

PROCESSUS D'APPRENTISSAGE EN COMPOSITION ASSISTÉE PAR ORDINATEUR (PACO)

Rapport scientifique

4/12/2018

UMR 9912 STMS : IRCAM – CNRS – Sorbonne Université

Responsable scientifique : **Jean Bresson**

Page web : →<http://repmus.ircam.fr/paco/>

Introduction – Contexte et objectifs. L'idée de produire des machines capables de composer est apparue tôt dans l'histoire de l'informatique musicale (Hiller and Isaacson, 1959) et se répercute de manière récurrente, aujourd'hui par exemple au sein communautés de recherche émergentes autour de thèmes tels que *Computational Creativity*, *Musical Metacreation* (Pasquier et al., 2016), etc. Cependant, la notion d'apprentissage a rarement été exploitée par les compositeurs dans une perspective d'aide à la création. On trouve ainsi de nombreuses applications de l'IA dans les domaines de l'analyse (Meredith, 2015) et de la fouille de données musicales (Illescas et al., 2008), dans les systèmes génératifs "autonomes" (Ghedini et al., 2015) ou les agents de (co-)improvisation (Assayag et al., 2009a), ou encore pour le suivi de geste et l'interaction temps réel (Françoise et al., 2013). Dans le domaine de la composition assistée par ordinateur (CAO), en revanche, et peut-être par défiance vis-à-vis d'une certaine déposssession de la créativité au profit du calcul (Dubnov et Surges, 2014), chercheurs et compositeurs¹ se sont tournés vers des démarches constructivistes et vers d'autres aspects des technologies de l'information, comme celui du *end-user programming* (soit le fait de donner à l'utilisateur final d'un système la capacité d'en être lui-même le programmeur) et des langages de programmation visuels (Assayag et al., 2009b).² En pointe de cette approche, l'environnement OpenMusic³ développé par notre équipe est un langage de programmation visuelle complet, lui-même basé sur le langage Common Lisp, permettant aux compositeurs de créer des programmes visuels (ou mixtes, intégrant programmation graphique et textuelle) pour concevoir des processus de génération ou de traitement des structures musicales (Bresson et al., 2011).

L'objectif de ce projet exploratoire était donc d'étudier des applications possibles de l'apprentissage automatique en composition musicale assistée par ordinateur. En contraste avec une approche consistant à produire des systèmes créatifs plus ou moins autonomes, nous nous sommes intéressés à l'apport potentiel de l'IA (et en particulier de techniques faisant appel à l'apprentissage) comme assistance à la composition dans la réalisation de tâches telles que la classification et le traitement de structures musicales, ou la production de structures et paramètres musicaux par recomposition de motifs. Dans ces différents cas de figures, nous souhaitons à terme proposer aux compositeurs des outils leur permettant de mettre en œuvre des cycles expérimentaux complets incluant la préparation des données d'apprentissage, l'entraînement des modèles, et l'ajustement de leurs paramètres pour la génération de nouvelles données, ou la résolution de problèmes.

Les structures musicales intervenant dans les systèmes et processus compositionnels (qu'ils soient réalisés par ordinateur ou pas) transportent des caractéristiques plus ou moins abstraites, qui peuvent être décelables « en surface » (sur une partition, une forme concrète, ou à l'écoute), mais parfois difficile à quantifier et à traiter explicitement, que ce soit dans une démarche d'analyse ou de création. Ces caractéristiques peuvent être porteuses de sens formel, ou simplement permettre de catégoriser ces structures dans une compréhension globale de l'œuvre. En ce sens également, l'apprentissage automatique pourrait susciter de nouvelles modalités de contrôle et de compréhension des objets musicaux.

1. A quelques exceptions près, cf. par exemple (Cope, 1996).

2. Un contre-exemple sont les travaux menés récemment à l'IRCAM pour l'aide à l'orchestration et la recherche de combinaisons de timbres instrumentaux, qui font un usage intensif des techniques d'apprentissage profond, cf. par exemple (Crestel, 2018).

