



Projet de Fin d'études : Sujet 2

Réalisé par :
Zaher Hamadeh
Nicolas Jbeyli

Encadré par :
Rima Kilany Chamoun

I- Introduction

II- Coup d'oeil sur l'application

III- Reconnaissance

III-1 Contexte

III-1.1 Le son musical

III-1.2 La gamme

III-2 La constitution du spectre

III-2.1 La lecture du fichier audio

III-2.2 La transformée de Fourier

III-3 La récupération de la fréquence fondamentale

III-3.1 Fréquence à puissance maximale

III-3.2 Ressemblance maximale

III-3.3 Maximums locaux

III-4 Des méthodes à implementer au futur

III-4.1 K-mean Cluster

III-4.2 Reconnaissance intelligente

III-4.3 Reconnaissance rythmique

IV- Synthèse

IV-1 Fichier MIDI

IV-1.1 Format

IV-1.2 General MIDI

IV-2 VST

IV-2.1 Instruments VST

IV-2.2 VST SDK

V-Références

Introduction:

Dans un monde où les réseaux sociaux sont en plein essor et semblent couvrir tout les mediums, nous avons trouvés une faille qui demeure inexploitée. Notre but est de créer un outil qui permet à un utilisateur de créer sa propre petite mélodie et de la publier pour qu'elle puisse être écoutée par son cercle social.

Il nous arrive souvent d'avoir une mélodie jouant dans notre tete. Une mélodie quelconque qui n'a pas d'origine spécifique. On se demande souvent que serait devenu de cette mélodie si nous avions un piano, une guitare ou même un saxophone pour la concrétiser, mais le seul instrument qu'on sait jouer est le triangle. Notefy permet à toute personne de jouer tout instrument.

La solution que nous avons trouvé est la suivante; une application mobile qui transforme le fredonnement de l'utilisateur en une série de notes musicales. Il pourra ensuite sélectionner un instrument pour jouer ces notes et en superposer d'autres. Une fois qu'il est satisfait de sa création, il pourra la partager sur son profil pour son plaisir personnel et le plaisir de ses amis, sur ce réseau social, qui pourrons applaudir sa création.

Ce projet consiste à crée un réseau social de toutes pièces et donc nous avons du pain sur la planche. Notre projet se decompose en plusieurs étapes :

