

ANALYSE DE SIGNAUX MUSICAUX MULTIPITCH

Mémoire de maîtrise es sciences appliquées
Spécialité : génie électrique

Laurent BONNET

RÉSUMÉ

Le problème global de l'analyse multipitch est d'extraire les paramètres, du signal étudié, qui permettraient de détecter le nombre de sons et d'estimer leur fréquence fondamentale. L'analyse multipitch voit ses applications aussi bien dans la séparation de source, la représentation structurelle des signaux audio, la synthèse du son et principalement la transcription automatique de la musique. C'est dans ce dernier domaine que s'insère la problématique que l'on veut résoudre. En effet, un système de transcription automatique de musique requiert entre autre, une détection des notes jouées par segmentation temporelle, une détection de rythme, une analyse multipitch, une reconnaissance de l'accord joué. Toutes ces étapes pour en ressortir une partition se rapprochant, idéalement, le plus possible de ce qui a été joué par le musicien.

Ce mémoire de maîtrise traite de la séparation multipitch dans un signal polyphonique (en particulier les accords de guitare). Le but est de déterminer le nombre et la valeur des fréquences fondamentales présentes dans un tel signal. Pour résoudre une telle problématique, on s'est basé sur des techniques de traitement de signal, mais aussi sur des connaissances acoustiques et musicales. Le problème se concentre sur l'extraction de paramètres tels que le début de la note ou de l'accord joué, sa longueur et sa composition, dans laquelle on va retrouver les informations sur les fréquences fondamentales.

La méthode développée est décomposée en 2 étages. Premièrement, à l'aide de la dérivée de l'enveloppe temporelle, le signal est segmenté dans le temps. Cela permet donc de séparer les différents accords joués. Ensuite, une procédure itérative basée sur la détection et la soustraction des raies spectrales est appliquée au spectre d'amplitude pour estimer les possibles fréquences fondamentales ou harmoniques. Afin d'améliorer la résolution fréquentielle de la transformée utilisée (*TFD*), une méthode d'interpolation fréquentielle est appliquée autour des harmoniques détectées dans le spectre. Pour tester les algorithmes, des signaux multipitch artificiels ainsi qu'une base de données d'accords de guitare ont été utilisés. La performance va dépendre de la complexité harmonique de chaque son.

REMERCIEMENTS

J'aimerais remercier mon directeur de recherche, M. Roch Lefebvre du département de génie électrique et informatique pour m'avoir donné la possibilité de réaliser ces travaux de recherche et pour les précieux conseils qu'il m'a apportés. Je remercie aussi tous les membres du Laboratoire de Recherche sur la Parole et l'Audio que j'ai eu l'occasion de côtoyer au cours de ce projet. Plus particulièrement Mylène Kwong et Thierry Gagnon avec qui j'ai eu la chance de travailler sur ce projet ainsi que Stephane Ragot pour ses multiples conseils.

Je remercie aussi tout le personnel du Département de génie électrique et informatique pour le support technique et personnel qu'ils ont pu m'apporter. Notamment les membres du groupe de support informatique, les secrétaires du département de génie électrique et informatique, ainsi que celles du secrétariat des études supérieures. Je tiens aussi à remercier M. Charles-Antoine Brunet pour ses précieux conseils concernant L^AT_EX.

Je tiens vivement à remercier mes parents, Chantal et Dominique sans qui je n'aurais pu avoir la possibilité de venir effectuer ma maîtrise au Québec, ainsi que ma soeur Virginie, mon frère Christophe pour tout le support qu'ils m'ont apporté. Je tiens également à remercier tous mes amis proches, notamment M. Jean-François Pratte, pour leur support inconditionnel durant tout le projet.

Je tiens encore à remercier mon directeur de recherche pour le soutien financier qu'il a su m'apporter au cours de ce projet.

