

Intelligence artificielle et musique, sous le haut patronage de Guillaume de Machaut



Francis Rousseaux,
Professeur des Universités spécialisé
en Intelligence Artificielle et Humanités
Numériques, en poste à l'Université Reims
Champagne Ardenne

**Et Musique est une science
Qui veut qu'on rie et chante et danse. [...]
Partout où elle est, joie y porte,
Les déconfortés réconforte, [...]
Fait-elle les gens réjouir. [...]
N'instrument qui ait souffle
ou touche ou corde
Qui par musique ne s'accorde.**

Guillaume de Machaut (1300-1377), Prologue V

Sage qui, tel le compositeur et écrivain Guillaume de Machaut, préfère mettre en avant les effets manifestes de la musique sur les gens plutôt que d'insister sur l'effort de codification et de perfectionnement qu'elle exige du compositeur et des interprètes. Mais le chanoine de Reims le savait aussi, qui rédigea un traité poétique sur son métier (son fameux *Prologue*) : pour produire, servir et transmettre la musique, tous les moyens innovants sont bons à explorer.

Au milieu des années 1950, d'autres chantres de la musique explorent et revisitent certaines pratiques musicales au moyen de techniques de synthèse et de traitement du signal sonore, offertes par le numérique et les ordinateurs naissants. C'est ainsi qu'un Max Mathews met au point la suite des logiciels MUSIC, qui permettent à des compositeurs pionniers de générer les premières pièces musicales à l'aide d'ordinateurs.

À la même époque, l'intelligence artificielle et l'apprentissage machine (*machine learning*) font leurs premiers pas, portés par des chercheurs souvent passionnés de musique, tel Marvin Minsky au MIT, qui ne manqueront pas d'imaginer à leur tour des dispositifs originaux conduisant à des bouleversements productifs de certaines pratiques musicales. Mais l'IA d'alors (c'est pourquoi elle est dite « symbolique », par opposition aux approches « connexionnistes » issues de la cybernétique, auxquelles elle s'oppose) a besoin, pour faire la preuve de ses performances, de représentations symboliques stables, abstraites et manipulables, moins fuyantes que l'insaisissable flux sonore. On cherche alors à isoler la question du timbre pour la confier aux spécialistes du traitement du signal, afin de se concentrer sur les structures organiques d'objets musicaux (notes, silences, et autres accords, tels que représentés dans les partitions musicales), en « suspendant » leur mise en son. Au début des années 1980, la norme d'interface MIDI (*Musical Instrument Digital Interface*), à la fois format de fichier et protocole de communication entre instruments électroniques, synthétiseurs et logiciels, vient à point nommé, qui propose un encodage numérique du geste instrumental permettant d'affranchir les questions proprement musicales des questions de timbre, et d'ouvrir à un mode d'existence « événementiel » idéalisé de la musique, produite par un pianiste surhumain possédant 127 doigts et jouant sur un instrument à la délicate sensibilité, autorisant le paramétrage dynamique de timbres différents pour chaque touche.

La séparation artificielle entre « son » et « musique hors son » (partitions musicales, liées à l'intention compositionnelle) a permis aux chercheurs d'encoder une pièce musicale comme un agencement d'événements simples, proches des notes de musique bien connues des musiciens occidentaux contemporains, et de représenter quantité d'autres objets et structures émanant du solfège académique. La restriction artistique était certes considérable, mais a ouvert la voie à de passionnantes recherches exploratoires (un peu comme des roulettes fixées provisoirement sur un vélo pour que l'enfant se familiarise dans la phase d'apprentissage) : pendant que certains se concen-

Apprentissage profond et génération de musique

par Jean-Pierre Briot, Directeur de recherche au CNRS.

Avec les chaînes de Markov, les réseaux de neurones sont un des deux principaux modèles d'apprentissage machine utilisés pour la musique. Les premiers travaux, préfigurant les travaux actuels¹, datent de la fin des années 1980. L'apprentissage profond (*deep learning*), outre des avancées majeures en termes de ressources (computationnelles, données et techniques), a récemment introduit des modèles (architectures) spécifiques, dits génératifs, ainsi dédiés, non plus à la classification ou la prédiction, mais à la génération de contenus (musique, mais aussi images ou texte)². Les principaux sont les auto-encodeurs variationnels (VAE) et les réseaux antagonistes génératifs (GAN), suivis récemment par le modèle *Transformer*. Il existe également des modèles dédiés au transfert de style (par exemple pour produire des accompagnements dans un style donné).