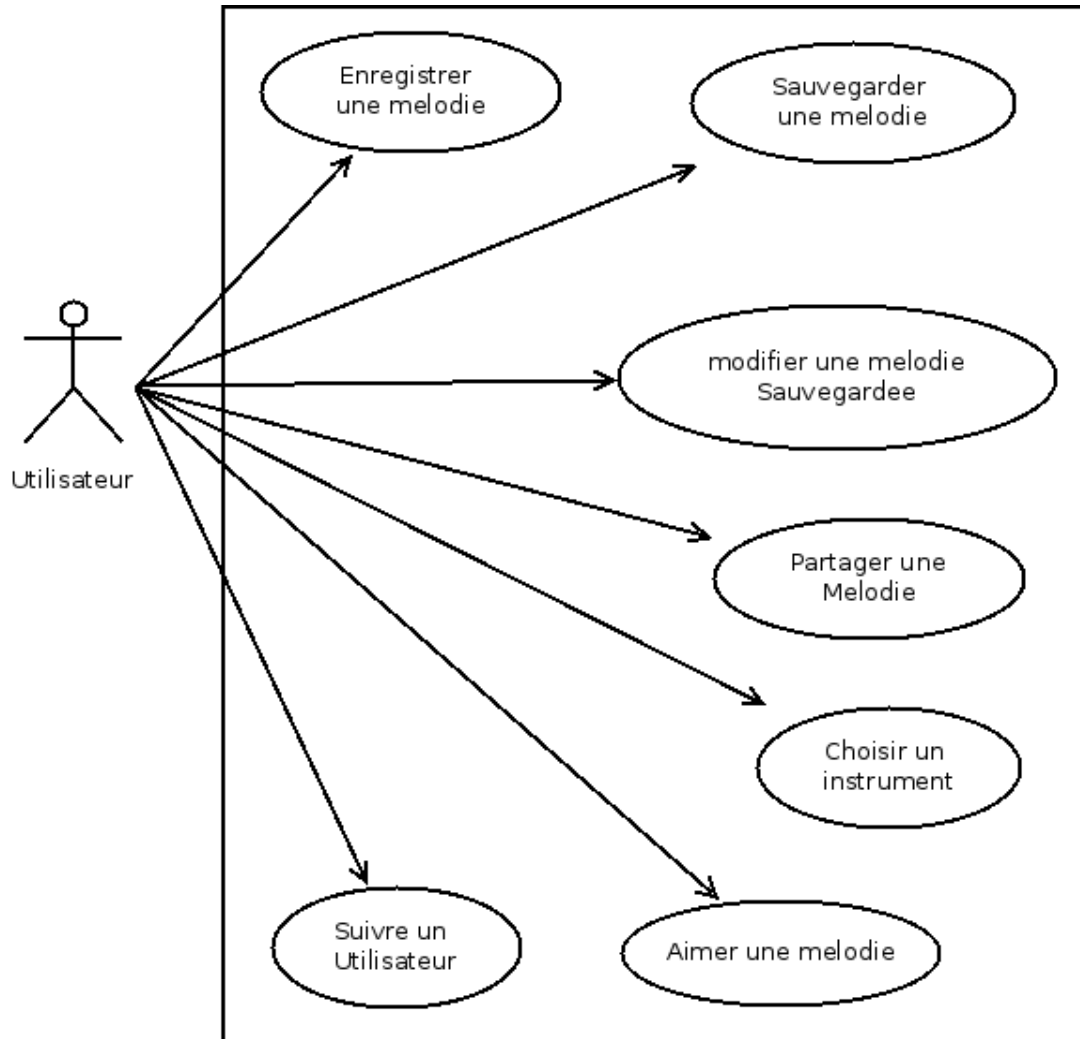
- Créer le module qui converti le fredonnement en notes musicales.
- Créer le module qui convertit les notes en un format audio
- Créer l'application pour iOS
- Créer les profils des utilisateurs et leurs bases de données
- Créer le réseau social et sa base de données

Notre Projet, a pour but de permettre à un utilisateur non professionnel de produire une mélodie de haute qualité.

II- Coup d'oeil sur l'application

Tout d'abord on a du définir les actions que pourra effectuer notre utilisateur pour créer la navigation dans l'application ainsi que les fonctionnalités de cette dernière.

Ce diagramme de cas d'utilisations illustre les actions qui pourront être accomplies par un utilisateur quelconque.



Ces actions sont divisées en deux categories :

- Locales : Enregistrer une mélodie, sauvegarder une mélodie, modifier une mélodie sauvegardée, choisir un instrument.
- Sur le réseau social : Partager une mélodie, aimer une mélodie, suivre un utilisateur.

La difference majeure entre ces deux categories est le fait que les actions sur le réseau social nécessitent une connection internet alors que les autres non.

La mélodie est composée de plusieurs morceaux, chaque morceau est joué avec un instrument. Une fois que l'utilisateur enregistre un morceau, il peut modifier l'instrument ou supprimer ce morceau. Une fois qu'il fini il peut soit sauvegarder sa mélodie en local ou la partager sur le réseau social.

Nous avons commencer par developper les fonctionnalités qui ne nécessitent pas un accès à internet. Jusqu'à present, l'utilisateur peut enregistrer un morceaux, modifier son instrument, sauvegarder une mélodie et modifier une mélodie sauvegardée. La prochaine étape sera l'integration de l'algorithme qui transforme un fichier sonore en un fichier MIDI.

III- Reconnaissance

III-1 Contexte :

Le traitement du signal musical semble constituer une partie mineure du large terrain de la reconnaissance vocale. Cependant, les signaux musicaux possèdent des caractéristiques acoustiques et structurelles qui les distinguent des autres signaux comme celui de la parole.

III-1.1 Le son musical

Le son musical possède quatre caractéristiques principales:

- La hauteur: la note perçue, scientifiquement c'est la fréquence fondamentale du signal.
- La durée: la longueur en temps de l'évènement musical.
- L'intensité: la puissance avec laquelle le son musical a été produit
- Le timbre: la couleur ou la qualité distinctive du son musical propre à chaque instrument qui produit le son. Scientifiquement, cela correspond à la distribution des harmoniques dans le spectre de fréquences du son.