TABLE DES MATIÈRES

1	INTRODUCTION	1
2	ACOUSTIQUES DES INSTRUMENTS A CORDES	3
2.1	Généralités	3
2.1.1	La hauteur	4
2.1.2	L'intensité	4
2.1.3	Le timbre	6
2.2	Notions physique de la guitare	6
2.2.1	Description de la partie vibrante	6
2.2.2	Caractéristiques	8
2.2.3	Cas de la corde pincée	10
2.3	Caractéristiques des signaux périodiques	13
2.3.1	Représentation	13
2.3.2	Relations harmoniques	15
2.4	Spécifications	17
2.4.1	Evolution de la fréquence fondamentale	17
2.4.2	Succession d'accords	20
2.4.3	Construction des accords	20
2.5	Conclusion	24
3	ANALYSE MULTIPITCH : ETAT DE L'ART	27
3.1	Méthodes de recherche	27
3.2	Estimation de la fréquence fondamentale	28
3.2.1	Problématique	28
3.2.2	Schéma général d'un algorithme	29

3.3	Algorithme de résolution de pitch	31
3.3.1	Al algorithmes temporels à court terme	31
3.3.2	Al algorithmes fréquentiels à court terme	33
3.4	Algorithme de résolution multipitch	37
3.4.1	Modèle basé sur la note	38
3.4.2	Modèle basé sur l'audition	39
3.4.3	Lissage spectral	39
3.4.4	Estimation de la fréquence fondamentale et séparation de source	41
3.4.5	Autocorrelation	43
3.4.6	Conclusion	45
4	ALGORITHMES DÉVELOPPÉS	47
4.1	Description	47
4.1.1	Caractéristiques tirées des autres algorithmes	48
4.1.2	Fondement de notre méthode	48
4.2	Différents algorithmes	49
4.2.1	Présentation	49
4.2.2	Premières itérations et évolutions	49
4.3	Segmentation temporelle	51
4.3.1	Par un calcul d'énergie	51
4.3.2	Par la dérivée de l'enveloppe temporelle	54
4.3.3	Par la méthode de Teager	58
4.3.4	Conclusion	61
4.4	Détermination des fréquences fondamentales	65
4.4.1	Transformée de Fourier	66
4.4.2	Détection des maximums	79
4.4.3	Approche logique pour la détection des fréquences fondamentales	80

<i>TABLE DES MATIÈRES</i>	vii
4.4.4 Interpolation	82
4.5 Conclusion	83
5 SIMULATIONS ET RÉSULTATS	85
5.1 Différents signaux traités	85
5.1.1 Signaux artificiels	85
5.1.2 Mélodie : 1 note	88
5.1.3 Accord	89
5.2 Résultats	94
5.2.1 Signaux artificiels	96
5.2.2 Accord	100
5.3 Conclusion	104
6 CONCLUSION	107
I ANNEXE 1	109
II ANNEXE 2	111
BIBLIOGRAPHIE	113

LISTE DES FIGURES

1.1	Schéma bloc général d'un modèle de transcription automatique de musique	2
2.1	Variation de l'amplitude selon l'intensité de la corde jouée	5
2.2	Variation de l'amplitude selon la corde jouée	5
2.3	Spectre de deux notes jouées en simultanées ($E2 = 82.42$ Hz et $A2 = 110$ Hz) de guitare	7
2.4	Schématization simple d'une guitare	8
2.5	Modélisation d'une infime partie de la corde	10
2.6	Valeurs possibles des modes de vibrations	12
2.7	Signal de guitare avec zoom sur une partie stationnaire	14
2.8	Signal mono-fréquence à 5 harmoniques	15
2.9	Signal polyphonique à 3 fréquences fondamentales	16
2.10	Evolution d'un signal acoustique au cours du temps	18
2.11	Evolution de la fréquence fondamentale d'une note	19
2.12	Segmentation par accord d'une partie d'une pièce musicale	21
2.13	Accords majeurs avec leurs compositions	22
2.14	Accord C avec sa composition	22
3.1	Schéma général d'un algorithme de détermination de la fréquence fondamentale	30
3.2	Schéma bloc du modèle de Tolonen	43
4.1	Schéma bloc de l'algorithme	50
4.2	Détection d'attaque par la méthode du calcul d'énergie dans le cas d'un accord seul	53
4.3	Détection d'attaque par la méthode du calcul d'énergie dans le cas d'une succession d'accords rapprochés	54
4.4	Détection d'attaque par la méthode du calcul d'énergie dans le cas d'une succession d'accords non rapprochés	55

