Dynamiques créatives de l'interaction improvisée

Table

1. RESUME DU PROJET

Le projet *Dynamiques créatives de l'interaction improvisée* porte sur la constitution, l'adaptation et la mise en œuvre effective de modèles performants d'écoute artificielle, d'apprentissage, d'interaction et de création automatique de contenus musicaux pour permettre la constitution d'agents musicaux numériques, autonomes, créatifs, capables de s'intégrer de façon interactive et artistiquement crédible dans des dispositifs humains variés tels que la scène vivante, la production, la pédagogie, l'écoute active ou de contribuer à terme aux compétences perceptives et communicatives de systèmes d'intelligence artificielle embarquée.

Le projet met en avant l'interaction improvisée, à la fois comme modèle cognitivement inspiré de l'action et de la décision individuelles et collectives, comme schéma de découverte et d'apprentissage non supervisé, et comme outil discursif pour l'échange humain - agent numérique, dans une perspective de modélisation du style et de l'interaction.

L'interaction improvisée entre des humains et des agents numériques est un domaine récent issu des études sur la créativité artificielle, qui repose sur l'observation que l'immense majorité des interactions humaines sont improvisées, et convoque plusieurs problématiques de recherche très actives : l'apprentissage interactif, dont les modèles se construisent dans le temps même de l'interaction et dont les résultats infléchissent les conditions de cette interaction; la perception artificielle, base de cette interaction; la modélisation de l'interaction sociale et expressive entre agents humains et/ou numériques, dans ses dimensions à la fois anthropologique, sociale, linguistique, et informatique. Ce type d'interaction met en jeu la boucle perception / action et engage l'apprentissage dans une conception renouvelée où l'agent apprend notamment par les réactions d'autres agents à ses propres productions. Le modèle d'apprentissage intégrant la démarche d'improvisation est donc aussi bien génératif et réflexif.

Intégrer écoute artificielle, apprentissage de comportements musicaux, modélisation temporelle des structures musicales et dynamiques d'interaction créative dans une architecture permettant l'expérimentation effective en temps réel constitue un défi ambitieux de la société de l'information et de la communication, riche de nombreuses applications potentielles susceptibles de changer la donne dans la relation entre humains et agents artificiels créatifs dans le contexte des industries culturelles, notamment pour la production et post-production audiovisuelles et musicales, le jeu vidéo, le spectacle vivant, les formats innovants renouvelant la diffusion et l'écoute active de la musique, les nouvelles formes narratives basées sur l'interaction.

Le projet articule entre elles trois grandes problématiques de recherche autour d'un environnement logiciel expérimental en tirant au mieux parti de l'expertise des partenaires et de leurs interactions. Ces trois thèmes, pour chacun desquels au moins deux des partenaires collaborent et co-encadrent des travaux correspondent aux principales compétences exercées de manière parallèle, compétitive et contributive, par un agent numérique « créatif » en situation d'interaction improvisée avec des humains et d'autres agents : l'écoute informée visant à analyser la scène sonore pour extrapoler la structure musicale en exploitant les similarités observées et les connaissances a priori disponibles ; l'apprentissage adaptatif de structures musicales visant à intégrer modélisation de séquences formelles et approches probabilistes pour rendre mieux compte de la complexité du discours musical à partir de données nécessairement limitées ; et les dynamiques d'interaction improvisée permettant d'envisager les architectures multi-agents et les modèles de connaissance et de décision permettant de mettre concrètement en jeu les scénarios de co-improvisation impliquant acteurs humains et numériques.

2. TABLEAU DES PERSONNES IMPLIQUEES DANS LE PROJET

Partenaire	Nom	Prénom	Emploi actuel	H. mois	Rôle et responsabilité dans le projet
Ircam	Assayag[1]	Gérard	DR	11	Coordinateur, apprentissage et modélisation des structures et des processus musicaux
Ircam	Bevilacqua[1]	Frédéric	DR	2	Scénarios d'interaction et IHM adaptatifs et multimodaux
Ircam	Bloch	Georges	MCF, CR associé	16	Expert musical et multimedia, cas d'usage, integration symbolique / audio-video
Ircam	Bresson[1]	Jean	CR1	1	Langages et environnements d'informatique musicale
Ircam	Nika	Jérôme	Doctorant	12	Improvisation guidée par un scénario
Ircam	Sanlaville	Kevin	Doctorant	18	Adaptation temporelle et dynamiques collectives de l'interaction
UBO	Marchand	Sylvain	PR	12	Écoute informée créative, séparation informée, perception artificielle
UBO	Koehl	Vincent	MCF	3	Écoute / perception artificielle
UBO	Paquier	Mathieu	MCF	3	Écoute / perception artificielle
Inria	Vincent	Emmanuel	CR1	9	Apprentissage des structures musicales
Inria	Liutkus	Antoine	CR2	9	Nouveaux contenus par séparation informée

3. CONTEXTE, POSITIONNEMENT ET OBJECTIF DE LA PROPOSITION DETAILLEE

3.1. OBJECTIF DU PROJET

Le projet *Dynamiques créatives de l'interaction improvisée* porte sur la constitution, l'adaptation et la mise en œuvre effective de modèles performants d'écoute artificielle, d'apprentissage, d'interaction et de création automatique de contenus musicaux pour permettre la constitution d'agents musicaux numériques, autonomes, créatifs, capables de s'intégrer de façon interactive et artistiquement crédible dans des dispositifs humains variés tels que la scène vivante, la (post-)production, la pédagogie, le jeu et les nouvelles formes narratives, ou de contribuer à terme aux compétences perceptives et communicatives de systèmes d'intelligence artificielle embarquée.

L'objectif est de constituer des agents créatifs autonomes par apprentissage direct résultant d'une exposition au jeu vivant de musiciens humains improvisants, en créant une boucle de rétro-action stylistique par l'exposition simultanée de l'humain aux productions improvisées des agents numériques eux-mêmes, donc à partir d'une situation de communication humain-agent numérique évoluant dans le temps et portée par une dynamique complexe. Un apprentissage hors-ligne sur des corpus musicaux sera

[1] *Ces participants Ircam cumulant 14 h.m. sont des des permanents employés de l'Ircam, les autres Ircam sont enseignant-chercheur universitaire associé à l'Ircam (G. Bloch) ou doctorants à l'Ircam sous contrat doctoral de l'université UPMC (J. Nika et K. Sanlaville), dont les H.M. n'apparaissent donc pas dans le document administratif et financier au titre de l'Ircam.*

aussi anticipé pour "colorer" stylistiquement l'individualité numérique des agents ou situer l'expérience dans un genre (jazz, classique, pop etc.). Outre l'application à la musique vivante dans l'interaction « live » entre agents numériques et humains, la situation de jeu pourra être étendue à des applications innovantes comme l'interaction d'utilisateurs de compétences variées avec des archives patrimoniales audio-visuelles ressuscitées dynamiquement dans des scénarios créatifs ou pédagogiques de co-improvisation, ainsi que dans le cadre général des nouvelles formes narratives génératives et interactives, des systèmes immersifs, et généralement des formes nouvelles d'interaction promues par les industries culturelles.

Le but est aussi bien de constituer une expertise artificielle de la pratique musicale (« machine musicanship ») par cette interaction que de susciter une riche expérience instantanée de communication humain-numérique, susceptible de renvoyer une satisfaction esthétique à l'utilisateur, d'enrichir sa production sonore et musicale, de "dialoguer" avec lui par imitation ou contradiction, et, en général, de stimuler et dynamiser l'expérience de jeu individuel et collectif. Cette interaction humain-agent numérique sera étendue à une interaction agent numérique-agent numérique dans des configurations riches (plusieurs humains, plusieurs agents numériques en interaction). Au fur et à mesure de l'expérience se formeront une ou des individualités musicales numériques autonomes, capables d'intervenir de manière crédible dans des situations d'interaction complexes avec des humains et d'autres agents.

Une entité créative artificielle dans un contexte audio-musical subsumera elle-même une collection d'agents élémentaires concurrents, contributifs et compétitifs, capables d'apprentissage interactif, qui prendront en charge des tâches d'écoute artificielle, de découverte de structures temporelles à court et long terme, de modélisation du style, de génération de séquences symboliques, de rendu audio temps-réel, mais aussi de visualisation et d'interface homme-machine (IHM).

Le projet produira deux nouvelles thèses (écoute informée, apprentissage de structures musicales), et bénéficiera des apports de deux thèses en cours (improvisation guidée et adaptation temporelle de l'interaction). Il fournira une suite logicielle et des données expérimentales mises à la disposition de la communauté, tout en se confrontant à plusieurs verrous, notamment : la constitution d'un modèle de connaissances et de décision pour les stratégies d'interaction synchrones et asynchrones d'agents pris dans une interaction collective et créative ; la constitution d'une écoute active et informée par l'exploitation des informations disponibles et la co-adaptation perception/action dans un cadre d'apprentissage interactif ; l'intégration des aspects multi-dimensions et multi-échelle dans l'apprentissage des structures musicales, et l'identification automatique des « dimensions sémantiques » portant le message musical émis par chaque agent.

3.2. CONTEXTE ET ENJEUX STRATEGIQUES

L'interaction *improvisée* entre des humains et des agents numériques est un domaine récent issu des études sur la créativité artificielle, qui convoque plusieurs problématiques de recherche très actives : l'apprentissage interactif, dont les modèles se construisent dans le temps même de l'interaction et dont les résultats infléchissent les conditions de cette interaction; la perception artificielle, base de cette interaction; la modélisation de l'interaction sociale et expressive entre agents humains et/ou numériques, dans ses dimensions à la fois anthropologique, sociale, linguistique, et informatique. Ce type d'interaction met en jeu la boucle perception / action et engage l'apprentissage dans une conception renouvelée où l'agent apprend notamment par les réactions d'autres agents à ses propres productions. La recherche sur la créativité est supportée au niveau européen par l'Objectif *ICT-2013.8.1 Technologies and scientific foundations in the field of creativity* dans ses actions *Intelligent computational environments stimulating and enhancing human creativity* et *Progress towards formal understanding of creativity.* Affronter la question de l'interaction improvisée prise comme moteur de créativité et présente au cœur de toutes les activités humaines constitue bien un défi central, répondant parfaitement à l'axe *Interactions des mondes physiques, de l'humain et du monde numérique* du défi *Société de l'information et de la communication* dans l'appel initial, pour lequel nous souhaitons intégrer écoute artificielle, apprentissage de comportements musicaux, modélisation temporelle des structures musicales et dynamiques d'interaction créative dans une architecture permettant l'expérimentation effective en temps réel.

Les applications envisagées se déclinent en des directions variées selon que les différents composants sont utilisés de manière interactive ou hors-ligne et sont susceptibles de changer la donne dans la relation artistique entre humains et agents artificiels créatifs. L'écoute et l'apprentissage hors-ligne de grandes

bases de données musicales peuvent alimenter des systèmes génératifs de composition et de performance et trouvent des débouchés dans l'ajout de fonctionnalités créatives dans les logiciels de production et post-production audio et multimedia, les jeux numériques, l'accès renouvelé aux patrimoines audiovisuels. L'écoute et l'apprentissage intégrés dans l'interaction permettent de programmer des agents capables de réagir en temps réel au jeu de musiciens humains dans les applications d'improvisation homme-machine, dans les installations interactives multimédia, dans la composition d'œuvres hybrides pour instruments acoustiques et ordinateur et dans la mise au point de formats de "musique variable" pour la production et la distribution musicales. L'architecture d'agents concurrents avec contraintes de scénario trouve des débouchés dans la générativité de musique contextuelle pour les jeux numériques, les applications innovantes du web, le cinéma génératif, en paramétrant la génération interactive de nouvelles séquences temporelles par déroulement du jeu, l'incarnation et/ou l'évolution d'un personnage, le comportement de l'utilisateur.

Enfin, la décomposition informée couplée à la recréation de séquences musicales inédites et crédibles par des agents créatifs bouleverse la notion même d'accès public à la musique enregistrée, entraînant des conséquences potentiellement majeures sur l'industrie culturelle et la valorisation des archives du patrimoine, avec un public qui passera de l'état passif à une véritable participation. De ce fait, le projet est aussi en relation avec l'axe *Le numérique au service des arts, du patrimoine, des industries culturelles et éditoriales*, ainsi que *Création, cultures et patrimoines* dans le défi *Sociétés innovantes, intégrantes et adaptatives* de l'appel initial.

3.3. ORIGINALITE ET DIFFICULTES DU PROJET

Une des originalités du projet est de mettre en avant l'interaction improvisée, à la fois comme modèle cognitivement inspiré de l'action et de la décision individuelle et collective, comme schéma de découverte et d'apprentissage non supervisé, et comme outil discursif pour l'échange humain-agent numérique. Un autre aspect met en avant la modélisation du style comme représentation caractéristique de l'identité des agents résultant de cette interaction improvisée.

Par rapport aux produits et prototypes existant dans ce domaine (*OMax* de l'Ircam, *Continuator* de Sony-CSL, Logiciel *Band-in-a-Box*, *GenJam* de A. Biles, *SongSmith* de MicroSoft, *EMI* de D. Cope, *HyperScore* de Farbood, etc.) la voie explorée par ce projet sera résolument originale en ce qu'elle promouvra *simultanément* :

- *l'improvisation* comme modèle général du dialogue humain-agents numériques
- *la modélisation stylistique* comme caractérisation et construction des individualités (celle de l'humain comme celle de l'agent)
- l*'adaptation temporelle de l'interaction*, c'est-à-dire l'évolution en temps-réel des modalités de l'interaction, en prenant en compte ses différentes échelles temporelles, des phénomènes à court-terme (réactions, synchronisation) aux stratégies organisées et émergentes (tours de parole, polyphonie, processus et formes)
- *l'apprentissage interactif et génératif,* c'est-à-dire la capacité d'interagir d'emblée avec une connaissance minimale et celle d'accumuler progressivement cette connaissance au cours de l'interaction improvisée
- *la dynamique réactive* synchrone ou asynchrone, qui combine l'échange dialogique humain/numérique, et la situation d'enrichissement / augmentation concomitants de l'humain par le numérique.