Ainsi, le son musical est un signal avec une fréquence fondamentale bien définie. Il est formé d'une série harmonique de sinusoïdes à des fréquences multiples de la fréquence fondamentale. L'être humain perçoit ce signal comme un évènement unique qui correspond à une note musicale.

Trouver une définition stricte de la musique peut être problématique. Cependant, la mélodie en est un élément omniprésent et essentiel. Une mélodie est une séquence de son, une succession de notes avec des hauteurs distinctes.

III-1.2 La gamme

La gamme est un aspect commun qui fait partie des conventions musicales. Une gamme est formée d'un ensemble de notes distinctes qui se répètent à chaque octave. L'octave d'une note est une note dont la fréquence fondamentale en est le double. La musique occidentale est basée sur la gamme au tempérament égale, qui divise l'octave en 12 pas égaux sur une échelle logarithmique. Chaque fréquence est $2^{1/12} \approx 1.06$ fois plus grande que la précédente, cette intervalle s'appelle un demi-ton.

La gamme majeure traverse l'octave utilisant sept notes (les touches blanches du piano), ce sont les fameux Do, Ré, Mi, Fa, Sol, La, Si. Les touches noires (5 touches) sont nommées d'après la note directement en-dessous (Do dièse est un demi-ton plus haut que la Do) ou en-dessus (Ré bémol est un demi-ton plus bas que la Ré). La note la plus basse du piano est la La0 (27.5 Hz), la plus haute est la Do8 (4186 Hz).

L'étendue vocale de l'être humain est connue. Elle s'étend en moyenne de la Do2 (65.41 Hz), à la Do6 (1046.50 Hz)

III-2 La constitution du spectre

Notre but principal est de pouvoir récupérer les fréquences fondamentales (f_0) du signal contenant la mélodie. Mais d'abord il faut établir le spectre du signal, et connaître la puissance de chaque fréquence présente dans le signal.

On ne s'intéresse pas au spectre du signal complet. Cela nous indique toutes les fréquences présentes dans la mélodie chantée, sans définir quand les notes sont chantées. Pour cela on divise le signal en des fenêtres suffisamment petites pour s'assurer de l'existence d'une note unique dans cette intervalle de temps et suffisamment grande pour obtenir un spectre précis. La manipulation se fait sur chaque fenêtre.

III-2.1 La lecture du fichier audio

Un fichier audio non comprimé contient les échantillons du signal. La fréquence d'échantillonnage (F_s) peut prendre les valeurs 16000 Hz, 24000 Hz, 44100 Hz, etc., chaque échantillon est codé sur 8, 16, 24, ou 32 bits. Un fichier audio peut sauvegarder le signal sur une ou plusieurs chaînes (mono, stéréo...), tous ces informations se trouve dans l'entête du fichier.

Après la lecture des échantillons, la valeur de ces derniers est normalisée et transformée en un nombre réel compris entre -1 et 1. Ensuite, on divise le signal en des fenêtres de taille 4096 ou 8192 échantillons (≈ 100 ms dépendant de la fréquence d'échantillonnage). Une taille puissance de 2 optimise la transformée de Fourier rapide. Ainsi, en appliquant Fourier sur chaque fenêtre on pourra obtenir le spectre de chaque intervalle.

III-2.2 La transformée de Fourier

L'outil le plus populaire pour obtenir un spectre du signal est la transformée de Fourier. Dans notre cas la transformée de Fourier discrète

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}}$$

X_k est complexe, son module nous indique la puissance de chaque fréquence dans le signal, et la phase précise l'alignement des fréquences dans le signal. Le module est seulement utilisé. X est un tableau et k prend les valeurs $[0-K]$ avec $k=N/2$ car X est symétrique, N étant la taille du signal. L'indice k représente la fréquence $k \cdot F_s/N$, F_s étant la fréquence d'échantillonnage.

Calculer cette définition directement requiert une complexité $O(N^2)$, pour cela on va implémenter l'algorithme de la transformée rapide (FFT) dont la complexité est réduite à $O(N \log(N))$.

