

PROGRAMME INTITUTS ET INITIATIVES

Appel à projet – campagne 2021

Proposition de projet de recherche doctoral (PRD)

SCAI - Sorbonne Center of Artificial Intelligence

Intitulé du projet de recherche doctoral (PRD): **Ecoute musicale artificielle : une approche unifiée IA symbolique / deep learning pour passer d'une écoute locale et réactive à une écoute globale et narrative**

Directeur.rice de thèse porteur.euse du projet (titulaire d'une HDR) :

NOM : **ASSAYAG** **Prénom :** **Gérard**
Titre : Directeur de Recherche ou
e-mail : Gerard.Assayag@ircam.fr
Adresse professionnelle : IRCAM, 1, place Igor Stravinsky 75004 Paris
(site, adresse, bât., bureau)

Unité de Recherche :

Intitulé : STMS
Code (ex. UMR xxxx) : UMR 9912

École Doctorale de rattachement de l'équipe (future école doctorale du/de la doctorant.e) : **ED130-EDITE**

Doctorant.e.s actuellement encadré.e.s par la.e directeur.rice de thèse (préciser le nombre de doctorant.e.s, leur année de 1^{er} inscription et la quotité d'encadrement) : 2 (2018 et 2020).

Nicolas OBIN passera son HDR prochainement avec une soutenance prévue fin 2021-début 2022, et reprendra la direction de la thèse à ce moment là.

Co-encadrant.e :

NOM : **NIKA** **Prénom :** **Jérôme**
Titre : Chargé de Recherche ou HDR
e-mail : Jerome.Nika@ircam.fr

Unité de Recherche :

Intitulé : STMS
Code (ex. UMR xxxx) : UMR 9912

École Doctorale de rattachement : **ED130-EDITE**
Ou si ED non Alliance SU :



Co-encadrant.e :

NOM : **OBIN** Prénom : **Nicolas**
Titre : Maître de Conférences des Universitésou HDR

e-mail : Nicolas.Obin@ircam.fr

Unité de Recherche :

Intitulé : STMS
Code (ex. UMR xxxx) : UMR 9912

ED130-EDITE

École Doctorale de rattachement : Ou si ED non Alliance SU :

Doctorant.e.s actuellement encadré.e.s par la.e co-directeur.rice de thèse (préciser le nombre de doctorant.e.s, leur année de 1^e inscription et la quotité d'encadrement) :3 (2017 - soutenance à venir, 2019, 2019).

Cotutelle internationale : Non Oui, précisez Pays et Université :

Selon vous, ce projet est-il susceptible d'intéresser une autre Initiative ou un autre Institut ?

Non Oui, précisez CM - Collegium Musicae

Description du projet de recherche doctoral (*en français ou en anglais*) :

Ce texte sera diffusé en ligne : il ne doit pas excéder 3 pages et est écrit en interligne simple.

Détailler le contexte, l'objectif scientifique, la justification de l'approche scientifique ainsi que l'adéquation à l'initiative/l'Institut.

Le cas échéant, préciser le rôle de chaque encadrant ainsi que les compétences scientifiques apportées. Indiquer les publications/productions des encadrants en lien avec le projet.

Préciser le profil d'étudiant(e) recherché.

Contexte

Les développements récents des recherches portant sur la co-créativité humain-machine ont donné naissance à de nombreux modèles d'agents génératifs capables de s'intégrer dans la production d'un discours musical collectif dans le périmètre d'un « rôle » établi (un exemple trivial étant par exemple « soliste » ou « accompagnateur »). La définition des mécanismes régissant la contribution d'un agent à ce discours associent :

- Une mémoire musicale (un modèle génératif construit sur une base de donnée musicale)
- Une perception (conversion acoustique-symbolique) d'un flux musical opéré par une machine d'écoute.
- Un comportement établi d'action/réaction (traitement symbolique-symbolique, et conversion symbolique-acoustique) traduisant des spécifications venant d'un utilisateur ou d'un module de perception en "intentions" : des requêtes au modèle de mémoire musicale dont l'exécution génère les contributions de l'agent au discours musical.

