Contexte et objectifs

Contexte

Depuis les débuts de l'informatique, la musique à su utiliser les nouveaux outils mis à sa disposition avec différents types de logiciel permettant d'assister le musicien, l'ordinateur constituant même de nos jours un instruments à part entière avec l'essor de la musique électronique et la MAO en général (Musique Assistée par Ordinateur).

En informatique musicale on distingue grossièrement deux types de logiciels. D'abords, les séquenceurs ou DAW (Digital Audio Workstation) qui permettent l'enregistrement, le montage, le séquencement et la gestion d'effets audio, parmi lesquels on peut citer Ableton, Cubase, GarageBand et FruitLoops. Ensuite, les logiciels d'édition de partition qui permettent au musicien de saisir de façon plus ou moins intuitive un morceau de musique et bien souvent de faire jouer cette partition par des banques de sons ou des sons synthétiques. Parmi ceux-ci, on peut citer Finale, Sibelius, LilyPond ou GuitarPro. Notons qu'une grande partie de ces logiciels est capable d'écrire une parition à partir d'un fichier MIDI importé.

Problème posé

Ces derniers nécessite systématiquement la saisie manuelle, par l'utilisateur - généralement un musicien - de toutes les notes de la partition. Cette saisie est souvent fastidieuse, d'autant plus quand la partition n'est pas triviale à écrire. Que ce soit dans le cas d'une composition (pour laquelle aucune partition n'existe encore) ou dans le cas d'une tentative de reproduction (pour laquelle aucune partition n'est accessible), de nombreux musicien souhaite créer une partition, soit de tête soit à partir d'un modèle audio qu'ils chercheront à reproduire, notamment grâce à la possibilité de jouer la partition.

Notre projet est donc de répondre à ce problème en facilitant le travail du musicien et en générant à partir d'un fichier audio, une partition aussi proche que possible de la réalité. Cette tâche est notamment connue sous le nom de « transcription automatique » d'un morceau de musique (*automatic music transcription*). L'enjeu est de détecter d'une part « quand » des notes sont jouées, que ce soit en déterminant le tempo, le début et la fin des notes. D'autre part, il faut détecter « quelles » notes ont été jouées (*pitch detection*).

Schéma fonctionnel

Le schéma ci-dessous représente la composition de notre système de transcription automatique de la musique. On y repère différents blocs qui seront détaillés dans ce rapport. Ces blocs sont les suivants :

* La détection d'onsets (*onset detection*) qui permet de déterminer le début d'une note. Quand deux notes se suivent consécutivement, le début de la seconde correspond à la fin de la première. Ce bloc inclus implicitement la détection d'offset (*offset detection*) c'est à dire la détection d'un silence entre deux notes jouées consécutivement.
* La segmentation du signal audio vis à vis des onsets ET offsets pris dans l'ordre chronologique.
* Le calcul du tempo à partir de la sortie de l'onset détection et faisant appel au bloc suivant :
* L'analyse de la composition rythmique, permettant d'assigner à chaque segment, la durée musicale la plus probable compte tenu du tempo.
* La construction de mesures régulière, en corrigeant ainsi d'éventuelles erreurs apparues dans le bloc précédent.
* La détermination des tons, c'est à dire de la note dans la gamme tempérée.
* La détermination des octaves, sans tenir compte du ton trouvé.

État de l'art

Le problème est traité depuis les années 70 environ, notamment grâce à des transferts de solutions techniques venant du traitement de la parole. Cependant, en cherchant à résoudre le problème, on se heurte rapidement à différents problèmes.

* Pour la musique polyphonique, ou la détection de plusieurs notes simultanées (*multipitch detection*), aucune solution proposée à ce jour ne propose de résultats convaincants.
* Concernant la détection de mélodie ou bien le cas de la musique monophonique, le timbre des instruments, la perception humaine des sons ou le bruit provoque encore des erreurs relativement inacceptables.
* Concernant la détection de tempo, l'interprétation humaine est relativement subjective et pose donc problème.
* Concernant la détection d'onsets, là encore l'oreille humaine est bien meilleure qu'un algorithme de nos jours et est capable de percevoir des nuances très subtiles dans l'apparition des notes que les algorithmes manqueront.

Globalement, de nombreuses méthodes existent pour tenter de résoudre tous ces problèmes. Notre travail a majoritairement consisté à chercher les méthodes pour lesquelles nous obtenions les meilleurs résultats, tentant d'améliorer une solution existante (dans le cas de l'estimation de tempo), voire en créant complètement notre méthode selon notre expérience et notre intuition (cas de l'analyse et la correction rythmique).

Les résultats que l'on peut espérer seront loin d'être parfait d'autant que les erreurs entre les algorithmes se combineront entre elles. En effet, une note ne sera considérée comme bien détectée que si on trouve le bon ton, à la bonne octave et correspondant à la bonne durée jouée à un tempo bien estimé !

Solutions analogues

Certains logiciels remplissant plus ou moins cette tâche sont diffusés à titre commercial ou gratuit et disponibles depuis les années 2000 environ. Parmi eux, on peut cité les logiciels/applications reconnaissant les accords : Chordec, AnySong Chord Recognition, Mémo musical (de Apple).

Ces logiciels permettent de transcrire ce que l'on peut décrire comme une partition simplifiée incluant la notion de tempo, de rythmique mais d'un point de vue harmonique, on se contente de détecter un accord c'est à dire une classe de notes jouées simultanément sans notion d'octave. Ces logiciels sont conçus pour la musique monophonique mais contenant plusieurs notes jouées en même temps (un instrument, plusieurs notes à la fois).

D'autres logiciels traitent plus directement notre problème. On peut citer de nombreuses solutions labellisées comme des « convertisseurs mp3 vers MIDI », ce qui correspond à ce que nous cherchons à faire, le format MIDI permettant bien de pouvoir obtenir une partition par la suite. On peut également citer les logiciels payants Intelliscore et Sibelius qui adressent directement notre problème au sein d'un éditeur de partition.

Objectifs et limites posées

L'objectif de ce projet de fin d'étude est de proposer un système de transcription automatique de bonne qualité, c'est à dire qu'il faut que la partition créée permette de reconnaître le morceau dont elle est issue.

On l'a dit, la transcription musicale automatique ne peut pas être réalisée au sens strict du terme dans l'état de l'art actuel. Pour s'en approcher, il faut réduire son cadre avec certaines limitations et optimiser le système pour obtenir les meilleurs résultats possibles dans ce cadre. Pour cela nous nous posons, dans ce projet les contrainte suivantes :

* L'algorithme de *pitch-detection*, que nous appellerons Analyse Harmonique, sera **optimisé pour la guitare**. Cela signifie que seule les notes allant du E2 (82.4Hz) au D#6 (1244.5Hz) pourront être détectées. Le choix de la guitare comme instrument exclusif aura également un impact sur la détection d'onset.
* Cet algorithme ne traitera également que le problème de la détermination d'**UNE SEULE** note jouée à la fois.
* L'estimation de tempo se fera dans l'intervalle de **55 bpm et 180 bpm.**
* La détermination des durées musicales, que nous appellerons Analyse Rythmique, se fera pour un tempo donné. La durée musicale **la plus courte possible sera la double croche**, toute les autres durées musicales possibles seront des multiples de la durée d'une double-croche (croche, croche pointée, noire, etc.)
* L'analyse rythmique sera également optimisée pour construire des **mesures 4/4** (soit l'équivalent de 4 noires par mesure)
* La détection d'offsets (de fin de notes, c'est à dire de l'apparition de silences) ne sera pas évaluée et ne fera pas l'objet d’exigences de qualité.