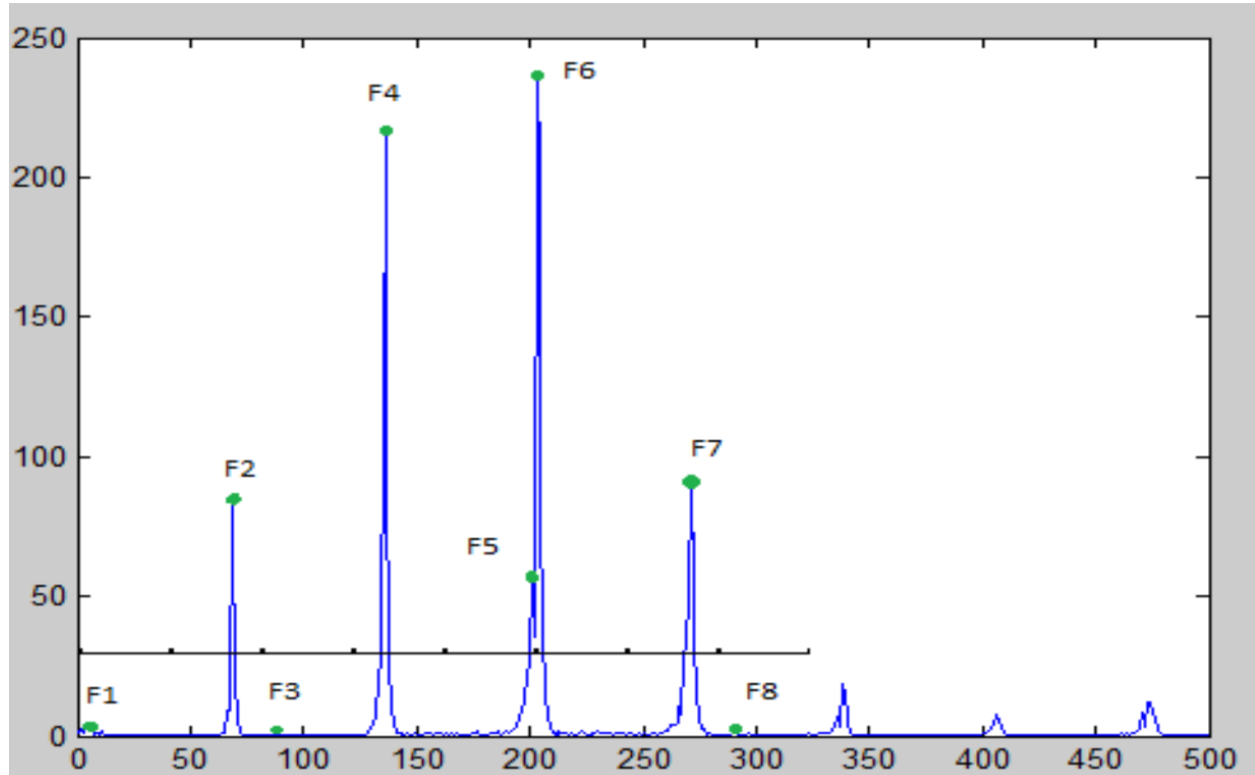


fig 1. Spectre d'une voix féminine chantant la note de fréquence fondamentale $66 \cdot F_s / N$