Les machines d'écoute permettent d'enrichir ou d'explorer la mémoire musicale de l'agent en temps réel, en relation à une structure formelle pré-établie permettant de développer des scénarios, à une influence « événementielle » issue du contexte musical dynamique, ou par la construction d'anticipations à moyen-terme, à partir de l'extraction de représentations acoustiques simples (hauteurs de note, « timbre », etc.). Le défi à relever réside désormais dans la réalisation d'une écoute à l'échelle "macro" de la dimension « narrative » d'un flux audio musical dans ses évolutions formelles ou texturales, ce qui permettrait à des agents génératifs de déployer des processus de décisions plus évolués, et de soutenir un discours musical sur le long-terme.

Par ailleurs, le « deep learning » (l'apprentissage par réseaux de neurones profonds) a permis des avancées extrêmement spectaculaires en traitement du signal audio musical, et notamment dans le domaine du MIR (Music Information Retrieval) : depuis le suivi de hauteurs dominante ou la



séparation des pistes instrumentales dans un flux audio polyphonique jusqu'à la transcription musicale automatique, l'analyse de structure musicale, ou la génération de musique. De manière générale, le deep learning permet d'extraire à partir du flux audio musical une représentation latente de l'information acoustique optimisée en fonction des tâches à accomplir (transcription, séparation, etc...) . En particulier, le « champ réceptif » d'un réseau de neurones convolutionnels permet de représenter l'information acoustique successivement sur différents horizons temporels, et les réseaux de convolutions dilatées permettent désormais d'apprendre efficacement des dépendances temporelles à long-terme. Si les exploitations du deep learning dans les machines d'écoute réactives ont pour l'heure uniquement abordé le champ de l'extraction et de la prédiction locale de structures formelles sous-jacentes , elles semblent tout indiquées relever le défi exposé plus tôt.

Objectif

Ce sujet de thèse transdisciplinaire se situe au carrefour de l'intelligence artificielle, du traitement du signal et de l'information, de la créativité computationnelle avec comme applications la génération automatique de musique et la musicologie computationnelle.

Ce projet cherchera à créer une machine d'écoute réactive capable de concilier les deux aspects de l'intelligence artificielle afin de générer un contenu musical cohérent et adapté à une performance musicale en temps-réel et guidé par un scénario musical pré-établi. D'une part, l'intelligence artificielle symbolique (comportement symbolique-symbolique et réaction symbolique-acoustique) sous la forme d'agents génératifs agissant à partir de représentations symbolique de la musique et intégrant des connaissances musicales a priori et des contraintes sur l'évolution du discours musical telles qu'elles sont formulées actuellement; d'autre part, une machine d'écoute permettant d'apprendre (écoute acoustique-symbolique) à un partir d'un flux audio un ensemble de représentations acoustiques latentes pertinentes pour l'agent génératif et la construction d'un discours musical structuré sur le long-terme. Par exemple, contrôler la corrélation entre la ligne mélodique et l'accompagnement harmonique, d'identifier la ligne mélodique principale, ou de déterminer les points de rupture dans le discours musical.

La thèse sera en particulier focalisée sur la réalisation d'un module de perception et décision par deep learning conférant aux agents la capacité de repérer quelle est la topologie d'évènements saillants ou pivots, et les dimensions audios pertinentes à écouter dans un signal avec lequel l'agent interagit, ainsi que les modalités de leurs évolutions. Il pourrait s'agir, par exemple, de détecter la transition entre un premier mouvement musical purement rythmique où l'énergie est la dimension dominante, à un second mouvement plus harmonique où l'attention doit donc être portée sur les hauteurs ou autres descripteurs mélodico-harmoniques. Cette compréhension informera enfin leurs processus de décisions, permettant ainsi de choisir automatiquement un « rôle » adapté.