4.5	Filtre passe bas de type sinuscardinal de longueur : 512, 1024 et 2048 échantillons	56
4.6	Détection d'attaque par la méthode de la dérivée de l'enveloppe temporelle avec une longueur de filtre de 512 échantillons	57
4.7	Détection d'attaque par la méthode de la dérivée de l'enveloppe temporelle avec une longueur de filtre de 1024 échantillons	58
4.8	Détection d'attaque par la méthode de la dérivée de l'enveloppe temporelle avec une longueur de filtre de 2048 échantillons	59
4.9	Détection d'attaque par la dérivée de l'enveloppe temporelle dans le cas d'une succession d'accords rapprochés	60
4.10	Détection d'attaque par la dérivée de l'enveloppe temporelle dans le cas d'une succession d'accords éloignés	61
4.11	Détection d'attaque par la méthodes de Teager dans le cas d'un accord isolé	62
4.12	Détection d'attaque par la méthodes de Teager dans le cas d'une succession d'accords rapprochés	63
4.13	Détection d'attaque par la méthodes de Teager dans le cas d'une succession d'accords éloignés	64
4.14	Schéma bloc de l'algorithme de segmentation temporelle	65
4.15	Schéma bloc de l'algorithme de détection des fréquences fondamentales	66
4.16	Transformée de Fourier Discrète d'une longueur de signal de 4096 échantillons	68
4.17	Transformée de Fourier Discrète d'une longueur de signal de 1024 échantillons	69
4.18	Transformée de Fourier Discrète pour des signaux ayant de 2 à 6 notes	70
4.19	Spectre et enveloppe spectrale d'une portion (2048 échantillons) de la note F#2	72
4.20	Spectre et enveloppe spectrale d'une portion (2048 échantillons) de la note G3	73
4.21	Spectre et enveloppe spectrale d'une portion (2048 échantillons) de la note C4	73
4.22	Spectre et enveloppe spectrale d'une portion (2048 échantillons) de la note E4	74
4.23	Spectre et enveloppe spectrale d'une portion (2048 échantillons) de l'accord A (La)	74
4.24	Spectre d'origine et spectre résultant de la division par l'enveloppe spectrale d'une portion (2048 échantillons) de la note F#2	75
4.25	Spectre d'origine et spectre résultant de la division par l'enveloppe spectrale d'une portion (2048 échantillons) de la note G3	76

4.26	Spectre d'origine et spectre résultant de la division par l'enveloppe spectrale d'une portion (2048 échantillons) de la note C4	76
4.27	Spectre d'origine et spectre résultant de la division par l'enveloppe spectrale d'une portion (2048 échantillons) de la note E4	77
4.28	Spectre d'origine et spectre résultant de la division par l'enveloppe spectrale d'une portion (2048 échantillons) de l'accord A	77
4.29	Spectre résultant de la division de l'enveloppe spectrale et spectre seuillé d'une portion (2048 échantillons) de l'accord A	78
4.30	Zoom sur le début du spectre (b) de la Figure 4.29	80
5.1	Signal temporel synthétisé composé de 3 fréquences fondamentales	87
5.2	Signal temporel synthétisé composé d'une fréquence fondamentale	89
5.3	Signaux temporels de 4 notes de guitare	90
5.4	Allure temporelle de 2 accords	94
5.5	NER par rapport à la polyphonie et aux longueurs de signal analysé	95
5.6	NER pour les signaux simulés avec le Δf_{min}	99
5.7	NER pour les accords de guitare avec le nombre d'accords détectés	104

LISTE DES TABLEAUX

2.1	Tableau des fréquences audibles	4
2.2	Table associant la fréquence de vibration à chaque corde d'une guitare	7
4.1	Fréquences fondamentales et harmoniques composant des différents notes testées	71
4.2	Maximum (en terme d'échantillon) du spectre de la Figure 4.30	79
4.3	Tableau des fréquences fondamentales détectées avant et après interpolation	83
5.1	Tableau des fréquences fondamentales (Hz) à vide et des 5 premiers frets d'une guitare	86
5.2	Tableau présentant la composition des notes testées pour un signal monophonique	91
5.3	Tableau présentant la composition des accords testés pour une polyphonie d'ordre 2	91
5.4	Tableau présentant la composition des accords testés pour une polyphonie d'ordre 3	92
5.5	Tableau présentant la composition des accords testés pour une polyphonie d'ordre 4	92
5.6	Tableau présentant la composition des accords testés pour une polyphonie d'ordre 5	93
5.7	Tableau présentant la composition des accords testés pour une polyphonie d'ordre 6	93
5.8	Tableau présentant les fréquences testées et les fréquences détectées pour un signal monophonique artificiel	96
5.9	Tableau présentant les fréquences testées et les fréquences détectées pour une polyphonie d'ordre 2 pour les signaux artificiels	97
5.10	Tableau présentant les fréquences testées et les fréquences détectées pour une polyphonie d'ordre 3 pour les signaux artificiels	97
5.11	Tableau présentant les fréquences testées et les fréquences détectées pour une polyphonie d'ordre 4 pour les signaux artificiels	98