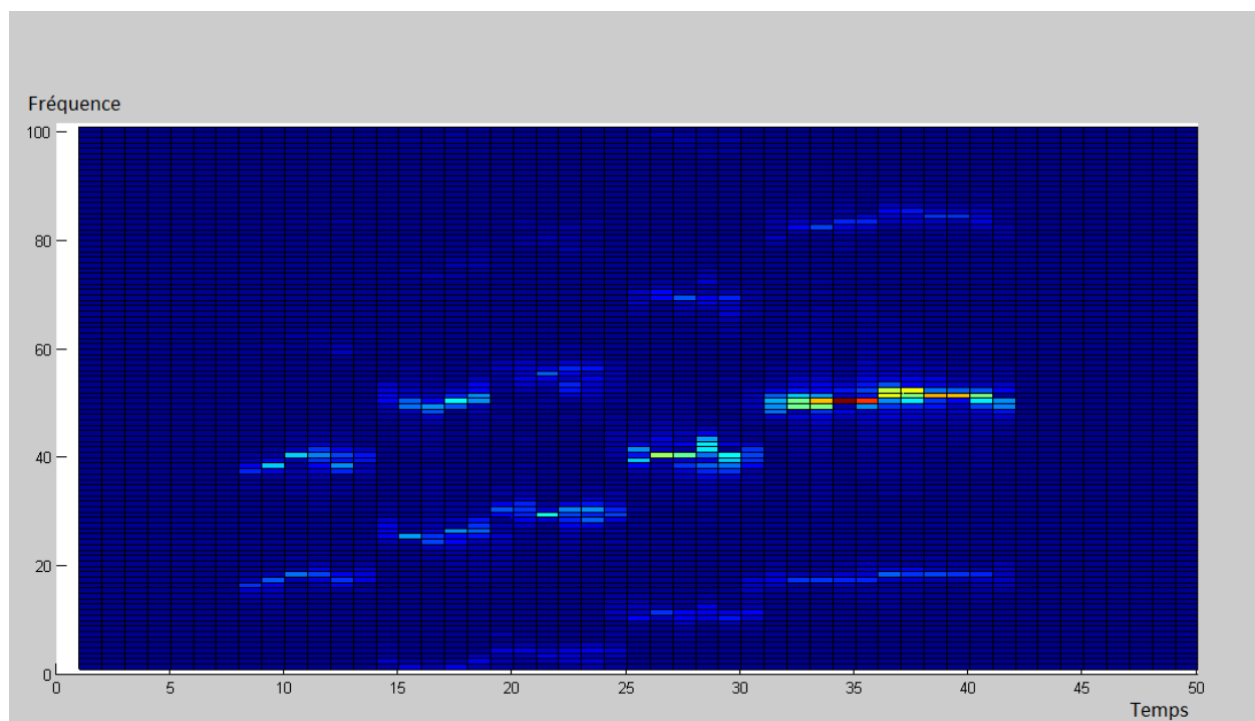


fig 2. Variation du spectre en fonction du temps

III-3 La récupération de la fréquence fondamentale

Pour reconnaître la fréquence fondamentale d'un spectre, nous avons mis en place différentes méthodes dans le but de trouver la solution la plus satisfaisante. Ce travail est toujours en œuvre.

III-3.1 Fréquence à puissance maximale

Une première méthode triviale est de choisir la fréquence dont la puissance est maximale. L'étendue vocale de l'être humain est connue. On cherche donc une fréquence appartenant à cette étendue. Cependant la fréquence fondamentale n'est pas nécessairement la plus présente, surtout si le son est riche en harmonique. Par exemple : le spectre de la note Do3 (130.8 Hz), contient la fréquence fondamentale $f_0=130.8$ Hz ainsi que les harmoniques $2f_0$, $3f_0$, $4f_0$... la fréquence fondamentale du Do4 (octave plus haut que Do3) est $f_0=261.6$ Hz, et appartient donc au spectre du Do3. Les deux notes (Do3 et Do4) se trouvent toutes les deux dans l'étendue vocale de l'être humain. La Do3 peut être confuse pour une Do4 si la puissance de la première harmonique est plus importante que celle de la fondamentale. Cela induit à des résultats non exacts.

III-3.2 Ressemblance maximale

Le spectre des notes musicales peut être connu. On crée des spectres à raies automatiquement pour chaque note, de telle sorte que leur puissance soit uniformisée. Ainsi, on reconnaît les notes du signal en cherchant la ressemblance maximale. Cependant, cette méthode s'est prouvée inefficace pour multiple raisons:

- La différence de timbre induit aux erreurs
- Des difficultés existent si on désire récupérer les silences dans le signal
- Le non professionnalisme de l'utilisateur lui permet de chanter des notes non standardisées. Un son de fréquence 250 Hz se trouve entre le Si3 (246.94 Hz) et le Do4 (261.63 Hz), dans la méthode précédente on récupère la fréquence 250 Hz et on la rapproche à une note standard, tandis que dans cette méthode l'intensité de la note chantée est perdue

III-3.3 Maximum locaux

Cette méthode est une amélioration de la première. Elle recherche les maximums locaux en espérant trouver la fréquence fondamentale et ses harmoniques. On choisit une fenêtre égale en fréquence à la note la plus basse que l'être humain peut chanter. On est sûr d'obtenir au plus un seul pic signifiant par fenêtre. On calcule la différence des fréquences à puissance maximale locale et on calcule le produit de leur puissance. Ainsi, on reconnaît la différence la plus présente qui correspond à la fréquence fondamentale.

