

# Gestes sonores et transfert de timbre : concevoir une interface expressive pour RAVE

Zachary Hardy  
zachary.hardy@umontreal.ca  
Université de Montréal  
Montréal, Québec, Canada

Mahault Sampy  
mahault.sampy@umontreal.ca  
Université de Montréal  
Montréal, Québec, Canada

## Abstract

Ce projet explore les enjeux d'accessibilité associés à l'utilisation d'outils musicaux intégrant l'intelligence artificielle. Bien que prometteurs sur le plan créatif, ces dispositifs demeurent souvent peu intuitifs et réservés à des utilisateurs experts. Nous proposons une interface de création sonore fondée sur le modèle RAVE de l'IRCAM, visant à concilier accessibilité, expressivité et contrôle intuitif. À travers une démarche de recherche-création, nous développons un outil qui facilite l'interaction significative avec l'espace latent de RAVE, en s'adressant à la fois aux musiciens expérimentés en musique électronique et aux artistes novices. Ce travail s'inscrit dans une perspective de démocratisation des technologies créatives et d'enrichissement du dialogue humain-machine.

## Keywords

RAVE, génération audio, espace latent, contrôle expressif, interface créative, intelligence artificielle

### ACM Reference Format:

Zachary Hardy and Mahault Sampy. 2025. Gestes sonores et transfert de timbre : concevoir une interface expressive pour RAVE. In *Proceedings of UDEM 2025*. ACM, New York, NY, USA, 4 pages.

## 1 Introduction

La dernière décennie a vu l'émergence de modèles génératifs audio de plus en plus sophistiqués grâce à l'arrivée de l'intelligence artificielle et du Machine Learning dans la sphère musicale, transformant ainsi profondément les possibilités de création sonore. Parmi ces modèles, RAVE (Robust Audio Variational autoEncoder) conçu par le laboratoire ACIDS de l'IRCAM se distingue par sa capacité à générer des sonorités complexes et riches à partir d'un espace latent multidimensionnel. Cependant, malgré ces avancées techniques remarquables, un obstacle majeur persiste : comment interagir de manière significative avec ces modèles dont la complexité mathématique et conceptuelle freine techniquement la plupart des utilisateurs ?

Nous commencerons par contextualiser les enjeux liés à l'intégration de l'intelligence artificielle dans la création musicale ainsi que ceux liés à l'interfaçage de modèles génératifs audio. Nous présenterons ensuite notre hypothèse de travail, la méthodologie adoptée pour développer l'outil, ainsi que les observations issues son utilisation en contexte de création. Enfin, nous conclurons en discutant des perspectives d'amélioration futures.



This work is licensed under a Creative Commons Attribution 4.0 International License.

UDEM 2025, MUS3329X

© 2025 Copyright held by the owner/author(s).

## 2 Contexte et revue de littérature

### 2.1 Introduction à la problématique de l'IA en musique

L'intégration de l'intelligence artificielle (IA) dans la création musicale constitue l'une des transformations majeures des pratiques artistiques contemporaines. Bown [1] souligne que cette évolution soulève des questions fondamentales sur la nature de la création musicale et ses valeurs associées. Les modèles génératifs audio comme RAVE (Robust Audio Variational autoEncoder) permettent désormais la synthèse de sons d'une qualité sans précédent, mais comme l'observe Gioti [8], ils représentent aussi une transformation profonde du processus créatif.

Malgré ces avancées, un défi persiste : Comment concevoir des interfaces permettant une interaction véritablement significative avec ces modèles complexes ? Notre recherche aborde précisément cette question : comment interfacier de manière plus significative des modèles de génération audio pour faciliter la création musicale tant pour les musiciens expérimentés que pour ceux novices dans l'utilisation de l'intelligence artificielle.

### 2.2 Transformations des pratiques artistiques avec l'IA

Tatar, Ericson et Cotton [10] identifient un changement paradigmatique dans la relation artiste-technologie, où l'IA devient un collaborateur actif dans le processus créatif. Cette reconfiguration modifie fondamentalement la position de l'artiste, qui adopte désormais des rôles variés du concepteur au curateur.

Esling et Devis [6] soulignent que l'IA nous oblige à repenser les fondements de la créativité humaine. La créativité devient un phénomène émergent issu de l'interaction humain-machine, remettant en question les conceptions traditionnelles de l'originalité artistique.