Les résultats sont assez impressionnants, malgré le manque de structure produite. Le relatif manque d'originalité, intrinsèque aux approches d'apprentissage malgré des astuces pour sortir de la distribution apprise, n'est pas un problème si l'on considère les techniques

actuelles comme outil d'assistance et non comme générateur autonome (sans capacité d'auto-évaluation esthétique). Pour ce faire, des approches de contrôle, fondées sur le conditionnement ou/et sur l'exploration d'un espace latent (dans le cas d'auto-encodeurs), sont proposées, ainsi que des approches de génération modulaire et incrémentale (par exemple pour la complétion de certains segments). Une tendance actuelle est la montée en puissance des modèles de génération d'audio

(son, par opposition à une représentation symbolique telle qu'une partition ou la norme MIDI).

1. Jean-Pierre Briot, "From artificial neural networks to deep learning for music generation – History, concepts and trends", *Neural Computing and Applications (NCAA)*, (33):39-65, janvier 2021. DOI: 10.1007/s00521-020-05399-0.
2. Jean-Pierre Briot, *Apprentissage profond et génération musicale*, Bibliothèque Tangente, Editions Pôle Paris, 2019, n° 68, Hors série Intelligence artificielle, pp. 30-37, hal-02267790v2..

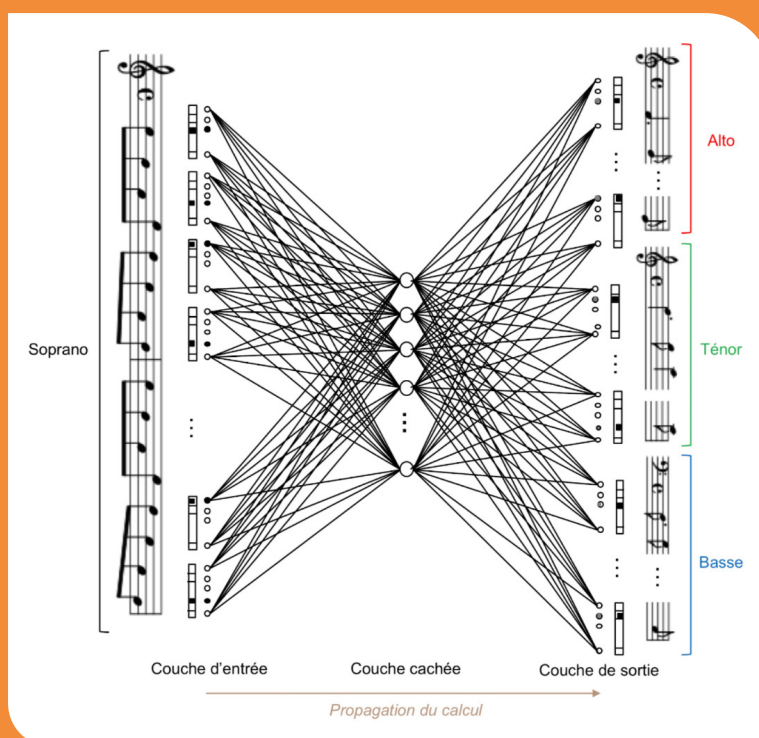


Illustration : Exemple d'architecture simple de réseau de neurones feedforward (à propagation avant) à une seule couche cachée générant un contrepoint à trois voix (dans le style des *Chorales de Bach*)

traient sur la synthèse ou l'échantillonnage sonore, d'autres développaient des séquenceurs, des éditeurs de partitions, des techniques et outils d'aide à la composition ou à l'orchestration, des suiveurs de partitions ou des inspirateurs pour l'improvisation. Fort heureusement, à partir des années 2000, la puissance de calcul des machines fut telle qu'il devint possible d'aller « de plain-pied » du son timbré jusqu'à la musique, et de retirer enfin les roulettes du vélo : le calcul de descripteurs

sonores et musicaux pertinents vint compléter les descriptions éditoriales, l'indexation et la fouille de sons et de musiques par les contenus devinrent possibles, ainsi que beaucoup d'autres applications intelligentes faisant enfin droit au son musical.