5.12	Tableau présentant les fréquences testées et les fréquences détectées pour une polyphonie d'ordre 5 pour les signaux artificiels	98
5.13	Tableau présentant les fréquences testées et les fréquences détectées pour une polyphonie d'ordre 6 pour les signaux artificiels	99
5.14	Tableau présentant les fréquences testées et les fréquences détectées pour des notes de guitares	101
5.15	Tableau présentant les fréquences testées et les fréquences détectées pour une combinaison de 2 notes de guitare	101
5.16	Tableau présentant les fréquences testées et les fréquences détectées pour une combinaison de 3 notes de guitare	102
5.17	Tableau présentant les fréquences testées et les fréquences détectées pour une combinaison de 4 notes de guitare	102
5.18	Tableau présentant les fréquences testées et les fréquences détectées pour une combinaison de 5 notes de guitare	103
5.19	Tableau présentant les fréquences testées et les fréquences détectées pour une combinaison de 6 notes de guitare	103

CHAPITRE 1

INTRODUCTION

La connaissance de la fréquence fondamentale des signaux sonores est utile dans de nombreux domaines du traitement des signaux acoustiques. La transcription automatique de partitions musicales est un exemple probant d'application de même que la séparation de source. Pour chaque application, le paramètre utilisé nécessite de connaître la fréquence fondamentale de façon plus ou moins précise, de l'estimer à des intervalles de temps réguliers ou non et plus ou moins rapprochés.

La détermination de la fréquence fondamentale a été étudiée assez largement, surtout dans le domaine de la parole [Hess, 1983]. Dans le domaine des signaux musicaux, c'est Rabiner [Rabiner et coll., 1976] qui a répertorié tous les algorithmes de détection monophonique qui vont être présentés dans le Chapitre 3. En ce qui concerne la détection polyphonique, les premiers travaux remontent à 1970 [Moorer, 1977], cependant il y avait de nombreuses limitations, mais cela a engendré de nombreux travaux dont les principaux sont [Lahat, 1987], [de Cheveigné, 1993], [Martin, 1996], [Klapuri, 1998], [Tolonen, 2000a] et [Klapuri, 2001]. Cependant, aucune de ces techniques n'est totalement fiable et implementable en temps réel. Il reste donc de nombreux travaux à effectuer pour être aussi efficace que dans les cas des signaux de parole ou des signaux monophoniques.

Les signaux musicaux ont généralement un comportement pseudo-stationnaire, mais présentent des variations plus brutales entre ces zones de relative stabilité. Par exemple, les caractéristiques de l'enveloppe spectrale, qui reflètent le timbre de l'instrument, seront relativement stables pour une note donnée. Par contre, l'évolution temporelle de la fréquence fondamentale, qui représente la mélodie, présente différents types de variations : variations brutales lors du passage d'une note à l'autre, variations continues quand la note est accompagnée d'un vibrato ou d'un glissando. De plus, la fréquence fondamentale des notes peut couvrir plusieurs octaves (la tessiture des instruments est rarement inférieure à trois octaves).

Le thème de recherche du travail présenté ici est celui de la détermination fiable des fréquences fondamentales à extraire d'un mélange de sons (Bloc 3 dans la Figure 1.1). Plus précisément, ce mélange de sons est un signal de guitare dont le degré de polyphonie peut varier de 1 à 6. En d'autres termes, le cas du problème de la détermination multipitch dans un accord de guitare sera traité. Cette problématique s'insère dans un projet ambitieux qui a pour but d'élaborer un système multimodal de transcription automatique de la musique. Le schéma bloc de ce système, qui représente les différents étapes de traitement, est présenté à la Figure 1.1. On peut voir que le système est composé de 6 blocs de complexité plus ou moins importante. Certains de ces blocs, comme le bloc 1 (Détection des transitoires) et le bloc 5 (Analyse rythmique) sont intimement liés, mais tous vont donner des paramètres essentiels à la résolution de la problématique.