Au delà de l'amélioration incrémentale de l'état de l'art déjà atteint par les partenaires et les autres équipes, l'objectif de ce projet vise en termes de ruptures :

- **la constitution d'une écoute active** entraînant l'adaptation continue des mécanismes génératifs en fonction du contexte perçu (densité, rythme et pulsation, organisation harmonique, orchestration), la compréhension de l'agent créatif se manifestant par son action, qui en retour bouleverse les conditions même de l'expérience et de l'apprentissage. Il y a encore peu d'études sur ces formes d'écoute artificielle couplées à de l'apprentissage interactif et génératif, clefs pour nous de la créativité artificielle. Cette écoute doit procéder à une décomposition musicale informée dans un cadre de modélisation perceptivement et cognitivement inspirée qui affronte notamment la question d'exploiter la structure d'une pièce musicale pour mener à bien la séparation des sources dans le cas ou cette

structure n'est pas strictement élémentaire (e.g. répétitive). Nous bénéficions pour ces questions de l'expertise du projet DReaM coordonné au Lab-STICC/UBO et de celle du projet OMax à l'Ircam.

- **l'intégration des aspects multi-dimensions et multi-échelles** dans l'apprentissage des structures musicales et l'identification automatique des « dimensions sémantiques » portant le message musical émis par chaque agent, en repérant dans les signaux d'entrée les innovations structurantes à fort potentiel sémantique qui méritent un haut niveau d'activation des processus d'apprentissage. Cette question encore peu traitée implique de prendre en compte le problème de la rareté des données dans la situation improvisée, et l'émergence dans le modèle de niveaux multiples d'organisation de la matière musicale rendant compte de sa complexité. Nous bénéficions pour traiter cette question de l'expertise issue des travaux en théorie de l'information musicale menés à l'Ircam avec l'UCSD et de ceux sur la modélisation bayésienne de séquences musicales menés par Inria.
- **la constitution d'un modèle de décision** pour les stratégies synchrones et asynchrones : les modèles sociaux sous-jacents de l'improvisation sont le dialogue (alternance) et l'enrichissement (polyphonie). Les prototypes déjà réalisés dans la famille de logiciels OMax explorent déjà indépendamment ces deux modalités, mais le mécanisme de prise de décision automatisé pour les articuler dans le temps constitue de fait un verrou, et fait pour l'instant l'objet de décisions humaines. Ce point est à rapprocher de la question générale des dynamiques collectives d'agents dans leur dimension sociales, linguistiques, ou expressives, pour laquelle nous bénéficierons des avancées d'une thèse sur l'adaptation temporelle de l'interaction qui commence à l'Ircam et au LTCI dans le projet SeNSE du Labex SMART à l'UPMC.

3.4. POSITIONNEMENT ET COMPLEMENTARITE DES PARTENAIRES

Les partenaires fournissent une combinaison idéale pour un tel sujet qui n'a pas d'équivalent ailleurs en Europe. L'ANR est donc particulièrement bien placée pour financer ces travaux.

Le laboratoire STMS de l'Ircam, équipe Représentations Musicales (resp. Gérard Assayag, coordinateur du projet), porte les thématiques d'informatique musicale, de modélisation de l'interaction improvisée, de créativité artificielle. Le contexte hautement pluridisciplinaire du laboratoire STMS à l'Ircam permet une forte interaction inter-équipes et l'équipe représentations musicales collabore de manière continue avec les équipes Analyse-synthèse (descripteurs du signal, fouille de données sonores), ISMM (architectures d'interaction, gestualité), ainsi que PDS (Perception et design sonore). Le département production et pédagogie de l'Ircam permettra de mettre en place l'expérimentation avec les musiciens, le travail de studio, et le retour d'usage. Les logiciels OpenMusic et OMax de l'équipe représentent l'état de l'art en matière de composition assistée par ordinateur et d'interaction improvisée musicien / machine. L'équipe est aussi un leader dans le domaine des formalisations algébriques des structures musicales et de la synchronisation musicale musiciens / machines. L'équipe est ou a été récemment associée à plusieurs programmes nationaux dont les résultats ont une forte visibilité : ANR Sample Orchestrator (classé projet phare) ; ANR SOR2 ; ANR ImproTech ; ANR INEDIT (qui pose des questions d'articulation entre modèles de calcul synchrone / asynchrone qui nous seront utiles) ; Labex SMART coordonné par l'Institut de robotique de l'UPMC (ISIR). Elle est impliquée dans des collaborations continues avec de grands pôles internationaux : California Institute for Telecommunication and Information Technology (Université de San Diego, Prof. Shlomo Dubnov, Machine Listening et Musical Information Dynamics), Centre for Interdisciplinary Research in Music Media and Technology de McGill (Prof. S. McAdams, Perception et la cognition musicales), équipe AVISPA (Prof. Camilo Rueda) du Colciencias (le CNRS colombien) en pointe sur les langages TCC (Timed Concurrent Constraints) et la formalisation de la mobilité en calcul de processus.

Le laboratoire Lab-STICC, CNRS UMR 6285, équipe IHSEV : Interaction Humain Système et Environnement Virtuel (thème Image / Son) à l'Université de Bretagne Occidentale - UBO (resp. scientifique Pr. Sylvain Marchand) dont une des spécialités est l'analyse de scènes auditives pour l'interaction entre humains et systèmes artificiels, est particulièrement bien placé pour prendre en charge les questions d'écoute artificielle et de reconnaissance des structures musicales. Il a coordonné le programme ANR DReaM, qui a créé une véritable rupture en matière de séparation de sources informée, à l'aide d'informations a priori sur le processus de production afin d'obtenir une structure musicale permettant de ré-interpréter la musique en direct. DReaM s'est limité à la musique produite sur support et aux derniers maillons de la chaîne de production : mixage et mastering. Plus précisément, c'est l'inversion

du mixage avec connaissances a priori qui a été centrale. Le défi de DYCI2 sera d'aborder, outre l'écoute artificielle proprement dite (extraction de descripteurs), les situations de séparation informée de scènes sonores vivantes avec une information partielle collectée ou inférée de l'apprentissage en amont. L'équipe coordonne aussi le projet ANR INGREDIBLE qui développe un acteur virtuel interagissant avec les humains par couplage gestuel et «affectif» dans le cadre du théâtre augmenté, et dispose donc d'une solide expérience en termes d'agents créatifs et de perception artificielle.

L'équipe-projet Inria Parole d'Inria Nancy - Grand Est est spécialisée dans la modélisation statistique de la parole et du langage parlé, domaine dont sont aussi issues plusieurs des méthodologies largement employées pour la modélisation de séquences musicales aux échelles du signal et des symboles. Emmanuel Vincent (CR1 Inria et responsable scientifique pour ce partenaire) y développe une recherche centrée sur la musique pour laquelle il déploie des méthodes pionnières de modélisation et d'estimation bayésienne. Ce partenaire est donc parfaitement équipé pour prendre en charge les questions fondamentales liées à l'apprentissage de séquences formelles dans les conditions propres à ce projet dans lesquelles les données accessibles lors de l'interaction sont nécessairement limitées. E. Vincent est l'un des fondateurs de MIREX, le Music Information Retrieval Evaluation eXchange, une compétition annuelle qui a largement contribué à établir internationalement le nouveau champ de recherche du MIR (*music information retrieval*). Il a participé aux projets européens Quaero et EUREKA Eurostars "i3DMusic", et il a piloté l'équipe associée Inria VERSAMUS (http://versamus.inria.fr/) avec l'Université de Tokyo sur la modélisation bayésienne de séquences musicales.

Les travaux sur la modélisation du style et l'interaction improvisée se sont beaucoup développés à l'Ircam avec le soutien de l'ANR (projets ImproTech et SOR2). La thèse de F. Maniatakos soutenue en 2012, « Graphs and Automata for the Control of Interaction in Computer Music Improvisation », aborde les aspects théoriques des modèles stylistiques de séquences. La thèse de B. Lévy soutenue en 2013, « Principles and Architectures for an Interactive Music Improvisation System with On the Fly Listening, Learning and Generative capabilities » explore les architectures d'interaction. Le logiciel OMax issu de ces travaux sert de plate-forme expérimentale à ces études et constitue une référence reconnue internationalement. Les projets ANR INEDIT sur l'unification des paradigmes de programmation synchrone et asynchrone et EFFICACE (JCJC) sur l'extension réactive des langages pour la composition, et l'ERC grant CREAM sur le code émotionnel de la musique, fournissent de nouveaux outils conceptuels et contribuent à créer un contexte favorable et une dynamique de recherche intenses à l'Ircam sur tout le spectre des phénomènes musicaux et sonores. Cette recherche occupe une place éminente et originale dans la compétition internationale (e.g. ERC Flow Machines Sony CSL, MIT Media-Lab) par son couplage unique avec la création contemporaine.

L'écoute artificielle a connu récemment des renouvellements intéressants avec l'introduction des méthodes de décomposition informée (projet ANR DReaM coordonné par S. Marchand, Lab-STICC, best paper award AES 2012), les approches en théorie de l'information musicale (Musical Information Dynamics) objet d'une collaboration Ircam / UCSD, la géométrie de l'information (l'équipe-projet Mutant Ircam / Inria a introduit pour la première fois cette année la question de l'écoute artificielle de flux audio dans ce domaine en organisant la branche *Geometry of Audio Processing* de la conférence GSI'13 (Geometric Sciences of Information).

En parallèle, la modélisation statistique de séquences musicales a émergé comme un sujet en plein essor dans la communauté MIR. L'équipe associée Inria VERSAMUS en est un élément fondateur en appliquant pour la première fois à des tâches musicales (harmonisation, transcription polyphonique) des outils de l'état de l'art du traitement automatique des langues, notamment pour la modélisation bayésienne non-paramétrique de séquences d'accords en reconnaissance du genre musical. Cette dynamique a récemment été reconnue par l'ajout officiel de "Music language processing" aux thématiques des IEEE/ACM Transactions on Audio, Speech, and Language Processing et de la IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), à l'initiative d'Inria.

3.5. ETAT DE L'ART

3.5.1 ECOUTE INFORMEE ET CREATION DE CORPUS D'APPRENTISSAGE

L'analyse computationnelle de scènes auditives (*Computational Auditory Scene Analysis* ou CASA) [Bregman 1990] est un vaste domaine de recherche dont le but est d'imiter le système auditif humain, en

identifiant dans ces scènes complexes des entités sonores via des critères perceptifs comme la localisation dans l'espace [Mouba and Marchand 2006], la structure harmonique, la synchronicité ou la corrélation des paramètres sonores [McAdams 1989]. Notons que ces derniers critères constituent un premier pas vers une structure musicale (niveau symbolique). Dans ce projet, nous souhaitons que des agents musicaux, réels (instrumentistes) ou virtuels (ordinateurs), soient capables de s'écouter les uns les autres afin de jouer ensemble. L'écoute artificielle proposée ici doit permettre une écoute séparée de chaque entité (chaque instrument), tout en restant créative : autoriser la ré-interprétation, en imitant ou en improvisant.

Toutefois l'approche CASA souffre de limitations en pratique : elle nécessite l'estimation de nombreux paramètres, ce qui constitue une source d'erreurs inévitable et souvent inacceptable pour les applications musicales réelles. Il s'agit du même problème que pour la séparation de sources [Comon and Jutten 2010], du moins dans son approche aveugle classique. Nous proposons d'étendre l'écoute artificielle avec l'approche informée [Marchand *et al.* 2012] introduite avec succès dans le projet ANR DReaM, afin de contrôler les erreurs produites.

Pour cela, il est nécessaire de constituer un corpus d'apprentissage où les paramètres sont connus avec exactitude. Après la phase d'apprentissage, l'information peut être ré-injectée lors de la phase d'estimation de l'écoute, en situation de performance musicale.

Un premier lien entre théorie de l'information et théorie de l'estimation a été fait dans la thèse [Fourer 2013] aussi bien pour des informations de nature audio (physique) que musicale (symbolique). De manière à obtenir une qualité élevée, il est nécessaire d'exploiter toute information ou connaissance disponible qui peut renseigner sur le contenu des signaux à écouter [Liutkus *et al.* 2013]. Ces informations peuvent être les partitions des différentes lignes instrumentales jouées [Ewert and Müller 2012], la position spatiale de ces sources dans le mélange [Duong *et al.* 2010], une imitation produite par des utilisateurs [Smaragdis and Mysore 2009] ou encore des reprises réalisées par d'autres musiciens [Gerber *et al.* 2012].

Récemment, il a été montré que la structure d'une pièce musicale pouvait être utilisée pour mener à bien la séparation [Liutkus *et al.* 2014]. Dans ces études, seules des structures musicales très simples ont été considérées, où la musique est soit localement répétitive, soit identique à elle-même dans le temps. Pour aller plus loin dans la modélisation, il est nécessaire de disposer de beaucoup de données : idéalement le nombre de morceaux qu'a écouté un improvisateur humain avant de pouvoir commencer lui-même à improviser… Or dans le domaine musical nous disposons de peu de données aujourd'hui, contrairement aux travaux en modélisation du langage naturel.

De plus, les enregistrements musicaux se présentent la plupart du temps sous la forme de mélanges dans lesquels différentes pistes sonores sont présentes simultanément. Un exemple usuel est le cas des enregistrements en stéréo ou au format 5.1, pour lesquels chaque piste instrumentale apparaît dans tous les canaux, donnant lieu à des scènes sonores spatiales complexes. Même dans le cas plus simple où un microphone est placé devant chaque musicien, du fait des propagations acoustiques qui ont lieu, chaque microphone enregistre également du signal parasite provenant des autres instruments (phénomène de "repisse").

3.5.2 MODELISATION DE SEQUENCES MUSICALES POUR L'INTERACTION

Les travaux sur la modélisation de séquences formelles pour l'appréhension des styles musicaux et l'interaction créative improvisée se sont beaucoup développés ces dernières années dans les communautés SMC (Sound and Music Computing) et informatique musicale (Computer Music). Les situations réalistes d'interaction (scène, studio, jeu) visées engageant de sérieuses contraintes de réactivité, de traitement temps-réel et de créativité des systèmes artificiels mis en jeu ont orienté les recherches vers certains types de problèmes et d'approches.