Cette évolution n'est pas sans débat. Kaila et Sturm [3] présentent une perspective agonistique mettant en lumière les tensions entre la promesse de démocratisation de la création musicale et les préoccupations concernant l'authenticité et l'identité artistique.

### 2.3 Défis d'interfaçage avec les modèles génératifs audio

L'interfaçage avec les modèles génératifs présente des défis considérables. Esling, dans son travail sur le Flow Synthesizer [2], identifie la difficulté de représenter intuitivement les espaces latents multidimensionnels qui encodent des caractéristiques sonores complexes et interdépendantes.

Cette complexité est exacerbée par le manque d'explicabilité. Tecks, Peschlow et Vigliensoni [7] soulignent que l'opacité des systèmes d'IA constitue un obstacle majeur à une pratique artistique significative pour les musiciens. De plus, Bryan-Kinns et

al. [4] ajoutent que l'explicabilité est une condition nécessaire à l'appropriation créative des systèmes d'IA musicale.

Puckette [9] offre une perspective sur l'équilibre entre puissance technique et accessibilité, suggérant que les interfaces musicales les plus réussies sont celles qui rendent la complexité manipulable à travers des métaphores soigneusement conçues.

## 2.4 Conception centrée sur l'utilisateur pour l'IA musicale

Oh et al. [5] proposent une approche de l'IA générative audio axée sur la demande, contrastant avec les démarches purement technologiques. Ils suggèrent que le développement devrait être guidé par les besoins réels des créateurs plutôt que par les seules possibilités algorithmiques.

L'explicabilité émerge comme un pilier central de cette approche. Bryan-Kinns et al. [4] identifient plusieurs stratégies pour rendre les systèmes d'IA musicale plus compréhensibles, notamment la visualisation des processus internes et l'utilisation de métaphores familières.

La diversité des profils d'utilisateurs constitue un autre aspect crucial. Les interfaces doivent s'adapter aux besoins variés, depuis les musiciens experts recherchant un contrôle précis jusqu'aux novices privilégiant une exploration intuitive.

## 2.5 Interfaces existantes et vers des interactions plus significatives

Le Flow Synthesizer d'Esling [2] représente une avancée significative, utilisant des flux normalisants pour créer un contrôle universel de synthétiseur audio. Sa force réside dans sa capacité à abstraire la complexité technique tout en préservant la richesse expressive.

D'autres approches privilégient des métaphores visuelles ou des contrôles gestuels, transformant la navigation dans l'espace latent en une expérience plus corporelle. Une tendance émergente consiste à combiner des paradigmes d'interaction traditionnels avec des visualisations dynamiques, créant un pont entre interfaces familières et nouveaux territoires d'exploration sonore.

Bown [1] suggère qu'une interaction devient significative lorsqu'elle permet à l'utilisateur d'exprimer ses intentions tout en maintenant un dialogue créatif avec le système. Esling et Devis [6] proposent que la signification émerge à l'intersection entre capacités techniques et intentions créatives, facilitant une forme de co-créativité.

L'équilibre entre contrôle précis et exploration intuitive émerge comme un axe central de conception. Les interfaces les plus significatives naviguent cette tension dialectique, offrant différents modes d'engagement adaptés aux divers moments du processus créatif.

## 2.6 Conclusion et positionnement de la recherche

Cette revue révèle un paysage riche mais fragmenté de recherches sur l'interfaçage avec les modèles génératifs audio. Malgré la reconnaissance de l'importance de l'explicabilité et de la diversité des utilisateurs, peu de recherches ont exploré l'application concrète de ces principes dans le contexte spécifique des modèles comme RAVE.

Notre recherche se positionne précisément à cette intersection. En développant une interface qui rend l'espace latent de RAVE

plus accessible tout en préservant sa richesse expressive, nous visons à créer un pont entre capacités techniques et besoins créatifs variés. Notre approche itérative, informée par les principes de conception centrée sur l'utilisateur, cherche à établir un nouveau paradigme d'interaction significative avec les modèles génératifs audio.

