

Bilan de thèse à mi-parcours

Benjamin Lévy

26 mars 2012

Introduction

La thèse de doctorat d'informatique (école doctorale 130, EDITE) dont le titre officiel est le suivant : *Avatars capables d'écoute, d'apprentissage et d'interaction audio-musicales* porte sur la constitution, l'adaptation et la mise en œuvres effective de modèles performants d'écoute artificielle, d'apprentissage et d'interaction appliqués à la musique vivante ou enregistrée pour concourir à la formation d'avatars musicaux numériques, réactifs et capables de s'intégrer de façon naturelle dans des dispositifs variés tels que la scène, la (post-)production, la pédagogie musicales. Le projet comporte trois aspects principaux :

- l'écoute artificielle
- l'apprentissage génératif de séquences formelles
- l'intégration dans l'interaction

Le présent rapport dresse un bilan du chemin parcouru sur différentes facettes du projet environ 18 mois après le début officiel de la thèse. Un premier volet concerne le développement du logiciel OMax, noyau du projet, nous décrirons les explorations engagées à partir du ce cœur logiciel ainsi quelques unes des expériences de terrain accumulées. Enfin nous rappellerons les écrits émanés de ce travail.

1 Développement

Le logiciel OMax constitue le centre de ce travail de thèse. Ce logiciel existant sous diverses formes et évolutions depuis 2004 a été entièrement rénové entre 2009 et 2010 afin de le rendre plus réactif, interactif et de le doter d'une interface visuelle.

1.1 OMax pour le forum

L'un des premiers travaux à effectuer en arrivant sur le sujet a été de finaliser une version fonctionnelle, cohérente et stable d'OMax afin de pouvoir le diffuser (à travers le portail de l'IRCAM). Cette étape de développement a mené à un travail très important de modularisation du système permettant ainsi une analyse fonctionnelle fine et structurée des différents éléments composant le logiciel. Cette décomposition organique a ouvert la porte à la fois à la compréhension détaillée du système et au repérage des parties abouties et des parties méritant une exploration plus approfondie.

1.2 Version expérimentale

La finalisation d'une version stable d'OMax a également ouvert la porte, par le repérage de ses défauts fonctionnels et/ou organisationnels, à la mise en place d'une version expérimentale du système. Cette version, décomposée en éléments essentiels plus petits et conceptuellement plus cohérents apporte de nombreuses possibilités concrètes d'exploration. Par exemple, le stockage de la séquence d'entrée a pu être décorrélé du modèle d'apprentissage permettant multiplication des analyses (voir 2.1 et 2.2). La décomposition de la partie générative (quasiment monolithique dans la version distribuée du logiciel) en éléments fonctionnels séparés a permis également l'ouverture vers des stratégies variées de navigation (voir 2.3, 2.4 et 2.5). Cette version expérimentale alimente également les réflexions sur la pertinence et la généralité des éléments stockés et appris.

2 Expérimentations

Grâce à la mise en place d'une version expérimentale d'OMax, de nombreuses explorations ont pu être réalisées concernant la partie *analyse* et la partie générative du système. Celles-ci correspondent principalement aux aspects 1 et 3 tel que décrits dans le sujet initial de la recherche (voir **Introduction**). Il s'agit d'expériences sur l'écoute et les alphabets ainsi que d'expérience de générations régies par des règles différentes de celle existant initialement.

2.1 Alphabets

2.1.1 Diversification

Le modèle de connaissance utilisé dans OMax est basé sur un graphe (l'oracle des facteurs) qui nécessite un alphabet, c'est à dire une classification stricte (classes d'équivalence) des événements constituant une séquence. Au départ totalement intégré au stockage de la séquence d'entrée, l'externalisation de ce processus de classification dans la nouvelle version expérimentale du système a permis une totale indépendance du travail d'extraction d'information puis de regroupement (clustering) préparatoire à la l'identification (comparaison stricte) des unités.

Une des premières direction d'expérimentation a donc été de multiplier (à volonté) les différents alphabet ainsi que leur constitution. Par exemple, le passage d'une échelle de notes (pitch) en demi-tons à une échelle en quart-de-tons ou bien la construction du modèle sur la durée des notes (quantifiée) auparavant (quasiment) inutilisée.

Plus finement encore, plusieurs jeux de paramètres de classification — grâce à une distance euclidienne pondérée — des vecteurs de dimension 10 décrivant le timbre du son (MFCCs) peuvent être testés en parallèle sur une même extraction afin d'obtenir plusieurs analyses (plus ou moins fines) de la séquence d'entrée. On peut ainsi obtenir un ensemble de modèles (construits avec le même algorithme) décrivant le même signal d'entrée avec différents niveaux de détails.

2.1.2 Alphabet évolutif

Une des améliorations principales de la dernière version d'OMax est l'implémentation d'un système de découverte de l'alphabet (spectral). Cet alphabet adaptatif diffère notoirement des autres alphabets précédemment utilisés qui prédefinissent une classification immuable pour la construction du modèle. Au contraire, l'alphabet adaptatif autorise la mutation de symboles en fonction des éléments déjà appris. Ce mécanisme et ses conséquences reste à étudier plus finement mais s'avère déjà très intéressant puisqu'il donne de très bon résultat et semble correspondre au processus cognitif de répétition et de mémorisation des éléments déjà énoncés.