Un premier problème concerne les modèles statistiques et formels de séquences musicales et leur capacité à capturer des styles musicaux. [Conklin 2003] a dressé une synthèse des procédés de génération musicale à partir des modèles statistiques. Les modèles à contexte (*context models*) qui en constituent la plus grande part assignent des probabilités aux événements sous condition des séquences d'événements passés (contextes). Ces modèles incluent les divers systèmes markoviens, les n-grammes et automates à états-finis, ils sont efficaces d'un point de vue computationnel, s'apprennent aisément à partir d'exemples et sont représentés par des structures formelles (e.g. arbres, automates) pour lesquels existe une vaste

littérature algorithmique. Une fusion des modèles statistiques et logiques de séquence a été tentée par [Pachet 2011]. A la suite des travaux séminaux concernant le codage optimal et les prédicteurs universels [Feder 1996] nous avons montré que les modèles à contextes pouvaient devenir adaptatifs et supporter des ordres markoviens variables en utilisant un compresseur optimal pour encoder les séquences [Assayag 1999; Dubnov 2003]. Cette optimalité n'est malheureusement assurée qu'asymptotiquement sur la longueur des séquences. Nous alors validé pour les séquences musicales puis amélioré une structure issue de la communauté des langages formels et de la bio-informatique, l' « oracle des facteurs » [Crochemore 2007], un graphe linéaire contenant à la fois une représentation temporelle et une cartographie des contextes d'ordre markovien arbitraire, bien adaptée à la génération de nouvelles séquences par simple marche aléatoire [Assayag 2004, Assayag 2010, Maniatakos 2012] et ne dépendant plus de la longueur des séquences. A l'appui de ces modélisation statistique et formelle, la dynamique d'information musicale (*musical information dynamics*), discipline inspirée de la théorie de l'information moderne apporte de nouveaux moyens d'analyse et de contrôle à l'aide de mesures de débit d'information (*information rate*). Cette mesure permet une sélection automatique des meilleurs descripteurs audio (*feature selection*) pour l'analyse qinsi que de leurs seuils de distance de sorte à maximiser l'information dans le système [Dubnov 2011, Surges 2013].

Un deuxième problème concerne les contraintes de l'interaction mettant parallèlement en jeu l'apprentissage et la génération dans une situation de jeu interactif improvisé. Ces contraintes ont été explorées notamment du point de vue de l'algorithmique et des architectures informatiques [Blackwell 2012], de la dynamique des interactions vues comme un système complexe [Canonne 2011], ou de la synchronisation entre musicien et agents informatiques par correspondance entre les descripteurs audio extraits du jeu live et ceux stockés dans le modèle [Moreira 2013] dans une approche de type MIR (Music Information Retrieval). L'apprentissage est en général conduit sur des corpus, en temps différé par rapport à l'interaction improvisée et l'interaction suit une grille rigide. Nous avons conduit des recherches qui ont abouti à l'environnement *OMax*, dont le modèle formel est l'oracle des facteurs, et dont l'architecture d'interaction [Levy 2013] et le schéma d'apprentissage et de génération [Assayag 2007, Assayag 2010] permettent l'acquisition à la volée pendant la performance sans modèle de connaissance à priori [Lévy 2012], ainsi qu'une synchronisation adaptative entre musicien et agent informatique sur les plans harmonique, textural et rythmique (travaux de post-doc de Laurent Bonnasse-Gahot). OMax est une référence internationalement reconnue en matière de logiciel pour la scène et l'improvisation, qui a donné lieu a de nombreux concerts publics, workshops et master-classes en France et à l'étranger[2]. Il constitue une base très solide pour ce projet dont un des enjeux est de doter ce type de systèmes, encore contrôlés en grande partie par un opérateur humain pour ce qui est des choix stratégiques dans l'interaction, des capacités d'autonomie, de décision et d'adaptation qui leur manquent encore pour affronter sans supervision humaine des situations de jeu complexes impliquant un nombre significatif d'agents.

3.5.3 MODELISATION BAYESIENNE POUR L'ANALYSE DE MUSIQUE

En parallèle des travaux des communautés SMC et informatique musicale, la modélisation statistique de la musique symbolique est devenue un sujet de recherche à part entière dans la communauté MIR (*music information retrieval*), que ce soit pour l'analyse harmonique, la transcription de mélodie, ou la segmentation couplet/refrain [Raphael 2004, Ryynänen 2008, Paulus 2008]. Alors que la plupart des travaux restent basés sur des modèles de Markov cachés d'ordre n=1, des modèles bayésiens plus évolués ont vu le jour ces dernières années. Ces modèles sont inspirés en particulier des techniques de l'état de l'art pour le traitement automatique de la langue écrite/parlée, qui nécessite une adaptation aux spécificités de la musique.

Un premier problème concerne l'estimation de la probabilité d'une séquence à partir d'un corpus d'apprentissage. La difficulté est d'estimer cette probabilité même lorsque la séquence n'apparaît qu'un petit nombre de fois voire aucune fois dans le corpus d'apprentissage, alors qu'elle peut se produire dans le corpus de test. Des techniques de lissage des probabilités des n-grammes (Kneser-Ney, Pitman-Yor process...) ont été développées dans la communauté de traitement de la parole [Chen 1996, Teh 2006] que

[2] Cf. Workshop ImproTech Paris New-York 2012 co-organisé à New-York par l'Ircam, l'EHESS, New York University et Columbia University (http://repmus.ircam.fr/improtechpny). OMax est utilisé comme partenaire par de nombreux musiciens de stature internationale (68 performances publiques visibles sur http://www.dailymotion.com/RepMus).

nous avons validées pour la modélisation de séquences d'accords [Scholz 2009, Raczyński 2014]. Ces techniques restent à étendre à des modèles de séquences plus avancés, tels que l'oracle des facteurs. D'autres chercheurs ont constaté que les facteurs de lissage souhaités dépendent du genre musical et proposé de les fixer manuellement selon le genre considéré [Pachet 2012].

Un deuxième problème concerne la modélisation simultanée de plusieurs dimensions musicales : mélodie, harmonie, rythme... [Conklin 2013]. L'approche la plus immédiate consiste à modéliser chaque symbole de la séquence comme un point de l'espace obtenu par produit cartésien des dimensions et à répertorier les séquences et leurs probabilités dans cet espace. En pratique, l'explosion combinatoire de la taille de l'espace modélisé implique que la plupart des séquences possibles ne sont jamais observées. Cette limitation fondamentale contraint encore aujourd'hui les systèmes d'improvisation automatique à apprendre une seule dimension des séquences (par exemple, la hauteur) et à répéter à l'identique les autres dimensions associées (par exemple, le rythme). Pour dépasser cette limitation et augmenter l'expressivité de l'agent improvisateur en lui permettant de générer de nouvelles séquences (par exemple, une séquence de hauteurs et de rythmes observés par le passé mais pas conjointement), il est nécessaire de réduire la dimension de l'espace par une paramétrisation appropriée. L'interpolation log-linéaire de modèles uni-dimensionnels, initialement proposée pour le traitement des langues dans [Klakow 1998], a ainsi été utilisée pour la modélisation conjointe de la hauteur et de l'harmonie dans [Mauch 2010, Raczyński 2013b] et pour la modélisation de la hauteur polyphonique dans [Raczyński 2013a] en interpolant un ensemble de modèles d'ordre n=1. La généralisation de telles approches à l'oracle des facteurs est un problème ouvert. L'interpolation, qui peut être vue comme une approximation de rang 1 du tenseur des probabilités conditionnelles, limite les couplages entre dimensions qui peuvent être modélisés. Des approximations de rang plus élevé restent à explorer.

Un troisième problème concerne l'adaptation des modèles au style musical. La probabilité d'une séquence musicale dépend en effet énormément du contexte musical dans lequel elle est jouée. À nouveau, l'approche la plus immédiate qui consiste à apprendre un modèle pour chaque style [Lee 2007] est confrontée au problème du nombre quasi-infini de styles possibles, chaque compositeur et interprète ayant son propre style. Ce problème, analogue à celui de l'adaptation des modèles de langage naturel au thème d'un document, a été traité par l'utilisation de modèles de thème tels que la *Latent Dirichlet Allocation* (LDA) [Blei 2003], qui modélisent un document comme l'interpolation de plusieurs thèmes. Nous avons généralisé la LDA à la modélisation de séquences d'accords en fonction du genre musical en proposant un nouveau modèle bayésien non-paramétrique de thème [Raczyński 2014]. Une fois de plus, ce modèle reste à généraliser à des modèles de séquence autres que les n-grammes. Des heuristiques apparaissent aussi nécessaires pour réduire sa complexité calculatoire très importante, due à la modélisation de tous les styles en parallèle.

Un quatrième problème concerne la modélisation de régularités à long terme dans l'organisation temporelle de la musique. Dans la plupart des genres musicaux, les répétitions de phrases musicales ne se produisent pas à n'importe quel instant mais elles suivent une certaine forme. Les morceaux de jazz sont par exemple le plus souvent composés de blocs de 32 mesures, eux-mêmes composés de 4 blocs de 8 mesures formant une séquence de type "aaba", "aabc" ou "abac" [David 1998]. Alors que l'oracle des facteurs a une structure "plate", la modélisation de la forme passe par un modèle hiérarchique [Paiement 2005]. La difficulté est que la forme varie d'un style musical à l'autre et qu'elle est rarement parfaitement régulière au cours d'un même morceau: les blocs de base peuvent être raccourcis, rallongés ou tuilés et des ponts plus courts peuvent s'insérer afin d'augmenter l'intérêt formel du morceau. Nous avons conçu un formalisme appelé "système-contraste" qui généralise les formes "aaba", "aabc" et "abac" et qui explique ainsi la majorité des formes rencontrées dans d'autres genres musicaux d'un point de vue musicologique [Bimbot 2010, Bimbot 2014]. L'exploitation de ce formalisme pour l'analyse et la génération automatiques de séquences musicales reste un problème totalement ouvert.

En résumé, de nombreux modèles bayésiens avancés ont récemment été développés pour l'analyse de contenus musicaux, mais ils ne sont pas directement utilisables pour la composition et l'improvisation. Par ailleurs, les algorithmes d'apprentissage sont conçus pour des corpus figés, qui ne reflètent pas la situation d'improvisation considérée dans ce projet. Le mariage entre ces deux familles de modèles (modèles de séquences et approche bayésienne) et l'apprentissage continu en situation d'improvisation restent des verrous majeurs.

3.5.4 DYNAMIQUE ET LOGIQUE DE L'INTERACTION IMPROVISEE

Entreprises par des chercheurs issus de la linguistique dès le début des années 1970, un certain nombre d'études se sont intéressées aux informations paralinguistiques véhiculées dans l'interaction d'agents, qu'il s'agisse de la prise de parole [Thórisson 2002], de l'émission de signaux paraverbaux et de leur contextualisation [Allwood 1976] ou de l'étude de signaux simples comme les hochements de tête pendant une conversation [Ishi 2014]. Ces signaux non-verbaux suivent une certaine évolution dynamique [Thórisson 2002] et comme pour la communication verbale dépendent énormément du contexte dans lequel ils sont émis et parfois des interactions précédentes pour être décodés correctement voire seulement perçus. Un exemple souvent utilisé est celui de l'ironie car dans ce cas des informations sont transmises par le biais de la prosodie qui inversent le sens transmis par la communication verbale [Fujie 2004]. Ces comportements non-verbaux ont été modélisés dans des agents artificiels [Kopp 2010] montrant qu'une plus grande richesse de l'interaction avec un système artificiel entraîne un meilleur engagement et donc une plus grande efficacité des utilisateurs, cela se traduisant concrètement par une plus grande richesse du vocabulaire, un plus fort taux de succès de l'interaction et un prolongement du temps d'interaction. [Novielli 2010]. L'interaction n'est plus perçue comme un simple échange de signaux mais comme une construction adaptative, empruntant de nombreuses modalités sur différentes échelles de temps [Knapp 2013]. Cela implique de concevoir des systèmes qui construisent au fur et à mesure leur représentation du monde et s'adaptent de fait au monde tel qu'ils le perçoivent et non plus comme un ensemble de règles logiques. C'est partiellement le cas l'agent numérique *OMax* [Assayag 2010] qui élabore progressivement une représentation de la musique perçue tout en intervenant lui-même dans l'échange musical, ou des *Talking Heads* [Steels 2003], où des agents parviennent à construire une représentation du monde via un jeu de langage en se désignant les objets de leur champ de vision.

Dans ce cadre de communication non-verbale inter-agents l'improvisation porte intrinsèquement les notions de spontanéité et de réactivité. Elle n'est cependant naturellement pas antagoniste avec celles de contraintes à long terme, de planification et d'organisation comme le soulignent les études cognitives à son sujet [Pressing 1984]. C'est le cas des grilles de jazz, de la basse écrite dans l'improvisation baroque, ou des prescriptions du raga indien qui constituent des séquences formalisées de contraintes. Cette conciliation se retrouve dans certains systèmes musicaux interactifs présentant une structure organisant l'improvisation sur le long terme. Dans cette catégorie, le concept de « control improvisation » [Donze 2013] propose un processus de génération d'une séquence d'événements de contrôle guidée par une séquence de référence et satisfaisant une spécification donnée. [Surges 2013] permet l'écriture de scripts déterminant en amont de la performance l'évolution des descripteurs d'analyse à prendre en compte et des paramètres de génération. Rejoignant des travaux sur l'utilisation des grilles de jazz dans l'improvisation [Chemillier2004] et dans la continuité des recherches initiées dans [Assayag 2004, 2007] sur la navigation dans une cartographie (oracle de facteurs) capturant partiellement la logique musicale séquentielle d'un improvisateur, nous avons ajouté des contraintes à long terme et des connaissances a priori en introduisant l'idée de « scénarios » pour guider l'improvisation [Nika 2012], dans le cadre d'une thèse co-dirigée par Marc Chemillier et Gérard Assayag à l'Ircam et l'EHESS.

La dynamique d'interaction improvisée puise donc aux sources de la communication non-verbale, de l'adaptation temporelle et multi-échelle des échanges, et des structures formelles de séquences et de scénarios venant rajouter une perspective à long terme.

