# Analyse harmonique

Cette partie intervient à la suite de la segmentation des onsets. Elle consiste à déterminer la note et l’octave de l’onset extrait (correspondant à une seule note). Il existe de nombreuses méthodes différentes que nous n’avons pas toutes implémentées et sur lesquelles il serait intéressant de travailler. Nous avons donc listé les principales méthodes, puis comme pour l’onset detection, nous avons essayé au fil de notre projet plusieurs méthodes, en essayant de les combiner, ou d’inventer une méthode par une approche intuitive.

Au fil des recherches que nous avons pu faire, nous avons constaté que les études portant sur la reconnaissance de parole englobent quasiment toutes les recherches sur la détection de la fréquence fondamentale.

## Time domain approaches

Ces méthodes consistent à regarder le signal d'entrée comme une amplitude fluctuante dans le domaine temporel et essayer de trouver des motifs de répétition dans la forme d'onde qui donnent des indices quant à sa périodicité. La forme d’onde qui représente cette modification de la pression dans l’air est alors analysée pour pouvoir dégager la fréquence fondamentale.

### Zero crossing method

Une technique simple qui consiste à compter le nombre de fois que le signal croise le niveau de référence 0 par unité de temps. Si la puissance spectrale de la forme d'onde est concentrée autour de f0, alors il va franchir la ligne zéro deux fois par cycle, comme sur la figure 1. Cependant, si la forme d'onde contient des composantes spectrales de fréquences plus élevées, comme dans la figure 2, alors le signal pourrait franchir la ligne zéro plus de deux fois par cycle. Une solution est de mettre un filtre qui atténue ces hautes fréquences qui détériorent la détection, mais la fréquence de coupure du filtre doit être bien choisie pour ne pas supprimer une fréquence fondamentale f0 qui se situerai dans les hautes fréquences.

|  |
| --- |
| /Users/apple/Desktop/Capture d’écran 2016-02-02 à 09.44.31.png |
| Figure 1 - Méthode zero crossing |

### Peak rate

Cette méthode compte le nombre de pics positifs par seconde dans le signal. En théorie, le signal aura une valeur maximale et minimale pour chaque cycle. On n’aurait qu’a compter ces valeurs maximales pour déterminer la fréquence du signal. Dans la pratique, un détecteur de crête locale doit être utilisé pour trouver l'endroit où la forme d'onde est localement le plus grand et le nombre de ces maxima locaux en une seconde est la fréquence de la forme d'onde, à moins que chaque période de la forme d'onde contient plus d'un maximum local.

**Inconvénient de ces méthodes :** Nous récupérons un signal audio de guitare comprenant normalement une seule note. Cependant, du fait du timbre de l’instrument, d’autres fréquences font leurs apparitions dans la représentation fréquentielle de la transformée de Fourier. Les méthodes décrient ci-dessus fonctionne plutôt bien sur des sons très purs (sans modifications), mais dans le cadre de notre projet, la guitare émet un spectre fréquentiel trop complexe pour que ces méthodes donnent de bons résultats.

### YIN estimator

La difficulté avec les techniques d'auto-corrélation réside dans le fait que les pics se produisent à des sous-harmonique, donc il est parfois difficile de déterminer quel pic représente la fréquence fondamentale et quels pics font références aux harmoniques et autres fréquences contenues dans le signal.

YIN tente de résoudre ces problèmes en plusieurs manières. Cet algorithme est basé sur la fonction de différence, qui, bien que similaire à l’autocorrélation, tente de minimiser la différence entre le signal et le signal retardé-décalé, au lieu de maximiser le produit (autocorrélation).

### Convolution

La méthode proposée ici est de générer une banque de sinus correspondant aux fréquences fondamentales que l’on souhaite détecter. On procède ensuite à un produit de convolution entre le signal x et l’ensemble des signaux constituants la banque de fréquences (les sinus et x doivent avoir la même longueur).

Le produit de convolution exprime la quantité de recouvrement d’une fonction y lorsqu’on la déplace sur une autre fonction x. Ainsi, le sinus qui possède la quantité de recouvrement la plus grande avec x (le sinus qui ressemblera le plus à notre signal) sera considérée comme celui qui se rapproche le plus de notre signal, et sa fréquence sera alors la fréquence fondamentale.