## 3 Hypothèse et méthodologie de recherche

### 3.1 Hypothèse

Nous supposons ainsi qu'en interfaçant RAVE avec des contrôles plus significatifs, il est possible de faciliter le jeu et la création sonore à partir de ces modèles, réduisant ainsi les barrières technologiques qui peuvent limiter l'accès, tant pour les musiciens expérimentés que pour ceux moins familiers avec les technologies de génération audio.

Nous avons ainsi développé une interface hybride combinant des contrôles inspirés des synthétiseurs traditionnels avec des visualisations de l'espace latent permettant une compréhension intuitive du modèle. La structuration de l'interface permettant différents niveaux d'engagement avec le modèle rend l'outil accessible aux débutants, notamment grâce à la présence de préréglages tout en offrant la profondeur nécessaire aux utilisateurs expérimentés. Enfin, la présence de retours visuels et sonores immédiats sur les actions de l'utilisateur renforce le sentiment d'agentivité et de contrôle, éléments essentiels à une expérience créative satisfaisante.

En validant cette hypothèse, notre recherche vise à contribuer non seulement au développement d'interfaces plus efficaces pour RAVE, mais également à l'établissement de principes de conception plus larges pour l'interaction significative avec les modèles génératifs audio. Nous espérons que cela informera une nouvelle génération d'outils de création sonore qui exploitent pleinement le potentiel des avancées en intelligence artificielle tout en restant accessibles à un public diversifié.

### 3.2 Méthodologie

Ce projet s'inscrit dans une démarche de recherche-création combinant conception itérative et réflexion théorique sur l'interaction significative avec les modèles génératifs audio. L'objectif était de développer un prototype d'interface facilitant la manipulation expressive du modèle RAVE, en conciliant accessibilité et profondeur de contrôle.

Le processus méthodologique s'est structuré en plusieurs étapes : Nous avons tout d'abord procédé à une analyse des paradigmes d'interaction existants afin de définir clairement nos conceptions d'expressivité artistique dans le but de clarifier nos propres objectifs pour notre outil. Nous avons ensuite développé plusieurs prototypes en faisant une rétroaction systématique en contexte de création musicale. Trois versions successives de l'outil ont été élaborées, chacune intégrant des ajustements issus d'observations internes : nous sommes ainsi passés d'une interface inspirée des synthétiseurs classiques à une interface hybride fondée sur la manipulation graphique d'enveloppes et le contrôle des dimensions latentes.

Le choix de cette approche itérative permettait d'ajuster le design au plus près de nos besoins créatifs réels, tandis que la recherche-création offrait un cadre propice à l'évaluation qualitative de l'impact artistique du dispositif. Toutefois, les évaluations ont été réalisées sur un échantillon restreint d'utilisateurs, limitant la portée généralisable des conclusions. Un retour de plus

d'utilisateurs issus de milieux musicaux plus variés nous permettrait d'affiner encore davantage l'outil et sa portée créative.

## 4 Développement de l'outil

### 4.1 Description de l'outil

L'outil développé repose sur une approche hybride visant à faciliter l'interaction expressive avec le modèle RAVE. Il se structure autour de plusieurs modules principaux :

- un contrôle direct de plusieurs dimensions de l'espace latent du modèle ;
- un module d'enveloppe permettant de sculpter l'évolution temporelle du son ;
- un module oscillateur agissant comme source d'inférence dynamique ;
- un multiplicateur de couches (layers) offrant la possibilité d'épaissir la texture sonore ;
- un contrôle de niveau de base (base level) garantissant une activité sonore constante ;
- la possibilité d'appliquer l'enveloppe uniquement à certaines couches spécifiques, pour affiner la manipulation sonore.

Ce design vise à offrir aux utilisateurs des outils de contrôle intuitifs tout en exploitant la richesse du modèle génératif. En combinant la manipulation de l'espace latent avec des principes issus de la synthèse sonore traditionnelle, l'interface permet une exploration fluide et nuancée du timbre.

Le développement de l'outil s'est appuyé sur une double ambition : rendre la création sonore par IA accessible aux musiciens expérimentés en musique électronique, tout en restant compréhensible et utilisable par des musiciens plus novices désireux d'explorer de nouveaux territoires sonores.

Ce double objectif a fortement influencé la sélection et la conception des paramètres disponibles. Un enjeu clé a été de déterminer le nombre minimal de contrôles nécessaires pour garantir une interaction efficace sans compromettre la liberté créative. Un trop grand nombre de paramètres aurait pu complexifier l'appropriation de l'outil, tandis qu'une simplification excessive risquait de réduire son potentiel expressif.