3. OpenMusic : →<http://repmus.ircam.fr/openmusic/>

Reconnaissance de « gestes musicaux » par apprentissage. Nous nous sommes intéressés dans un premier temps à la notion de « geste », en nous basant sur la bibliothèque xmm⁴ développée dans l'équipe ISMM (Interaction Son Musique Mouvement) du laboratoire STMS pour le suivi et de reconnaissance de mouvements en temps réel (Françoise et al., 2013). Une adaptation de cette bibliothèque pour l'environnement OpenMusic (OM-xmm) a fait l'objet du stage d'ingénieur de Paul Best (ECE Paris), encadré par Jean Bresson et Diemo Schwarz (ISMM). xmm utilise des modèles probabilistes incluant notamment des HHMM (*Hierarchical Hidden Markov Models*). Entre autres caractéristiques (Lee et Kim, 1999), ceux-ci ont la particularité de permettre la reconnaissance efficace des séries temporelles (une abstraction pertinente du point de vue musical) à partir de petits ensembles de données d'entraînement — une hypothèse également adaptée dans le cas d'utilisation que nous envisageons, où ces données d'entraînement seront généralement saisies ou transférées par les utilisateurs, et non pas extraites automatiquement à partir de bases de données ou du web.

Dans un contexte de composition assistée par ordinateur, nous prolongeons la compréhension initiale de la notion de geste avec celle de « geste musical » (Godøy et Leman, 2010), que nous essaierons de déceler et catégoriser dans des extraits sonores à partir des données issues d'analyses du signal. Nous utilisons pour cela la bibliothèque IAE (également développée par l'équipe ISMM), qui offre des outils efficaces d'extraction de descripteurs audio (Schnell et al., 2017). Étendu à cette dimension sonore et musicale, un « geste » pourra ainsi recouvrir des notions diverses liées à l'évolution des descripteurs temporels du son, des motifs mélodiques, dynamiques, ou autres caractéristiques sonores se distinguant dans l'espace multidimensionnel des descripteurs.⁵ Suivant cette approche, Alireza Farhang, compositeur en résidence à l'IRCAM (2018), utilise OM-xmm pour la reconnaissance de gestes musicaux appris sur un enregistrement audio de sa pièce *Zymiad*, dont la partition, puis des segments de l'enregistrement ont été préalablement annotés (Bresson et al., 2018). L'analyse par descripteurs et l'entraînement d'un modèle xmm permet au compositeur de classer des nouveaux extraits suivant les types de gestes qu'il a préalablement définis.⁶

Dans le cadre de son stage, Paul Best a également compilé des résultats préliminaires de cette expérience, et mis en place des méthodes et outils permettant aux utilisateurs de la bibliothèque OM-xmm de tester la validité des modèles et de les optimiser (choix des descripteurs audio pertinents, paramètre du modèle d'apprentissage, etc.) à l'aide de techniques de recherche opérationnelle (Best et al., 2018).

Boîte à outil de modèles d'apprentissage. Suivant notre objectif visant à impliquer les compositeurs dans la conception des modèles d'apprentissage, ce deuxième volet de nos travaux va s'attacher à proposer des outils et algorithmes intégrés dans l'environnement OpenMusic, personnalisables et applicables librement à tous types de données musicales. Nous reprenons donc les bases du domaine, avec des algorithmes élémentaires permettant de classer des objets représentés par des vecteurs de propriétés (*features*) dans un espace multidimensionnel (Manning et al., 2009).

Nous avons développé en collaboration avec Anders Vinjar (chercheur et compositeur indépendant basé à Oslo) une bibliothèque OpenMusic (OM-AI) articulant ces différents algorithmes autour d'une structure appelée VECTORSPACE. Cette bibliothèque contient également une base extensible d'algorithmes de description permettant d'extraire des caractéristiques (*features*) et de constituer des vecteurs personnalisés représentés dans cet espace à partir d'objets musicaux. Dans une première approche, une classification « non supervisée » (par exemple avec l'algorithme dit *k-means*) permet de regrouper en *clusters*, ou catégories, un ensemble de vecteurs. Une notion de contexte ou de proximité peut également être inférée lors de la phase d'apprentissage sur des séquences ordonnées (les objets sont-ils « proches », ont-ils été vus dans un même contexte ?). Enfin, des approches supervisées permettent de réaliser une classification à partir d'exemples, par comparaison de vecteurs entre de nouvelles données et les caractéristiques des différentes classes déterminées lors de l'apprentissage, ou encore par similarité avec un voisinage établi dans l'espace multi-dimensionnel (approche *k-nearest neighbors*). Une interface graphique permet la visualisation de l'espace de vecteurs en 2D ou 3D, suivant des dimensions choisies. Un ensemble de modules (également extensible et personnalisable) permet d'utiliser ces informations de classification ou de similarité dans les programmes visuels développés dans OpenMusic.