À partir des années 2010 une nouvelle donne apparaît lorsque les techniques prédictives basées sur les réseaux de neurones (*deep learning*), nourris d'océans de données numériques

toujours plus vastes, réalisent des prouesses spectaculaires, d'abord en reconnaissance d'image et en traduction automatique, et bientôt dans les domaines de prédilection de la musique. La controverse scientifique « approches symboliques » vs « approches néo-connexionnistes » est productive, et ses retombées aux intersections de l'IA et de la musique sont légions. Car même si, en dépit de leurs prouesses, les techniques de *deep learning* sont loin de satisfaire le programme de l'intelli-

gence artificielle générale (comme ne cessent de leur reprocher les « symbo- listes » remarquant que « *c'est l'agenda intellectuel de Wiener qui domine aujourd'hui sous la bannière de la terminologie de McCarthy* » (Michael Jordan, cité dans [Cardon, Cointet et Mazières 2018]), les applications qu'elles permettent s'avèrent très stimulantes.

Qu'en pensent « les gens », arbitres ultimes chers à Guillaume de Machaut? Dansent-ils, chantent-ils et rient-ils, se réconfortent-ils et se réjouissent-ils sous l'effet d'une musique profondément renouvelée par sa confrontation avec l'IA? Sans doute pas directement, mais l'IA joue d'ores et déjà un rôle de premier plan dans les modes d'existence de la musique, à travers nos manières de naviguer dans ses espaces, d'utiliser de recommandations artificielles, et d'apprendre à en jouer. Demain, des travaux sur des problématiques de style et de créativité pourraient démultiplier ces impacts.

Un détour par les humanités et les arts numériques nous permettra d'étayer cette prophétie. L'informatique ayant essentiellement pour objet l'automatisation du calcul numérique, ses méthodes et ses technologies se sont développées prioritairement dans les domaines des sciences exactes et de

l'ingénierie, d'une part parce que le format des connaissances disponibles s'y prêtait d'emblée, d'autre part parce que la formation scientifique des acteurs facilitait naturellement leur appropriation des outils.

C'est pourquoi le développement d'une intelligence artificielle capable d'intervenir dans le champ « réservé » de l'interprétation humaine s'est ramifié en une myriade de champs d'investigation faisant la part belle aux sciences de l'ingénieur, qui ont structuré l'essentiel de la recherche et des développements en informatique et en IA durant un demi-siècle, dans la foulée de la pensée de Leibniz [Leibniz 1734] : « *La musique est une pratique cachée de l'arithmétique, l'esprit n'ayant pas conscience qu'il compte.* »

À partir de la fin du XXe siècle pourtant, les sciences et technologies du numérique ont entrepris la conquête des Sciences humaines et sociales (SHS, les anciennes Humanités), pour donner naissance aux humanités et aux arts numériques (les *Digital Humanities/Arts* des Anglo-Saxons). C'est que les chercheurs en SHS se sont peu à peu formés au numérique et ont cherché à confronter leurs disciplines à ces nouveaux outils. Les connaissances développées par les SHS n'étant sou-

vent pas spontanément représentables et opérationnalisables dans/par des systèmes informatisés, leur démarche a consisté à déplacer les objets d'études, les méthodes et les approches pour permettre au numérique de s'immerger dans les modélisations.

Les sciences de l'ingénieur étant aujourd'hui largement mondialisées, les acteurs de l'industrie du numérique, les chercheurs et les politiques ont compris que les véritables enjeux culturels et stratégiques de la numérisation consistent désormais dans l'élaboration d'humanités et d'arts numériques performants. Gageons que la rencontre entre l'IA et la musique ne fait que commencer à fructifier. En l'espèce Guillaume de Machaut reste notre meilleur guide, pourvu qu'on actualise à la marge son *Prologue* (nous ajoutons le texte en caractères orange) :

**N'instrument qui ait soufflé ou touché
ou corde ou algorithme d'IA
Qui par musique ne s'accorde**

D. Cardon, J.-P. Cointet et A. Mazières, « *La revanche des neurones : L'invention des machines inductives et la controverse de l'intelligence artificielle* », Réseaux, La Découverte, 2018, 5 (211), pp.173-220. hal-02005537.
G-W. Leibniz, « *Epistolae ad diversos*, herausgegeben von Chr. Kortholt », vol.1/2, Leipzig, 1734.