4. PROGRAMME SCIENTIFIQUE ET TECHNIQUE, ORGANISATION DU PROJET

4.1. PROGRAMME SCIENTIFIQUE ET STRUCTURATION DU PROJET

Le projet articule entre elles trois tâches de recherche autour d'un environnement expérimental :

1. **l'écoute informée créative** (resp. Sylvain Marchand, UBO) vise à donner les moyens à un agent numérique d'analyser les scènes sonores en temps réel pour extrapoler la structure musicale en exploitant les similarités observées et des connaissances disponibles a priori. Cette recherche se situe dans le cadre de la «décomposition musicale informée» et souhaite dépasser l'état de l'art en partant d'une scène sonore complexe vivante, par exemple produite dans le cas des musiques mixtes ou improvisées en direct. De l'information a priori est toujours disponible : partition des parties écrites, prises de son lors de répétitions, annotations, et il est possible d'inférer aussi une information partielle

à partir de l'apprentissage de corpus stylistiquement proches. Il s'agit donc de retrouver la structure musicale, notamment dans sa décomposition polyphonique, en direct.

2. **l'apprentissage interactif de structures musicales** (resp. Emmanuel Vincent, Inria) vise, à partir des données séquentielles fournies par les processus d'écoute, à apprendre des modèles symboliques susceptibles de capturer les structures musicales de haut niveau, multi-dimensionnelles et multi-échelles émergeant dans un certain contexte de performance musicale et d'improvisation. Nous proposons une approche nouvelle consistant en l'intégration entre modélisation de séquences formelles (plus spécifiquement, l'oracle des facteurs) et approches bayésiennes : le développement de modèles et d'algorithmes rendant mieux compte de la complexité du discours musical pose le problème d'apprendre de tels modèles à partir d'une quantité de données nécessairement limitée dans le scénario visé. L'approche bayésienne offre une solution théorique à ce problème dont l'efficacité a été largement démontrée pour la modélisation du langage parlé. L'intégration entre modèles de séquences formelles et approches bayésiennes dans un contexte musical offre de nombreuses perspectives non explorées à ce jour. Le défi sera en particulier de maintenir l'efficacité calculatoire des premiers tout en bénéficiant de la robustesse au sur-apprentissage des secondes.
3. **les dynamiques d'interaction improvisée** (resp. Gérard Assayag, Ircam) permettent une interaction improvisée riche et créative entre agents artificiels et humains et posent les questions de l'adaptation temporelle et collective de l'interaction à plusieurs échelles. Cette tâche profite de l'écoute informée et des capacités d'analyse et d'apprentissage des deux premières en articulant un système adaptatif et anticipatif de gestion de l'interaction improvisée en se posant la question du modèle de mémoire, de connaissance et de contrôle interne des agents. Il s'agit de dépasser les approches classiques statiques et déterminées préalablement et de pouvoir adapter en temps-réel les modèles, représentations et modalités d'apprentissage de l'interaction, en prenant en compte ses différentes échelles temporelles et les dynamiques collectives susceptibles d'y être attachées.

Ces trois tâches, pour chacune desquelles au moins deux des partenaires collaborent et mettent en avant une expertise reconnue, correspondent aussi aux principales compétences exercées de manière parallèle, compétitive et contributive par un agent créatif autonome en situation d'interaction improvisée avec des humains ou d'autres agents. Elles sont complétées par une tâche de gestion du projet et par une tâche d'intégration, d'expérimentation, de validation et de retour d'usage intégrant aussi la dissémination.

Les deux premières tâches de recherche sont structurées autour de deux nouvelles thèses encadrées à l'UBO et à Inria (en co-encadrement avec l'Ircam) et explorent l'écoute informée et la modélisation des structures musicales en incorporant progressivement les aspects les plus créatifs (de la séparation des sources à l'écoute structurante et l'écoute creative par décomposition et recomposition pour l'UBO, de l'apprentissage sur les descriptions multi-dimensionnelles de la musique à la détection de dimensions sémantiques et leur mise en jeu dans les scénarios d'interaction pour Inria).

La troisième tâche, nourrie des deux premières pour ce qui concerne les capacités d'écoute et d'analyse affronte la question de la structure interne des agents numériques et de leur dynamique d'interaction improvisée, en prenant en compte l'adaptation temporelle et collective de leurs comportements à travers des modèles d'attention, de décision et de connaissance. Elle bénéficie de l'expertise issue de deux thèses en cours à l'Ircam en collaboration avec l'EHESS et le LTCI sur le guidage de l'improvisation et l'adaptation temporelle de l'interaction, et s'appuiera sur un nouveau post-doc pour l'étude de la structure de mémoire, de connaissance et de contrôle des agents créatifs numériques.

Les outils (algorithmes, maquettes logicielles, corpus) issus de ces trois tâches seront intégrés en fin de projet dans une suite logicielle qui sera mise à la disposition en tout ou partie de la communauté scientifique et artistique. L'utilisation commune envisagée du langage Python et de la plate-forme temps-réel Max/MSP et/ou Pure Data facilitera les échanges. À travers cette suite logicielle et son instanciation en différentes applications spécifiques il sera alors possible de construire des systèmes multi-agents intelligents susceptibles d'affronter de manière crédible des situations d'interaction improvisée plus complexes en mettant en jeu une variété de modèles d'écoute, d'apprentissage et de génération.

Dans la quatrième tâche, outre l'intégration logicielle prise en charge par un ingénieur d'étude recruté pour le projet, des sessions d'expérimentations avec des musiciens professionnels de niveau international sont prévues dès la première année et au long du projet à l'Ircam et UBO, pour évaluer les algorithmes et les maquettes logicielles à divers points d'aboutissement. Ces sessions seront complétées par deux résidences artistiques à l'Ircam à partir de la deuxième année, pour confronter les résultats de la recherche

à des projets artistiques ou culturels d'une certaine ambition. Une importante collaboration avec l'EPFL à Lausanne permettra aussi, dans cette tâche, d'intégrer une série d'expériences validantes autour des archives numérisées du Montreux Jazz Festival (5000 heures, plus grande collection live au monde, classée au patrimoine mondial par l'Unesco) qui contiennent un grand nombre de séquences d'improvisation soliste et polyphonique. Ces expériences mettront en jeu l'écoute informée pour la séparation de sources, l'apprentissage des structures pour la reconnaissance et l'organisation temporelle et formelle des composantes musicales multi-dimensionnelles, et la dynamique d'interaction pour proposer des scénarios inédits de re-création dynamique des concerts par la génération temps-réel de nouvelles séquences, puisant dans le matériau audio-visuel enregistré, et en interaction (guidage, contrepoint, solo/accompagnement) avec l'utilisateur humain et/ou des agents numériques. La quatrième tâche comprend aussi la dissémination avec notamment l'organisation prévue d'un colloque international et d'un workshop dans une grande conférence du domaine.

4.2. ORGANISATION DU PROJET EN TACHES

4.2.1 TACHE WP0 : GESTION DU PROJET

Responsable : G. Assayag (Ircam)

Partenaires impliqués (4 H.Mo) : Ircam : 2 (G. Assayag), Inria : 1 (E. Vincent), UBO : 1 (S. Marchand)

Durée : 36 mois de T0 à T0+36

Objectifs généraux de la tâche

Coordination du projet. La tâche WP0 est dédiée à la coordination du projet assurée par l'Ircam. Le coordinateur veillera à l'animation inter-partenaires et à la remise des livrables selon le calendrier prévu. Le coordinateur sera assisté dans sa tâche par un *comité de suivi* composé de lui même et des responsables scientifiques de chaque partenaire. Le comité de suivi se réunira 3 fois par an et établira un tableau de bord collectif d'indicateurs d'avancement des tâches (algorithmes, rapport, publications). Il discutera et préparera les accords de propriété intellectuelle et de licences si nécessaire (voir § Stratégie de valorisation) et mettra en place un site web collaboratif (wiki) permettant la planification et le partage des diverses productions du projet (rapports, logiciels, documentations, corpus, études de cas). La partie publique du wiki permettra de diffuser les développements et d'illustrer leurs utilisations lors de performances et d'actions pédagogiques. Le coordinateur aura aussi la responsabilité de déterminer avec les partenaires, dès les premières réunions, les caractéristiques communes de la plate-forme expérimentale (langage, architecture, API, environnements temps-réel expérimentaux) de manière à assurer l'intégrabilité tout au long du projet.

Comité scientifique. Nous compléterons le comité de suivi par un comité scientifique constitué de 4 experts étrangers. Ce comité sera constitué pendant les six premiers mois du projet. Le rôle de ces experts est de donner un avis scientifique et technique, en particulier sur les choix faits en amont et de participer au comité de thèse des doctorants associés au projet. Sont pressentis pour ce comité Camillo Rueda, directeur du Département d'ingénierie de l'Universidad Javeriana de Cali, Colombie, et spécialiste des contraintes temporelles, Elaine Chew, Professor of Digital Media, Centre for Digital Music, Queen Mary University of London, spécialiste des approches computationnelles de la tonalité et de la performance, Shlomo Dubnov, Directeur du Center for Research In Entertainment and Learning, California Institute for Telecommunications and Information Technology (CALIT2), UCSD, spécialiste de théorie de l'information et dynamique de l'information audio-musicale, David Wessel, directeur du Center for New Music and Audio Technologies (CNMAT) de l'université de Berkeley, spécialiste des interactions homme-machine et des nouveaux instruments.

Livrables

L0.1 (M3) Mise en place du Wiki du projet
L0.2 (M12) Rapport d'avancement annuel
L0.3 (M24) Rapport d'avancement annuel
L0.4 (M36) Rapport final

4.2.2 TACHE WP1: ECOUTE INFORMEE CREATIVE

Responsable : S. Marchand, UBO

Partenaires impliqués (55 H.Mois) : Ircam : 4 (G. Assayag, G. Bloch), Inria : 6 (A. Liutkus), UBO : 45 (S. Marchand, V. Koehl, M. Paquier, doctorant)

Durée : 36 mois de T0 à T0+36

Objectifs généraux de la tâche

Dans ce projet, nous souhaitons que des agents musicaux, réels (instrumentistes) ou virtuels (ordinateurs), soient capables de s'écouter les uns les autres afin de jouer ensemble de manière satisfaisante en conditions d'improvisation (et donc en direct). Un ordinateur doit par conséquent être capable d'analyser la scène sonore [Bregman 1990], pour extrapoler la structure musicale, et ce en temps réel et à la manière d'un musicien. Le système auditif humain retrouve à partir du signal sonore des paramètres musicaux et les organise, par exemple verticalement ou horizontalement dans un plan temps-fréquence. Verticalement, il s'agit alors de regrouper à un instant donné les paramètres appartenant à la même entité perceptive, en général produite par la même source sonore. Horizontalement, il s'agit d'organiser les paramètres dans le temps, en retrouvant la structure musicale. Dans les deux cas il s'agit de structurer les informations musicales, en exploitant les similarités observées et des connaissances a priori. Car sans information a priori, en pratique cette structuration reste imparfaite et incomplète. Imparfaite car une erreur d'estimation est inévitable, et pourtant souvent inacceptable dans les conditions musicales réelles. Et incomplète, car les paramètres estimés sont classiquement de simples "descripteurs" ne permettant pas la resynthèse du son musical, pourtant nécessaire en situation de concert. Dans cette tâche nous proposons de reprendre le formalisme de l'approche informée pour l'écoute active [Marchand *et al.* 2012, Liutkus *et al.* 2013]. Il s'agit de généraliser cette approche pour la séparation de sources aux autres canaux d'information (sous-tâche 1.1), de prendre en compte la structure temporelle de la musique (sous-tâche 1.2), et de considérer une information approximative, non exacte (sous-tâche 1.3), avec toujours comme objectif de pouvoir regénerer le son musical de manière plus ou moins fidèle ou créative : de l'imitation à l'improvisation.

- *Sous-tâche WP1.1 : Séparation informée multi-canal*

Responsable : S. Marchand, UBO

Partenaires impliqués (18 H.Mois) : Inria : 2 (A. Liutkus), UBO : 16 (S. Marchand, V. Koehl, M. Paquier, doctorant)

Durée : 12 mois de T0 à T0+12

Description

La structuration verticale du son musical se confond en général avec la "séparation de sources" sonores. En effet, les sons produits par les différents objets sonores (ou sources) se mélangent à un instant donné, et sans information a priori il est impossible de les reséparer avec une qualité suffisante. Il est donc souhaitable d'exploiter des informations provenant de différents canaux. Ces informations peuvent être de type audio et provenir du son de chaque objet sonore (capté par un microphone spécifique) ou des voies de l'image spatiale du son global (lors d'un enregistrement stéréophonique, soit 2 voies, ou multi-canal - souvent en format 5.1, soit 6 voies). Le projet ANR DReaM s'est principalement limité à l'exploitation de l'image (sonore) spatiale globale, avec la connaissance de la position et des signaux des objets de la scène sonore, pris en conditions de studio. Mais en conditions "*live*", ces positions sont souvent inconnues et ces signaux comportent des interférences, chaque microphone spécifique captant un peu les signaux des autres sources qui jouent en même temps. Ce phénomène est connu sous le nom de "repisse". Les signaux sources sont alors corrélés. Nous proposons de nettoyer la repisse dans les corpus d'enregistrements *live*, et ce afin de disposer du son produit par chacun des instruments avec la meilleure qualité possible. Une approche envisagée consiste à exploiter les connaissances a priori importantes dans de tels enregistrements, et en particulier le fait que pour chaque microphone, une source est prépondérante. Et les informations ne se limitent pas au type audio. En présence d'information de type vidéo, les positions des objets sonores pourraient être estimées à partir des images (graphiques). Ces informations peuvent également être de type musical, comme les paramètres de hauteur (problématique de l'estimation de hauteurs dans le cas polyphonique), cf. [Fourer 2013].

Les bases théoriques ont dans l'ensemble déjà été posées, et il convient de les appliquer à des corpus et des cas pratiques. Par exemple, côté UBO l'approche informée combinant théorie de l'estimation et théorie de l'information a été proposée à des niveaux aussi bien signal (son) que symbolique (musique) [Fourer 2013]. Aussi, les connaissances a priori peuvent-elles être efficacement encodées dans un

formalisme bayésien, s'intègrant harmonieusement avec les récents travaux d'Inria sur le modèle local Gaussien [Duong et al., 2010]. Compte tenu de l'importante masse de données à traiter, cette tâche pourra de plus bénéficier d'une modélisation non paramétrique locale des spectrogrammes audio [Liutkus et al., 2014], efficace d'un point de vue calculatoire.

