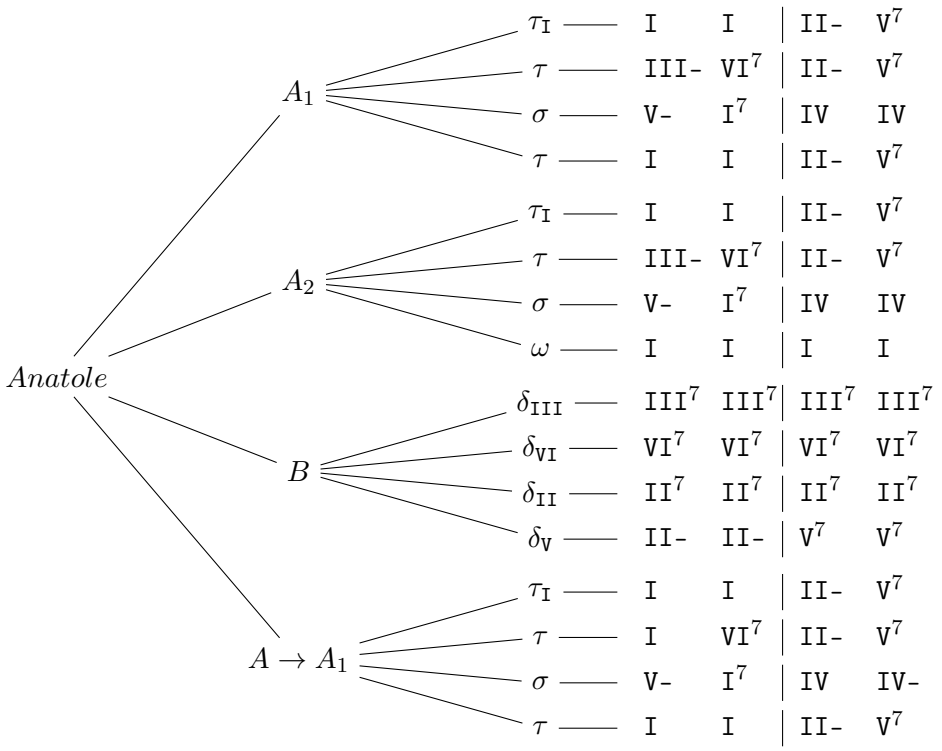

FIGURE 6.3 – Diagrame de la dérivation d'anatole du thème *Celerity* de Charlie Parker (issu de l'*Omnibook*. Chaque accord dure deux temps. $A \rightarrow A_1$ représente l'application de la règle 4 de la grammaire 2.

6.1.3 Évaluation de la grammaire

Afin d'évaluer la Grammaire 2 pour la représentation des anatoles, nous avons généré avec elle 30 dérivations d'anatoles. Les progressions d'accords multi-niveaux ont été analysées par deux musiciens de jazz professionnels :

— Pascal Mabit, avec qui cette grammaire a été construite,
— Louis Bourhis, contrebassiste de jazz qui n'était pas impliqué dans le processus et qui par conséquent a analysé les progressions d'accords d'un point de vue purement musicologique.

La Figure 6.4 montre un exemple d'anatole généré par la grammaire syntagmatique. D'autres dérivations d'anatoles sont disponibles sur : `members.loria.fr/KDeguernel/equivalent-chord-progressions`.

Tous les anatoles générés avec la grammaire syntagmatique ont été validés par les deux musiciens comme des anatoles valides. Cependant, les progressions d'accords générées doivent être vues comme des réalisations d'anatoles effectuées en pratique lors de l'accompagnement d'une improvisation. Ceci est cohérent avec le corpus d'apprentissage utilisé dans le sens où les progressions d'accords notées dans l'*Omnibook* essaient de se rapprocher à chaque grille à ce qui a été joué en accompagnement des solos de Charlie Parker. Pour la composition automatique de grilles d'anatoles pour un thème de jazz, il serait judicieux d'utiliser des contraintes de parallélisme [Giraud et al., 2014] pour

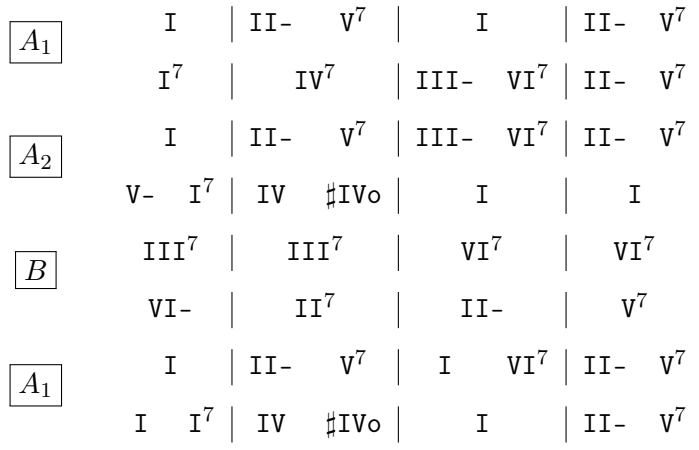

FIGURE 6.4 – Exemple de progression d'accords générée par la grammaire syntagmatique pour les anatoles.

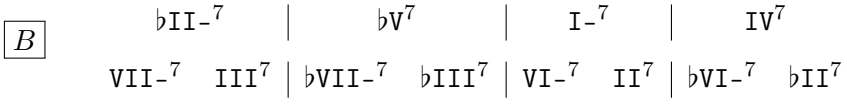

FIGURE 6.5 – Section B de l'anatole *The Eternal Triangle* de Sonny Stitt.

assurer des symétries entre les différents A d'une même grille.

De plus, les anatoles générés avec cette grammaire couvrent l'ensemble des possibilités des anatoles traditionnels de type *bebop*, ce qui correspond au corpus d'apprentissage utilisé. Aucune variation importante d'anatole de ce style n'a été jugée manquante, validant la capacité de modélisation du style de l'apprentissage des symboles terminaux de la grammaire. Cependant, cette grammaire ne permet pas de générer des formes plus modernes d'anatoles comme par exemple celle de *The Eternal Triangle* par Sonny Stitt qui possède une structure différente pour la section B, ou la version de *The Theme* par Lee Morgan dans l'album *Straight Ahead* qui possède des substitutions d'accords *post-bebop*. La figure 6.5 montre la section B de *The Eternal Triangle*. Cette section conserve l'aspect contrastant du pont avec la section A et la notion de déplacement tonaux sur le cycle des quintes, mais sa structure diffère du pont de l'anatole traditionnel et utilise beaucoup de substitutions tritoniques. Il serait intéressant d'étendre la grammaire pour les anatoles pour des progressions d'accords de styles plus variés.

Ayant montré qu'une grammaire syntagmatique était une représentation efficace de l'organisation structurelle hiérarchique d'un scénario, nous souhaitons effectuer un apprentissage automatique de cette structure.

6.2 Apprentissage de structures multi-niveaux

Dans cette partie, nous proposons une méthode pour effectuer un apprentissage automatique de structures multi-niveaux à partir d'un corpus d'appren-

tissage (dans notre cas des séquences d'accords). Pour cela, nous souhaitons utiliser des méthodes d'induction de grammaire (voir partie 3.2.2). Notre idée se base sur les travaux de Zitouni et al. [2000] sur la sélection de séquences de mots pour effectuer une compression de l'information d'une séquence pour créer une grammaire représentant des éléments structurels de la musique. Contrairement aux travaux présentés par Guichaoua [2017] sur l'utilisation de grammaire à dérivation unique [Gallé, 2011] pour la description de structure musicale dans un morceau, nous proposons une méthode probabiliste utilisant un corpus d'apprentissage plus large. Ceci permet de pouvoir déterminer des structures communes dans des scénarios équivalents et permet également d'éviter les problèmes de modélisation lors de l'apparition de variations locales d'un même motif.

La première étape de notre méthode est d'effectuer une segmentation du scénario en unifiant en un symbole les paires de symboles partageant l'information mutuelle la plus forte. Pour deux symboles a et b consécutifs, l'information mutuelle $J(a, b)$ est définie par

$$J(a,b) = \log \frac{\texttt{count}(a.b)T}{\texttt{count}(a)\texttt{count}(b)} \ , \tag{6.3}$$

où `count` est la fonction de comptage et T est la taille du corpus d'apprentissage. Une grande valeur d'information mutuelle signifie que les symboles a et b apparaissent consécutivement de manière bien plus fréquente qu'ils ne le seraient par pur hasard. Le regroupement itératif des symboles partageant l'information mutuelle la plus forte permet alors de former une structure hiérarchique du scénario.

Nous avons alors appliqué cette méthode sur le sous-corpus d'anatoles de l'*Omnibook*, constituant 26 variations différentes d'anatoles. Les résultats obtenus sont mitigés et plusieurs limites ont été détectées :

- L'application itérative de cette méthode a tendance à provoquer un phénomène d'agglutination autour des symboles rares, limitant les regroupements des symboles par petits groupes.
- Aucune relation autre que l'identité n'existe entre les différents symboles. Par exemple, toutes les variations de cellules-anatoles sont considérées strictement différentes (`I VI- II- V`^{7} est symboliquement différent de `III- VI`^{7} `II- V`^{7} alors que ces séquences ont la même fonction tonale).
- Bien que l'organisation aux plus bas niveaux semble raisonnable, lorsque l'on considère des niveaux d'organisation plus élevée, les résultats obtenus sont très mauvais. Ceci est lié au problème précédent, car à plus hauts niveaux, nous obtenons un alphabet trop grand pour trop peu de données.

Tout d'abord, pour réduire le premier problème, nous introduisons la notion de durée des symboles, correspondant à la durée des accords qu'ils représentent. Pour un symbole a, nous notons sa durée $l(a)$. Ainsi, pour éviter le problème d'agglutination autour des symboles rares, nous souhaitons privilégier l'unification des paires des symboles de courtes durées. Pour cela, nous proposons

une normalisation de l'information mutuelle par la durée des symboles :

$$\tilde{J}(a,b) = \frac{1}{l(a)+l(b)} \log \frac{\texttt{count}(a.b)T}{\texttt{count}(a)\texttt{count}(b)} \ . \tag{6.4}$$

Ensuite, pour réduire le second problème, nous proposons lors de la création d'un nouveau symbole de vérifier s'il peut être considéré comme une variation d'un symbole d'une même durée en utilisant la structure séquentielle. Nous souhaitons considérer comme équivalent deux symboles d'une même durée avec des voisinages similaires en terme d'information mutuelle. Deux symboles a et b sont considérés équivalents si

$$\Psi(a,b) = \frac{1}{K} \sum_{u,v} (J(u,a) - J(u,b))^2 + (J(a,v) - J(b,v))^2 \leq \xi \ , \tag{6.5}$$

avec K la taille du vocabulaire, u la liste des symboles à gauche de a et b, v la liste des symboles à droite de a et b et ξ un seuil à déterminer expérimentalement en fonction du corpus.

Nous proposons alors l'Algorithme 5 pour effectuer l'induction de grammaire à partir d'un corpus de scénario.

Algorithme 5 Induction de grammaire sur un corpus de scenario

Entrée : Corpus de scénarios.
Sortie : Liste de règles de réécriture.
Répéter
  Trouver a et b tels que $\tilde{J}(a,b) = \max_{x,y} \tilde{J}(x,y)$.
  Créer la règle de réécriture $X_{ab} \rightarrow a + b$.
  $l(X_{ab}) \leftarrow l(a) + l(b)$.
  Remplacer toutes les occurrences de $a + b$ par X_{ab} dans le corpus.
  Si $\exists$ un symbole Y tel que $l(Y) = l(X_{ab})$ **et** $\Psi(Y, X_{ab}) < \xi$ **alors**
    Créer la règle de réécriture $Y \rightarrow X_{ab}$.
    Remplacer toutes les occurrences de X_{ab} par Y dans le corpus.
  Fin Si

Nous avons appliqué cet algorithme sur le sous-corpus d'anatoles de l'*Omnibook*. La Figure 6.6 montre l'analyse hiérarchique obtenue sur un exemple d'anatole. L'analyse automatique de 10 anatoles a été verifiée et validée par Pascal Mabit. Premièrement, les différentes fonctions harmoniques (cellules-anatoles, cellules-christophe, etc.) sont retrouvées correctement et l'organisation structurelle est correcte aux différentes échelles souhaitées : accords, fonctions tonales et sections. Deuxièmement, les différentes variations d'une même fonction harmonique sont correctement identifiées comme équivalents. Par exemple, `I VI`^{7} `II- V`^{7} et `III- ♭III`^{7} `II- V`^{7} sont jugés équivalent, de même pour `V- I`^{7} `IV IV-` et `I`^{7} `I`^{7} `IV ♭VII`^{7}.

Cependant, cette analyse est moins précise que celle obtenue par la grammaire créée dans la partie précédente sur certains points :

— Les différentes types de sections A (appelés A_1 et A_2 dans la Grammaire 2) sont considérées comme strictement différentes.