### Autocorrelation

La corrélation entre deux formes d'onde est une mesure de la similitude. Les formes d'onde sont comparées à intervalles de temps différents, et leur "identité" est calculé à chaque intervalle. Le résultat d'une corrélation est une mesure de similarité en fonction du décalage dans le temps entre les débuts des deux formes d'onde.

Pour des signaux périodiques, une caractéristique d'auto-corrélation intéressante est que la fonction d'auto-corrélation est elle-même périodique. Comme le décalage augmente à la moitié de la période de la forme d'onde, la corrélation atteint un minimum. En effet, la forme d'onde est hors de phase avec sa copie retardée dans le temps. Comme le décalage augmente à nouveau à la longueur d'une période, l'auto-corrélation atteint un maximum, car la forme d'onde et sa copie retardée dans le temps sont en phase. Le premier pic de l'auto-corrélation indique la période de la forme d'onde (lorsque les harmoniques ne sont pas trop présentes).

Certains problèmes surviennent avec cette méthode, notamment quand on fait l’autocorrélation d’un signal pseudoperiodique de nature harmonique complexe (figure 2).

Le premier pic ne correspond pas forcément à la période de la forme d'onde complète, mais à la période d’une des harmoniques (ou d’une autre fréquence composant le timbre de l’instrument) du signal. La difficulté est alors de pouvoir distinguer entre "grands" et "petits" pics, selon que les harmoniques et autres fréquences sont plus ou moins présentes.

|  |
| --- |
|  |
| Figure 2 - Signal et résultat de l'autocorrélation du signal |

## Frequency-Domain Methods

Dans ce mode de fonctionnement, nous avons utilisé pour certaines méthodes la STFT (Short Time Fourier Transform) qui est une transformée de Fourier particulière pour déterminer la fréquence sinusoïdale et la phase d'une section locale d'un signal.

Donc pour chaque segment d’une chanson (un segment correspondant à un onset), on applique la STFT.

On en déduit la fréquence fondamentale par les méthodes présentées ci-dessous. Grâce à cette fréquence, on peut en déduire la hauteur de la note et donc son octave, en veillant à ce que ces fréquences soient comprises dans le spectre fréquentiel de la guitare (on recherche grâce à une fonction la valeur de la fréquence la plus proche de celle trouvée correspondant à une fréquence de la guitare).

### Harmonic product spectrum (HPS)

La fréquence fondamentale peut être déterminée en mesurant les fréquences des composantes harmoniques supérieures et en calculant le plus grand dénominateur commun de ces fréquences harmoniques.

Le plus grand dénominateur commun peut être déterminé en découpant le spectre.

Pour réaliser le découpage, il suffit de diminuer la fréquence d'échantillonnage du spectre par 2,3,4,5 etc.. à partir du premier échantillon. Ensuite on recherche le pic le plus élevé dans le résultat de la multiplication des ces spectres pour retrouver la fréquence fondamentale (les pics se décalent).

Cette méthode est peu coûteuse en temps de calcul, résistante au bruit (additif ou multiplicatif), réglable à différents types d'entrées (en changeant le nombre d'harmoniques à envisager ou à remplacer les multiplications par des additions).

|  |
| --- |
| Capture%20d’écran%202016-02-02%20à%2014.51.41.png |
| Figure 3 - Harmonic product spectrum |

### Méthodes intuitives

L’idée était de prendre les pics les plus importants grâce à un seuil et de les étudier. Nous avons remarqué que sur certains morceaux, les spectres fréquentiels correspondant aux notes jouées présentaient comme plus fort pic la fréquence de l’harmonique inférieure. Les fréquences de la note fondamentale, de la tierce ainsi que la quinte étaient également présentes de façon régulière.

Un fois ces pics isolés, nous pouvions alors déterminer la fréquence fondamentale. Cependant, ce schéma ne se présentait pas toujours et parfois la fréquence fondamentale était le plus fort pic sur la transformée de Fourier locale. En prenant ces changements en compte, cet algorithme fut performant mais que sur certains morceaux. Il n’est malheureusement pas adaptable dans la majorité des morceaux que nous avons testés.

Une autre méthode intuitive, assez similaire à la première, était de prendre les pics de plus forte amplitude et de regarder la somme des amplitudes ayant les mêmes fréquences multiples. Ainsi nous aurions pu dégager les fréquences harmoniques de la note la plus présente et déterminer la note jouée. Cependant, il se trouve que cette règle n’est pas toujours respectée. En effet, la fréquence (et ses multiples) la plus présente dans le spectre fréquentiel n’est pas toujours celle de la fréquence fondamentale dans le cas de notre guitare.