Indicateurs de succès, risques et solutions de repli

L'évaluation de la qualité de la séparation se fait principalement en comparant, avec des mesures objectives ou des tests d'écoute subjectifs, les résultats produits aux signaux de référence utilisés pour des exemples de test. Ces signaux de référence sont suffisamment simples pour que l'extraction des informations se fasse avec une erreur faible. Le principal risque est de faire face à des signaux trop complexes (pour l'extraction fiable d'informations, et a fortiori pour la séparation), mais nous avons toujours la possibilité de limiter notre étude (sons d'instruments plus simples - par exemple monophoniques et harmoniques, style de jeu plus limité, etc.).

Livrables

L1.1 (M12) Séparation informée multi-canal (rapport + maquette logicielle)

- *Sous-tâche WP1.2 : Écoute structurante*

Responsable : S. Marchand, UBO

Partenaires impliqués (19 H.Mois) : Ircam : 3 (G. Assayag, G. Bloch), Inria : 2 (A. Liutkus), UBO : 14 (S. Marchand, V. Koehl, M. Paquier, doctorant)

Durée : 15 mois de T0+9 à T0+24

Description

Pour aller plus loin dans la décomposition du son musical, il convient de prendre en compte le temps. Les résultats produits dans le passé peuvent s'avérer être une information utile pour une séparation dans le futur. Par exemple, il est courant que les instruments de musique entrent en jeu au fur et à mesure. Le mélange devient alors de plus en plus complexe, mais la prise en compte des informations du passé (où la situation était plus simple) peut permettre d'affiner l'estimation à l'instant présent. On peut voir cela comme une séparation informée par le passé, avec exploitation de la structure musicale. À plus long terme, cela peut permettre la compression du son musical à un niveau d'abstraction supérieur à celui du MP3 (qui lui ne tient pas compte de la structure musicale).

L'analyse structurelle d'une pièce musicale produit entre autres une description de son contenu en termes de similarités. Il est fréquent qu'une analyse de contenu donne lieu à la production d'une matrice de similarité, permettant d'identifier différentes concordances au sein d'une pièce musicale. L'utilisation de telles matrices de similarité dans un contexte de séparation a fait l'objet de travaux prometteurs récemment [Liutkus *et al.*, 2013]. Dans le cadre de cette tâche, ces travaux seront étendus au cas où les structures qui se retrouvent à différents endroits du mélange ne sont pas parfaitement identiques, mais ont subi des déformations particulières, qu'elles soient harmoniques ou rythmiques. Le formalisme des modèles additifs à noyaux [Liutkus *et al.*, 2014] est particulièrement bien adapté pour aborder cette problématique.

Indicateurs de succès, risques et solutions de repli

L'analyse structurelle d'une musique peut se faire à l'oreille, avec un expert humain. Nous disposons d'ailleurs d'un bon nombre de pièces déjà analysées, et qui pourront servir de références lors de la comparaison avec les résultats produits. Il est aussi possible de générer des exemples synthétiques de musiques pour les tests. Ensuite, la notion de similarité entre structures musicales est un sujet complexe, mais de nombreux travaux existent déjà sur ce sujet dans le domaine MIR *(Music Information Retrieval)*. Il serait trop risqué de s'attaquer d'emblée aux pièces musicales trop complexes. La solution est encore une fois de se limiter dans un premier temps aux cas simples (music pop, répétitive), puis de monter progressivement en complexité (jusqu'au jazz et aux musiques contemporaines improvisées).

Livrables

L1.2 (M24) Ecoute structurante (rapport + maquette logicielle)

- *Sous-tâche WP1.3 : Décomposition / recomposition par imitation*

Responsable : S. Marchand, UBO

Partenaires impliqués (18 H.Mois) : Ircam : 1 (G. Bloch), Inria : 2 (A. Liutkus), UBO : 15 (S. Marchand, doctorant)

Durée : 18 mois de T0+18 à T0+36

Description

Jusqu'à présent les informations utilisées pour la séparation / structuration étaient partielles mais exactes. Le but de cette sous-tâche est d'étendre la problématique des sous-tâches précédentes à une information s'éloignant de l'enregistrement d'origine, mais restant "semblable", par exemple une autre réalisation d'un même modèle statistique. Cette information peut alors provenir par exemple d'une imitation ou d'une variante. En pratique, l'imitation peut être une reprise du même morceau par d'autres musiciens (assez fidèle à l'original), et la variante peut être le résultat d'une improvisation (plus éloignée de l'original). On peut alors imaginer, à partir de cette imitation ou improvisation, le renforcement (solo) ou au contraire l'atténuation ("*music minus one*") d'une composante (source) de la musique initialement enregistrée. Par exemple, une imitation fidèle d'un des objets sonores (sources) peut permettre sa séparation de la musique enregistrée (mélange) avec une qualité acceptable. Le défi est de pouvoir maintenir une bonne qualité avec une imitation de moins en moins proche de l'original. Cela permettrait de faire le lien entre musique produite (sur support) et musique *live* (en direct) : une musique mixte, réinterprétable en direct, en fonction de l'imitation / improvisation. Cela permettrait également de pouvoir transformer des enregistrements anciens à l'aide de reprises modernes (problématique du "*back catalog*" posée par les *majors* de l'industrie musicale).

Dans le but de séparer un morceau de musique, il est possible qu'on dispose d'enregistrements qui lui sont similaires et en particulier qu'on dispose des pistes séparées d'une reprise. Si certaines études ont porté sur ce problème de la séparation informée par des reprises [Gerber *et al.*, 2012], elles ne considèrent pas le problème fondamental de l'alignement de ces reprises avec le mélange à séparer. Dans cette tâche, nous aborderons cette problématique à la fois par l'utilisation de modèles algébriques et de factorisation tensorielle [Liutkus *et al.* 2013]. On peut également penser à un algorithme d'alignement temporel fonctionnant en parallèle sur chaque piste, avec une mesure de similarité sur le mélange obtenu (sorte d'algorithme "*Dynamic Time Warping*" [Müller 2007] ou DTW parallèle). Mais les fluctuations du temps ne sont pas les seules à considérer. En effet, il faudra également considérer les changements de hauteur (transpositions), intensité (amplifications / atténuations), timbre (filtrages), etc.

Indicateurs de succès, risques et solutions de repli

La stratégie est de partir d'une information exacte, cas dans lequel nous disposons déjà de nombreux résultats, et de "dégrader" cette information (en la rendant inexacte), et ce de manière progressive. Par exemple, si on connaît exactement le signal d'une composante (source) de la musique, la renforcer ou l'atténuer est trivial. Des premiers résultats existent pour une information légèrement dégradée. Mais réaliser la même tâche avec imitation éloignée de l'original est un véritable défi. Heureusement, il est toujours possible de se limiter à certaines dégradations de l'information (seulement la mise en espace, ou seulement le timbre, ou bien des transpositions, etc.). Avec en plus la prise en compte de l'alignement temporel de chaque entité du son musical, un risque majeur est que les algorithmes produits soient trop complexes pour tourner en temps réel. Mais une version en temps différé serait déjà utile (notamment pour le "*back catalog*").

Livrables

L1.3 (M36) Décomposition / recomposition par imitation (rapport + maquette logicielle)

4.2.3 TACHE WP2 : APPRENTISSAGE INTERACTIF DE STRUCTURES MUSICALES

Responsable : E. Vincent, Inria

Partenaires impliqués (55 H.Mois) : Ircam : 16 (G. Assayag, G. Bloch, post-doc), Inria : 39 (E. Vincent, doctorant)

Durée : 36 mois de T0 à T0+36

Objectifs généraux de la tâche

Cette tâche, bâtie sur les fonctionnalités d'écoute artificielle de la tâche 1, vise à apprendre des modèles symboliques susceptibles de capturer les structures musicales multi-dimensionnelles et multi-échelles émergeant dans un certain contexte d'improvisation. Pour cela, nous proposons une approche nouvelle

consistant en l'intégration entre les modèles de séquences formelles (plus spécifiquement, l'oracle des facteurs), qui forment le coeur des systèmes de composition et d'improvisation automatique actuels, et l'approche bayésienne, qui a montré son intérêt pour la représentation de dépendances plus complexes et qui permet l'apprentissage à partir d'une faible quantité de données grâce au lissage des modèles.

- *Sous-tâche WP2.1 : Apprentissage de structures multi-dimensionnelles*

Responsable : E. Vincent, Inria

Partenaires impliqués (18 H.Mois) : Ircam : 5 (G. Assayag, G. Bloch), Inria : 13 (E. Vincent, doctorant)

Durée : 24 mois de T0 à T0+24

Description :

Cette tâche vise à concevoir des modèles et des algorithmes d'apprentissage capables de représenter les corrélations entre plusieurs dimensions musicales (durées, hauteurs, intensités, timbres, harmonie) en évitant l'explosion combinatoire inhérente à la mesure de ces corrélations. Ce problème étudié dans le cas des n-grammes [Raczyński2013a] est exacerbé avec l'oracle des facteurs en raison de la longueur accrue des sous-séquences modélisées. De plus, ces sous-séquences n'ont pas nécessairement la même longueur pour chaque dimension, ce qui rend la représentation des corrélations plus complexe et nécessite des heuristiques afin de contrôler le coût de calcul. On généralisera dans un premier temps la technique d'interpolation log-linéaire de modèles uni-dimensionnels [Raczyński2013a] au cas de l'oracle des facteurs, en cherchant une heuristique appropriée pour le parcours rapide du modèle ainsi appris. Dans un deuxième temps, on pourra s'intéresser à une représentation d'ordre plus élevé du tenseur des probabilités conditionnelles analogue à celle fournie par les modèles de Markov faiblement couplés [Saul1999] dans le cas des n-grammes.

Indicateurs de succès, risques et solutions de repli:

La nature des systèmes de composition et d'improvisation automatique nécessite une évaluation subjective par des humains. Le succès de l'algorithme développé sera jugé à la fois par des auditeurs experts et par des improvisateurs. Le risque principal porte sur la pertinence musicale des séquences générées. Le paramètre de lissage sera réglable, de sorte à réduire le cas échéant la probabilité des sous-séquences multi-dimensionnelles non observées dans le corpus d'apprentissage et à retomber dans le pire des cas sur le système actuel (qui ne génère que des sous-séquences multi-dimensionnelles observées dans le corpus d'apprentissage). L'évaluation automatique de la pertinence dans la tâche 2.2 devrait permettre in fine d'effectuer ce réglage de façon automatique.

Livrables

L2.1.1 (M12) : Apprentissage de structures multi-dimensionnelles, première version algorithme (rapport + maquette logicielle)

L2.1.2 (M24) : Apprentissage de structures multi-dimensionnelles, version finale algorithme (rapport + maquette logicielle)

- *Sous-tâche WP2.2: Sélection de dimensions et apprentissage par renforcement*

Responsable : E. Vincent, Inria

Partenaires impliqués (18 H.Mois) : Ircam : 5 (G. Assayag, G. Bloch, post-doc), Inria : 13 (Emmanuel Vincent, doctorant)

Durée : 24 mois de T0+12 à T0+36

Description :

Le message musical émis par un improvisateur est porté par une ou plusieurs dimensions "sémantiques". Par exemple, dans les styles traditionnels, la hauteur peut être porteuse d'information (ni trop déterministe ni trop aléatoire) alors que les timbres sont non informatifs (très déterministes ou très aléatoires). Dans une improvisation contemporaine ou dans certains styles des musiques du monde le timbre peut au contraire devenir prépondérant. La ou les dimensions sémantiques dépendent du style de l'improvisateur et elles varient au cours d'un même morceau. La capacité d'un improvisateur à interagir passe par sa capacité à identifier la ou les dimensions musicales sémantiques utilisées par les autres improvisateurs, à détecter l'innovation stylistique résultant du changement de dimension et à observer leurs réactions positives ou négatives à ses propres improvisations. Cette tâche vise à identifier automatiquement ces

dimensions et ces réactions et à modifier l'apprentissage dans une boucle perception / action. Pour cela, nous proposons d'utiliser la mesure de débit d'information de [Dubnov 2011] dans le but nouveau d'identifier les dimensions porteuses d'information. La réaction positive ou négative des autres agents sera évaluée par des indicateurs objectifs (l'évolution de leur quantité d'information, le nombre de sous-séquences potentiellement générables à un instant donné, etc). Un algorithme d'apprentissage par renforcement [Sutton1998] sera alors proposé afin d'augmenter ou de diminuer la probabilité des sous-séquences de façon adaptative en fonction des réactions générées.

Indicateurs de succès, risques et solutions de repli:

Le succès de l'algorithme développé sera mesuré par des auditeurs experts et par les indicateurs objectifs de réaction positive des autres improvisateurs. En raison de la complexité du matériau musical, il est possible que la réaction des autres improvisateurs ne soit pas mesurable objectivement de façon suffisamment précise. Dans ce cas, nous utiliserons un dispositif de réaction explicite (par exemple, une pédale) par lequel les improvisateurs humains pourront indiquer leur niveau de satisfaction, de façon à valider l'algorithme d'apprentissage par renforcement proposé.

Livrables

L2.2.1 (M24) : Sélection de dimensions et apprentissage par renforcement, première version algorithme (rapport + maquette logicielle)

L2.2.2 (M36) : Sélection de dimensions et apprentissage par renforcement, version finale algorithme (rapport + maquette logicielle)

- *Sous-tâche WP2.3: Apprentissage de structures multi-échelles*

Responsable : E. Vincent, Inria

Partenaires impliqués (19 H.Mois) : Ircam : 6 (G. Assayag, G. Bloch, post-doc), Inria : 13 (E. Vincent, doctorant)

Durée : 12 mois de T0+24 à T0+36

Description :

Une dernière tâche plus exploratoire porte sur la modélisation de la forme par une structure multi-échelle. Cette tâche vise à rendre compte des phénomènes cognitifs de plus haut niveau impliqués dans la mémorisation à long terme et la restitution innovante de phrases musicales dans l'improvisation musicale et la composition. Ce problème reste très difficile dans le cas général [Bimbot2014]. Dans le cadre du projet DYCI2, nous nous focaliserons sur le sous-problème de modélisation de la forme connaissant les frontières des blocs. Dans le cas du jazz, cela signifie par exemple être capable d'identifier la forme de chaque bloc ("aaba", "aabc" ou "abac") en connaissant sa durée (32 mesures) et de générer de nouvelles séquences suivant cette forme contrainte. L'objectif sera de développer une version hiérarchique à deux couches de l'oracle des facteurs, afin de modéliser à la fois les séquences "a", "b" ou "c" possibles et les transitions entre ces séquences. Un autre problème concerne l'estimation d'une forme (e.g. aa'bcacb) émergeant en improvisation libre en dehors du cas idiomatique (jazz, classique), qui peut être évaluée en termes de mesures informationnelles (débit d'information, densité de corrélation...) indépendamment des formes canoniques (aaba). Cette question rejoint celle des dimensions sémantiques de WP2.2 , qui peuvent constituer un indice de segmentation formelle supplémentaire.