Nous avons donc cherché à atteindre un équilibre entre accessibilité et richesse de jeu. L'interface propose ainsi un ensemble de contrôles suffisamment réduits pour favoriser une prise en main rapide, mais suffisamment articulés pour permettre une réelle exploration artistique. Cette approche vise à encourager l'engagement des utilisateurs dans un processus créatif intuitif, tout en leur offrant les moyens d'affiner progressivement leur maîtrise de l'outil.



Figure 1: Capture d'écran de l'interface développée pour l'interaction expressive avec RAVE.

### 4.2 Développement

#### 4.2.1 Processus d'itération et ajustements pendant la phase de conception

#### 4.2.2 Explication du processus créatif qui a mené à la création des œuvres proposées

Afin d'évaluer l'outil développé, nous devons faire de notre mieux pour explorer le plein potentiel de l'outil. Il nous est alors apparu assez évident que nous devions adopter une attitude compositionnelle basée dans la pratique des "séquences-jeux" afin de procéder à cette évaluation. Nous avons ainsi chacun réalisé une courte étude composée uniquement de matériaux sonores provenant de ces fameuses séquences-jeu. À partir de là, notre travail de compositeur a été d'aller chercher au sein de ces minutes, des articulations et moments que nous jugions intéressants d'un point de vue musical et compositionnel.

Ayant tous les deux des sensibilités pour la musique acoustique du fait de notre parcours académique, cela s'est reflété assez rapidement dans l'esthétique musicale que nous avons adoptée dans nos compositions.

De plus, l'outil permettant de construire des séquences entières d'articulations très serrées se rapprochant de la pratique du micro-montage, cela a naturellement alimenté cette inclination déjà présente vers la musique électroacoustique.

## 5 Résultats en Recherche-Création/Discussion

### 5.1 Résultats trouvés

L'outil développé répond à plusieurs enjeux identifiés dans notre démarche de conception. Tout d'abord, il accélère nettement la création sonore en permettant la génération quasi instantanée d'une grande variété d'articulations. L'utilisateur bénéficie d'un réel contrôle sur le matériau généré, notamment grâce à la précision offerte par la modulation de l'enveloppe.

De plus, l'outil propose une solution pratique à un problème concret de stockage : en considérant le modèle RAVE comme une "banque de sons" dynamique, il devient possible de générer une infinité de textures sonores à partir d'un ensemble de données initiales réduit, limitant ainsi les besoins en espace de stockage, ce qui représente un avantage important dans un contexte de création musicale numérique.

Enfin, l'aspect ludique de l'interface graphique contribue à rendre le processus de création sonore plus engageant et intuitif, en favorisant l'exploration active et la prise en main rapide par l'utilisateur.

### 5.2 Discussion et pistes d'amélioration

Plusieurs pistes d'amélioration restent cependant à envisager. En effet, l'outil n'est pas autonome et nécessite l'utilisation de Max/MSP, ce qui peut représenter une barrière financière pour certains utilisateurs, notamment en l'absence d'une licence Max ou Ableton Live avec Max for Live mais aussi une barrière technique. Une implémentation sous forme de *plugin* VST constituerait une option intéressante pour toucher un public plus large.

La diversité sonore accessible reste également limitée par le nombre de modèles préentraînés disponibles, ainsi que par la difficulté, pour les utilisateurs non spécialistes, d'entraîner de nouveaux modèles personnalisés. Cela limite ainsi la personnalisation de l'expérience des utilisateurs et pourrait devenir une source de frustration sur le long terme.

Enfin, sur le plan de l'interface, bien que de nombreux paramètres aient déjà été supprimés, une abstraction encore plus poussée pourrait être envisagée afin de rendre l'outil encore plus intuitif. Il serait également pertinent de questionner la lisibilité de

certain modules spécifiques, tels que les sections "encode" et "decode" qui peuvent être opaques pour des utilisateurs novices: leur compréhension par l'utilisateur est-elle indispensable, ou doit-on privilégier une approche davantage orientée vers l'expérience et l'exploration intuitive plutôt qu'une explicabilité technique exhaustive ?