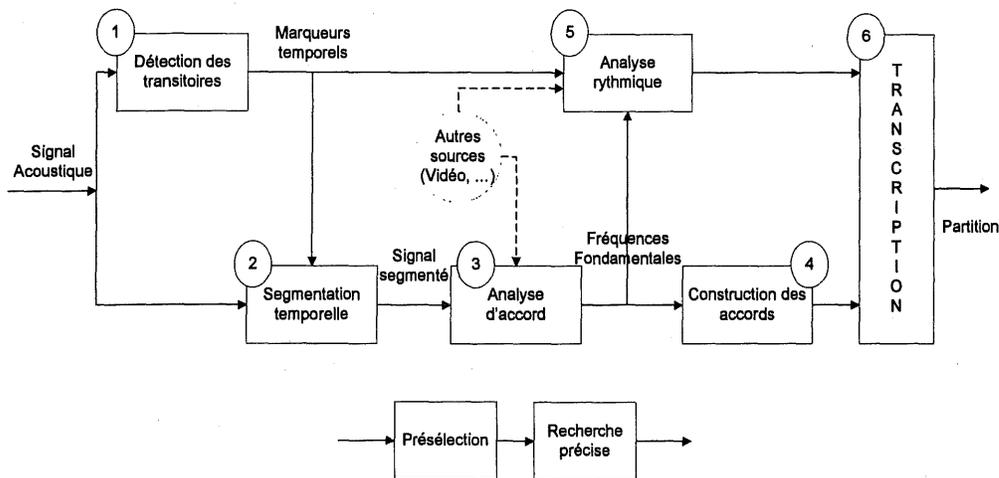


Figure 1.1 – Schéma bloc général d'un modèle de transcription automatique de musique

Le mémoire est organisé comme suit. Au chapitre 2, une étude sur l'acoustique des instruments à cordes puis particulièrement sur les caractéristiques de la guitare a été réalisée. La recherche bibliographique portant principalement sur les différents méthodes d'analyse multipitch est présentée au chapitre 3. Les chapitres 4 et 5 constituent les parties principales de ce mémoire qui sont les différents algorithmes développés permettant de résoudre la problématique posée et les résultats expérimentaux obtenus par les tests.

CHAPITRE 2

ACOUSTIQUES DES INSTRUMENTS A CORDES

Les instruments à cordes ont certaines caractéristiques qui les distinguent des autres instruments de musique. Le phénomène de la corde vibrante qui est la particularité des instruments à cordes a été largement étudié et a permis de mieux comprendre le fonctionnement de ces instruments.

Au cours de ce projet, on étudie l'instrument à corde qu'est la guitare. Dans ce chapitre, on en fait une brève description, tant au point de vue composition que fonctionnement [Fletecher, 1991]. Ce dont on va se soucier le plus, ce sont les sons que le musicien est capable de produire avec une guitare. La complexité est très variée, on va aller d'un son ne comportant qu'une seule note jusqu'à un son en comportant au maximum 6. Il est aussi primordial de comprendre comment les accords sont construits et quelles sont les relations interharmoniques que l'on va rencontrer entre les différents accords joués aux points de vue des fréquences fondamentales.

L'étude de ces caractéristiques est essentielle afin d'analyser de la meilleure façon possible la note, l'accord ou la pièce musicale qui a été jouée et donc d'avoir l'estimation qui se rapproche le plus possible de celle-ci.

2.1 Généralités

Un son est déterminé par quatre facteurs : sa hauteur (fréquence grave ou aiguë), son intensité (amplitude), sa durée (court ou long, croche ou blanche) et son timbre.