Indicateurs de succès, risques et solutions de repli:

Comme précédemment, le succès de l'algorithme développé sera jugé par des auditeurs experts et des improvisateurs humains. En raison de la focalisation sur deux sous-problèmes plus simples, cette tâche ne présente pas de risque particulier.

Livrables

L2.3 (M36) : Apprentissage de structures multi-échelles (rapport + maquette logicielle)

4.2.4 TACHE WP3 : DYNAMIQUES D'INTERACTION IMPROVISEE

Responsable : G. Assayag, Ircam

Partenaires impliqués (55 H.Mois) : Ircam : 49 (G. Assayag, F. Bevilacqua, G. Bloch, J. Bresson, J. Nika, K. Sanlaville, post-doc), Inria : 3 (E. Vincent), UBO : 3 (S. Marchand)

Durée : 36 mois de T0 à T0+36

Objectifs généraux de la tâche

Cette tâche explore les conditions dans lesquelles peut se produire une interaction improvisée riche et créative entre agents artificiels et humains en examinant les questions de dynamiques collectives, d'adaptation aux différentes échelles temporelles, de structure cognitive interne permettant différents régimes de connaissance, et de guidage adaptatif fonction de l'écoute. Elle profite de l'écoute informée du WP1 et des capacités d'analyse et d'apprentissage du WP2.

- *Sous-tâche WP3.1 : Guidage de l'interaction improvisée*

Responsable : G. Assayag (Ircam)

Partenaires impliqués (17 H.Mois) : Ircam : 16 (G. Assayag, G. Bloch, J. Nika, post-doc), Inria : 1 (E. Vincent)

Durée : 24 mois de T0 à T0+24

Description :

Cette tâche étudie l'interaction improvisée d'un agent dans son aspect contraint et/ou planifié. On veut par exemple pouvoir improviser dans un contexte harmonique et rythmique explicite en exploitant les informations a priori structurant d'une part le corpus d'apprentissage et d'autre part le contexte d'improvisation supposé étiquetés par un vocabulaire symbolique commun. Le processus d'improvisation guidé par une séquence de référence sera modélisé ici comme l'articulation entre un *scénario* à suivre et une *mémoire* structurée et annotée à partir de laquelle on reconstituera dynamiquement des séquences musicales pour créer les nouvelles improvisations compatibles dans leur organisation temporelle avec la séquence imposée, selon les dimensions choisies, notamment harmoniques. Un exemple est la génération imposée de réalisations d'une grille harmonique A à partir d'un modèle entraîné sur une ou plusieurs réalisations de grilles harmoniques B, C, D etc.

Lorsqu'il n'y a pas de connaissance a priori du futur associé à un scénario, le processus génératif peut néanmoins être aiguillé de manière causale par un flux d'entrée, produit typiquement un musicien ou un autre agent numérique improvisant librement. Ce processus entretient alors une "synchronisation flottante" entre le modèle acquis en temps-réel du musicien et une mémoire structurée et annotée à partir de laquelle on reconstitue dynamiquement des séquences musicales cohérentes avec le flux d'entrée selon les dimensions choisies. Un exemple simple d'une telle interaction est l'accompagnement automatique d'une improvisation mélodique, un autre la génération de solos sur un enchaînement d'accord improvisés ; dans le cas général, un nombre arbitraire d'agents entraînés sur des corpus peuvent co-improviser en s'écoutant et s'adaptant souplement les uns aux autres sur les dimensions de hauteurs, de rythme, d'harmonie, comme le feraient des improvisateurs experts humains.

Modèles de scénarios et modèles de flux peuvent être combinés de manière à tirer simultanément parti de la capacité des premiers à gérer des plans d'improvisation et de celle des seconds à s'adapter avec souplesse aux contingences de l'écoute et de l'interaction, en tenant compte de l'organisation formelle ou des dimensions sémantiques caractéristiques du style étudiées aux WP2.3 et WP2.2.

Indicateurs de succès, risques et solutions de repli:

La conformité proprement dite des séquences engendrées aux scénarios ou aux flux imposés ne pose pas de risque particulier, à la différence de la pertinence et la créativité manifestées dans le choix des solutions. Celles-ci devront être jugées pour diverses configurations de paramètres par des auditeurs experts et des improvisateurs humains qui proposeront un classement susceptible de nous aider à pondérer les variables du modèle.

Livrables

L3.1.1 (M12) Guidage de l'interaction improvisée I , par scénario (rapport + maquette logicielle)
L3.1.2 (M24) Guidage de l'interaction improvisée II, par flux d'entrée (rapport + maquette logicielle)

- *Sous-tâche WP3.2: Adaptation temporelle de l'interaction*

Responsable : F. Bevilacqua (Ircam)

Partenaires impliqués (18 H.Mois) : Ircam : 17 (F. Bevilacqua, K. Sanlaville, post-doc), UBO : 1 (S. Marchand)

Durée : 24 mois de T0+6 à T0+30

Description :

L'objectif de cette tâche est de développer un système adaptatif et anticipatif de la gestion de l'interaction improvisée notamment dans le cas de collectivités d'agents impliquant humains et agents numériques. Il s'agit de dépasser les approches classiques statiques et déterminées préalablement et de pouvoir adapter en temps-réel les modèles, représentations et modalités d'apprentissage de l'interaction, en prenant en compte ses différentes échelles temporelles et les dynamiques collectives susceptibles d'y être attachées. Cela implique de reconnaître et adapter des phénomènes à court-terme (réactions, synchronisation) ainsi que des phénomènes à long-terme (vocabulaire, émergence de formes de plus haut niveau) en mettant à jour des représentations multi-échelles et hiérarchiques de l'interaction (impliquant des représentations cognitives, ou état mentaux). Ce thème articulera des connaissances, modèles et systèmes qui se développent pour l'instant dans des domaines séparés : apprentissage, agents conversationnels, interaction musicale improvisée. La formalisation d'un modèle cohérent d'interaction permettant de traiter ces différents domaines représente un défi qui nécessite le développement d'une approche globale de l'interaction permettant d'envisager de nombreuses autres applications. Cette tâche se focalisera sur divers aspects de l'« expressivité » et du « style » mise en œuvre dans l'interaction, à partir d'un modèle générique réemployable pour d'autres champs d'application.

Indicateurs de succès, risques et solutions de repli:

Un modèle générique de l'interaction décliné ensuite en domaines spécialisés tels que la musique et le langage risque d'être sous-optimal par rapport à des approches ad-hoc, ce qui se ressentirait sur la pertinence des résultats. Ce risque est compensé par la possibilité de mettre en place une architecture multi-niveaux dans laquelle les catégories les plus générales (turn taking, émergence d'organisation collectives, émotions) sont prises en charges par le système principal, et les catégories les plus spécifiques (synchronisation, adaptation, particularités idiomatiques des langages cibles) par des sous-systèmes ad-hoc.

Livrables

L3.2.1 (M18) Adaptation temporelle de l'interaction I (rapport + maquette logicielle)
L3.2.2 (M30) Adaptation temporelle de l'interaction II (rapport + maquette logicielle)

- *Sous-tâche WP3.3: Mémoire, connaissance et contrôle des agents créatifs*

Responsable : G. Assayag (Ircam)

Partenaires impliqués (20 H.Mois) : Ircam : 16 (G. Assayag, J. Bresson, K. Sanlaville, post-doc), Inria : 2 (E. Vincent), UBO : 2 (S. Marchand)

Durée : 27 mois de T0+9 à T0+36

Description :

Un agent artificiel créatif autonome capable d'écoute, d'apprentissage et d'interaction musicale improvisée avec d'autres agents humains et artificiels doit être équipé d'une structure "cognitive" située minimale lui permettant d'exercer ces fonctions dans un environnement complexe qui intègre ses propres productions, impliquant une certaine réflexivité sur son propre fonctionnement. Cette tâche vise à identifier les représentations et processus les mieux adaptés pour modéliser cette structure interne et son activation à partir des données de captation multi-dimensionnelles et multi-échelles des WP2.1 et WP2.3. Il s'agit de constituer une mémoire et un réseau de connaissances à différentes échelles de temps (mémoire échoïque, mémoire long-terme) et à différents niveaux d'activation et de contrôle, que ces derniers soient implicites (procéduraux ou réflexes) ou explicites (mémoire épisodique ou sémantique). Les structures couplées de mémoire et de contrôle, actuellement réalisées avec un oracle des facteurs, seront étendues notamment à l'aide de SOM (*self organizing maps*) pour la configuration automatique d'une topologie des objets musicaux acquis, contribuant à la formation d'une mémoire sémantique, et de procédures auto-réflexives de simulation des qualités d'éveil (curiosité déclenchée par un stimulus) d'attention (écouter ou pas les autres agents), de motivation (vouloir apprendre ou pas) et d'initiative (décider de jouer ou pas) en relation avec des modèles computationnels du soi et de l'intentionnalité (*self-model theory*).

Indicateurs de succès, risques et solutions de repli:

Ce module au cœur du dispositif bénéficie des résultats acquis au WP1 et WP2 en matière de perception artificielle et d'organisation interne des connaissances ce qui en limite le risque. Sa crédibilité est mise en jeu dans l'interaction avec des musiciens humains aussi bien dans l'apprentissage immédiat que dans l'exploitation des connaissances antérieures mais il a la possibilité d'appliquer « volontairement » des tactiques simples de repli (silence, discrétion, accompagnement simplifié) en cas de difficulté à atteindre son but.

Livrables

L3.3.1 (M24) Mémoire, connaissance et contrôle des agents créatifs I (rapport + maquette logicielle)
L3.3.2 (M36) Mémoire, connaissance et contrôle des agents créatifs II (rapport + maquette logicielle)

4.2.5 TACHE WP4 : INTEGRATION, EXPERIMENTATION, VALIDATION ET RETOUR D'USAGE

Responsable : G. Assayag (Ircam)

Partenaires impliqués (23 H.Mois) : Ircam : 13 (G. Assayag, G. Bloch, F. Bevilacqua, post-doc, ingénieur), Inria : 5 (E. Vincent, A. Liutkus, doctorant), UBO : 5 (S. Marchand, V. Koehl, M. Paquier, doctorant)

Durée : 36 mois de T0 à T0+36

Objectifs généraux de la tâche

Les défis scientifiques de ce projet visent les nouveaux besoins en matière de création et de performance musicale, ainsi que les nouveaux scénarios d'interaction créative des utilisateurs avec les objets multimédia, notamment les archives de performances vivantes. Les nouveaux outils proposés par les partenaires dans les WP1, 2 et 3 seront mis en jeu expérimentalement et confrontés aux musiciens experts tout au long du projet et alimenteront une plate-forme commune. Cette plateforme sera testée, étudiée sous l'angle des cas d'usage, et validée avec les utilisateurs finaux, c'est-à-dire les improvisateurs humains. Les résultats seront largement diffusés à la communauté scientifique, musicale, et au grand public.

- *Sous-tâche WP4.1 : Expérimentation, évaluation, dissémination*

Responsable : G. Assayag (Ircam)

Partenaires impliqués (13 H.Mois) : Ircam : 7 (G. Assayag, G. Bloch, F. Bevilacqua, post-doc, ingénieur), Inria : 3 (E. Vincent, A. Liutkus, doctorant), UBO : 3 (S. Marchand, V. Koehl, M. Paquier, doctorant)

Durée : 36 mois de T0+6à T0+36

Description :

Une première tâche consiste à expérimenter et évaluer les outils proposés avec des musiciens. L'Ircam est particulièrement bien placé étant labellisé "Centre de recherche musicale" et abritant deux formations supérieures (Master ATIAM et Cursus Jeunes Compositeurs), ayant la possibilité d'héberger des artistes en résidence dans un environnement art-sciences exceptionnel via la sélection par la plate-forme européenne Ulysse[3], de donner des performances publiques dans une salle de concert à acoustique variable unique au monde (l'espace de projection) et d'organiser des colloques internationaux. L'UBO héberge une formation supérieure aux métiers de l'image et du son reconnue, constituant elle aussi un milieu favorable pour l'expérimentation avec des musiciens professionnels et des étudiants. Enfin le centre MétaMédia de l'Ecole Polytechnique Fédérale de Lausanne (EPFL) dépendant de sa direction Innovation et Valorisation, dans le cadre de son partenariat avec l'Ircam, met à la disposition de ce projet la collection d'archives unique du Montreux Jazz Festival (MJF), inscrite au patrimoine documentaire mondial par l'Unesco[4] et intégralement digitalisée à l'EPFL. Les expérimentations comporteront extraction/analyse de structures musicales (en lien avec WP2), écoute, décomposition et recomposition créative (en lien avec WP1), nouveaux modes d'accès et re-création musicale interactive des archives (en lien avec le WP3) permettant à un musicien de "jouer" en interaction avec l'archive elle-même "ré-

[3] http://www.ulysses-network.eu

[4] http://www.unesco.org/new/fr/communication-and-information/flagship-project-activities/memory-of-the-world/register/full-list-of-registered-heritage/registered-heritage-page-8/the-montreux-jazz-festival-legacy/

improvisée" (en audio et remontage vidéo temps-réel) par des agents créatifs autonomes. Ces données sont immédiatement disponibles pour l'expérimentation.

Les résultats scientifiques et artistiques seront diffusés par l'organisation d'une manifestation internationale sur l'improvisation et les nouvelles technologies à l'Ircam et par l'organisation d'une session spéciale ou d'un atelier dédié dans l'une des grandes conférences du domaine (SMC, ICMC, DAFx, ACM Multimedia). Les résultats scientifiques seront par ailleurs publiés dans ces conférences ainsi que dans les meilleures revues du domaine (Computer Music Journal, Journal of New Music Research, IEEE/ACM Transactions on Audio, Speech, and Language Processing).

Indicateurs de succès, risques et solutions de repli:

Le succès sera mesuré par l'organisation effective des expérimentations et des événements scientifiques prévus. Cette tâche ne présente pas de risque particulier.