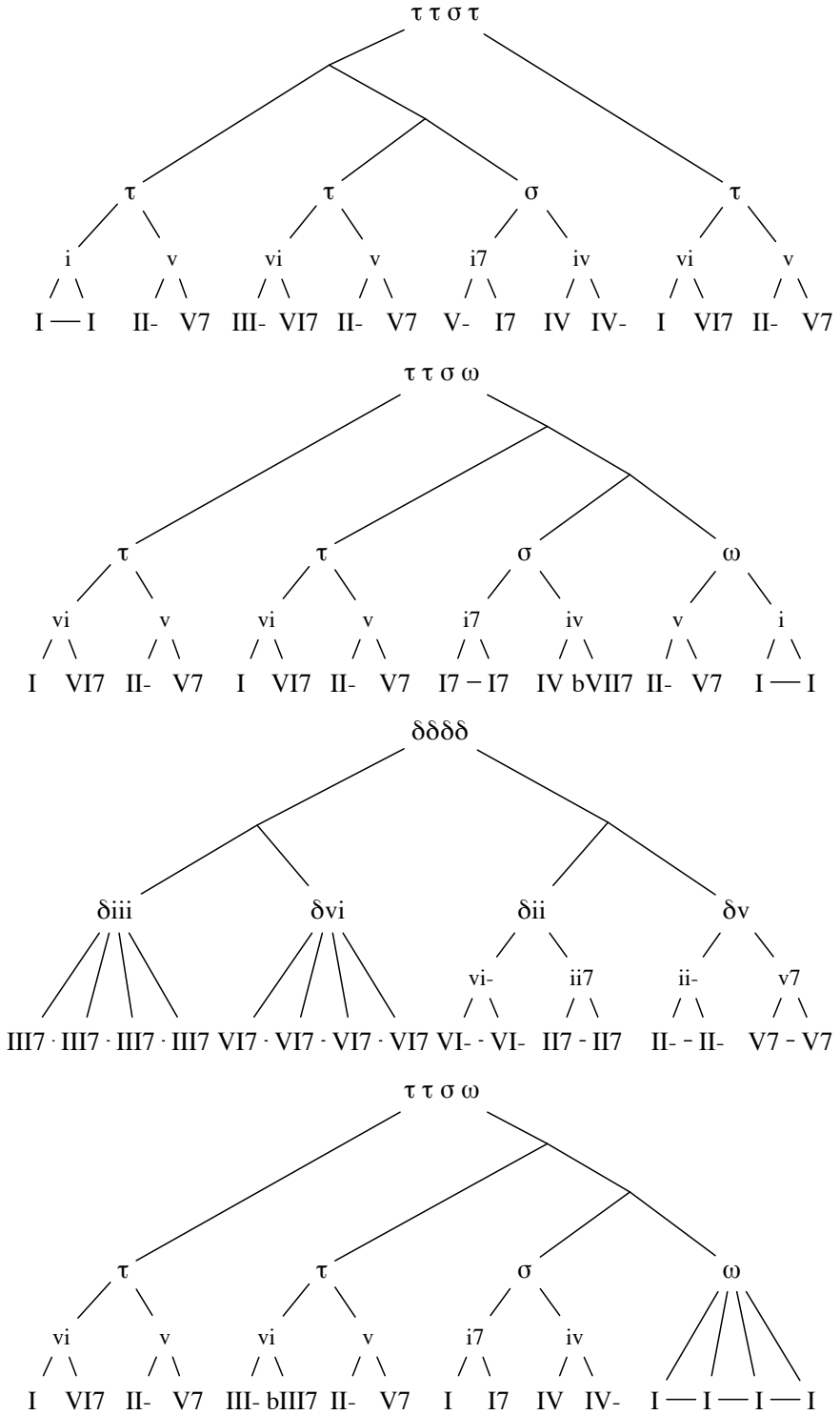

FIGURE 6.6 – Analyse hiérarchique obtenu à partir de l'Algorithme 4 d'un anatole. Le dernier niveau d'organisation n'est pas représenté ici.

— L'ensemble des variations de cellules-anatoles ont été jugés équivalentes effaçant alors le fait que la première cellule-anatole d'un A doit commencer par un accord de degré `I` (ce sous-ensemble de cellule-anatoles est appelé $\tau_{\texttt{I}}$ dans la Grammaire 2).

Finalement, les résultats de cette méthode sont encourageants. Les structures multi-niveaux obtenues sont très proches de celles obtenues par la grammaire créée avec un musicien professionnel. Cependant, nous avons utilisé ici un corpus homogène de scénario. Il serait intéressant de voir si des résultats satisfaisants pourraient être obtenus sur des corpus plus larges, mais aussi plus variés, comme par exemple celui du *Realbook*.

6.3 Exploiter une structure multi-niveaux dans l'improvisation guidée

Dans cette partie, nous présentons comment utiliser l'ensemble de l'information d'une scénario multi-niveaux pour enrichir la méthode d'improvisation de scénario temporel d'*ImproteK* présentée par Nika et al. [2017a] en prenant en compte la structure hiérarchique du scénario fournie par la grammaire syntagmatique. Nous rappelons d'abord les principes utilisés par Nika et al. pour générer des improvisations sur un scénario, puis nous proposons une méthode pour utiliser l'information multi-niveaux pour étendre les possibilités de génération.

6.3.1 Improvisation sur un scénario temporel

L'idée principale d'*ImproteK* est d'utiliser des connaissances *a priori* d'un scénario temporel afin d'introduire des mouvements d'anticipation dans le guidage de la génération d'improvisations. Dans le cas d'*ImproteK*, les scénarios sont représentés par une séquence de symboles appelés *étiquettes*. Contrairement à *OMax* et aux modèles que nous avons présentés dans les chapitres précédents, au cours de l'apprentissage, la mémoire du système n'est pas organisée selon la dimension qui est improvisée, mais est une séquence de contenus musicaux de la dimension improvisée organisée à partir des étiquettes de scénario. Pour reprendre l'exemple du jazz, les étiquettes peuvent être les accords d'une progression harmonique et les contenus les notes de la mélodie jouée par un improvisateur. Lors du processus de génération, pour un scénario donné, on cherche alors à combiner par rapport aux étiquettes une anticipation des éléments du scénario futur avec une navigation de la mémoire de manière similaire à *OMax* pour conserver une cohérence stylistique. Les contenus musicaux des états choisis sont alors joués pour générer l'improvisation.

Le processus de génération d'une improvisation se déroule alors en deux étapes successives d'anticipation puis de navigation (voir [Nika et al., 2017a] pour une description plus détaillée des algorithmes). Notons $S = S_1...S_s$ le scénario, T la position courante du scénario et considérons une mémoire construite avec un oracle des facteurs constitué des états $0...m$ avec les étiquettes $\Lambda_0...\Lambda_m$.

1. L'étape d'anticipation consiste à chercher des évènements dans la mémoire partageant un futur commun avec la position courante du scénario

tout en garantissant une continuité avec le passé de la mémoire. Cette étape est effectuée en indexant dans la mémoire les préfixes des suffixes de la position courante du scénario. On construit d'abord l'ensemble des états de la mémoire partageant un futur commun avec la position courante du scénario $S_T...S_s$:

$$\text{Futur}(T) = \{j \in [0...m] \mid \exists c_{\text{futur}} \in \mathbb{N}, \Lambda_j...\Lambda_{j+c_{\text{futur}}-1} \in \text{Pref}(S_T...S_s)\} \ . \tag{6.6}$$

Les $j \in \text{Futur}(T)$ sont alors les positions de début des facteurs de la mémoire correspondant à un préfixe du scénario à partir de la position courante. La taille de ces facteurs est mesurée par c_{futur}.

Puis, on construit l'ensemble des états de la mémoire partageant un passé commun avec l'état courant i de la mémoire :

$$\text{Passé}(i) = \{j \in [0...m] \mid \exists c_{\text{passé}} \in [1, j], \Lambda_{j-c_{\text{passé}}+1}...\Lambda_j \in \text{Suff}(0...i)\} \ . \tag{6.7}$$

Les $j \in \text{Passé}(i)$ sont les positions de fin des facteurs de la mémoire correspondant à un suffixe de l'évènement courant dans la mémoire. La taille de ces facteurs est mesurée par $c_{\text{passé}}$. La construction de cet ensemble est en pratique effectué en utilisant les propriétés sur les liens suffixiels de l'oracle des facteurs.

Pour l'étape d'anticipation, on cherche alors les positions j dans la mémoire telles que :

$$\text{Ant}(T, i) = \{j \in [0...m] \mid j \in \text{Futur}(T) \wedge j - 1 \in \text{Passé}(i)\} \ . \tag{6.8}$$

La Figure 6.7 illustre cette étape d'anticipation.

2. L'étape de navigation consiste à chercher des évènements de la mémoire partageant un contexte commun avec la position courante du scénario tout en respectant le scénario. On cherche les positions j dans la mémoire telles que :

$$\text{Nav}(T, i) = \{j \in [0...m] \mid \Lambda_j = S_T \wedge j - 1 \in \text{Passé}(i)\} \ . \tag{6.9}$$

Cette étape permet alors d'effectuer des chemins non-linéaires dans la mémoire permettant de créer des nouvelles phrases musicales, de manière similaire à *OMax* ou des modèles présentés dans les chapitres précédents. Cela permet au systèmes des variations plus locales que si l'on utilisait l'étape d'anticipation seule.

En pratique des valeurs minimale et maximale pour c_{futur} et $c_{\text{passé}}$ sont fixées pour éviter l'utilisation de fragments de la mémoire trop courts ou trop longs lors de la génération de l'improvisation.

6.3.2 Utiliser l'information multi-niveaux

Nous voulons à présent prendre en considération un scénario multi-niveaux pour la génération d'improvisations. Nous nous mettons alors dans le cas où le scénario n'est plus une simple séquence de symboles, mais une séquence de listes de symboles correspondant à chaque niveau. Pendant la construction

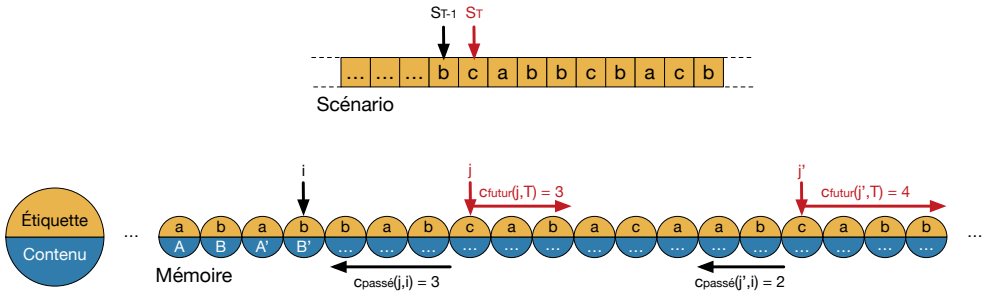

FIGURE 6.7 – Construction de l'ensemble des éléments de la mémoire partageant un futur commun avec la position courante du scénario T et un passé commun avec l'évènement courant de la mémoire i_{T-1} pour l'étape d'anticipation du processus de génération d'improvisation sur un scénario temporel [Nika et al., 2017a].

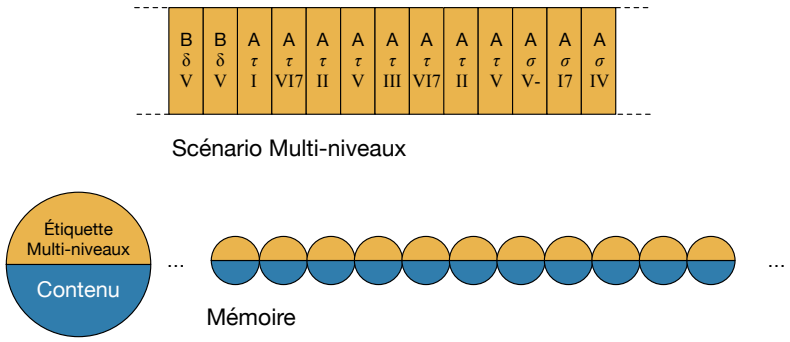

FIGURE 6.8 – Représentation d'un scénario et d'une mémoire multi-niveaux.

de la mémoire, celle-ci est alors organisée à partir d'étiquettes multi-niveaux correspondant aux scénarios multi-niveaux. La Figure 6.8 montre un exemple d'un scénario multi-niveaux avec trois niveaux d'organisation temporelle.

Nous adaptons alors la méthode de génération présentée dans la partie précédente pour prendre en compte l'information ajoutée par l'aspect multi-niveaux à la fois pour l'étape d'anticipation et pour l'étape de navigation. Pour chaque étape nous souhaitons étendre les possibilités de positions dans la mémoire à des états partageant des étiquettes multi-niveaux équivalentes à celles du scénario, c'est-à-dire des étiquettes partageant des informations communes à certains niveaux mais différents à d'autres niveaux. Par exemple, dans le cas du jazz, nous pouvons vouloir accepter des positions dans la mémoire qui n'ont pas le même accord que celui dans le scénario, mais qui possèdent la même fonction tonale et la même section. De cette manière, il est possible de générer des improvisations sur un scénario avec un accord n'existant pas dans la mémoire, mais partageant la même fonction que des accords connus. Nous voulons également être capables de favoriser les positions de la mémoire partageant une information commune forte avec le futur du scénario et le passé de la mémoire. Ainsi la génération rend mieux compte de la structure hiérarchique

du scénario pour guider l'improvisation.

Ce processus peut se formaliser comme suit. Notons $S = S_1...S_s$ le scénario multi-niveaux de longueur s avec Q niveaux, avec $\forall p \in [1...S], S_p = \{S_p^1...S_p^Q\}$. Soit T la position courante du scénario et considérons une mémoire construite avec un oracle des facteurs constitué des états $0...m$ avec les étiquettes multi-niveaux $\Lambda_0...\Lambda_m$, avec $\forall p \in [0...m], \Lambda_p = \{\Lambda_p^1...\Lambda_p^Q\}$.