4. xmm : →<http://ircam-rnd.github.io/xmm/>

5. Dans une optique similaire, voir aussi par exemple les travaux du laboratoire Musique et Informatique de Marseille sur les Unités Sémiotiques Temporelles (Favory, 2007).

6. Alireza Farhang : *Traces of Expressivity : high-level score of data-stream for interdisciplinary works.*
→http://www.alirezafarhang.com/alirezafarhang/texts_traces_of_expressivity_en

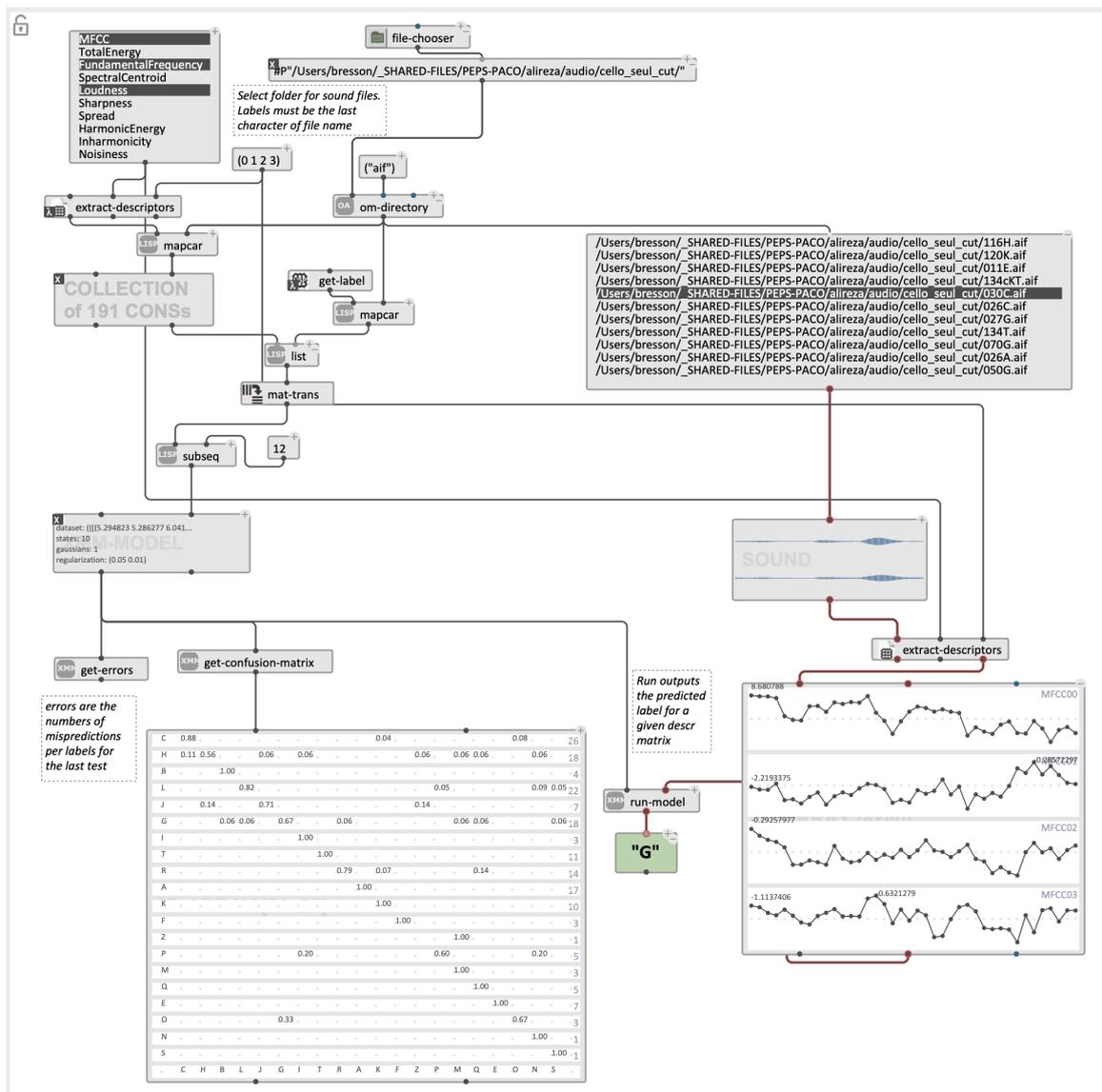


Figure 1 – OM-XMM : Reconnaissance de geste musicaux à partir de descripteurs audio.

