

Le langage des partitions musicales face à l'intelligence artificielle

Dans le domaine de l'informatique musicale, l'analyse et la génération automatique de partitions font l'objet de multiples recherches, en grande partie inspirées par des algorithmes d'Intelligence artificielle (IA) conçus à l'origine pour le traitement de contenus textuels. Le projet Music NLP, collaboration entre le Studio de création et de recherche en informatique et musiques expérimentales (SCRIME) à l'Université de Bordeaux, l'Institut national de recherche en sciences et technologies du numérique (INRIA) et le Centre de recherche en informatique signal et automatique de Lille (CRISAL) à l'Université de Lille, mène une réflexion sur ce détournement d'outils, ses limites techniques et la manière dont il nous interroge à un plus haut niveau sur le parallèle entre musique et langage naturel.

LOUIS BIGO

Professeur des universités, directeur du Studio de création et de recherche en informatique et musiques expérimentales (SCRIME) – Université de Bordeaux, CNRS, Institut polytechnique de Bordeaux, Laboratoire bordelais de recherche en informatique (LaBRI), UMR 5800

MIKAELA KELLER

Maîtresse de conférences, Centre de recherche en informatique, signal et automatique de Lille (CRISAL), UMR CNRS 9189 – Université de Lille, Centrale-Lille, INRIA, Institut Mines-Télécom Nord Europe

DINH-VIET-TOAN LE

Doctorant en informatique musicale, Centre de recherche en informatique, signal et automatique de Lille (CRISAL), UMR CNRS 9189 – Université de Lille, Centrale-Lille, INRIA, Institut Mines-Télécom Nord Europe

L'Intelligence artificielle (IA) a un impact croissant dans les différents rapports que nous entretenons à la musique. Elle intervient dans la conception d'outils destinés à enrichir la composition, la performance, l'écoute, l'analyse ou encore la pédagogie. Au-delà du développement d'outils pour les musiciens, mettre l'IA à l'épreuve des données musicales offre un cadre original pour étudier et améliorer notre compréhension des langages de la musique, aussi variés et multiples soient-ils.

Intelligences textuelles et musicales

Une large partie des recherches actuelles dédiées à l'analyse et à la génération automatique de partitions musicales emploient des techniques d'IA conçues à l'origine pour le traitement de textes¹, notamment les réseaux de neurones *Transformers*, à la base de programmes tels que ChatGPT. Cet emprunt est largement motivé par la performance remarquable de ces modèles dans le domaine du texte, mais aussi par l'assimilation courante de la musique à une sorte de *langage*. Bien que produisant des résultats de plus en plus spectaculaires, les compétences musicales des *Transformers* demeurent cependant rarement explicables à cause de la complexité de ces algorithmes, souvent comparés à des « boîtes noires » et dont la performance repose sur un accès à des volumes massifs de données de référence. Les recherches du projet Music NLP visent à améliorer l'explicabilité de ces modèles dans le domaine musical et à proposer ainsi un cadre original pour l'étude des différences entre musique et langage naturel.

« Langages » dans les partitions

Le terme de « langage musical » semble largement induit par l'utilisation de formules codifiées représentatives d'un « style musical ». Le répertoire classique abonde de ces éléments, parmi lesquels la « tonalité », la « texture », les « cadences » et la « structure », qui font l'objet de nombreuses recherches en analyse musicale computationnelle. L'idée de « langage musical » se manifeste également dans l'emprunt par les musiciens de termes propres au langage naturel. La partition est par exemple associée à un « discours musical », structuré en phrases, incluant même des mécanismes de « questions/réponses ». Touchant au concept de sémantique, propre au domaine du langage naturel et non à celui de la musique, ces analogies atteignent cependant rapidement leurs limites². La tendance du musicien à emprunter des termes au domaine du langage naturel illustre probablement davantage une complexité à décrire la musique qu'une réelle proximité entre les deux domaines.

Relations entre « mots », relations entre « notes »

Les algorithmes *Transformers* ont été conçus pour identifier des relations entre les mots d'un texte, via un mécanisme appelé « auto-attention », avec l'objectif de reconstituer des représentations sémantiques absentes. La transposition de ce procédé aux éléments constituant une partition musicale, typiquement les notes et les accords, nous interroge sur la nature de l'information véhiculée par la partition et de la manière

1. Dinh-Viet-Toan Le, Louis Bigo, Mikaela Keller et Dorien Herremans, « Natural language processing methods for symbolic music generation and information retrieval: A survey », *arXiv*, 2024 (voir aussi : <https://hal.science/hal-04621444/file/main.pdf>).