On dit qu'une séquence de symboles multi-niveaux $u = u_1...u_n$ avec $\forall p \in [1...n], u_p = \{u_p^1...u_p^Q\}$ est un préfixe équivalent à une séquence de symboles multi-niveaux $v = v_1...v_m$ avec $\forall p \in [1...m], v_p = \{v_p^1...v_p^Q\}$ par :

$$u = u_1...u_n \in \text{Pref\_eq}(v = v_1...v_m) \text{ ssi } \forall p \in [1...n], \exists q; u_p^q = v_p^q \ . \tag{6.10}$$

De même, on dit qu'une séquence de symboles multi-niveaux $u = u_1...u_n$ avec $\forall p \in [1...n], u_p = \{u_p^1...u_p^Q\}$ est un suffixe équivalent à une séquence de symboles multi-niveaux $v = v_1...v_m$ avec $\forall p \in [1...m], v_p = \{v_p^1...v_p^Q\}$ par :

$$u = u_1...u_n \in \text{Suff\_eq}(v = v_1...v_m) \text{ ssi } \forall p \in [1...n], \exists q; u_p^q = v_{p+m-n}^q \ . \tag{6.11}$$

Les deux étapes de génération sont alors modifiées pour prendre en compte l'information multi-niveaux et les étiquettes équivalentes.

1. Pour l'étape d'anticipation, on remplace l'ensemble Futur(T) par l'ensemble Futur_eq(T) définit par :

   $$\text{Futur\_eq}(T) = \{j \in [0...m] \mid \exists c_{\text{futur}} \in \mathbb{N}, \Lambda_j...\Lambda_{j+c_{\text{futur}}-1} \in \text{Pref\_eq}(S_T...S_s)\} \ . \tag{6.12}$$

   Les $j \in$ Futur_eq(T) sont les positions de début des facteurs de la mémoire correspondant à un préfixe équivalent du scénario à partir de la position courante.

   On remplace également l'ensemble Passé(i) par l'ensemble Passé_eq(i) définit par :

   $$\text{Passé\_eq}(i) = \{j \in [0...m] \mid \exists c_{\text{passé}} \in [1, j], \Lambda_{j-c_{\text{passé}}+1}...\Lambda_j \in \text{Suff\_eq}(0...i)\} \ . \tag{6.13}$$

   Les $j \in$ Passé_eq(i) sont les positions de fin des facteurs de la mémoire correspondant à un suffixe équivalent de l'évènement courant dans la mémoire.

   Finalement, pour l'étape d'anticipation, on considère toutes les positions j dans la mémoire telles que :

   $$\text{Ant\_eq}(T, i) = \{j \in [0...m] \mid j \in \text{Futur\_eq}(T) \wedge j-1 \in \text{Passé\_eq}(i)\} \ . \tag{6.14}$$

2. Les modifications pour l'étape de navigation sont similaires ; on étend les possibilités grâce aux étiquettes équivalentes. Ainsi, on cherche les positions j dans la mémoire telles que :

   $$\text{Nav\_eq}(T, i) = \{j \in [0...m] \mid \Lambda_j \sim S_T \wedge j-1 \in \text{Passé\_eq}(i)\} \tag{6.15}$$

   où $\Lambda_j \sim S_T$ si et seulement si $\exists q; \Lambda_j^q = S_T^q$.

Comme dans la partie précédente, en pratique des valeurs minimale et maximale pour c_{futur} et $c_{\text{passé}}$ sont fixées pour éviter l'utilisation de fragments de la mémoire trop courts ou trop longs lors de la génération de l'improvisation.

Afin de privilégier à chaque étape les positions j partageant le plus d'information commune avec le futur du scénario et le passé de la mémoire, un score de similarité entre étiquettes multi-niveaux est défini. L'utilisateur peut attribuer *a priori* à chaque niveau temporel q considéré un poids W_q tel que

$$\sum_q W_q = 1 \ . \tag{6.16}$$

Par exemple, dans le cas du jazz, on peut considérer que le niveau de la fonction tonale est plus important que celui des accords ou que celui des sections.

La similarité entre deux étiquettes multi-niveaux Λ_i et Λ_j est alors égale à

$$\varsigma(\Lambda_i, \Lambda_j) = \sum_q W_q \delta_{\Lambda_i^q, \Lambda_j^q} \ , \tag{6.17}$$

où δ est le symbole de Kronecker.

Chaque élément j dans $\text{Ant\_eq}(T, i)$ peut alors obtenir un score défini par :

$$\sum_{k=0}^{c_{\text{futur}}(j,T)} \varsigma(\Lambda_{j+k}, S_{T+k}) + \sum_{k=0}^{c_{\text{passé}}(j,i)} \varsigma(\Lambda_{j-k}, \Lambda_{i-k}) \tag{6.18}$$

et chaque élément j dans $\text{Nav\_eq}(T, i)$ peut obtenir un score défini par :

$$\varsigma(\Lambda_j, S_T) + \sum_{k=0}^{c_{\text{passé}}(j,i)} \varsigma(\Lambda_{j-k}, \Lambda_{i-k}) \ . \tag{6.19}$$

À partir de ces scores, différentes stratégies peuvent être appliquées pour privilégier les éléments de $\text{Ant\_eq}(T, i)$ et $\text{Nav\_eq}(T, i)$ partageant une information commune forte avec le futur du scénario et le passé de la mémoire. L'élément avec le plus haut score peut être choisi, ou on peut limiter le choix aux éléments avec un score supérieur à un certain seuil. Dans nos expériences, nous normaliserons ces scores par la somme des scores de tous les évènements possibles et nous choisirons un élément de manière aléatoire en suivant les probabilités obtenues.

6.4 Évaluation et retours des musiciens

6.4.1 Choix des paramètres

Pour évaluer l'utilisation de scénarios multi-niveaux pour la génération d'improvisation, nous avons réalisé des tests d'écoute avec des musiciens professionnels afin d'obtenir une évaluation qualitative des générations. Nous avons fait appel à deux musiciens et improvisateurs professionnels :

— Louis Bourhis, contrebassiste de jazz, diplômé de la Haute École de Musique de Lausanne,

— Joël Gauvrit, pianiste et professeur de jazz et de musique classique, diplômé du Conservatoire National Supérieur de Musique et de Danse de Lyon. [9]

Les tests d'écoute ont été faits séparément.

Nous nous sommes intéressés au cas des anatoles issus du corpus de l'*Omnibook* de Charlie Parker. Nous avons tout d'abord utilisé ce sous-corpus pour apprendre les symboles terminaux de la grammaire syntagmatique présentée dans la partie 6.1.2. Cette grammaire a ensuite été utilisée pour générer des scénarios multi-niveaux d'anatoles avec trois niveaux d'organisation : les accords, les fonctions tonales et les sections. Sur ces scénarios, nous avons généré des improvisations en suivant deux méthodes :

1. La méthode de génération basée sur *ImproteK* présentée dans la partie 6.3.1. L'anticipation du scénario et la navigation dans la mémoire sont alors effectuées ici seulement sur le niveau des accords du scénario (les autres niveaux ne sont pas pris en compte). Lorsqu'un accord de scénario n'apparaît pas dans la mémoire, des principes de transposition sont appliqués.
2. La méthode de génération basée sur un scénario multi-niveaux présentée dans la partie 6.3.2. L'anticipation du scénario et la navigation dans la mémoire sont effectuées avec l'ensemble de l'information multi-niveaux. Nous avons considéré ici, avec l'appui d'un musicien, que pour les anatoles, le niveau le plus important est le niveau des fonctions tonales et nous avons attribué un poids de 0,5 à ce niveau. Puis nous avons attribué un poids de 0,3 au niveau des accords et un poids de 0,2 au niveau des sections.

Dans les deux cas, les contenus de la mémoire sont basés sur un thème et improvisation issu d'un anatole de l'*Omnibook*. Contrairement aux expériences précédentes, le découpage dans la mémoire n'est pas effectué note-à-note mais temps-à-temps pour pouvoir faire correspondre plus facilement les étiquettes de la mémoire avec les étiquettes du scénario. La Figure 6.9 résume le processus complet de génération utilisé dans le cas multi-niveau. Des exemples d'improvisations générées par les deux méthodes sont disponibles sur : `repmus.ircam.fr/dyci2/ressources`. Les Figures 6.11 et 6.12 montrent également des relevés d'improvisations générées dans le cas multi-niveaux respectivement avec la mémoire apprise sur *Thriving from a Riff* et sur *An Oscar for Treadwell*.

6.4.2 Résultats et analyses

Chaque musicien a écouté au total une douzaine d'improvisations, plus ou moins, selon son vouloir. À chaque fois, pour une même progression d'accords générée par la grammaire syntagmatique, une improvisation était générée par chacune des deux méthodes. La mémoire de l'oracle des facteurs a été construite à partir de trois morceaux issus du sous-corpus d'anatole de l'*Omnibook* : *Thriving from a Riff*, *An Oscar for Treadwell*, et *Anthropology*.

9. Des biographies plus détaillées des musiciens sont fournies en Annexe B.

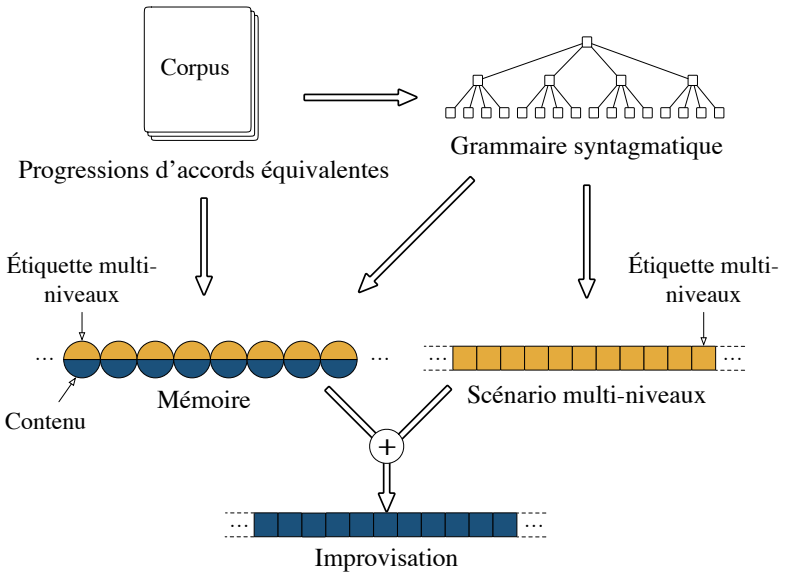

FIGURE 6.9 – Processus de génération d'improvisation sur un scénario multi-niveaux. Tout d'abord, une grammaire syntagmatique est construite avec un musicien à partir d'un corpus, puis les symboles terminaux de cette grammaire sont appris sur ce corpus. Cette grammaire est utilisée pour générer des scénarios multi-niveaux et pour fournir les informations multi-niveaux lors de l'apprentissage de la mémoire avec des étiquettes multi-niveaux. Finalement, les improvisations sont générées en anticipant le scénario multi-niveaux et en navigant la mémoire avec les étiquettes multi-niveaux.

Dès les premiers exemples, les deux musiciens ont remarqué une différence très nette dans la génération des deux méthodes. L'amélioration majeure constatée par les musiciens est la construction globale de l'improvisation sur l'ensemble de la grille. Les improvisations générées par la deuxième méthode fournissent des résultats jugés plus réalistes grâce à une meilleure prise en compte des structures harmoniques. Gauvrit dit :

> « C'est incroyable la différence. Ah ouais ! Bah oui là on a l'impression qu'il construit. Ah bah là c'est super. C'est un super chorus. J'ai un élève qui fait ça je suis content. C'est-à-dire que même rythmiquement, la façon dont il fonctionne à la structure harmonique. J'entends ça moi. Que l'autre non, du tout. »

Bourhis précise cette idée en faisant remarquer qu'avec la deuxième méthode, l'improvisation s'axe autour des fonctions et des sections, formant des phrases (ou ensembles de phrases) avec une logique à plus long terme. Cela peut être vu en particulier sur la Figure 6.11 :

> « L'architecture globale est beaucoup plus réfléchie. On sent qu'il est beaucoup plus à l'aise. Enfin l'ordinateur n'est pas à l'aise, mais nous on a l'impression d'écouter quelqu'un qui est beaucoup plus avancé. [...] C'est-à-dire, quand il commence une partie, il commence avec moins de notes et il va fournir dessus. Je trouve que

c'est plus humain dans ce sens-là. »

Les mélodies des improvisations générées par la deuxième méthode ont également été jugées généralement moins décousues. Cela peut s'expliquer par le fait que la deuxième méthode ouvre des possibilités dans la navigation permettant de faire moins de sauts dans la mémoire lors de progressions harmoniques locales difficiles. Bourhis dit :

> « C'est vrai qu'il y a les trucs, tu sais, du tout début où on a un peu peur, que l'ordi il prenne ce fragment, ce fragment et ce fragment et qu'il les pose les uns à la suite des autres sans trop les relier. Et du coup, ce que je t'ai expliqué tout à l'heure sur l'autre version, je trouve que ça fait une très grosse amélioration, au niveau de la crédibilité, du réalisme de Charlie Parker. »