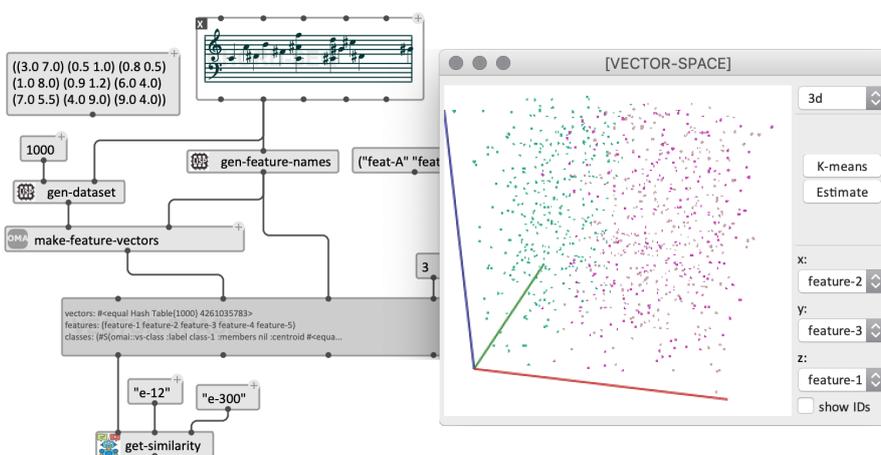


Figure 2 – OM-AI : Représentation et classification automatique d'objets musicaux dans un espace de vecteurs.

Modèles génératifs. Dans ce troisième volet, nous nous intéressons à l'utilisation de modèles génératifs basés sur l'apprentissage et la modélisation stylistique (Assayag et al., 1999) dans le cadre de la composition assistée par ordinateur. De tels modèles ont déjà été abordés et développés dans de précédents travaux de l'équipe Représentations Musicales, avec des bibliothèques comme LZ (encodage/décodage de séquences musicales avec reproduction des caractéristiques stylistiques), et l'*oracle des facteurs* au cœur du projet OMax (Assayag et Dubnov, 2004), utilisés notamment dans OpenMusic par le compositeur Shlomo Dubnov (Dubnov, 2008).

Nous nous concentrons à présent sur la dernière génération de ces systèmes génératifs (parfois qualifiés de systèmes « d'improvisation automatique ») à travers une connexion avec la bibliothèque DYCI2 (Nika et al., 2017b). DYCI2 est une bibliothèque écrite en Python, communiquant avec des environnements temps réel via OSC dans le but d'établir des situations de co-improvisation musicale humain-machine. Sur le modèle des systèmes qui l'ont précédé, un générateur DYCI2 apprend en construisant un oracle à partir de séquences musicales, étiquetées par un alphabet donné (un ensemble fini de *labels*), puis génère des nouvelles séquences en navigant dans cet oracle, guidé par des requêtes spécifiques exprimées sous forme d'une suite de *labels* (appelée « scénario ») (Nika et al., 2017a).

Dans le cadre d'un stage de L3 encadré par Jérôme Nika et Jean Bresson, Victoire Siguret (ENS de Lyon) a développé une bibliothèque C permettant d'invoquer les fonctionnalités principales de DYCI2, ainsi que l'interface entre Common Lisp et C correspondante dans une nouvelle bibliothèque OpenMusic (OM-DYCI2). Cette bibliothèque permet ainsi de réaliser un apprentissage de séquences *offline*, et de produire des nouvelles séquences suivant des requêtes dans OpenMusic — ces requêtes pouvant elles-même être issues d'autres processus génératifs, produits dans l'environnement de CAO et/ou avec DYCI2. Selon la base d'apprentissage (entrées MIDI, fichiers audio...), il est ensuite possible de reconstituer les séquences générées par extraction et concaténation de segments localisés dans les séquences MIDI ou audio originales (donnés comme « réponse » de l'agent de génération aux requêtes formulées).

Le couplage de cet outil avec les outils de segmentation et d'annotation automatique développés par l'équipe Analyse et Synthèse des Sons du laboratoire STMS permet de produire des bases de données d'apprentissage importantes à partir d'enregistrements sonores. Ce processus est en œuvre pour la production du projet *Lullaby Experience*⁷ du compositeur Pascal Dusapin (réfèrent scientifique IRCAM : Jérôme Nika), dont le matériau sonore sera produit en grande majorité grâce à la bibliothèque OM-DYCI2.