Livrables

L4.1.1 (M12) Expérimentation des WP1, 2, 3, sessions de studio Ircam et UBO (rapport) avec experts musiciens, premières expérimentations avec les archives du MJF / EPFL

L4.1.2 (M18) organisation à l'Ircam d'une manifestation internationale sur improvisation et nouvelles technologies (colloque)

L4.1.3 (M24) organisation d'un track ou d'un workshop dédié dans une des grandes conférences du domaine (SMC, ICMC, DAFx, ACM Multimedia) (workshop)

L4.1.4 (M24) Projets artiste en résidence à l'Ircam sur les nouvelles dynamiques d'interaction improvisée, deuxième train d'expérimentations avec les archives du MJF/EPFL (rapports)

L4.1.5 (M 30) Maquette logicielle de recréation improvisée et interactive d'archives du MJF/EPFL, incluant écoute créative, extraction, analyse des données et interaction improvisée avec l'utilisateur (rapport et logiciel)

- *Sous-tâche WP4.2 : Intégration suite logicielle*

Responsable : G. Assayag (Ircam)

Partenaires impliqués (10 H.Mois) : Ircam : 6 (G. Assayag, G. Bloch, post-doc, ingénieur), Inria : 2 (E. Vincent, A. Liutkus, doctorant), UBO : 2 (S. Marchand, V. Koehl, M. Paquier, doctorant)

Durée : 9 mois de T0+27 à T0+36

Description :

Les outils développés par les partenaires alimenteront une plate-forme expérimentale commune qui intègrera les différents modules logiciels d'écoute, de modélisation des structures musicales et d'interaction générative. L'utilisation partagée du langage Python et de la plate-forme d'expérimentation temps-réel Max/MSP et/ou Pure Data ainsi que la définition d'une API commune entre tous, convenues dès les premières réunions de coordination du WP0 et réactualisées tout au long du projet de manière à permettre l'échange continu de modules de code entre les partenaires, faciliteront les échanges. Une phase finale d'intégration permettra de packager ces modules afin de les distribuer sous licence LGPL.

Indicateurs de succès, risques et solutions de repli:

Le succès sera mesuré par la mise à disposition des différents modules développés. La nature modulaire des tâches et l'utilisation d'une API commune limitent les risques en cas de non-remise d'un module: ce module sera alors remplacé par un module existant, en particulier dans le cas où un traitement expérimental n'est pas encore disponible en temps-réel. Les résultats seront évalués dans différents scénarios incluant des processus d'interaction temps-réel et des phases de traitement hors-ligne (analyses de corpus).

Livrables

L4.2 (M36) Intégration suite logicielle des outils de WP1, 2 et 3 (logiciel)

4.3. CALENDRIER DES TACHES, LIVRABLES ET JALONS

4.3.1 PLANNING DES TACHES

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

4.3.2 RECAPITULATIF DE L'EFFORT EN H.MOIS

	Total H.Mois	**WP0**	**WP1**				**WP2**				**WP3**				**WP4**
				1.1	1.2	1.3		2.1	2.2	2.3		3.1	3.2	3.3	
Ircam	**84**	**2**	4	0	3	1	16	5	5	6	**49**	16	17	16	**13**
Inria	**54**	1	6	2	2	2	**39**	13	13	13	3	1	0	2	5
UBO	**54**	1	**45**	16	14	15	0	0	0	0	3	0	1	2	5
Total DYCI2	***192***	***4***	***55***	*18*	*19*	*18*	***55***	*18*	*18*	*19*	***55***	*17*	*18*	*20*	***23***

4.3.3 LISTE DES LIVRABLES

Échéance	Livrable	Tâche	Description
M3	L0.1	WP0	Mise en place du Wiki du projet
M12	L0.2	WP0	Rapport d'avancement annuel
	L1.1	WP1	Séparation informée multi-canal (rapport + maquette)
	L2.1.1	WP2	Apprentissage de structures multi-dimensionnelles, algorithme v1 (rapport + maquette)
	L3.1.1	WP3	Guidage de l'interaction improvisée I par scénario (rapport et maquette)
	L4.1.1	WP4	Expérimentations WP1,2,3 avec experts musiciens, premières expérimentations avec les archives du MJF / EPFL (rapport)
M18	L3.2.1	WP3	Adaptation temporelle de l'interaction I (rapport et maquette)
	L4.1.2	WP4	Colloque International Improvisation et nouvelles technologies (Colloque)
M24	L0.3	WP0	Rapport d'avancement annuel
	L1.2	WP1	Ecoute structurante (rapport + maquette)
	L2.1.2	WP2	Apprentissage de structures multi-dimensionnelles, vers. finale algorithme (rapport + maquette)
	L2.2.1	WP2	Sélection de dimensions et apprentissage par renforcement, algorithme v1 (rapport + maquette)
	L3.1.2	WP3	Guidage de l'interaction improvisée II par flux d'entrée (rapport et maquette)
	L3.3.1	WP3	Mémoire, connaissance et contrôle des agents créatifs I (rapport et maquette)
	L4.1.3	WP4	Track/Workshop dans une conférence internationale (Workshop)
	L4.1.4	WP4	Artiste en résidence, deuxième train d'expérimentations avec les archives du MJF/EPFL (rapports)
M30	L3.2.2	WP3	Adaptation temporelle de l'interaction II (rapport et maquette)
	L4.1.5	WP4	Recréation improvisée et interactive d'archives du MJF/EPFL (rapport et logiciel)
M36	L0.4	WP0	Rapport final
	L1.3	WP1	Décomposition / recomposition par imitation (rapport + maquette)
	L2.2.2	WP2	Sélection de dimensions et apprentissage par renforcement, vers. finale algorithme (rapport + maquette)
	L2.3	WP2	Apprentissage de structures multi-échelles, algorithme (rapport + maquettet)
	L3.3.2	WP3	Mémoire, connaissance et contrôle des agents créatifs II (rapport et maquette)
	L4.2	WP4	Intégration suite logicielle des outils de WP1,2 et 3 (logiciel)

4.4. JUSTIFICATION SCIENTIFIQUE ET TECHNIQUE DES MOYENS DEMANDES

4.4.1 IRCAM

Montant global de l'aide demandée 200 428€

Personnel

La totalité du personnel impliqué dans le projet pour ce partenaire représente 84 h.m. dont 14 h.m chercheurs propres Ircam, 16 h.m. chercheur associé Ircam, 30 h.m. doctorants Ircam (sur contrat doctoral de l'UPMC), ainsi que *18 h.m. post-doc et 6 h.m. ingénieur à recruter sur le projet*. Le cofinancement en coût complet est demandé pour les 14 h.m. chercheurs propres Ircam et les 24 h.m. post-doc et ingénieur à recruter.

Le personnel permanent impliqué dans le projet représente 30 H.Mois répartis comme suit

- Gérard Assayag DR Ircam 11 H.Mois
- Frédéric Bevilacqua DR Ircam 2 H.Mois
- Jean Bresson CR Ircam 1 H.Mois
- Georges Bloch, MCF U. Strasbourg, Chercheur associé Ircam, 16 H.Mois

Le personnel non-permanent sans co-financement ANR demandé (doctorants) représente 30 H.Mois répartis comme suit:

- Jérôme Nika, doctorant, 12 H.Mois (2012-2015 sur contrat doctoral UPMC, co-direction Gérard Assayag et Marc Chemillier EHESS) travaillera principalement sur WP3.1
- Kevin Sanlaville, doctorant 18 H.Mois (2013-2016, co-direction F. Bevilacqua, Ircam, et C. Pelachaud, Telecom ParisTech, dans le Labex SMART, contrat doctoral UPMC) contribue au WP3.2., WP3.1.

Le personnel non-permanent pour lequel un co-financement est demandé représente 24 H.Mois répartis comme suit :

- Un post-doc sur 18 H.Mois qui travaillera principalement sur les tâches WP3.3 et WP4 (structure interne de mémoire, de connaissance et de de contrôle des agents créatifs, intégration des résultats de WP2 et WP3, expérimentation) avec des contributions en WP2.2, 2.3 et 3.2.
- Un ingénieur d'étude pour 6 H.Mois qui travaillera principalement sur la tâche WP4 (intégration logicielle, expérimentation)

Note : l'aide de co-financement demandée en cdd pour tout le projet *hors doctorants* porte sur 24 H.M Le total en H.M affecté au projet *hors doctorants* est de 54 H.Mois pour l'Ircam, 18 H.Mois pour Inria et 18 H.Mois pour UBO, soit 90 H.Mois pour la totalité du projet. *La proportion de 24 / 90 = 26,66% (< 30%) respecte donc les recommandations de l'ANR.*

Prestations de service externe : 15.000€ pour les frais d'organisation colloques et workshops et la mise en place des résidences de recherche à l'Ircam et des sessions d'expérimentations sous forme de prestations.

Missions :15.000€. Un nombre significatif de missions est prévu pour les actions d'expérimentation et de valorisation des résultats sur les archives de Montreux à l'EPFL, pour la coordination des partenaires, pour la dissémination internationale.

4.4.2 UBO

Montant global de l'aide demandée = 149968€ (dont 5768€ de frais de structure)

Equipement : 1 station de travail audio puissante + stockage + licence MaxMSP 4000€

Personnel

Personnel permanent impliqué dans le projet (18 H.Mois):

- Sylvain Marchand PR UBO 12 H.Mois
- Vincent Koehl MCF UBO 3 H.Mois
- Mathieu Paquier MCF UBO 3 H.Mois

Etant donné les besoins pour le projet, la demande en personnel auprès de l'ANR est d'une bourse de thèse financée à 100% sur 3 ans (115 200€). Le profil de ce poste est le suivant :

Sujet de thèse : Écoute informée créative

La thèse portera sur la tâche 1 du projet détaillée ci-dessus, en commençant par la généralisation de l'approche informée pour la séparation de sources à d'autres canaux d'information, pour poursuivre par la prise en compte de la dimension temporelle et finalement des informations de moins en moins exactes, afin de s'adapter au jeu improvisé. Elle sera encadrée par Sylvain Marchand (UBO).

Missions : 25000€ décomposés comme suit

- déplacements nationaux pour réunions d'avancement trimestrielles (6000€)
- séjours doctorant à Paris pour travail collectif à Paris et à Nancy (4000€)
- présentation des travaux à 6 conférences internationales (15000€)

4.4.3 INRIA

Montant global de l'aide demandée = 149604€ (dont 5754€ de frais de structure)

Equipement : 1 station de travail audio + licence MaxMSP 2850€

Personnel

Personnel permanent impliqué dans le projet (18 H.Mois):

- Emmanuel Vincent CR1 Inria 9 H.Mois
- Antoine Liutkus CR2 Inria 9 H.Mois

Etant donné les besoins pour le projet, la demande en personnel auprès de l'ANR est d'une bourse de thèse financée à 100% sur 3 ans (115200€). Le profil de ce poste est le suivant :

Sujet de thèse : Apprentissage de structures musicales en situation d'improvisation

La thèse portera sur la tâche 2 du projet détaillée ci-dessus, en commençant par l'apprentissage de structures multi-dimensionnelles pour poursuivre par l'apprentissage guidé par l'improvisation et l'apprentissage de structures multi-échelles. Elle sera co-encadrée par Emmanuel Vincent (Inria) et Gérard Assayag (Ircam).

Missions : 25800€ décomposés comme suit

- déplacements à Paris pour réunions d'avancement (4800€)
- séjours longs du doctorant à Paris pour co-encadrement, 1 mois par an (6000€)
- présentation des travaux à 6 conférences internationales (15000€)

5. STRATEGIE DE VALORISATION, DE PROTECTION ET D'EXPLOITATION DES RESULTATS, IMPACT GLOBAL DE LA PROPOSITION

La valorisation scientifique, artistique et sociétale de ce projet fait l'objet des livrables de la tâche WP4, incluant l'organisation de colloques internationaux et de tracks et workshops dans les grandes conférences du domaine, l'accueil de jeunes artistes de niveau international en résidence dans plusieurs des centres associés au projet, de séminaires et de cours régulièrement organisés dans les formations hébergées par les partenaires (Master ATIAM à l'Ircam, filière image & son de l'UBO).

Les prototypes expérimentaux issus du projets seront particulièrement mis à l'épreuve dans le cadre du projet de valorisation des archives du Montreux Jazz Festival (MJF) gérées par l'EPFL au MetaMedia Center[5]. Ces archives ont été captées dans des formats de toutes sortes selon les époques (près de 50 années), de la stéréo au multicanal pour l'audio, du VHS au 4K pour la vidéo. Cette collaboration ouvre des possibilités de travail sur place avec les étudiants et les musiciens (workshops du festival, nouveaux lieux dédiés de l'EPFL pour l'écoute active et l'expérimentation autour des archives). Cette collection à laquelle nous pourrons accéder librement dans un cadre de recherche représente 5,000 heures d'enregistrement audio-vidéo provenant de 4,000 concerts. C'est la plus grande collection systématique au monde d'enregistrements de concerts, classée au patrimoine mondial par l'Unesco.

Les résultats attendus des expérimentations autour du MJF/EPFL auront une très grande visibilité internationale, le MetaMedia Center ayant pour vocation la mise en valeur pratique des résultats scientifiques récents dans de nouveaux usages autour des objets multimedia. Ce partenariat culturel,

[5] http://metamedia.epfl.ch/

scientifique et technique aura une incidence forte sur les modes de valorisation créative des patrimoines audiovisuels musicaux et participera à la création d'activités nouvelles technico-artistiques autour de ces derniers dans l'esprit des recommandations européennes pour le programme cadre H2020.

Les retombées potentielles de ce projet sont riches et multiformes. Les partenaires fournissent une combinaison idéale pour un tel sujet dans une combinaison thématique qui n'a pas d'équivalent ailleurs en Europe. Les avancées scientifiques de DYCI2 auront donc un impact important et seront susceptibles de nourrir et de stimuler les recherches des partenaires et des autres équipes de la communauté pendant plusieurs années. En effet plusieurs des thématiques qui le structurent sont nouvelles ou constituent un regard nouveau, tout en puisant dans des compétences et des recherches antérieures des partenaires qui en limitent les risques. En particulier, l'écoute informée créative dépasse le cadre classique de la séparation des sources en intégrant à la fois la composante active et la connaissance musicale et ouvre sur des technologies inédites de re-création susceptibles de bouleverser l'accès à la musique enregistrée tel que nous le connaissons, en impliquant l'auditeur dans un processus créatif. Les résultats attendus en modélisation des comportements d'agents créatifs autonomes dans des dynamiques d'interaction complexe, associés aux facultés d'analyse fine et structurelle des flux audio-musicaux par l'écoute artificielle, la modélisation et l'apprentissage, ouvriront des avenues nouvelles en créativité artificielle et en intelligence artificielle embarquée. L'articulation des trois thèmes portés par les partenaires en un projet cohérent convergeant vers la constitution d'agents autonomes créatifs capables d'interaction improvisée est résolument originale et en mesure de marquer durablement l'état de l'art.