Lors des passages où le scénario possède des accords n'apparaissant pas dans la mémoire, la liberté fournie par la deuxième méthode de pouvoir jouer sur d'autres accords partageant la même fonction tonale (et également la même section) permet également des improvisations avec moins de fragmentations, créant une cohérence et une sensation de fluidité plus forte. Par exemple, cela a été entendu sur les improvisations générées avec la mémoire de *An Oscar for Treadwell* où les accords du pont étaient $\texttt{III}^7$ $\texttt{III}^7$ $\texttt{VI}^7$ $\texttt{VI}^7$ $\texttt{II}^7$ $\texttt{II}^7$ $\texttt{V}^7$ $\texttt{V}^7$ et le scénario était constitué d'une série de sous-dominantes / dominantes : $\texttt{VII-}^7$ $\texttt{III}^7$ $\texttt{III-}^7$ $\texttt{VI}^7$ $\texttt{VI-}^7$ $\texttt{II}^7$ $\texttt{II-}^7$ $\texttt{V}^7$. Un exemple de cette situation peut être vu dans la Figure 6.12. Bourhis dit :

> « C'est juste que la première version du coup, il ne sait pas comment gérer les sous-dominantes. Avec la deuxième méthode, il va moins dans les détails, mais ça lui permet d'être plus cohérent à grande échelle. »

La Figure 6.10 propose une analyse du début de la section *B* de l'improvisation présentée dans la Figure 6.12. Une remarque similaire a été faite par Gauvrit : dans certaines sections *A*, la cellule-christophe contenait un accord diminué, qui est une nature d'accord n'apparaissant pas dans la mémoire d'*Anthropology* ; dans cette situation, nous n'avons pas pu générer d'improvisation avec la première méthode, mais nous avons pu avec la deuxième méthode qui fournit des résultats convainquants :

> « Il a jamais vraiment clashé quand même sur les $\sharp\texttt{IVo}$. Il s'est bien débrouillé. »

De plus, pouvoir jouer sur des accords de scénarios différents des accords de la mémoire apporte une forme de créativité, les notes guides et les extensions faisant sonner des nouvelles couleurs lors de l'improvisation.

De manière plus surprenante, une autre amélioration notée par les deux musiciens est le fait que les improvisations générées par la deuxième méthode ont une meilleure cohérence rythmique. Le développement des improvisations en terme de densité et de débit rythmique a été jugé plus réaliste. Gauvrit compare les deux méthodes à ses élèves de jazz :

> « La construction a du sens. T'as l'impression qu'il y a un mec qui pense derrière. Carrément, c'est impressionnant... C'est super intéressant pédagogiquement. J'ai l'impression de voir des stades

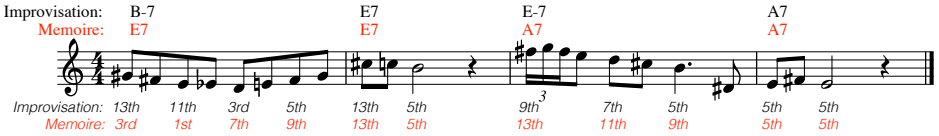

FIGURE 6.10 – Analyse des quatre premières mesures d'une improvisation avec la mémoire de *An Oscar for Treadwell* sur un scénario multi-niveaux généré par la grammaire syntagmatique. Les lignes marquées « Improvisation » montrent les accords et les degrés utilisés par la mélodie dans l'improvisation générée. Les lignes marquées « Mémoire » montrent les accords et les degrés des différentes parties de la mémoire utilisées. On remarque que l'improvisation joue une mélodie qui avait été apprise sur des accords différents des accords du scénario. Cependant, la mélodie a du sens dans sa continuité et harmoniquement, car ces accords partagent les mêmes fonctions tonales.

> différents chez mes élèves de compréhension tu vois, de ce qui est écrit, analytique et tout... Et c'est marrant la perception spontanément, parce qu'après quand t'y penses, c'est naturel, ce que ça engendre rythmiquement, le fait d'avoir conscience des structures harmoniques, ça engendre rythmiquement des choses qui sont beaucoup plus naturelles. »

Généralement, la deuxième méthode a été très peu critiquée par les musiciens. Bourhis a noté encore l'existence de sauts dans la mémoire un peu perturbants (avec par exemple des sauts d'octave peu réalistes), mais leur fréquence est bien moindre avec la deuxième méthode qu'avec la première :

> « Après, il y a toujours tous ces trucs tordus, mais tout est renforcé avec la première méthode quand c'est décousu, les sauts d'octaves principalement. »

Finalement, Gauvrit étend par la suite l'intérêt pédagogique que ces différentes méthodes peuvent illustrer, en ré-expliquant que l'intégration des structures harmoniques dans le processus de génération apporte un réalisme notable :

> « Et moi j'y vois une application pédagogique géniale, parce que tu montres ça à des élèves. Je pense que c'est absolument frappant. Puisque finalement, ça correspond à la réalité... Et puis c'est marrant de voir que le stade... Enfin moi je trouve que le déclic du moment où vraiment on y croit, c'est au moment où il y a une compréhension analytique. C'est là où on s'en rend compte... Ça sert ma crèmerie, c'est pour ça que ça m'intéresse. »

6.4.3 Résumé des résultats

Les retours des musiciens sur cette expérience sont très satisfaisants. La prise en considération de la structure multi-niveaux d'une progression harmonique permet la génération d'improvisations beaucoup plus réalistes. En plus

de pouvoir improviser de manière plus libre et variée (même sur des accords non vus auparavant), celle-ci permet une amélioration de la cohérence mélodique et rythmique. Généralement, la génération « ne se préoccupe pas trop du détail au service de ce qui peut se passer à plus grande échelle », se rapprochant alors des processus d'improvisation employés par des musiciens humains, permettant une meilleure considération des espaces harmoniques et un meilleur phrasé.

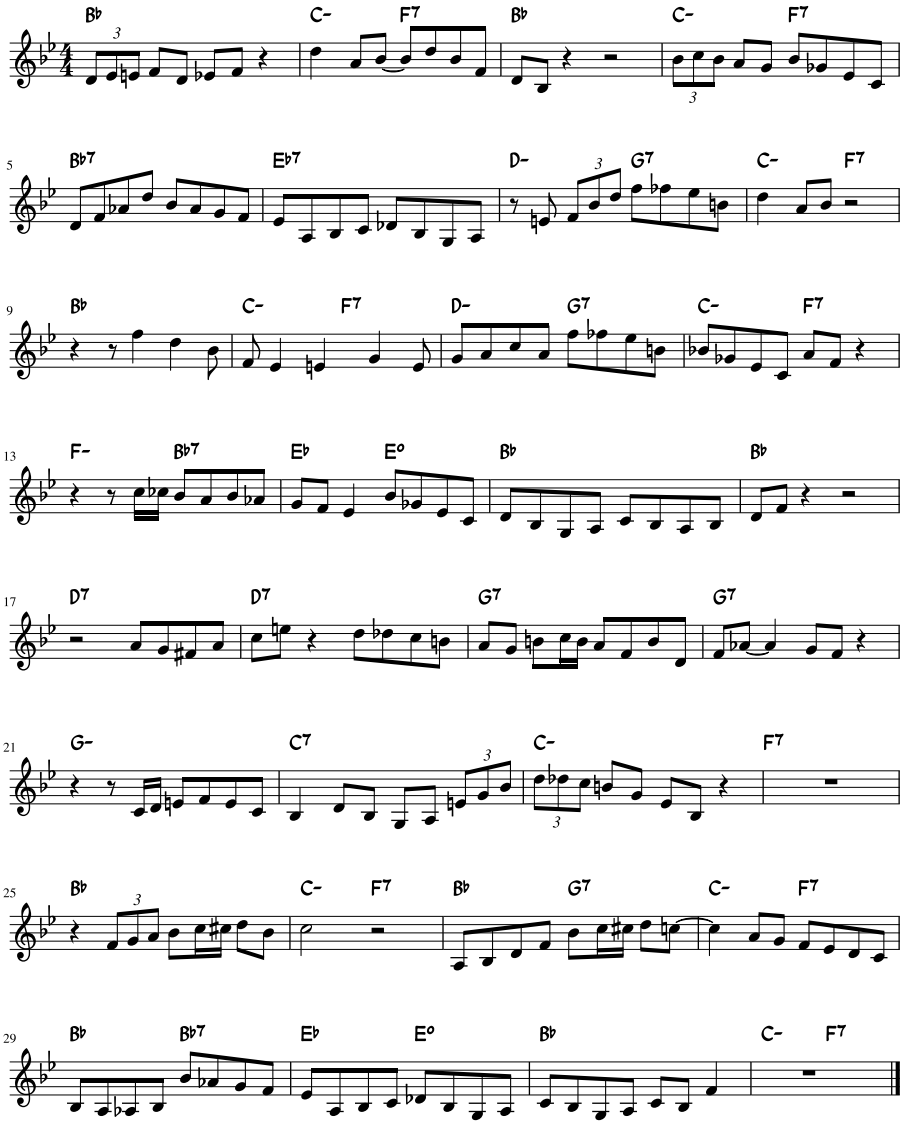

FIGURE 6.11 – Improvisation avec la mémoire de *Thriving from a Riff*. Progression d'accords générée avec la grammaire syntagmatique pour anatoles.

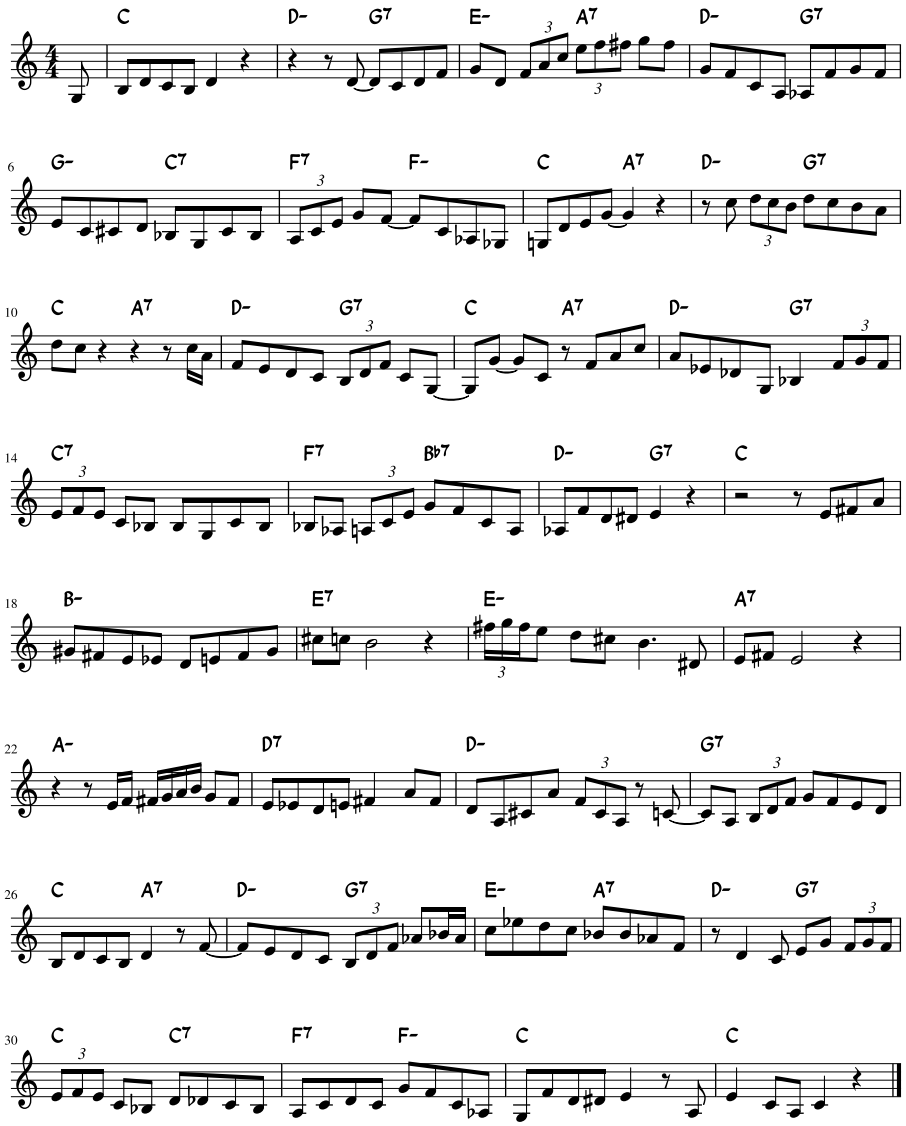

FIGURE 6.12 – Improvisation avec la mémoire de *An Oscar For Treadwell*. Progression d'accords générée avec la grammaire syntagmatique pour anatoles.

7

Conclusion

"We're not on the outside looking in, we're on the outside looking out."

– John Zorn

7.1 Résumé des contributions

Dans cette thèse, nous nous sommes intéressés à l'apprentissage de structures musicales *multidimensionnelles*, c'est-à-dire à une prise en considération de plusieurs données musicales porteuses d'informations sémantiques différentes mais corrélées comme la mélodie, l'harmonie, la densité, le rythme, etc., et nous nous sommes également intéressés aux structures *multi-niveaux*, c'est-à-dire à l'organisation du discours musical sur plusieurs échelles temporelles comme le point sonore, la mesure, la section, etc. Dans ce cadre, nous avons proposé trois contributions au domaine de l'improvisation musicale automatique.