2.1.1 La hauteur

La hauteur d'un son correspond à sa fréquence de vibration exprimée en Hertz. La fréquence est le nombre d'oscillations complètes d'un signal périodique sur une seconde. Les fréquences que l'on va étudier sont celles générées par les cordes d'une guitare et on appellera fondamentale celle qui correspond à la fréquence de la note jouée. Le Tableau 2.1 donne les fréquences de vibrations des notes audibles les plus utilisées en musique en fonction des différents octaves.

octave	0	1	2	3	4	5	6
Do	32.7	65.41	130.81	261.63	523.26	1046.51	2093.02
Do#	34.65	69.30	138.59	277.19	554.37	1108.74	2217.48
Ré	36.71	73.42	146.83	293.67	587.34	1174.67	2349.34
Ré#	38.89	77.78	155.56	311.13	622.26	1244.52	2489.04
Mi	41.20	82.41	164.82	329.63	659.26	1318.52	2637.04
Fa	53.65	87.31	174.62	349.23	698.46	1396.83	2793.85
Fa#	46.25	92.50	185.00	370.00	740.00	1480.00	2960.00
Sol	49.00	98.00	196.00	392.00	784.00	1568.00	3136.00
Sol#	51.91	103.83	207.65	415.31	830.62	1661.23	3322.47
La	55.00	110.00	220.00	440.00	880.00	1760.00	3520.00
La#	58.20	116.54	233.08	466.17	932.34	1864.67	3729.34
Si	61.74	123.47	246.94	493.89	987.78	1975.55	3951.10

TABLEAU 2.1 – Tableau des fréquences (Hz) audibles les plus utilisées en musique

2.1.2 L'intensité

On caractérise l'intensité d'un son par son amplitude. L'amplitude est définie comme étant la valeur du signal entre deux maximums; elle correspond au "volume" du son perçu. La Figure [?], montre que selon la façon dont la corde est jouée (douce ou forte) alors l'amplitude varie. Cela paraît évident, mais on voulait voir si la différence d'amplitude était importante,

et si on devait en tenir compte dans les algorithmes.

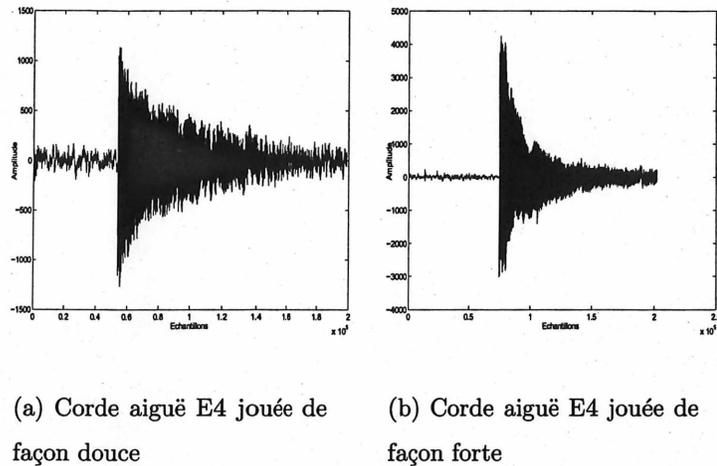


Figure 2.1 – Variation de l'amplitude selon l'intensité de la corde jouée

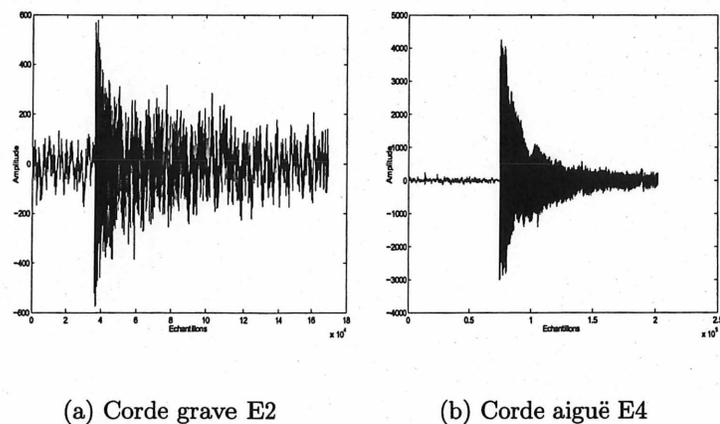


Figure 2.2 – Variation de l'amplitude selon la corde jouée

Dans la Figure 2.2, on compare les amplitudes de deux notes jouées avec environ la même intensité (d'après le musicien) mais ayant des fréquences fondamentales très éloignées en terme de fréquence ($E2 = 82.41$ Hz et $E4 = 329.63$ Hz). On remarque que selon la corde qui est jouée (et ceci avec environ la même intensité, (a) et (b)), plus la corde à une fréquence fondamentale aiguë, plus son amplitude va être grande. Il va donc être primordial de prendre cela en compte lors de la mise en oeuvre des algorithmes.