Nous avons également essayé de combiner ces deux méthodes mais les résultats finaux n’étaient pas aboutissants. Il serait intéressant de tester ces méthodes sur des instruments ayant des timbres spécifiques et différents de celui de notre guitare.

|  |
| --- |
|  |
| Figure 4 - Transformée de Fourier rapide du Mi-3 de la guitare |

Dans cet exemple de STFT, on voit bien que la note Mi-3 de la guitare possède un spectre fréquentiel composé : du Mi-3 pour le premier pic, du Mi-4 pour le second pic, de la tierce pour le 3ème pic et enfin du Mi-5 pour le quatrième.

### Cepstral method

La transformée de Fourier d’un signal a un nombre régulier de pics, représentant le spectre harmonique. Lorsque l’on prend le logarithme du spectre, les amplitudes des pics changent d’échelle. On obtient alors une forme d’onde périodique dans le domaine fréquentiel, où la période du signal est liée à la fréquence fondamentale du signal d’origine. Une transformation de Fourier inverse de ce signal possède alors un pic représentant la fréquence recherchée.

|  |
| --- |
|  |
| Figure 5 - Cesptral method |

On constate que le signal est bien périodique Avec la présence des pics à égal distance dans le spectre

Le cepstre est une transformation de ce signal du domaine temporel vers un autre domaine analogue au domaine temporel.

La formule correspondant à ce nouveau domaine temporel est :



## Combinaison de méthodes

Après avoir codé et testé la plupart de ces méthodes, les meilleurs résultats sont obtenus par la méthode de convolution pour la reconnaissance des tons (notes). L’autre méthode retenue pour la reconnaissance des octaves est l’Harmonic product spectrum.

## Résultats

Pour finir, nous obtenons comme résultats de bonnes détections : Tons: 93,08

Octaves: 92,7%

|  |
| --- |
|  |
| Figure 6 - Evolution des résultats de reconnaissance des tons et octaves au long du projet |

Les différents pics représentent les différentes méthodes testées. La baisse à certains niveaux peut aussi correspondre à un ajout de données pour les tests (nouveaux morceaux dans la base de données).

**Bibliographie :**

Page dénombrant plusieurs méthodes en expliquant grossièrement :

<https://ccrma.stanford.edu/~pdelac/154/m154paper.htm>

**YIN, a fundamental frequency estimator for speech and music) -**

Alain de Cheveigne - Ircam-CNRS, 1 place Igor Stravinsky, 75004 Paris, France

Hideki Kawahara - Wakayama University - 9 January 2002􏰀

**Joint Multi-pitch Detection using Harmonic envelope Estimation for Polyphonic Music**

**Transcription -** Emmanouil Benetos, Student Member, IEEE and Simon Dixon – 2011

**Pitch Extraction and Fundamental Frequency: History and Current Techniques**

David Gerhard Technical Report TR-CS 2003-06 November, 2003

**Multipitch Analysis of Polyphonic Music and Speech Signals Using an Auditory Model** - Anssi Klapuri, Member - IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 16, NO. 2, FEBRUARY 2008

**Comparison of pitch trackers for real-time guitar effects -**

Adrian von dem Knesebeck,Udo Zölzer - Dept. of Signal Processing and Communications, Helmut Schmidt University Hamburg, Germany

Liens intéressants :

**An efficient f0 determination algorithm based on the implicit calculation of the autocorrelation of the temporal excitation signal -** Joseph Di Martino, Yves Laprie

<http://kom.aau.dk/group/04gr742/pdf/pitch_worksheet.pdf>

**Efficient Pitch Detection Techniques for Interactive Music** - Patricio de la Cuadra, Aaron Master, Craig Sapp - Center for Computer Research in Music and Acoustics, Stanford University

**Multiple Fundamental Frequency Estimation by Summing Harmonic Amplitudes** - Anssi Klapuri Institute of Signal Processing, Tampere, University of Technology – Finland

<http://www.fit.vutbr.cz/~grezl/ZRE/lectures/05_pitch_en.pdf>

**Fundamental Frequency Detection -** Jan Cernocky, Valentina Hubeika

<http://note.sonots.com/SciSoftware/Pitch.html>