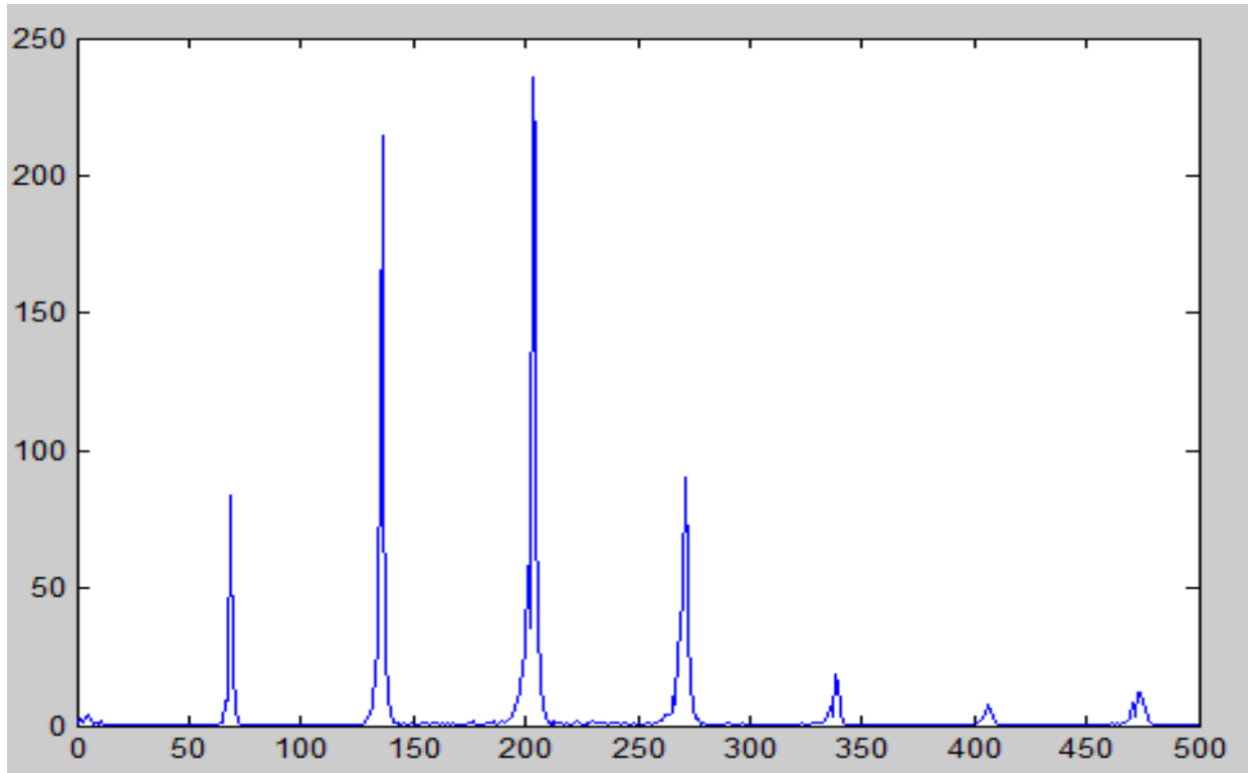


fig 3. Spectre et maximums locaux

D'après la figure 3. Supposons $F_{21}=F_2-F_1$

Et $P_{21}=P_2 \cdot P_1$ (avec P_1 la puissance de la fréquence F_1)

Et C_{21} le couple (F_{21}, P_{21}) ,

On additionne les P dont leur F est proche (moins qu'un demi-ton de différence), et on trouve quel F est le plus présent.