2. Ray Jackendoff, « Parallels and nonparallels between language and music », *Music Perception: An Interdisciplinary Journal*, vol. 26, n° 3, 2009, p. 195-204.

Les 8 dernières mesures de la partie d'exposition du premier mouvement de la *Sonate facile* (sonate pour piano n° 16, K. 545) de Wolfgang Amadeus Mozart avec plusieurs éléments typiques du langage classique annotés. Les textures comprennent un motif de « question et réponse » entre les deux mains (1.a), un accord d'accompagnement répété (1.b), une « basse d'Alberti » (1.c), des accords *sparses* (1.d), une « cadence authentique parfaite » (2), suivie d'une répétition cadentielle V/I et d'un « triple coup de marteau » concluant l'extrait (3) https://hal.science/tel-04433299/file/hdr_bigo.pdf (figure 2.1. p. 10).

dont elle se distingue de la sémantique d'un texte. Les travaux que nous menons analysent les raisons pour lesquelles l'auto-attention parvient malgré ces différences à apprendre dans les partitions des phénomènes musicaux abstraits tels que l'harmonie fonctionnelle, pour laquelle nous constatons par exemple une surprenante tendance de l'auto-attention à se référer au futur dans la partition pour expliquer le contexte harmonique présent.

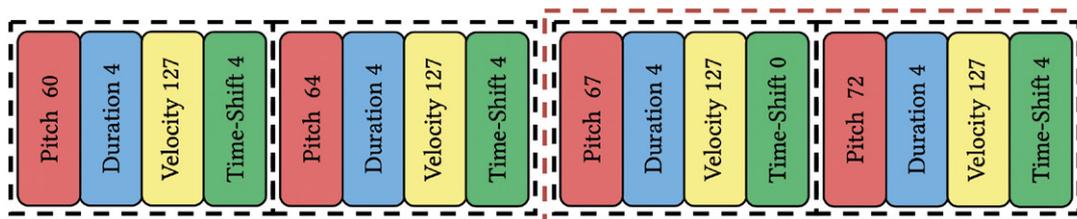
Structures séquentielles

La nature temporelle de la musique favorise sa représentation informatique sous la forme de « séquences d'éléments », typiquement des notes ou des accords, comparables aux séquences de caractères ou de mots constituant un texte. Bien que cette similarité de structure facilite le transfert d'algorithmes entre les deux domaines, représenter la musique sous la forme de simples séquences limite l'expression de deux caractéristiques qui lui sont propres. La première est la simultanéité, inhérente à la notion d'harmonie. La seconde est le rythme, induit par des distances temporelles précises entre les notes. Une variété de stratégies a été proposée pour exprimer ces concepts

dans des séquences, généralement *via* l'introduction d'éléments additionnels qui rendent compte de la temporalité musicale. Une partie des recherches du projet Music NLP vise à améliorer l'expressivité musicale de ces représentations, en substituant par exemple les hauteurs de notes par les intervalles qui les séparent, que ce soit entre des notes simultanées ou successives, ou en élargissant le vocabulaire en y ajoutant des « super-valeurs » représentant des sous-séquences de valeurs courantes dans le répertoire.

L'adaptation d'algorithmes de texte pour le traitement de partitions nous invite à une réflexion sur la spécificité de la musique et la complexité de sa structure. L'élaboration d'architectures de réseaux de neurones imaginés précisément pour la modélisation des éléments constituant une partition représente un défi majeur pour notre communauté dans les années à venir. Cet objectif s'ajoute par ailleurs à celui de constituer des modèles « explicables » contribuant à améliorer notre compréhension du répertoire musical, et dont l'entraînement requiert des ressources énergétiques raisonnables³. ■

3. Lasse F. Wolff Anthony, Benjamin Kanding et Raghavendra Selvan, « Carbontracker: Tracking and predicting the carbon footprint of training deep learning models », *arXiv*, 2020.



Représentation d'un extrait de partition par une séquence de valeurs. La temporalité des notes est encodée grâce à l'ajout de valeurs dédiées (*Time-Shift* et *Duration*). <https://hal.science/hal-04621444/file/main.pdf> (figure 5 p. 8).