Les recherches en intelligence artificielle à l'IRCAM

par Hugues Vinet, Directeur de l'Innovation et des Moyens de la Recherche, Ircam.

Les recherches en IA ont débuté à l'IRCAM dans les années 80, avec le développement de systèmes de synthèse sonore basés sur des règles puis d'environnements de composition assistée par ordinateur en LISP (PatchWork, OpenMusic). Des méthodes d'apprentissage, à base de descripteurs audio, sont utilisées au service de la création artistique pour l'orchestration assistée à partir de bases de données d'échantillons d'instruments ou en temps réel en situation de concert : suivi

de partition, de gestes, improvisation et co-créativité homme-ordinateur (systèmes OMax et DYCI2 et projet ERC REACH en cours). Un autre champ de recherche appliqué à la distribution de la musique est l'extraction automatique de métadonnées de contenu pour lequel l'IRCAM a produit une technologie de référence industrialisée auprès de leaders du domaine.

Les méthodes à base de *deep learning* surpassent et supplantent aujourd'hui les approches symboliques pour l'analyse et progressivement les applications temps réel ou génératives comme par exemple le traitement de timbres pour la production de la pièce *Convergence* d'Alexander Schubert combinant auto-encodeur variationnel et réseau antagoniste génératif. Plus globalement, le *deep learning* induit une rupture paradigmatique

en traitement de signal audio, en permettant l'inclusion dans l'apprentissage de caractéristiques sur lesquelles nous ne disposons de modèles explicites suffisamment performants, comme la synthèse vocale multi-locuteurs, ou le passage contextualisé du symbolique à des variations sonores continues relevant d'une performance humaine experte.



Vidéo de la performance de *Convergence*



Vidéos de recherches de l'Ircam sur l'improvisation et la co-créativité interprète-ordinateur.

Bobby Few, Turing et la musique improvisée

par **Olivier Koechlin**, Musicien et réalisateur d'applications musicales

Dans le domaine de la pratique musicale que l'on nomme « musique improvisée », un exercice radical consiste à engager la conversation avec un ou plusieurs autres musiciens, en séance de travail ou en public, sans aucun préalable sur ce qui va être joué, sans annoncer un titre du répertoire ni aucune structure connue, et de tenter de développer un discours commun sensible et cohérent. L'exercice est plus difficile qu'il n'y paraît, entre vertige de la page blanche, bavardages ego-centrés et bazar ambiant.

Peut-on aussi envisager cette démarche comme le « test de Turing » de la musique artificiellement intelligente ? Comment programmer un agent musicien improvisateur capable de donner l'illusion d'un comporte-



ment humain dans toutes les situations ?

Depuis le tournant du siècle, les recherches en IA ont donné des réponses plus convaincantes à ce problème, mais je voudrais décrire et donner à

écouter ici deux exemples déjà anciens, réalisés à une époque où les chaînes de Markov, les réseaux de neurones et l'apprentissage profond n'étaient pas encore mis en pratique dans ce domaine.

Le première expérimentation fut réalisée à l'Ircam en 1987 avec Bobby Few, pianiste de jazz et immense improvisateur, acteur remarqué du mouvement free jazz américain à Paris depuis les années 1970, qui nous a quitté récemment.

La deuxième expérience est plus ordinaire a priori mais elle éclaire bien la précédente. C'est la rencontre de quatre musiciens peu causants dans une répétition improbable le 6 juin 1996, selon les principes assumés d'improvisation totale présentés plus haut.

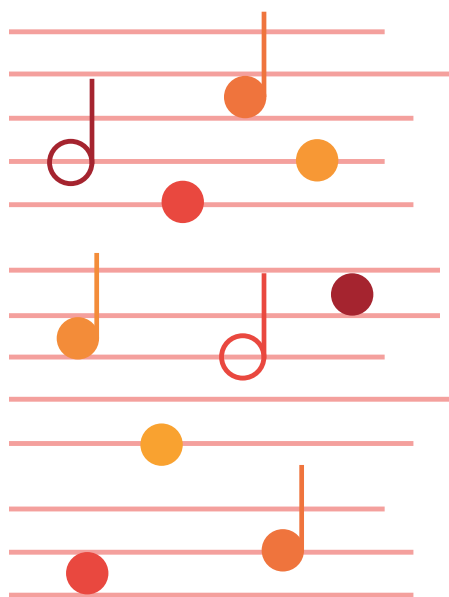


Écouter les extraits...