On constate que $F_{42} \approx F_{54} \approx F_{64} \approx F_{75} \approx F_{76} \approx f_0$

Et que évidemment la somme $P_{42} + P_{54} + P_{64} + P_{75} + P_{76}$ est beaucoup plus importante que tous les autres sommes donc f_0 est la différence la plus présente donc la fréquence fondamentale.

Actuellement nous avons implémentés cette méthode et nous sommes entrain de l'utilisée. Elle nous donne des résultats plus-ou-moins satisfaisante.

III-4 Des méthodes à implémenter au futur

On est toujours dans en quête de méthodes et d'algorithmes plus efficace, plus performant et plus satisfaisant.

III-4.1 K-mean Cluster

C'est une amélioration de la troisième méthode. Elle consiste à diminuer la taille de la fenêtre et améliorer la recherche de la différence à puissance maximale grâce à la méthode de K-mean Cluster. Cela peut rendre les résultats plus exactes.

III-4.2 Reconnaissance intelligente

C'est une amélioration de la deuxième méthode. Elle consiste à reconnaître le spectre en le comparant à celui de la voix de l'utilisateur. L'application apprendra aussi à reconnaître de mieux en mieux la voix de son utilisateur.

III-4.3 Reconnaissance rythmique

Pour permettre la reconnaissance des sons à attribuer aux instruments non tonals (batterie, percussion).

IV- Synthèse

On a récupéré la mélodie. Il nous reste à jouer la mélodie par un instrument. Nous verrons dans cette partie les techniques pour synthétiser une mélodie à partir d'une suite de notes et un instrument.

IV-1 Fichier MIDI

MIDI ou "Musical Instrument Digital Interface" est un protocole de communication et de commande permettant l'échange de données entre instruments de musique électronique, un ou plusieurs de ces « instruments » pouvant être des ordinateurs.

IV-1.1 Format

Les échanges se font en forme de messages numérique. Un message peut indiquer le déclenchement d'une note ou le changement d'instrument. Un fichier MIDI regroupe dans un fichier une succession de messages MIDI. Ces messages sont regroupés en canaux pour permettre à plusieurs mélodies d'être jouées simultanément de façon synchrone.

Ce format est standard et a été défini en 1988. Il est très compact, pour cela il a souvent été utilisé comme bande de son dans les anciens jeux vidéo. Il stock les commandes MIDI (le déclenchement des notes) en y ajoutant des informations temporelles. La majorité des lecteurs audio supporte ce format.

IV-1.2 General MIDI

La norme General MIDI a été une tentative de normaliser les sonorités et les instruments utilisés par le format MIDI. Un message MIDI qui indique l'instrument utilisé par un canal contient un entier lié à un instrument. Par exemple 1 correspond au piano, 48 au violon.

IV-2 VST

Les instruments du General MIDI ont un son très électronique et non réel. Le choix d'utilisation des VST a pour but d'assurer une qualité sonore satisfaisante pour l'utilisateur.

VST (Virtual Studio Technology) est un protocole créé par la société Steinberg. Les plugins VST sont des modules de conversion de données audio et midi en temps réel.

IV-2.1 Instruments VST

Ces instruments génèrent de l'audio à partir d'un fichier MIDI et des échantillons réels. Ils sont utilisés par une application VST hôte et simulent d'une manière réaliste n'importe quel instrument.

La société a développé des SDK pour permettre aux développeurs (de la société ou tierce personne) de pouvoir créer de nouveaux instruments et effet VST et des hôtes VST.

IV-2.2 VST SDK

Ce SDK est constitué de fichiers C++ qui offrent des outils pour manipuler les fichiers MIDI et audio.

Il existe un grand nombre de projet open source qui utilisent ce SDK dont on s'est inspiré pour développer notre premier synthétiseur MIDI-audio .

V- Référence

IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, VOL. 0, NO. 0,
2011

<http://www.csw2.co.uk/tech/midi2.htm>

Audio Interchange File Format AIFF-C, Apple Computer, Inc.

Foundations and Trends in
Computer Graphics and Vision
Vol. 4, No. 1 (2008) 1–73

<http://www.steinberg.net/en/company/developers.html>

<https://codeload.github.com/teragonaudio/MrsWatson/zip/master>