La première contribution étudie l'utilisation d'un apprentissage de structures multidimensionnelles pour le guidage d'une improvisation unidimensionnelle. Nous avons proposé un paradigme d'improvisation que nous appelons Intuition / Connaissance où des connaissances d'un passé culturel guident une intuition sur un contexte local. Cela permet de profiter d'un apprentissage sur un large corpus tout en pouvant s'adapter au style particulier d'une improvisation permettant la création d'un vocabulaire musical personnel du système. Nous avons implémenté ce paradigme en utilisant des connaissances multidimensionnelles représentées par des modèles probabilistes interpolés appris sur un large corpus et une intuition unidimensionnelle représentée par un oracle des facteurs appris sur un petit corpus ou en direct à partir du jeu d'un musicien. Dans ce système, la génération d'improvisation consiste à naviguer dans l'oracle des facteurs en utilisant les modèles probabilistes pour obtenir des probabilités de transition. Ce système peut également se placer dans une situation d'improvisation guidée par un guidage réactif, en utilisant des informations issues de l'environnement pour s'adapter à des événements musicaux locaux. Ce système a été évalué par des improvisateurs professionnels pour obtenir une évaluation qualitative des improvisations générées. Nous avons montré qu'un

tel paradigme permet la génération d'improvisations plus réalistes et mieux organisées vis-à-vis des dimensions considérées et permet également la création de musiciens hybrides en combinant des connaissances et des intuitions de styles différents.

Avec la deuxième contribution, nous avons proposé une méthode pour générer des improvisations musicales multidimensionnelles. Ce système se base sur un principe multi-agents, inspiré par les interactions entre musiciens dans une situation de jeu libre, où les dimensions partagent leurs connaissances pour pouvoir prendre une décision sur leur improvisation nourrie de leurs connaissances internes et des connaissances externes des autres dimensions. Nous avons représenté chaque dimension par un système Intuition / Connaissance, encore une fois implémenté par la combinaison d'un oracle des facteurs et de modèles probabilistes. Les communications entre les dimensions s'effectuent à l'aide d'un algorithme de propagation de croyance sur un graphe de *clusters*. La génération d'improvisations multidimensionnelles consiste à naviguer dans les oracles des facteurs construits sur chaque dimension, en utilisant les connaissances internes modifiées par la propagation de croyance des autres agents pour obtenir les probabilités de transition. Ce système a été évalué par des improvisateurs professionnels pour obtenir une évaluation qualitative des improvisations générées. Nous avons montré que les improvisations générées par ce système sont cohérentes à la fois pour chaque dimension individuellement et pour les relations entre dimensions ; les dimensions semblant avoir une volonté commune pour le développement de l'improvisation.

Enfin, la troisième contribution porte sur l'utilisation de structures multi-niveaux pour la génération dans la situation d'une improvisation avec guidage structurel, c'est-à-dire avec l'utilisation d'un scénario temporel comme dans le système *ImproteK*. Pour cela nous avons proposé d'utiliser une grammaire syntagmatique pour modéliser l'organisation hiérarchique en plusieurs niveaux temporels d'un scénario pour créer des scénarios multi-niveaux. Nous avons également proposé une méthode d'apprentissage automatique de cette structure à partir d'un corpus. Pour la génération, nous avons proposé de nouvelles heuristiques utilisant l'information de tous les niveaux temporels pour guider l'improvisation avec des principes d'anticipation de scénario et de navigation dans un oracle des facteurs. L'importance du suivi des différents niveaux peut-être déterminé par l'utilisateur. Ce système a été évalué par des improvisateurs professionnels pour obtenir une évaluation qualitative des improvisations générées. Nous avons montré que la prise en considération de la structure hiérarchique d'un scénario permet d'obtenir des improvisations plus réalistes et mieux organisées, ayant une cohérence aussi bien au niveau du point sonore qu'aux niveaux de la phrase mélodique et du récit.

7.2 Perspectives

7.2.1 Perspectives théoriques

D'un point de vue théorique, nous proposons plusieurs extensions possibles aux travaux présentés dans cette thèse. Nous avons étudié les avantages obtenus par l'apprentissage de structures multidimensionnelles et multi-niveaux

dans des contextes séparés, mais il serait intéressant de combiner ces pratiques. En particulier, nous pourrions nous intéresser à l'utilisation du paradigme Intuition / Connaissance dans une situation d'improvisation avec un guidage structurel. Par exemple, on pourrait considérer une mémoire mélodique avec connaissances multidimensionnelles guidée par un scénario harmonique multi-niveaux. D'un côté l'anticipation d'un scénario futur avec des étiquettes multi-niveaux permettrait d'obtenir une liste d'états de la mémoire permettant une cohérence avec le scénario, de l'autre côté l'intuition et les connaissances sur les dimensions improvisées permettraient une meilleure construction séquentielle de l'improvisation. Ceci pourrait potentiellement permettre de diminuer les sauts dans la mémoire jugés étranges (par exemple, les sauts d'octaves si cette dimension est considérée).

Une autre piste serait d'inférer des principes d'anticipation directement dans le paradigme Intuition / Connaissance en utilisant des modèles probabilistes tournés vers le futur. L'utilisation d'un oracle des facteurs sur une dimension permet d'obtenir des informations sur sa navigation future notamment grâce à l'utilisation du facteur de continuité. Cette extension serait particulièrement intéressante dans le cas du modèle interactif entre dimensions, le futur de l'ensemble des dimensions pouvant alors être anticipé, permettant potentiellement l'apparition de conduites mélodiques ayant été jugées manquantes par les musiciens pour ce modèle.

Il serait également intéressant d'évaluer des méthodes d'implémentation différentes du paradigme Intuition / Connaissance. Par exemple, les principes d'interpolation de sous-modèles probabilistes appris sur des séquences multidimensionnelles pourraient être généralisés avec l'utilisation de réseaux de neurones profonds et/ou récurrents permettant de représenter des relations non-linéaires entre les dimensions.

Une limite du système d'interaction entre dimensions remarquée par les musiciens est le fait que la cohésion entre les dimensions est parfois trop forte, limitant alors les possibilités d'exploration musicale ; les interactions manquaient alors de réactivité. Il serait intéressant d'étudier comment inférer cette notion de réactivité dans cette situation. Une possibilité pourrait être de donner des rôles fluides (par exemple de meneur et de suiveurs) aux différentes dimensions [Canonne & Garnier, 2011] et d'intégrer ses rôles lors des passages de messages entre les dimensions.

De plus, nous avons utilisé comme méthode de passage de messages un algorithme de propagation de croyance. D'autres méthodes de passage de messages pourraient être étudiées et comparées comme, par exemple, les *gossip algorithms* [Vanhaesebrouck et al., 2017].

Nous avons vu qu'avec l'apprentissage de structures multi-niveaux, nous pouvons générer plusieurs scénarios équivalents en utilisant leur structure hiérarchique commune. Nous pourrions alors imaginer un système où l'improvisation s'adapte au scénario, mais où les symboles terminaux du scénario (par exemple, les accords dans le cas d'une progression harmonique) s'adaptent également à ce qui est improvisé. La génération d'un scénario pourrait être vue comme une dimension particulière et les principes d'interaction entre dimensions pourraient s'appliquer. Une piste de généralisation de cette idée est présentée dans [Nika et al., 2017b] avec la notion d'inférence de scénario où

une écoute active de ce qui est improvisé par un musicien humain permet de définir un scénario à court-terme. Il serait intéressant d'utiliser la notion de structure multi-niveaux pour réaliser des inférences de scénarios à plus long terme.

Finalement, nous avons vu qu'utiliser un scénario multi-niveaux permet la génération d'improvisations bien plus réalistes. Cependant, la méthode que nous avons proposée demande une analyse *a priori* d'un scénario fait en amont de la génération. Il serait intéressant de développer des techniques d'apprentissage automatique permettant une analyse de structures multi-niveaux en direct lors de l'écoute d'un musicien, y compris dans des situations de jeu sans guidage structurel.

7.2.2 Perspectives expérimentales

D'un point de vue expérimental, les différents tests d'écoute que nous avons effectués se sont basés uniquement sur des apprentissages dans des corpus à la fois pour l'intuition et pour les connaissances. Pour le système Intuition / Connaissance, il serait intéressant d'effectuer des expériences utilisant le jeu d'un musicien humain appris en temps réel pour représenter l'intuition du système et d'obtenir l'avis des musiciens sur les interactions qu'ils ont avec le système.

De manière similaire, il serait également intéressant d'insérer un musicien humain dans notre système interactif multi-agents. Cela pourrait être effectué en considérant le jeu du musicien pour effectuer du guidage réactif à la *SoMax*, en utilisant pour chaque agent des modèles probabilistes adaptés.

De plus, nous avons testé nos systèmes pour des improvisations dans le style *bebop*. Il serait intéressant d'expérimenter sur ces systèmes avec des styles différents et utilisant des intuitions et des connaissances apprises sur des dimensions différentes. On peut par exemple imaginer une improvisation libre de musique contemporaine utilisant comme dimensions des descripteurs audio.

Dans la même direction, il serait également intéressant d'effectuer des études ethnomusicologiques pour tenter de créer des avatars plus réalistes de musiciens célèbres en déterminant les dimensions pertinentes à modéliser et en constituant des corpus d'apprentissage appropriés à la fois pour les connaissances et l'intuition du système.

Par rapport à l'apprentissage de structure multi-niveaux, une des limites des systèmes d'improvisation avec guidage structurel est la capacité de générer des improvisations qui se développent de manière cohérente sur plusieurs itérations d'un même scénario. Il serait intéressant de tester si les méthodes de modélisation de scénarios multi-niveaux pourraient être une méthode pour résoudre ce problème, en ajoutant un niveau d'organisation temporel supérieur au scénario.

Finalement, il serait intéressant de développer avec des musiciens professionnels les travaux sur la méta-composition de scénarios commencés par Nika [2016] pour les étendre à la méta-composition de scénarios multi-niveaux.

Annexe A

Écoutes recommandées

Quelques albums/pièces des musiciens mentionnés dans cette thèse.

Ian Anderson
Aqualung Chrysalis, 1971
Thick as a Brick Island, 1972
Live at Montreux 2003 Eagle Records, 2007
Homo Erraticus Kscope, 2014

Derek Bailey
Solo Guitar vol 2 Incus, 1972
Ballads Tzadik, 2002
Standards Tzadik, 2007

Anthony Braxton
For Alto Delmark, 1968
Five Pieces 1975 Arista, 1975
Six Compositions : Quartet Antilles, 1981
Anthony Braxton's Charlie Parker Project 1993 HarART, 1995
Beyond Quantum Tzadik, 2008

John Cage
Music of Changes Peters, 1951
Cheap Imitation Peters, 1969
Roaratorio : An Irish Circus on Finnegans Wake Peters, 1980

Steve Coleman
Rhythm People BMG, 1990
Curves of Life BMG, 1995
The Sonic Language of Myth BMG, 1998
Lucidarium Label Bleu, 2003
The Mancy of Sound Pi Recordings, 2011

John Coltrane
My Favorite Things Atlantic Records, 1961
A Love Supreme Impulse !, 1964
Ascension Impulse !, 1966
Meditations Impulse !, 1966

Marilyn Crispell
Circles Les Disques Victo, 1991
Altered Spaces Leo Records, 1993
Cascades Music and Arts, 1995
Vignettes ECM, 2007
Azure ECM, 2013

Art Farmer
What Happens ? Campi, 1968
On the Road Contemporary, 1976
A Work of Art Concord, 1981

Fred Frith
Legend Virgin, 1973
Western Culture Broadcast, 1978
Gravity RecRec Music, 1980
Speechless Ralph Records, 1981
Pacifica Tzadik, 1998

Keith Jarrett
Solo Concerts : Bremen/Lausanne ECM, 1973
Standards Live ECM, 1985
Still Live ECM, 1986
Whisper Not ECM, 2000
Yesterdays ECM, 2009

Steve Lacy
Reflections New Jazz, 1958
Evidence New Jazz, 1961
Anthem Arista/Novus, 1990

Georges E. Lewis
Changing With The Times New World Records, 1993
Voyager Avant, 1993
Triangulation Nine Winds, 1996

Charles Mingus
Blues & Roots Atlantic, 1959
Oh Yeah Atlantic, 1962
The Complete Town Hall Concert Blue Note, 1962

Larry Ochs
The Secret Magritte Black Saint, 1995
The Works : Vol. I Black Saint, 1995
The Fields Black Saint, 1996

John Oswald
Plunderphonic Mystery Lab, 1989
Plexure Avant, 1993
Grayfolded Swell/Artifacts, 1994

Charlie Parker
Bird and Diz Verve, 1952
Jazz at Massey Hall Debut Records, 1956
Yardbird in Lotus Land Phoenix Jazz Records, 1977

Ralph Peterson Jr.
Volition Blue Note, 1989
Fo'tet Augmented Criss Cross, 2004
Triangular III Onyx, 2016

David Rosenboom
Brainwave Music ARC Records, 1975
Systems of Judgement Centaur, 1991
Two Lines Lovely Music, 1996

David Shea
Hsi-Yu-Chi Tzadik, 1995
Satyricon Sub Rosa, 1997
Rituals Room40, 2014

Cecil Taylor
Nefertiti, the Beautiful One Has Come Revenant, 1962
Unit Structures Blue Note, 1966

Edgard Varèse
Intégrales Ricordi, 1925
Ionisation Ricordi, 1929
Poème électronique Ricordi, 1958

Iannis Xenakis
Metastaseis Boosey & Hawkes, 1953
Duel Salabert, 1959
Pléiades Salabert, 1979

Frank Zappa
The Grand Wazoo Bizarre, 1972
Apostrophe (') DiscReet, 1974
Roxy & Elsewhere DiscReet, 1974
Joe's Garage Zappa Records, 1979
The Best Band You Never Heard in Your Life Barking Pumpkin, 1991

John Zorn
Naked City Nonesuch Records, 1990
Radio Avant, 1993
At the Mountains of Madness Tzadik, 2005
The Dreamers Tzadik, 2008
Nova Express Tzadik, 2011

Annexe B

Biographies des musiciens interviewés

Louis Bourhis

Louis Bourhis a débuté son parcours musical au *Conservatoire de Lisieux* où il y étudie la guitare classique. Il y découvre par la suite également les musiques actuelles et le jazz qu'il étudie avec passion. En parallèle, il débute la pratique du saxophone et intègre rapidement des bigbands et diverses formations de jazz locaux. C'est à l'âge de dix-sept ans qu'il se tourne vers la contrebasse qui deviendra son instrument de prédilection. Il l'étudie au Conservatoire de Rouen jusqu'à l'obtention simultanée de son *Diplôme d'Études Musicales de Jazz* et de son admission au sein de la *Haute École de Musique de Lausanne* en 2014 dans la classe de Bänz Oester. Au sein de cette institution, il a eu l'opportunité de travailler avec Lee Konitz, Gerald Clayton, Larry Grenadier et Eric Harland et de se produire à diverse reprise pour le Montreux Jazz Festival, notamment à Rio de Janeiro dans le cadre des cinquante ans de ce festival. Depuis, il a intégré le *Conservatorium Van Amsterdam* et effectue son Master dans la classe de Frans van der Hoeven.