Pour ce faire, les recherches s'articuleront autour des trois axes suivants :

- L'apprentissage de représentations latentes permettant d'encoder des relations musicales à long-terme à partir d'un flux audio polyphonique (en pistes solistes séparées ou mélangées). Ces représentations seront typiquement utilisées pour améliorer la prise de décision permettant de contrôler la corrélation entre la ligne mélodique et l'accompagnement harmonique, d'identifier la ligne soliste et d'accompagnement, ou de déterminer les points de rupture dans le discours musical.
- L'utilisation d'un formalisme issu de la théorie de l'information permettant de quantifier les notions de nouveau et de connu en termes de quantité d'information , et de exploiter l'information mutuelle entre des flux audio séparés (e.g. soliste, accompagnement) pour induire et apprendre leur stratégies de co-adaptation.
- L'intégration du module d'écoute profonde dans la machine d'écoute réactive pour l'enrichir



de facultés d'attention à l'échelle narrative et piloter des modèles intentionnels.

Les apprentissages seront réalisés à partir de bases de données disponibles avec les pistes séparées de différents instrumentistes. Pour cela, nous pourrons compter sur la base de données multipistes de duo d'improvisations libres collectées dans le contexte du projet ANR MICA (Clément Canonne, Ircam) ainsi que d'enregistrements multipistes de concerts du festival de Jazz de Montreux (EPFL Meta Media Center / Montreux Jazz Heritage Lab, Resp. Alain Dufaux). Des collaborations artistiques seront menées tout au long de la thèse pour expérimenter les architectures proposées et les valider par un retour expert avec des musiciens professionnels en situation de performance. Cette thèse sera l'objet d'une collaboration entre l'équipe Représentation Musicale (RepMus) représentée par Gérard ASSAYAG et Jérôme NIKA et spécialisée dans l'écoute musicale artificielle réactive et l'intelligence artificielle symbolique et l'équipe Analyse et Synthèse des Sons (AS) représentée par Nicolas OBIN et spécialisé en traitement du signal audio et apprentissage par réseaux de neurones.

Assayag, Gérard, et al. "Omax brothers: a dynamic topology of agents for improvisation learning." Proceedings of the 1st ACM workshop on Audio and music computing multimedia. 2006.

Jérôme Nika, Marc Chemillier, Gérard Assayag. ImproteK: introducing scenarios into human-computer music improvisation. ACM Computers in Entertainment, 2017.

Laurent Bonnasse-Gahot, An update on the SOMax project, Technical report, Ircam, 2016.

Joakim Borg, Somax 2: A Real-time Framework for Human-Machine Improvisation, Technical report, Ircam, 2019, Dynamic Classification Models for Human-Machine Improvisation and, Composition Master report Ircam, Aalborg University, 2020

Jérôme Nika, Ken Déguernel, Axel Chemla—Romeu—Santos, Emmanuel Vincent, Gérard Assayag. DYCI2 agents: merging the "free", "reactive", and "scenario-based" music generation paradigms. International Computer Music Conference, Oct 2017, Shanghai, China.

Boulanger-Lewandowski, N., Bengio, Y., & Vincent, P. (2012). Modeling temporal dependencies in high-dimensional sequences: application to polyphonic music generation and transcription. Paper presented at the Proc. of the 29th Int. Conf. on Machine Learning.

Jansson, A., Humphrey, E., Montecchio, N., Bittner, R., Kumar, A., & Weyde, T. (2017). Singing voice separation with deep U-Net convolutional networks. in proc. of the International Society for Music Information Retrieval Conference (ISMIR).

Rachel M. Bittner, Brian McFee, Justin Salamon, Juan Pablo Bello (2017). Deep Saliency Representations for f0 Estimation in Polyphonic Music, in proc. of the International Society for Music Information Retrieval Conference (ISMIR).

Tristan Carsault. Introduction of musical knowledge and qualitative analysis in chord extraction and prediction tasks with machine learning. Machine Learning [stat.ML]. Sorbonne Universités, UPMC University of Paris 6, 2020.

Dubnov et al. "Audio Oracle Analysis of Musical Information Rate", Proc. of IEEE Semantic Computing Conference, ICSC2011, Palo Alto, CA, 2011

**Merci d'enregistrer votre fichier au format PDF et de le nommer :
«ACRONYME de l'initiative/institut – AAP 2021 – NOM Porteur.euse Projet »**

***Fichier envoyer simultanément par e-mail à l'ED de rattachement et au programme :
cd_instituts_et_initiatives@listes.upmc.fr avant le 20 février.***