Garantir l'intelligibilité du processus de composition avec l'IA symbolique

Startup tarnaise créée en septembre 2019, AIbstract développe une solution logicielle visant à offrir les services d'un compositeur virtuel. Permettant la création de musique personnalisée sur la base de critères fournis par l'utilisateur, cette solution fait appel à des technologies d'intelligence artificielle symbolique, lesquelles ont été privilégiées en comparaison

d'une approche connexionniste. En effet, celles-ci garantissent l'intelligibilité du processus de composition et donnent accès à un potentiel de contrôle maximal sur ce dernier. Les systèmes d'AIbstract ne résultent pas d'un processus d'apprentissage automatique, mais sont le fruit d'une transcription de connaissances musicologiques collectées et restructurées par les experts au sein de l'équipe. À la manière d'un outil d'optimisation, ces systèmes interviennent sur un ensemble de niveaux décisionnels (structures, gammes, accords, mélodie, rythme...) d'une façon comparable au processus mis en œuvre par un compositeur humain.



La création musicale et l'IA

Entretien avec André Manoukian, Compositeur et co-fondateur de Matchtune



André Manoukian et Philippe Guillaud, fondateurs de Matchtune

Le compositeur André Manoukian a lancé, avec l'ingénieur Philippe Guillaud, la startup MatchTune dont les solutions sont basées sur l'intelligence artificielle. Il nous parle de l'évolution de cet art face ou avec l'essor de l'IA.

La musique repose essentiellement sur les mathématiques où tout est affaire de nombres et de fréquences. L'ordinateur semble donc doté de toutes les capacités nécessaires à la création et l'édition automatisée de musique. Quels sont les principaux défis liés à cette création ?

La musique, c'est des maths et de l'éros, de la science et de la spiritualité, du calcul et de l'intuition. Ce mélange paradoxal la rend unique. Il est évident que l'ordinateur ne possède que la moitié de la formule à chaque fois. Il

lui manque l'éros, la spiritualité et l'intuition.

Le jour où un ordinateur éprouvera la douleur d'un chagrin d'amour, il pourra composer une mélodie digne de ce nom, en attendant, sans l'humain, ne restent que les chiffres secs.

Malgré le caractère extrêmement normé de la musique, l'attrait d'un morceau réside souvent dans la transgression des règles musicales. Les imperfections, qui créent des tensions et ajoutent du relief, ne sont-elles pas paradoxalement les plus difficiles à produire pour un ordinateur ? Est-il capable de faire le distinguo entre ces imperfections qui viennent caresser les oreilles des auditeurs et les fautes de goût ?

Le jazz, avec ses « dissonances », répond à des règles harmoniques strictes et faciles à programmer. Par contre, une fausse note est un imprévu, un accident sur lequel le musicien expérimenté rebondit comme un skieur de freestyle rebondit sur une bosse.

Programmer l'imprévu, l'aléatoire, est bien sûr possible, cependant la magie opère sur scène par interaction mystérieuse entre le musicien et le public. Cette interaction quantique (une expérience n'a lieu que si elle est observée) n'est pas encore à l'ordre du jour dans la programmation binaire.

Nous voyons émerger en peinture des algorithmes GAN capables d'imiter avec des résultats de plus en plus réalistes le style de grands maîtres. Comment voyez-vous évoluer la création musicale dans les

années à venir ?

L'avantage de l'IA est qu'elle assiste le compositeur dans ses tâches de recherche de sons, de documentation sonore, d'algorithmes rythmiques, de nouvelles propositions harmoniques. Elle assiste le musicien en augmentant ses capacités. La musique générée par l'IA est toujours bricolée par ceux qui s'en servent et c'est de ce bricolage parfois maladroit que naissent les titres présentant un intérêt. On n'a pas attendu les ordinateurs pour faire de la musique de merde, il n'y a qu'à voir les productions de masse de Stock Aitken Waterman des années 1980...

Le cycle de vie des morceaux risque-t-il d'être réduit par la musique à la volée, capable de créer une bande son permanente, en temps réel ? L'avènement de l'IA dans la musique participe-t-il à une désacralisation de cet art ?

L'IA sera toujours sabotée et détournée par de joyeux subversifs. C'est ce qui nous sauvera. L'art utilisera toujours la technique pour arriver à ses fins, sinon ce n'est pas de l'art.