Ces réflexions nourriront des pistes pour de futures itérations de l'outil, dans le but de le rendre encore plus accessible, performant et adapté aux pratiques musicales d'aujourd'hui.

## 6 Conclusion

Ce projet visait à explorer les enjeux liés à l'interaction significative avec les modèles de génération audio, en particulier à travers le développement d'un outil proposant une interface plus intuitive pour interagir avec le modèle RAVE. L'approche adoptée, mêlant exploration de l'espace latent, création de gestes sonores et transfert de timbre, a permis de concevoir une interface expressive conciliant accessibilité et richesse créative.

Les résultats obtenus mettent en évidence la capacité de l'outil à accélérer le processus de création sonore, à renforcer le contrôle artistique des utilisateurs, et à ouvrir de nouvelles perspectives en matière de génération audio dynamique. L'intégration d'éléments graphiques a également contribué à rendre l'expérience de création plus ludique et intuitive.

Plusieurs axes d'amélioration subsistent néanmoins, notamment en ce qui concerne l'accessibilité technologique, la diversité des modèles disponibles, et la simplification de l'interface pour en faciliter encore davantage l'usage. Le développement futur d'une version VST pourrait notamment étendre considérablement l'accessibilité de l'outil à un public plus large.

Par ailleurs, il serait intéressant d'évaluer si cet outil peut trouver sa place au sein de démarches artistiques au-delà du champ de la musique électroacoustique. Une phase de test auprès d'une palette d'utilisateurs plus variée permettrait d'évaluer plus finement son potentiel en contexte de création musicale.

En s'inscrivant dans une démarche de recherche-crédation, ce projet contribue à enrichir les réflexions sur l'interfaçage des modèles génératifs audio et ouvre des pistes pour le développement d'outils de création intégrant l'intelligence artificielle de manière plus intuitive, expressive et démocratique.

## References

- [1] Oliver Bown. 2021. Sociocultural and Design Perspectives on AI-Based Music Production: Why Do We Make Music and What Changes if AI Makes It for Us? In *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Springer International Publishing, 1–20. [https://doi.org/10.1007/978-3-030-72116-9\\_1](https://doi.org/10.1007/978-3-030-72116-9_1)
- [2] Philippe Esling, Naotake Masuda, Adrien Bardet, Romeo Despres, and Axel Chemla-Romeu-Santos. 2020. Flow Synthesizer: Universal Audio Synthesizer Control with Normalizing Flows. *Applied Sciences* 10 (2020). <https://www.mdpi.com/2076-3417/10/1/302>
- [3] Bob L. T. Sturm et Anna-Kaisa Kailia. 2024. Agonistic Dialogue on the Value and Impact of AI Music Applications. <https://aimc2024.pubpub.org/pub/p3rww87r/release/1>.
- [4] Nick Bryan-Kinns et Berker Banar et Corey Ford et Courtney N. Reed et Yixiao Zhang et Jack Armitage. 2024. Explainable AI and Music. In *Artificial Intelligence for Art Creation and Understanding*. CRC Press, 1–29.
- [5] Sangshin Oh et Minsung Kang et Hyeonggi Moon et Keunwoo Choi et Ben Sangbae Chon. 2023. A Demand-Driven Perspective on Generative Audio AI. <http://arxiv.org/abs/2307.04292>.
- [6] Philippe Esling et Ninon Devis. 2020. Creativity in the Era of Artificial Intelligence. <http://arxiv.org/abs/2008.05959>.
- [7] Austin Tecks et Thomas Peschlow and Gabriel Vigliani. 2024. Explainability Paths for Sustained Artistic Practice with AI. <http://arxiv.org/abs/2407.15216>.
- [8] Artemi-Maria Gioti. 2021. Artificial Intelligence for Music Composition. In *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Springer International Publishing.

- [9] Puckette Miller. 2022. The Developer: What Do Music Software Developers Do? In *Artificial Intelligence and Music Ecosystem*. Focal Press, 24–34.
- [10] Kivanç Tatar, Petter Ericson, Kelsey Cotton, Paola Torres Núñez Del Prado, Roser Batlle-Roca, Beatriz Cabrero-Daniel, et al. 2024. A Shift in Artistic Practices through Artificial Intelligence. *Leonardo* (2024), 293–297. [https://doi.org/10.1162/leon\\_a\\_02523](https://doi.org/10.1162/leon_a_02523)