DYCI2 est un projet de recherche exploratoire, s'appuyant sur des outils dont tous ne sont pas en open source et sous licence LGPL. Les partenaires conviennent de mettre à disposition des autres partenaires ces outils dans le cadre du projet. Les archives MJF sont mises à disposition par l'EPFL dans le cadre de recherche uniquement, et en démonstration publique dans des circonstances définies pour ce qui concerne les droits des musiciens (dans l'enceinte du festival, dans celle des aménagements dédiés de l'EPFL, et, pour les concerts récents, dans l'enceinte des MJF cafés sous licence comme il en existe maintenant dans le monde entier et notamment à Paris).

Tous les développements logiciels prévus pour permettre l'interconnexion et l'interopérabilité des outils seront organisés, autant que possible, sous forme de composants logiciels modulaires, qui seront distribués sous licence LGPL, de façon à faciliter la réutilisation de ces technologies.

Les partenaires s'engagent à signer un accord de consortium précisant ces points avant la fin de la première année.

6. REFERENCES BIBLIOGRAPHIQUES

Les auteurs membres des partenaires de ce projet sont soulignés.

6.1. ECOUTE INFORMEE ET CREATION DE CORPUS D'APPRENTISSAGE

[Bregman 1990] Bregman, A. Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press, 1990.

[Clifford and Reiss 2011] Clifford, A., and Reiss, J. Microphone interference reduction in live sound. In Proceedings of the International Conference on Digital Audio Effects (DAFx), 2011.

[Comon and Jutten 2010] Comon, P. and Jutten, C., editors. Handbook of Blind Source Separation: Independent Component Analysis and Blind Deconvolution. Academic Press. 2010.

[Duong *et al.* 2010] Duong, N., Vincent, E., and Gribonval, R. Under-determined reverberant audio source separation using a full-rank spatial covariance model. IEEE Transactions on Audio, Speech, and Language Processing, 18(7):1830-1840, 2010.

[Ewert and Müller 2012] Ewert, S. and Müller, M. Using score-informed constraints for NMF-based source separation. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Kyoto, Japan, 2012.

[Fourer 2013] Fourer, D. Approche informée appliquée à l'analyse du son et de la musique. Thèse de doctorat de l'Université Bordeaux 1, décembre 2013.

[Gerber *et al.* 2012] Gerber, T., Dutasta, M., Girin, L., and Févotte, C. Professionally-produced music separation guided by covers. In Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR), 2012.

[Kokkinis and Mourjopoulos 2010] Kokkinis, E. and Mourjopoulos, J. Unmixing acoustic sources in real reverberant environments for close-microphone applications. Journal of the Audio Engineering Society, 58(11):907-922, 2010.

[Liutkus *et al.* 2013] Liutkus, A., Durrieu, J., Richard, G., and Daudet, L. An overview in informed audio source separation. In Proceedings of the 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), 2013.

[Liutkus *et al.*, 2014] Liutkus, A., Rafii, Z., Pardo, B., Fitzgerald, D., and Daudet, L. Kernel spectrogram models for source separation. In HSCMA, Nancy, France, 2014.

[Marchand *et al.* 2012] Marchand, S., Badeau, R., Baras, C., Daudet, L., Fourer, D., Girin, L., Gorlow, S., Liutkus, A., Pinel, J., Richard, G., Sturmel, N., and Zhang, S. DReaM: a novel system for joint source separation and multi-track coding. 133rd Audio Engineering Society (AES) Convention, San Francisco, California, USA, October 2012.

[McAdams 1989] McAdams, S. Segregation of concurrent sounds: effects of frequency modulation coherence. Journal of the Acoustical Society of America, 86(6):2148–2159, 1989.

[Mouba and Marchand 2006] Mouba, J., and Marchand, S. A source localization / separation / respatialization system based on unsupervised classification of interaural cues. In Proceedings of the International Conference on Digital Audio Effects (DAFx), pages 233-238, Montreal, Quebec, Canada, September 2006.

[Müller 2007] Müller, M. Information Retrieval for Music and Motion, pages 69-84, Springer, 2007.

6.2. MODELISATION DE SEQUENCES MUSICALES POUR L'INTERACTION

[Assayag 1999] G. Assayag, S. Dubnov, and O. Delerue. Guessing the composer's mind: applying universal prediction to musical style. In Proceedings of the International Computer Music Conference, Bejing, China, 1999. International Computer Music Association.

[Assayag 2004] G. Assayag, S. Dubnov « Using Factor Oracles for machine Improvisation », Soft Computing, vol. 8, n° 9, Septembre, 2004

[Assayag 2007] Assayag, G. , Bloch, G. « Navigating the Oracle: a Heuristic Approach », Proc. ICMC'07, The In. Comp. Music Association, Copenhagen 2007.

[Assayag 2010] Assayag, G., Bloch, G., Dubnov, S., Cont, A., Interaction with Machine Improvisation, in The Structure of Style, Springer Verlag, Eds: Kevin Burns, Shlomo Argamon, Shlomo Dubnov (Eds), pp. 219-246, 2010

[Blackwell 2012] Tim Blackwell, Oliver Bown & Michael Young. Live Algorithms: Towards Autonomous Computer Improvisers. In Jon McCormack & Mark d'Inverno, editeurs, Computers and Creativity, pages 147–174. Springer Berlin Heidelberg, 2012.

[Dubnov 2011] Dubnov, S., Assayag, G., Cont, A., « Audio Oracle Analysis of Musical Information Rate », Proceedings of IEEE Semantic Computing Conference, ICSC2011, Palo Alto, CA, 2011, pp. 567-571

[Canonne 2011] Canonne, C., Garnier, N., « A Model for Collective Free Improvisation », Mathematics and Computation in Music. Third International Conference MCM 2011, IRCAM, Paris, France, June 15-17, 2011. Proceedings, Springer, 2011.

[Conklin 2003] Conklin, D. Music Generation from Statistical Models , Proceedings of the AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences, Aberystwyth, Wales, 30– 35, 2003.

[Crochemore 2007] M. Crochemore, Lucian Ilie, Emine Seid-Hilmi - The Structure of Factor Oracles, Int. J. Found. Comput. Sci. 18(4), 781–797 (2007).

[Dubnov 2003] S. Dubnov, G. Assayag, O. Lartillot, G. Bejerano « Using Machine-Learning Methods for Musical Style Modeling », IEEE Computer, vol. 10, n° 38, Octobre, 2003

[Feder 1992] M. Feder, N. Merhav, and M. Gutman, "Universal Prediction of Individual Sequences," IEEE Trans. Information Theory, vol. 38, 1992, pp. 1258-1270.

[Lévy 2012] Lévy, B., Bloch, G., Assayag, G., « OMaxist Dialectics : capturing, Visualizing and Expanding Improvisations », NIME 2012, Ann Arbor, 2012, pp. 137-140

[Levy 2013] Lévy, B., « Principles and Architectures for an Interactive and Agnostic Music Improvisation System », Thèse de doctorat de l'UPMC, Ircam, Equipe Représentations Musicales, 2013

[Maniatakos12] Maniatakos, F., « Graphs and Automata for the Control of Interaction in Computer Music Improvisation. », Thèse de Doctorat de l'Université Pierre et Marie Curie, Ircam, Equipe Représentations Musicales, 2012

[Moreira 2013] Moreira, J., Roy, P., Pachet, F. VirtualBand: Interacting with Stylistically Consistent Agents.ISMIR, pages 341-346, Curitiba (Brazil), 2013

[Pachet 2011] Pachet, F. and Roy, P. Markov constraints: steerable generation of Markov sequences. Constraints, 16(2):148-172, March 2011

[Pachet 2012] F. Pachet. Musical virtuosity and creativity. *Computers & Creativity*, 2012.

[Surges 2013] Surges, G. and Dubnov, S. "Feature Selection and Composition using PyOracle." Workshop on Musical Metacreation, Ninth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. Northeastern University, Boston, MA. October 14-15, 2013.

6.3. MODELISATION BAYESIENNE ET ANALYSE DE MUSIQUE

[Bimbot 2010] F. Bimbot, E. Deruty, G. Sargent and E. Vincent. Semiotic structure labeling of music pieces: Concepts, methods and annotation conventions. *Proc. ISMIR*, 2012.

[Bimbot 2014] F. Bimbot, G. Sargent, E. Deruty, C. Guichaoua, E. Vincent. Semiotic description of music structure: An introduction to the Quaero/Metiss structural annotations. *Proc. AES 53rd Int. Conf. on Semantic Audio*, 2014.

[Blei 2003] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 2003.

[Chen 1996] S. Chen and J. Goodman. An empirical study of smoothing techniques for language modeling. *Proc. ACL*, 1996.

[Conklin 2013] D. Conklin. Multiple viewpoint systems for music classification. *Journal of New Music Research*, Volume 42, Number 1, pp. 19-26, March 2013

[David 1998] N. David. *Jazz arranging*, Ardsley House, 1998.

[Klakow 1998] D. Klakow. Log-linear interpolation of language models. *Proc. ICSLP*, 1998.

[Lee 2007] K. Lee. A system for automatic chord transcription using genre-specific hidden Markov models. *Proc. AMR*, 2007.

[Mauch 2010] M. Mauch and S. Dixon. Simultaneous estimation of chords and musical context from audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 2010.

[Paiement 2005] J.-F. Paiement, D. Eck and S. Bengio. A probabilistic model for chord progressions. *Proc. ISMIR*, 2005.

[Paulus 2008] J. Paulus and A. Klapuri. Music structure analysis using a probabilistic fitness measure and an integrated musicological model. *Proc. ISMIR*, 2008.

[Raczyński 2013a] S.A. Raczyński, E. Vincent and S. Sagayama. Dynamic Bayesian networks for symbolic polyphonic pitch modeling. *IEEE Transactions on Audio, Speech and Language Processing*, 2013.

[Raczyński 2013b] S.A. Raczyński, S. Fukayama and E. Vincent. Melody harmonisation with interpolated probabilistic models. *Journal of New Music Research*, 2013.

[Raczyński 2014] S.A. Raczyński and E. Vincent. Genre-based music language modelling with latent hierarchical Pitman-Yor process allocation. *IEEE Transactions on Audio, Speech and Language Processing*, to appear.

[Raphael 2004] C. Raphael and J. Stoddard. Harmonic analysis with probabilistic graphical models. *Computer Music Journal*, 2004.

[Ryynänen 2008] M. P. Ryynänen and A. P. Klapuri. Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal*, 2008.

[Saul 1999] L.K. Saul and M.I. Jordan. Mixed memory Markov models: Decomposing complex stochastic processes as mixtures of simpler ones. *Machine Learning*, 1999.

[Scholz 2009] R. Scholz, E. Vincent and F. Bimbot. Robust modeling of musical chord sequences using probabilistic N-grams. *Proc. ICASSP*, 2009.

[Vincent 2010] E. Vincent, S. Raczyński, N. Ono and S. Sagayama. A roadmap towards versatile MIR. *Proc. ISMIR, special session "The future of music information retrieval"*, 2010.

[Sutton 1998] R.S. Sutton and A.G. Barto. *Reinforcement learning: An introduction*. MIT Press, 1998.

[Teh 2006] Y.-W. Teh. A hierarchical Bayesian language model based on Pitman-Yor processes. *Proc. ACL*, 2006.

6.4. ADAPTATION TEMPORELLE, DYNAMIQUE ET LOGIQUE DE L'INTERACTION

[Allwood 1976] ALLWOOD, Jens. Linguistic communication as action and cooperation. Gothenburg monographs in linguistics, 1976, vol. 2, p. 637-663.

[Chemillier 2004] M. Chemillier, Toward a formal study of jazz chord sequences generated by Steedman's grammar. Soft Computing, Vol. 8, No. 9, pp. 617–622, 2004.

[Donze 2013] A. Donze, S. Libkind, S.A. Seshia, D. Wessel, Control improvisation with application to music. Technical report No. UCB/EECS-2013-183. EECS Department, University of California, Berkeley, 2013.

[Fujie 2004] FUJIE, Shinya, YAGI, Daizo, MATSUSAKA, Yosuke, et al. Spoken dialogue system using prosody as para-linguistic information. In : Speech Prosody 2004, International Conference. 2004.

[Ishi 2014] ISHI, Carlos Toshinori, ISHIGURO, Hiroshi, et HAGITA, Norihiro. Analysis of relationship between head motion events and speech in dialogue conversations. Speech Communication, 2014, vol. 57, p. 233-243.

[Knapp 2013] KNAPP, M., HALL, J., Nonverbal Communication in human interaction, Cengage Learning, 2013.

[Kopp 2010] KOPP, Stefan. Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. Speech Communication, 2010, vol. 52, no 6, p. 587-597.

[Nika 2012] J. Nika, M. Chemillier, "ImproteK, integrating harmonic controls into improvisation in the filiation of OMax," in Proceedings of the International Computer Music Conference, pp. 180–187, 2012.

[Novielli 2010] NOVIELLI, Nicole, DE ROSIS, Fiorella, et MAZZOTTA, Irene. User attitude towards an embodied conversational agent: Effects of the interaction mode. Journal of Pragmatics, 2010, vol. 42, no 9, p. 2385-2397.

[Pressing 1984] J. Pressing, Cognitive processes in improvisation. Advances in Psychology, Vol. 19, pp. 345–363, 1984.

[Steels 2003] STEELS, Luc. Evolving grounded communication for robots. Trends in cognitive sciences, 2003, vol. 7, no 7, p. 308-312.

[Thórisson 2002] THÓRISSON, Kristinn R. Natural turn-taking needs no manual: Computational theory and model, from perception to action. In : Multimodality in language and speech systems. Springer Netherlands, 2002. p. 173-207.