2.1.3 Enrichissement

Dans cette direction également, mais considérant l'apprentissage a priori d'un corpus, complété par la suite ; l'adaptation de l'alphabet peut être également contrôlée afin de favoriser les hybridations entre le corpus et le complément (temps-réel). Ceci s'est révélé très efficace pour augmenter les possibilités de croisement avec le corpus.

2.2 Écoutes parallèles

Le travail sur les différents alphabets utilisés et utilisables et la multiplication des natures d'alphabet pousse la recherche dans le sens de l'augmentation des possibilités d'écoute du logiciel. En dehors des écoutes menant à la construction d'un alphabet et d'un modèle de séquence, l'écoute de paramètres additionnels tels que la dynamique (nuance) ou le rythme commence à être explorée. Ces informations "complémentaires" peuvent servir au guidage ou à l'expression de contrainte pour la génération (voir 2.3).

2.3 Génération avec contrainte

L'une des possibilités de contrôle (musical) d'un système génératif tel qu'OMax est l'introduction de guidage — ou plus informatiquement parlant, de contraintes — permettant de maîtriser plus finement le discours (musical) du système. Ces contraintes peuvent s'exprimer sur des paramètres "internes", c'est à dire de nature identique à ceux qui ont permis de construire le modèle, par exemple en réduisant les motifs de notes "autorisés" lorsque le graphe (facteur d'oracle) a été construit sur les notes. Des expériences ont été réalisées en ce sens mais elles semblent moins prometteuses que les expériences sur les contraintes externes.

En effet, les contraintes pour guider le discours musical peuvent également s'exprimer sur des paramètres "externes" au modèle, c'est à dire des descriptions différentes de celle(s) utilisée(s) comme alphabet. Et l'on peut ainsi tout en générant des suites de notes ou de timbres pousser le discours musical vers un registre particulier (grave/aiguë) ou dans une dynamique choisie (nuance). De nombreuses expériences dans cette direction donnent déjà des résultats très encourageants.

2.4 Génération multi-descriptive

La multiplication des écoutes décrites au paragraphe 2.1.1 sert également la partie générative d'OMax. Des expériences ont été entreprises dans lesquelles la génération ne repose plus sur une seule des descriptions (menant à un modèle de séquence) mais plus plusieurs d'entre elles. De nombreuses stratégies sont alors possibles. Le choix du modèle peut être laissé à l'utilisateur, il peut également être défini par un "profil" de génération, solutions dans lesquelles on priviliege l'alternance entre les modèles. Mais l'utilisation des différents modèles sous-jacents peut aussi être concurrente. Ceci peut mener à une vision du système proche d'une base de données dans laquelle, à la demande, les différentes descriptions fournissent des possibilités de génération qui sont ensuite rassemblées et examinées.

2.5 Explorations parallèles

Enfin, une autre voie d'expérimentation a été abordée pour la partie générative d'OMax, celle de l'exploration parallèle de plusieurs chemins dans le modèle. Cette possibilité ouvre de vastes domaines comme la recherche de chemins particuliers dans un graphe ou l'apprentissage par renforcement (pondération de chemins explorés et intéressants). Quelques premières expériences

ont été menées en donnant un objectif particulier aux agents chargés de l'exploration de chemins puis en collectant le résultat de tous les agents mis ainsi en concurrence.

3 Expériences musicales

3.1 Collectives

→Nine Spirit

3.2 Modales

→Coleman

3.3 Théâtrales

→Lazar

3.4 Autonomes

→Kimura

4 Écrits

Depuis le début de la thèse, hormis les nombreuses présentations formelles ou informelles du travail en cours, à diverses occasions, trois principaux documents écrits ont été rédigés.

4.1 Documentation d'OMax

Le développement d'une version finalisée et stable d'OMax a nécessité également la rédaction d'une documentation pour l'utilisateur. À la fois vulgarisé et minutieux, ce texte de plus d'une vingtaine de page a également permis de mettre en lumière les points clairs et formalisés du système ainsi que les points moins explorés ou plus expérimentaux.

4.2 Plan de thèse

Un an après le début de la thèse, une première ébauche de plan de manuscrit a été préparée. Cet écrit sera bien entendu amener à évoluer au cours du travail. Il a permis cependant de structurer et organiser les directions prises au cours des expérimentations.

4.3 Article

Un article décrivant les profondes améliorations présentes dans la dernière version d'OMax (architecture, visualisation, description et alphabet spectral) a également été rédigé et accepté à la conférence internationale *New Interfaces for Musical Expression* 2012 se déroulant à Ann Arbor (Michigan). Voici le résumé (en anglais de l'article) :

OMax is an improvisation software based on a graph representation encoding the pattern repetitions and structures of a sequence, built incrementally and in real-time from a live Midi or Audio source. We present in this paper a totally rewritten version of the software. The new design leads to refine the spectral listening of OMax and to consider different methods to build the symbolic alphabet labeling our symbolic units. The very modular and versatile architecture makes possible new musical configurations and we tried the software with different styles and

musical situations. A novel visualization is proposed, which displays the current state of the learnt knowledge and allows to notice, both on the fly and a posteriori, points of musical interest and higher level structures. These new developments raise numerous and interesting research directions around OMax.