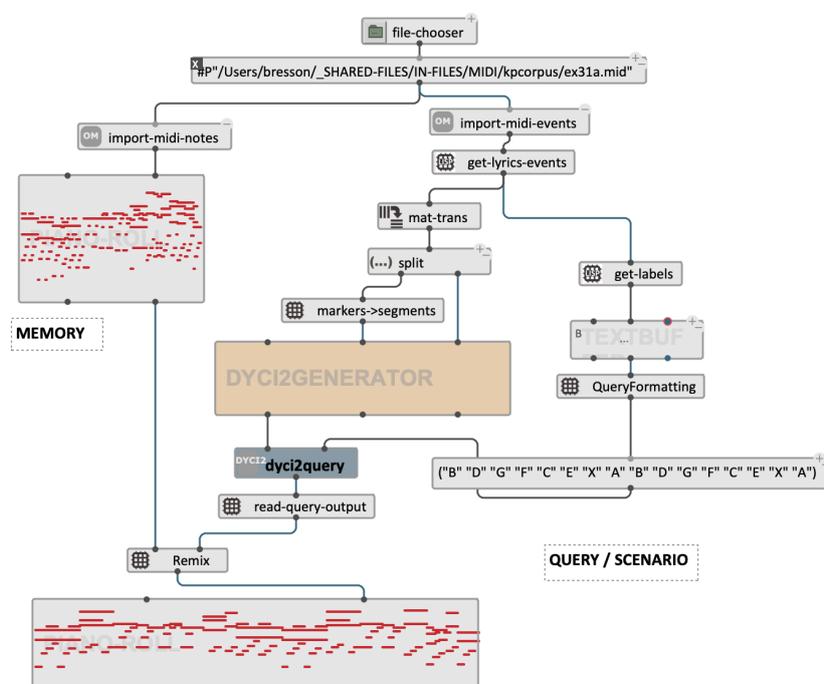


Figure 3 – OM-DYCI2 : Génération de séquences par apprentissage et guidage par scénario.

7. →<https://www.ircam.fr/article/detail/lullaby-experience/>

Workshop : Music Composition and Creative Interaction with Machine Learning

SMC'18 : International Sound and Music Computing conference
Cyprus University of Technology, Limassol, 6/07/18.

Programme :

- Diemo Schwarz (IRCAM) : Learning gestures from audio descriptors.
- Jerome Nika (Univ. La Rochelle / IRCAM) : Symbolic human-machine musical interaction
- Tetsuro Kitahara (Nihon University, Japan) : An improvisation support system with melody generation based on user-drawn melodic outlines
- Jean Bresson (IRCAM) : Machine learning in the compositional workflow.
- Anders Vinjar (Norway) : An exploration AI techniques for computer-aided composition
- Hugo Scurto (IRCAM) : Creating Human-AI Collaboration in Music Applications.
- Philippe Pasquier (Simon Fraser University, Canada) : Computer assisted composition and sound design at the Metacreation Lab

Page web du workshop : →<http://repmus.ircam.fr/paco/workshop-smc>



Figure 4 – Workshop *Music Composition and Creative Interaction with Machine Learning*, SMC'18, 6/07/2018.

Publications :

- J. Bresson, P. Best, D. Schwarz, A. Farhang : From Motion to Musical Gesture : Experiments with Machine Learning in Computer-Aided Composition. *International Workshop on Musical Metacreation (MuMe)*. Co-located with the *International Conference on Computational Creativity (ICCC'18)*, Salamanca, Espagne, 2018.
→<https://hal.archives-ouvertes.fr/hal-01815988>
- P. Best, J. Bresson, D. Schwarz : Musical Gesture Recognition Using Machine Learning and Audio Descriptors. *Proceedings of the International Conference on Content-Based Multimedia Indexing (CBMI'18)*, La Rochelle, France, 2018.
→<https://hal.archives-ouvertes.fr/hal-01839050>

Production logicielle (bibliothèques OpenMusic) :

- OM-XMM. Modèles probabilistes pour le suivi et la reconnaissance de gestes.
→<https://github.com/openmusic-project/om-xmm>
Documentation : →<https://github.com/openmusic-project/om-xmm/wiki>
- OM-AI. Boîte à outils pour l'apprentissage automatique et supervisé.
→<https://github.com/openmusic-project/omai>
- OM-DYCI2. Modèles génératifs guidés par scénario.
→<https://github.com/DYCI2/om-dyci2>
Documentation : →[https://github.com/DYCI2/om-dyci2/wiki/\[en\]](https://github.com/DYCI2/om-dyci2/wiki/[en])