Joël Gauvrit

Après un *Premier Prix* de piano du *Conservatoire National Supérieur de Musique de Lyon* dans la classe d'Eric Heidsieck, Joël Gauvrit s'oriente vers le jazz : il fonde un trio puis un quintette, avec lequel il joue ses propres compositions et remporte différentes récompenses, dont le *Grand prix du Tremplin international de Jazz d'Avignon*. Il retourne à ses premiers amours par le biais de la pratique d'instruments historiques : il se forme au clavecin avec Noëlle Spieth, au pianoforte et au clavicorde auprès de Patrick Cohen et suit les Masterclasses de Malcom Bilson, Alexei Lubimov, Bart Van Oort, Tom Beghin et Claire Chevalier. Joël Gauvrit crée en 2006 l'ensemble Il Mondo della Luna, dont l'ambition est d'explorer le répertoire de trio avec pianoforte de la seconde moitié du XVIIIe siècle. Il Mondo della Luna s'est produit à Paris et dans divers lieux en Ile de France, ainsi que dans les festivals Sinfonia en Périgord, Les Chants de la Dore et Les Journées musicales d'Automne de Souvigny, dont le concert a été retransmis sur la chaîne Mezzo et fera l'objet d'un DVD. Avec

l'Ensemble, il a également enregistré un programme autour des trois derniers trios de Mozart. Joël Gauvrit se produit en récital (Festival de Saint-Yrieix, Festival des Chants de la Dore...), en soliste (avec la chanteuse Gerlinde Sämann et l'Orchestre de Bordeaux-Aquitaine) et comme continuiste et pianiste d'orchestre avec La Philharmonie du Luxembourg et Arsys Bourgogne, ainsi que l'Orchestre de Chambre du Luxembourg, sous la direction de Pierre Cao ou Nicolas Brochot. Comme chambriste il joue régulièrement avec les ensembles FA7 (musique contemporaine), Les Folies du Temps et Galuppi (musique ancienne), et il participe avec l'ensemble Zellig à une création de Thierry Pécou. Joël Gauvrit a enregistré les Inventions et Sinfonias de J. S. Bach au clavicorde. Il enseigne le clavecin, le piano et le jazz au *Conservatoire de Lisieux*.

Pascal Mabit

Après des études aux *Conservatoires de Caen et Versailles* avec Thierry Lhiver et Sylvain Beuf, Pascal Mabit rentre au département Jazz et Musiques Improvisées du *Conservatoire National Supérieur de Musique et de Danse de Paris* en 2012, travaillant ainsi avec Riccardo del Fra, Hervé Sellin, François Théberge, Dré Pallemaerts, Pierre de Bethmann, Pierre Bertrand et Glenn Ferris, mais aussi, par le biais de master-classes et stages, avec Danilo Perez, Eddie Henderson, Itibere Zwarg, Larry Grenadier, Joachim Kühm, Pierre Jodlowsky, Jimmy Cobb, Marc Ducret, Billy Hart, Guillermo Klein et Fabrizio Cassol. Il y a obtenu le *DNSPM de Jazz et Musiques Improvisées* en 2014, le *Prix d'Improvisation Générative* en 2015 et a achevé son parcours en 2016, récompensé d'un Master. Il y a également décroché le *Diplôme d'État de Jazz et Musiques Improvisées*. En plus de cette formation prestigieuse, il a eu l'occasion de partager la scène avec des musiciens de Jazz tels que Emmanuel Bex, Jean-Benoît Culot, Laurent Bataille, Magic Malik et des orchestres tels que le Bibendum Big-Band et le Dedication Ensemble de Philippe Maniez dont il est membre régulier et actif depuis sa création, ou encore l'ONJ. Il est également présent dans le domaine des musiques transversales, notamment avec l'ensemble Contraste, pour un spectacle autour de « The Fairy Queen ». Très influencé par les musiques de Aka Moon et de Steve Coleman, il fonde en 2015 le trio Mobius Ring avec ses amis (et camarades du *CNSM*) Emmanuel Forster et Kevin Lucchetti, pour lequel il compose activement et dont un disque est sorti en février 2018. Enfin, également passionné par la pédagogie, il a l'occasion de transmettre ses savoirs au sein du département Jazz du *CRD Arthur Honegger du Havre* depuis septembre 2016 et lors du stage annuel des *Jazzitudes* du Pays d'Auge depuis 2013.

Bibliographie

Philip Agre (1997). Toward a critical technical practice : Lessons learned in trying to reform AI. In G. Bowker, L. Gasser, L. Star, & B. Turner (Eds.), *Social Science, Technical Systems and Cooperative Work : Beyond the Great Divide* (pp. 131–158). Erlbaum.

Ilge Akkaya, Daniel J. Fremont, Rafael Valle, Alexandre Donzé, Edward A. Lee, & Sanjit A. Seshia (2016). Control improvisation with probabilistic temporal specifications. In *Proceedings of the IEEE 1st International Conference on the Internet-of-Things Design and Implementation* (pp. 187–198).

Cyril Allauzen, Maxime Crochemore, & Mathieu Raffinot (1999). Factor oracle : a new structure for pattern matching. In *Proceedings of SOFSEM'99, Theory and Practice of Informatics* (pp. 291–306).

Charles Ames (1989). The Markov process as a compositional model : a survey and tutorial. *Leonardo*, 22(2), 175–187.

Diego Ardila, Cinjon Resnick, Adam Roberts, & Douglas Eck (2016). Audio DeepDream : optimizing raw audio with convolutional networks. In *Proceedings of the 17th International Society for Music Information Retrieval Conference.*

Christopher Ariza (2005). *An open design for computer-aided algorithmic music composition : athenaCL.* PhD thesis, New York University.

Simha Arom (1987). Les musiques traditionnelles d'Afrique Centrale : conception / perception. In *Actes du Symposium International Composition et Perception Musicales* (pp. 177–198).

Gérard Assayag (1998). Computer assisted composition today. In *Proceedings of the 1st Symposium on Music and Computers.*

Gérard Assayag (2016). Improvising in creative symbolic interaction. In J. B. L. Smith, E. Chew, & G. Assayag (Eds.), *Mathematical Conversations : Mathematics and Computation in Music Performance and Composition* (pp. 61–74). World Scientific ; Imperial College Press.

Gérard Assayag & Georges Bloch (2007). Navigating the oracle : a heuristic approach. In *Proceedings of the 33rd International Computer Music Conference* (pp. 405–412).

Gérard Assayag, Georges Bloch, & Marc Chemillier (2006a). OMax-ofon. In *Proceedings of the 3rd Sound and Music Computing Conference.*

Gérard Assayag, Georges Bloch, Marc Chemillier, Arshia Cont, & Shlomo Dubnov (2006b). OMax Brothers : a dynamic topology of agents for impro-

vization learning. In *Proceedings of the 1st ACM Workshop on Audio and Music Computing for Multimedia* (pp. 125–132).

Gérard Assayag & Shlomo Dubnov (2004). Using factor oracles for machine improvisation. *Soft Computing*, 8-9, 604–610.

Gérard Assayag, Shlomo Dubnov, & Olivier Delerue (1999). Guessing the composer's mind : applying universal prediction to musical style. In *Proceedings of the 25th International Computer Music Conference* (pp. 496–499).

Christel Baier & Joost-Pieter Katoen (2008). *Principles of model checking*. MIT Press.

Derek Bailey (1980). *Improvisation, its nature and practice in music*. Mootland Publishing.

Eric Battenberg & David Wessel (2012). Analyzing drum patterns using conditional deep belief networks. In *Proceedings of the 13th International Society for Music Information Retrieval Conference* (pp. 37–42).

Matthew I. Bellgard & Chi-Ping Tsang (1999). Harmonizing music the Boltzmann way. In N. Griffith & P. M. Todd (Eds.), *Musical Networks* (pp. 261–277). MIT Press.

Paul Berliner (1994). *Thinking in jazz : The infinite art of improvisation*. Chicago University Press.

Greg Bickerman, Sam Bosley, Peter Swire, & Robert M. Keller (2010). Learning to create jazz melodies using deep belief nets. In *Proceedings of the International Conference on Computational Creativity* (pp. 228–236).

Louis Bigo, Mathieu Giraud, Richard Groult, Nicolas Guiomard-Kagan, & Florence Levé (2017). Sketching sonata form structure in selected classical string quartets. In *Proceedings of the International Society for Music Information Retrieval Conference*.

Frédéric Bimbot, Emmanuel Deruty, Gabriel Sargent, & Emmanuel Vincent (2016). System & Contrast : a polymorphous model of the inner organization of structural segments within music pieces. *Music Perception*, 33(5), 631–661.

Frédéric Bimbot, Gabriel Sargent, Emmanuel Deruty, Corentin Guichaoua, & Emmanuel Vincent (2014). Semiotic description of music structure : an introduction to the Quaero/Metiss structural annotations. In *Proceedings of the AES 53rd International Conference on Semantic Audio* (pp. 32–43).

Laurent Bonasse-Gahot (2014). *An update on the SoMax project*. Technical report, IRCAM.

Dimitri Bouche, Jérôme Nika, Alex Chechile, & Jean Bresson (2017). Computer-aided composition of musical processes. *Journal of New Music Research*, 46(1), 3–14.

John Bowers (2002). Improvising machines ethnographically informed design for improvised electro-acoustic music. Master's thesis, University of East Anglia, Norwich, UK.

Peter F. Brown, Peter V. deSouza, Robert L. Mercer, Vincent J. Della Pietra, & Jenifer C. Lai (1992). Class-based n-gram models of natural language. *Computational linguistics*, 18(4), 467–479.

Benjamin Börschinger, Bevan K. Jones, & Mark Johnson (2011). Reducing grounded learning tasks to grammatical inference. In *Proceedings of the Empirical Methods in Natural Language Processing Conference* (pp. 1416–1425).

John Cage (1973). *Silence : lectures and writings*. Wesleyan University Press.

Jay Campbell (2014). Basic organization of pitch and time in Pierre Boulez's 'Messagesquisse'. In J. Zorn (Ed.), *Arcana VII : Musicians on Music* (pp. 25–36). Hips Road/Tzadik.

Philippe Canguilhem (2015). Improvisation as concept and musical practice in the 15th century. In A. M. Busse Berger & J. Rodin (Eds.), *The Cambridge History of Fifteenth-Century Music* (pp. 149–163). Cambridge University Press.

Clément Canonne & Nicolas Garnier (2011). A model for collective free improvisation. In *Proceedings of the 3rd International Conference on Mathematics and Computation in Music* (pp. 29–41).

Marc Chemillier (2001). Improviser des séquences d'accords de jazz avec des grammaires formelles. In *Actes des Journées d'Informatique Musicale* (pp. 121–126).

Marc Chemillier (2004). Toward a formal study of jazz chord sequences generated by Steedman's grammar. *Soft Computing*, 9(8), 617–622.

Axel Chemla-Romeu-Santos (2015). Guidages de l'improvisation. Master's thesis, Ircam, UPMC.

Stanley F. Chen (2009). Performance prediction for exponential language models. In *NAACL'09 Proceedings of Human Language Technologies* (pp. 450–458).

Stanley F. Chen & Joshua Goodman (1998). *An empirical study of smoothing techniques for language modeling*. Technical Report TR-10-98, Harvard University.

Noam Chomsky (1956). Three models for the description of language. *IEEE Transactions on Information Theory*, 3(2), 113–124.

Noam Chomsky (1965). *Aspects of the Theory of Syntax*. MIT Press.

Noam Chomsky (1972). *Studies on semantics in generative grammar*. Walter de Gruyter.

Noam Chomsky (1975). *Reflections on Language*. Pantheon Books.

Noam Chomsky & Morris Halle (1968). *The Sound Pattern of English*. Harper & Row.

Jerry Coker (1964). *Improvising jazz*. Fireside.

Darrell Conklin (2002). Representation and discovery of vertical patterns in music. In *Proceedings of the 2nd International Conference of Music and Artificial Intelligence* (pp. 32–42).

Darrell Conklin (2003). Music generation from statistical models. In *Proceedings of the AISB Symposium on Artificial Intelligence and Creativity in the Arts and Sciences* (pp. 30–35).

Darrell Conklin (2013). Multiple viewpoint systems for music classification. *Journal of New Music Research*, 1(42), 19–26.

Darrell Conklin & Christina Anagnostopoulou (2001). Representation and discovery of multiple viewpoint patterns. In *Proceedings of the International Computer Music Conference* (pp. 479–485).

Darrell Conklin & John G. Cleary (1988). Modelling and generating music using multiple viewpoints. In *Proceedings of the 1st workshop on AI and Music* (pp. 125–137).

Darrell Conklin & Ian H. Witten (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 1(24), 51–73.

Arshia Cont (2008). Antescofo : anticipatory synchronization and control of interactive parameters in computer music. In *Proceedings of the International Computer Music Conference* (pp. 33–40).

Jacques Coursil (2008). Hidden principles of improvisation. In J. Zorn (Ed.), *Arcana III : Musicians on Music* (pp. 58–65). Hips Road/Tzadik.

Marilyn Crispell (2000). Elements of improvisation. In J. Zorn (Ed.), *Arcana : Musicians on Music* (pp. 190–192). Hips Road/Tzadik.

Maxime Crochemore, Christophe Hancart, & Thierry Lecroq (2007a). *Algorithms on strings*. Cambridge University Press.

Maxime Crochemore, Lucian Ilie, & Emine Seid-Hilmi (2007b). The structure of factor oracle. *International Journal of Foundations of Computer Science*, 18(4), 781–797.

Maxime Crochemore & Wojciech Rytter (1994). *Text algorithms*. Oxford University Press.

Maxime Crochemore & Renaud Vérin (1997). On compact directed acyclic word graphs. In J. Mycielski, G. Rozenberg, & A. Salomaa (Eds.), *Structures in Logic and Computer Science*, volume 1261 of *LNCS* (pp. 192–211). Springer-Verlag.

Alain De Cheveigné & Hideki Kawahara (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4), 1917–1930.

W. Bas de Haas, Martin Rohrmeier, Remco C. Veltkamp, & Frans Wiering (2009). Modeling harmonic similarity using a generative grammar of tonal harmony. In *Proceedings of the 10th International Society for Music Information Retrieval Conference* (pp. 549–554).

Carl de Marcken (2015). Lexical heads, phrase structure and the induction of grammar. In *Proceedings of the 3rd Workshop on Very Large Corpora* (pp. 14–26).

Diego Di Carlo, Ken Déguernel, & Antoine Liutkus (2017). Gaussian framework for interference reduction in live recordings. In *Proceedings of the Audio Engineering Society Conference on Semantic Audio*.

Diego Di Carlo, Antoine Liutkus, & Ken Déguernel (2018). Interference reduction on full-length live recordings. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*.

Alexandre Donzé, Rafael Valle, Ilge Akkaya, Sophie Libkind, Sanjit A. Seshia, & David Wessel (2014). Machine improvisation with formal specifications. In

Proceedings of the 40th International Computer Music Conference (pp. 1277–1284).

Shlomo Dubnov, Gérard Assayag, & Arshia Cont (2007). Audio oracle : a new algorithm for fast learning of audio structures. In *Proceedings of the International Computer Music Conference* (pp. 224–227).

Shlomo Dubnov, Gérard Assayag, & Ran El-Yaniv (1998). Universal classification applied to musical sequences. In *Proceedings of the International Computer Music Conference* (pp. 332–340).

Shlomo Dubnov, Gérard Assayag, Olivier Lartillot, & Gill Bejerano (2003). Using machine-learning methods for musical style modeling. *IEEE Computer*, 10(38), 73–80.

Alan Durant (1989). Improvisation in the political music since 1960. In C. Norris (Ed.), *Music and the politics of culture* (pp. 252–282). New York : St. Martin's.

Ken Déguernel, Nathan Libermann, & Emmanuel Vincent (2017a). La musique comme une langue. In M. Artigue, J. Germoni, E. Ghys, & E. Godlewski (Eds.), *Mathématiques et langages* (pp. 18–19). Commission française pour l'enseignement des mathématiques.

Ken Déguernel, Jérôme Nika, Emmanuel Vincent, & Gérard Assayag (2017b). Generating equivalent chord progressions to enrich guided improvisation : application to rhythm changes. In *Proceedings of the the 14th Sound and Music Computing Conference* (pp. 399–406).

Ken Déguernel, Emmanuel Vincent, & Gérard Assayag (2016). Using multidimensional sequences for improvisation in the OMax paradigm. In *Proceedings of the 13th Sound and Music Computing Conference* (pp. 117–122).

Ken Déguernel, Emmanuel Vincent, & Gérard Assayag (2018). Probabilistic factor oracles for multidimensional machine improvisation. *Computer Music Journal*, 42(2).

Douglas Eck & Jasmin Lapalme (2008). *Learning musical structure directly from sequences of music.* Technical report, University of Montreal, Department of Computer Science.

Tuomas Eerola (2004). *The dynamics of musical expectancy : cross-cultural and statistical approaches to melodic expectations.* PhD thesis, University of Jyväskylä.

Aaron Einbond, Diemo Schwarz, Riccardo Borghesi, & Norbert Schnell (2016). Introducing CatOracle : corpus-based concatenative improvisation with the audio oracle algorithm. In *Proceedings of the 42nd International Computer Music Conference* (pp. 140–147).

Miguel Ferrand, Peter Nelson, & Geraint Wiggins (2002). A probabilistic model for melody segmentation. In *Proceedings of the 2nd International Conference on Music and Artificial Intelligence.*

Giuseppe Fiorentino (2013). *"Folia". El origen de los esquemas armónicos entre tradición oral y transmisión escrita.* Reichenberger.

Michel Foucault (1966). *Les mots et les choses.* Gallimard.

Alexandre R. J. François, Isaac Schankler, & Elaine Chew (2010). Mimi4x : an interactive audio-visual installation for high-level structural improvisation. *International Journal of Arts and Technology*, 6(2), 138–151.

William A. Gale & Kenneth W. Church (1994). What's wrong with adding one ? In N. Oostdijk & P. de Hann (Eds.), *Corpus-Based Research into Language* (pp. 189–200). Rodolpi.

Matthias Gallé (2011). *Searching for compact hierarchical structures in DNA by means of the smallest grammar problem*. PhD thesis, Université Rennes 1.

Harold Garfinkel (1967). *Studies in ethnomethodology*. Prentice Hall.

Fiammetta Ghedini, François Pachet, & Pierre Roy (2016). Creating music and texts with flow machines. In *Multidisciplinary Contributions to the Science of Creative Thinking* (pp. 325–343). Springer.

Toby Gifford & Andrew R. Brown (2011). Beyond reflexivity : mediating between imitative and intelligent action in an interactive music system. In *Proceedings of the 25th BCS Conference on Human-Computer Interaction*.

Mathieu Giraud, Ken Déguernel, & Emilios Cambouroupoulos (2014). Fragmentations with pitch, rhythm and parallelism constraints for variation matching. In *Proceedings of the 10th International Symposium of Computer Music Multidisciplinary Research* (pp. 298–312).

Mathieu Giraud, Richard Groult, Emmanuel Leguy, & Florence Levé (2015). Computational fugue analysis. *Computer Music Journal*, 39(2), 77–96.

Irving J. Good (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, 40(3–4), 237–264.

Corentin Guichaoua (2017). *Modèles de compression et critères de complexité pour la description et l'inférence de structure musicale*. PhD thesis, Université Rennes 1.

Mark A. Hall & Lloyd Smith (1996). A computer model of blues music and its evaluation. *Journal of the Acoustical Society of America*, 100(2), 1163–1167.

Jean-Paul Haton, Christophe Cerisara, Dominique Fohr, Yves Laprie, & Kamel Smaïli (2006). *Reconnaissance automatique de la parole*. Dunod.

Lejaren A. Hiller & Leonard M. Isaacson (1959). *Experimental music : composition with an electronic computer*. McGraw-Hill.

John E. Hopcroft & Jeffrey D. Ullman (1979). *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley.

Ray Jackendoff & Fred Lerdahl (2006). The capacity for music : what is it, and what's special about it ? *Cognition*, 100(1), 33–72.

Barnabé Janin (2014). *Chanter sur le livre, manuel pratique d'improvisation polyphonique de la Renaissance*. Symétrie.

Berit Janssen, W. Bas de Haas, Anja Volk, & Peter van Kranenburg (2013). Finding repeated patterns in music : state of knowledge, challenges, perspectives. In *Proceedings of the International Symposium on Computer Music Modeling and Retrieval* (pp. 277–297).

Tony Jebara (2004). *Machine learning : discriminative and generative*. Springer.

Harrold Jeffreys (1948). *Theory of probability. 2nd edition.* Clarendon Press, Oxford.

Frederick Jelinek & Robert L. Mercer (1980). Interpolated estimation of Markov source parameters from sparse data. In *Pattern Recognition in Practice* (pp. 381–397).

Steven Johnson (2012). *The New York schools of music and the visual arts.* Routledge.

Stefano Kalonaris (2016). Markov networks for free improvisers. In *Proceedings of the 42nd International Computer Music Conference* (pp. 181–185).

Slava M. Katz (1987). Estimation of probabilities from sparse data for the language model component of a speech recognizer. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(3), 400–401.

Dietrich Klakow (1998). Log-linear interpolation of language models. In *Proceedings of the 5th International Conference on Spoken Language Processing* (pp. 1695–1698).

Dan Klein & Christopher D. Manning (2004). Corpus-based induction of syntactic structure. In *Proceedings of the Association for Computational Linguistics* (pp. 479–486).

Stefan Koelsch, Martin Rohrmeier, Renzo Torrecuso, & Sebastian Jentschke (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences*, 110(38), 15443–15448.

Daphne Koller & Nir Friedman (2009). *Probabilistic graphical models : principles and techniques.* MIT Press.

Alfred Korzybski (1994). *Science and sanity. An introduction to non-aristotelian systems and general semantics. Fifth edition.* Institute of General Semantics.

Kenichi Kurihara & Taisuke Sato (2004). An application of the variational Bayesian approach to probabilistic context-free grammar. In *Proceedings of the International Joint Conference on Natural Language Processing.*

Tom Kwiatkowski, Sharon Goldwater, Luke Zettlemoyer, & Mark Steedman (2012). A probabilistic model of syntactic and semantic acquisition from child-directed utterances and their meanings. In *Proceedings of the European Chapter of the Association for Computational Linguistics* (pp. 234–244).

Arnaud Lefebvre & Thierry Lecroq (2000). Computing repeated factors with a factor oracle. In *Proceedings of the 11th Australasian Workshop On Combinatorial Algorithms* (pp. 145–158).

Fred Lerdahl & Ray Jackendoff (1983). *A Generative Theory of Tonal Music.* MIT Press.

Mark Levine (1995). *The jazz theory book.* Sher Music.

George E. Lewis (2000a). Teaching improvised music : An ethnographic memoir. In J. Zorn (Ed.), *Arcana : Musicians on Music* (pp. 78–109). Hips Road/Tzadik.

George E. Lewis (2000b). Too many notes : computers, complexity and culture in Voyager. *Leonardo Music Journal*, 10, 33–39.

Wenchao Li, Alessandro Forin, & Sanjit A. Seshia (2010). Scalable specification mining for verification and diagnosis. In *Proceedings of the 47th Design Automation Conference* (pp. 755–760).

Benjamin Lévy, Georges Bloch, & Gérard Assayag (2012). OMaxist dialectics : capturing, visualizing and expanding improvisations. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 137–140).

Fivos Maniatakos, Gérard Assayag, Fréderic Bevilacqua, & Carlos Agon (2010). On architecture and formalisms for computer-assisted improvisation. In *Proceedings of the Sound and Music Computing Conference*.

Tomáš Mikolov, Anoop Deoras, Stefan Kombrink, Lukáš Burget, & Jan 'Honza' Černocký (2011). Empirical evaluation and combination of advanced language modeling techniques. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association* (pp. 605–608).

Marcel Mongeau & David Sankoff (1990). Comparison of musical sequences. *Computer and the Humanities*, 24, 161–175.

Ingrid Monson (1996). *Saying something : jazz, improvisation and interaction*. University of Chicago Press.

Julian Moreira, Pierre Roy, & François Pachet (2013). Virtualband : interacting with stylistically consistent agents. In *Proceedings of the International Society for Music Information Retrieval* (pp. 341–346).

Michael C. Mozer (1994). Neural network music composition by prediction : exploring the benefits of psychoacoustic constraints and multi-scale processing. *Connection Science*, 6(2–3), 247–280.

Hermann Ney, Ute Essen, & Reinhard Kneser (1994). On structuring probabilistic dependences in stochastic language modeling. *Computer, Speech and Languages*, 8, 1–38.

Jérôme Nika (2016). *Guiding human-computer music improvisation : introducing authoring and control with temporal scenarios*. PhD thesis, UPMC - Université Paris 6 Pierre et Marie Curie.