Références

- Assayag, G., Dubnov, S. "Using Factor Oracles for machine Improvisation". *Soft Computing*, 8(9), 2004.
- Assayag, G. ; Dubnov, S. ; and Delerue, O. "Guessing the Composer's Mind : Applying Universal Prediction to Musical Style". *Proceedings of the International Computer Music Conference (ICMC)*, Beijing, China, 1999(a).
- Assayag, G., Rueda, C., Laurson, M., Agon, C., Delerue, O. "Computer Assisted Composition at IRCAM : From PatchWork to OpenMusic". *Computer Music Journal*, 23(3), 1999(b).
- Bresson, J., Agon, C., Assayag, G. "OpenMusic – Visual Programming Environment for Music Composition, Analysis and Research". *ACM MultiMedia – OpenSource Software Competition*, Scottsdale, USA, 2011
- Briot, J.-P., Hadjeres, G., Pachet, F. *Deep Learning Techniques for Music Generation*. Computational Synthesis and Creative Systems series, Springer, 2018.
- Cope, D. *Experiments in Musical Intelligence*. A-R Editions, Madison, 1996.
- Crestel, L. *Réseaux de neurones pour l'orchestration musicale automatique*. Thèse de doctorat, Sorbonne Université, Paris, 2018.
- Dubnov, S. "Memex and Composer Duets : Computer-Aided Composition Using Style Modeling and Mixing". In Bresson, J., Agon, C., Assayag, G. (Eds.) *The OM Composer's Book 2*. Editions Delatour / Ircam, 2008.
- Dubnov, S., Surges, G. "Delegating Creativity : Use of Musical Algorithms in Machine Listening and Composition". In Lee, N. (Ed.) *Digital Da Vinci : Computers in Music*. Springer, 2014.
- Favory, J. "Les unités sémiotiques temporelles". *Mathématiques et sciences humaines*, 178, 2007 [online].
- Françoise, J., Schnell, N., Bevilacqua, F. "A Multimodal Probabilistic Model for Gesture-based Control of Sound Synthesis". *ACM MultiMedia*, Barcelona, Spain, 2013.
- Ghedini, F., Pachet, F., Roy, P. "Creating Music and Texts with Flow Machines". In *Multidisciplinary Contributions to the Science of Creative Thinking*. Springer, 2015.
- Godøy, R. I., Leman, M. (Eds.) *Musical Gestures : Sound, Movement, and Meaning*. Routledge, 2010.
- Hiller, L., Isaacson, L. *Experimental Music : Composition With an Electronic Computer*. McGraw-Hill, 1959.
- Illescas, P. R., Rizo, D., Iñesta., J. M. "Learning to Analyse Tonal Music". *Proceedings of the International Workshop on Machine Learning and Music*, Helsinki, Finland, 2008.
- Lee, H.-K., Kim, J. H. "An HMM-based threshold model approach for gesture recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10), 1999.
- Manning, C. D., Raghavan P., Schütze, H. *An Introduction to Information Retrieval*. Cambridge University Press, 2009.
- Meredith, D. (Ed.) *Computational Music Analysis*. Springer, 2015.
- Nika, J., Chemillier M., Assayag, G. "ImproteK : introducing scenarios into human-computer music improvisation". *ACM Computers in Entertainment*, 14(2). Special issue on Musical metacreation, Part 1, 2017(a).
- Nika, J., Déguernel, K., Chemla–Romeu Santos, A., Vincent, E., Assayag, G. "DYCI2 agents : merging the "free", "reactive", and "scenario-based" music generation paradigms". *Proceedings of the International Computer Music Conference (ICMC)*, Shanghai, China, 2017(b).
- Pasquier, P., Eigenfeldt, A., Bown, O., Dubnov, S. "An Introduction to Musical Metacreation". *ACM Computers in Entertainment* 14(2), Special issue on Musical Metacreation, 2016.
- Schnell, N., Schwarz, D., Larralde, J., Borghesi, R. "PiPo, A Plugin Interface for Afferent Data Stream Processing Modules". *International Symposium on Music Information Retrieval*, Suzhou, China, 2017.