Jérôme Nika, Dimitri Bouche, Jean Bresson, Marc Chemillier, & Gérard Assayag (2015). Guided improvisation as dynamic calls to an offline model. In *Proceedings of the 12th Sound and Music Computing Conference*.

Jérôme Nika & Marc Chemillier (2014). Improvisation musicale homme-machine guidée par un scénario temporel. *Technique et Science Informatiques*, 7–8(33), 651–684.

Jérôme Nika, Marc Chemillier, & Gérard Assayag (2017a). ImproteK : introducing scenarios into human-computer music improvisation. *ACM Computers in Entertainment*, 4(2), 4 :1–27.

Jérôme Nika, Ken Déguernel, Axel Chemla-Romeu-Santos, Emmanuel Vincent, & Gérard Assayag (2017b). DYCI2 agents : merging the "free", "reactive" and "scenario-based" music generation paradigms. In *Proceedings of the 43rd International Computer Music Conference*.

Larry Ochs (2000). Devices and strategies for structured improvisation. In J. Zorn (Ed.), *Arcana : Musicians on Music* (pp. 325–335). Hips Road/Tzadik.

François Pachet (2002). The Continuator : musical interaction with style. In *Proceedings of the International Computer Music Conference* (pp. 211–218).

François Pachet & Pierre Roy (2011). Markov constraints : steerable generation of Markov sequences. *Constraints*, 16(2), 148–172.

François Pachet, Pierre Roy, Julian Moreira, & Mark d'Inverno (2013). Reflexive loopers for solo musical improvisation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2205–2208).

Víctor Padilla & Darrell Conklin (2016). Statistical generation of two-voice florid counterpoint. In *Proceedings of the 13th Sound and Music Computing Conference* (pp. 380–387).

Alexandre Papadopoulos, Pierre Roy, & François Pachet (2014). Avoiding plagiarism in Markov sequence generation. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence* (pp. 2731–2737).

Alexandre Papadopoulos, Pierre Roy, & François Pachet (2016). Assisted lead sheet composition using FlowComposer. In *Proceedings of the 22nd International Conference on Principles and Practice of Constraint Programming* (pp. 769–785).

Charlie Parker & Jamey Aebersold (1978). *Charlie Parker Omnibook*. Alfred Music Publishing.

John K. Pate & Mark Johnson (2016). Grammar induction from (lots of) words alone. In *Proceedings of the 26th International Conference on Computational Linguistics* (pp. 23–32).

Marcus T. Pearce (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition*. PhD thesis, City University, London.

Marcus T. Pearce & Geraint A. Wiggins (2012). Auditory expectation : the information dynamics of music perception and cognition. *Topics in Cognitive Science*, 4(4), 625–652.

Fuchun Peng, Dale Schuurmans, & Shaojun Wang (2004). Augmenting naive Bayes classifiers with statistical language models. *Information Retrieval*, 7(3), 317–345.

Jeremy Pickens (2000). A comparison of language modeling and probabilistic text information retrieval approaches to monophonic music retrieval. In *Proceeding of the 1st International Symposium on Music Information Retrieval Conference*.

Jeremy Pickens, Juan Pablo Bello, Giuliano Monti, Tim Crawford, Matthew Dovey, Mark Sandler, & Don Byrd (2002). Polyphonic score retrieval using polyphonic audio queries : a harmonic modeling approach. *Journal of New Music Research*, 32(2), 223–236.

Richard C. Pinkerton (1956). Information theory and melody. *Scientific American*, 194(2), 77–86.

Dan Ponsford, Geraint Wiggins, & Chris Mellish (1999). Statistical learning of harmonic movement. *Artificial Intelligence*, 46(1), 77–105.

Carlos Pérez-Sancho, David Rizo, & José M. Iñesta (2008). Stochastic text models for music categorization. In *Proceedings of the Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition and Structural and Syntactic Pattern Recognition* (pp. 55–64).

Stanisław A. Raczyński, Satoru Fukayama, & Emmanuel Vincent (2013a). Melody harmonisation with interpolated probabilistic models. *Journal of New Music Research*, 42(3), 223–235.

Stanisław A. Raczyński & Emmanuel Vincent (2014). Genre-based music language modelling with latent hierarchical Pitman-Yor process allocation. *IEEE Transactions on Audio, Speech and Language Processing*, 22(3), 672–681.

Stanisław A. Raczyński, Emmanuel Vincent, & Shigeki Sagayama (2013b). Dynamic Bayesian networks for symbolic polyphonic pitch modeling. *IEEE Transactions on Audio, Speech and Language Processing*, 21(9), 1830–1840.

Brian Ravenet, Angelo Cafaro, Beatrice Biancardi, Magalie Ochs, & Catherine Pelachaud (2015). Conversational behavior reflecting interpersonal attitudes in small group interactions. In *Proceedings of the 15th International Conference on Intelligent Virtual Agents* (pp. 375–388).

Martin Rohrmeier (2007). A generative grammar approach to diatonic harmonic structure. In *Proceedings of the 4th Sound and Music Computing Conference* (pp. 97–100).

David Rosenboom (1996). Improvisation and composition - synthesis and integration into the music curriculum. In *Proceedings of the 71st Annual Meeting, National Association of Schools of Music.* (pp. 19–31).

David Rosenboom (2000). Propositional music : on emergent properties in morphogenesis and the evolution of music. In J. Zorn (Ed.), *Arcana : Musicians on Music* (pp. 203–232). Hips Road/Tzadik.

Robert Rowe (1992). *Interactive Music Systems : Machine Listening and Composing.* MIT Press.

Pierre Roy & François Pachet (2013). Enforcing meter in finite-length Markov sequences. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence* (pp. 854–861).

Nicholas Ruozi (2012). The Bethe partition of log-supermodular graphical models. In *Advances in Neural Information Processing Systems.*

Giorgio Sanguinetti (2012). *The art of partimento. History, theory and practice.* Oxford University Press.

Kevin Sanlaville, Gérard Assayag, Frédéric Bevilacqua, & Catherine Pelachaud (2015). Emergence of synchrony in an adaptive interaction model. In *Intelligent Virtual Agents Doctoral Consortium.*

Kevin Sanlaville, Gérard Assayag, Frédéric Bevilacqua, & Catherine Pelachaud (2016). Modèles probabilistes pour l'interaction entre agents. In *Actes du Workshop Affect, Compagnon Artificiel, Interaction.*

Isaac Schankler, Jordan B.L. Smith, Alexandre R.J. François, & Elaine Chew (2011). Emergent formal structures of factor oracle-driven musical improvisations. *Mathematics and Computation in Music*, 6726, 241–254.

Erik M. Schmidt & Youngmoo E. Kim (2013). Learning rhythm and melody features with deep belief networks. In *Proceedings of the International Society for Music Information Retrieval* (pp. 21–26).

John Schott (2000). "We are revealing a hand that will later reveal us" – notes on form and harmony in Coltrane's work. In J. Zorn (Ed.), *Arcana : Musicians on Music* (pp. 345–366). Hips Road/Tzadik.

Diemo Schwarz (2007). Corpus-based concatenative synthesis. *IEEE Signal Processing Magazine*, 24(2), 92–104.

Claude Shannon (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.

David Shea (2000). One/two. In J. Zorn (Ed.), *Arcana : Musicians on Music* (pp. 145–152). Hips Road/Tzadik.

Jacques Siron (2015). *La partition intérieure – jazz, musiques improvisées. 9ème édition revue et corrigée.* Outre mesure.

Jordan B. L. Smith (2014). *Explaining listener differences in the perception of musical structure.* PhD thesis, Queen Mary University of London.

Jordan B. L. Smith, Isaac Schankler, & Elaine Chew (2013). Why do listeners disagree about large-scale formal structure ? A case study. In *Proceedings of the Society for Music Perception and Cognition Conference.*

Bob Snyder (2000). *Music and Memory.* MIT Press.

Fei Song & W. Bruce Croft (1999). A general language model for information retrieval. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 279–280).

Mark Steedman (1984). A generative grammar for jazz chord sequences. *Music Perception*, 2(1), 52–77.

Mark Steedman (1996). The blues and the abstract truth : music and mentals models. In J. Oakhill & A. Garnham (Eds.), *Mental Models in Cognitive Science* (pp. 305–318). Erlbaum.

Andreas Stolcke, Jing Zheng, Wen Wang, & Victor Abrash (2011). SRILM at sixteen : update and outlook. In *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop.*

David Sudnow (1978). *Ways of the hand : The Organization of Improvised Conduct.* MIT Press.

Greg Surges & Shlomo Dubnov (2013). Feature selection and composition using PyOracle. In *Proceedings of the 2nd International Workshop on Musical Metacreation* (pp. 114–121).

David Temperley (2008). A probabilistic model of melody perception. *Cognitive Science*, 32(2), 418–444.

Chi-Ping Tsang & Matthew I. Bellgard (1990). Sequence generation using a network of Boltzmann machines. In *Proceedings of the 4th Australian Joint Conference on Artificial Intelligence* (pp. 224–233).

Rafael Valle, Alexandre Donzé, Daniel J. Fremont, Ilge Akkaya, Sanjit A. Seshia, Adrian Freed, & David Wessel (2016). Specification mining for machine improvisation with formal specifications. *Computers in Entertainement – Special Issue on Musical Metacreation, Part II*, 14(3), 6 :1–20.

Aäron Van Den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, & Koray Kavukcuoglu (2016). Wavenet : a generative model for raw audio. *arXiv preprint*, arXiv :1609.03499.

Paul Vanhaesebrouck, Aurélien Bellet, & Marc Tommasi (2017). Decentralized collaborative learning of personalized models over networks. In *Proceedings of the International Conference on Artificial Intelligence and Statistics* (pp. 509–517).

Cheng-i Wang & Shlomo Dubnov (2014). Guided music synthesis with variable markov oracle. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (pp. 55–62).

Raymond P. Whorley, Geraint A. Wiggins, Christophe Rhodes, & Marcus T. Pearce (2013). Multiple viewpoint systems : time complexity and the construction of domains for complex musical viewpoints in the harmonisation problem. *Journal of New Music Research*, 42(3), 237–266.

Ian H. Witten & Timothy C. Bell (1991). The zero-frequency problem : estimating the probabilities of novel events in adaptive text compression. *IEEE Transactions of Information Theory*, 37(4), 1085–1094.

Li-Chia Yang, Szu-Yu Chou, & Yi-Hsuan Yang (2017). MidiNet : a convolutional generative adversarial network for symbolic-domain music generation. In *Proceedings of the 18th International Society for Music Information Retrieval Conference* (pp. 324–331).

Chengxiang Zhai & John Lafferty (2004). A study of smoothing methods for language models applied to information retrieval. *ACM Transactions on Information Systems*, 22(2), 179–214.

Imed Zitouni, Kamel Smaïli, & Jean-Paul Haton (2000). Beyond the conventional statistical language models : the variable-length sequences approach. In *Interspeech* (pp. 562–565).

Jacob Ziv & Abraham Lempel (1978). Compression of individual sequences via variable-rate coding. *IEEE Transactions of Information Theory*, 24(5), 530–536.

Résumé

Les systèmes actuels d'improvisation musicales sont capables de générer des séquences musicales unidimensionnelles par recombinaison du matériel musical. Cependant, la prise en compte de plusieurs dimensions (mélodie, harmonie...) et la modélisation de plusieurs niveaux temporels sont des problèmes difficiles. Dans cette thèse, nous proposons de combiner des approches probabilistes et des méthodes issues de la théorie des langages formels afin de mieux apprécier la complexité du discours musical à la fois d'un point de vue multidimensionnel et multi-niveaux dans le cadre de l'improvisation où la quantité de données est limitée.
Dans un premier temps, nous présentons un système capable de suivre la logique contextuelle d'une improvisation représentée par un oracle des facteurs tout en enrichissant son discours musical à l'aide de connaissances multidimensionnelles représentées par des modèles probabilistes interpolés. Ensuite, ces travaux sont étendus pour modéliser l'interaction entre plusieurs musiciens ou entre plusieurs dimensions par un algorithme de propagation de croyance afin de générer des improvisations multidimensionnelles. Enfin, nous proposons un système capable d'improviser sur un scénario temporel avec des informations multi-niveaux représenté par une grammaire hiérarchique. Nous proposons également une méthode d'apprentissage pour l'analyse automatique de structures temporelles hiérarchiques.
Tous les systèmes sont évalués par des musiciens et improvisateurs experts lors de sessions d'écoute.

Abstract

Current musical improvisation systems are able to generate unidimensional musical sequences by recombining their musical contents. However, considering several dimensions (melody, harmony...) and several temporal levels are difficult issues. In this thesis, we propose to combine probabilistic approaches with formal language theory in order to better assess the complexity of a musical discourse, both from a multidimensional and multi-level point of view in the context of improvisation where the amount of data is limited.
First, we present a system able to follow the contextual logic of an improvisation modelled by a factor oracle whilst enriching its musical discourse with multidimensional knowledge represented by interpolated probabilistic models. Then, this work is extended to create another system using a belief propagation algorithm representing the interaction between several musicians, or between several dimensions, in order to generate multidimensional improvisations. Finally, we propose a system able to improvise on a temporal scenario with multi-level information modelled with a hierarchical grammar. We also propose a learning method for the automatic analysis of hierarchical temporal structures.
Every system is evaluated by professional musicians and improvisers during listening sessions.