2.1.3 Le timbre

Le timbre est la composition harmonique d'un son. Dans cette composition harmonique, on retrouve les harmoniques de la fréquence fondamentale ainsi que l'information donnée par l'attaque de la corde (grattée, pincée, à quel endroit). Une corde mise en vibration dans un mode X (médiator, doigt, archet...) va produire certains harmoniques qui vont traverser les systèmes acoustico-mécaniques composés par le chevalet, la table de l'instrument et d'éventuels autres éléments qui vont jouer un rôle de filtre. Au final, la table transmettra à l'air un signal typique de l'instrument qui permettra à l'oreille de l'identifier. Donc, dans un son, on a une fréquence fondamentale (qui donne sa hauteur) et des fréquences secondaires (partiels) lui donnant sa couleur. Ces fréquences secondaires s'organisent principalement autour des harmoniques de la fréquence de base, les autres composantes (les partiels) correspondent aux informations parasites qui peuvent être dues à l'attaque ou alors à un bruit lors de l'enregistrement. Étant donné que d'autres raies spectrales (que celles composant la note) forment le spectre, on peut dire qu'à cause du timbre, un son ne correspond jamais totalement à la note jouée. En effet, comme on peut le voir sur la Figure 2.3, les raies spectrales de plus fortes amplitudes correspondent aux harmoniques des deux notes jouées, mais on retrouve aussi d'autres raies spectrales (notamment dans les fréquences en-dessous de 500 Hz) ne correspondant pas à la note jouée et donc pouvant fausser l'analyse qui serait faite.

2.2 Notions physique de la guitare

2.2.1 Description de la partie vibrante

Pour étudier la corde vibrante, il faut commencer par connaître le système dans lequel elle évolue et quels sont les agents extérieurs qui agissent sur sa vibration. La chaîne dans laquelle se trouve la corde commence par le musicien qui exerce une action sur la corde par l'intermédiaire d'un système excitateur qui peut être un doigt, un médiator, un archet [Bouillaud et Gognon, 1997]. Celui-ci fournit à la corde une énergie mécanique qui lui permet de se mettre en vibration, c'est la phase d'attaque. La corde est donc le siège d'un

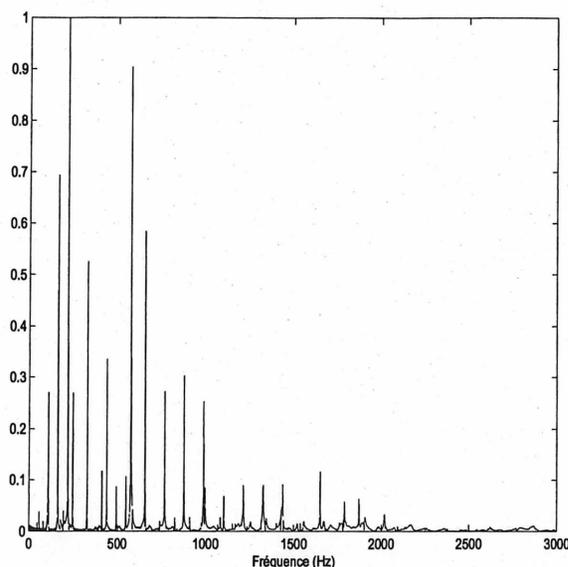


Figure 2.3 – Spectre d'une note (G3 = 196 Hz) de guitare

régime vibratoire qui, transformée en ondes sonores est perceptible par l'homme avec une hauteur et un timbre précis.

Les six cordes de la guitare sont tendues entre la tête et le chevalet, lui-même solidaire du corps de la guitare. On présente ici le chiffrage des notes selon la représentation européenne et américaine (le plus souvent utilisé). Elles sont accordées pour vibrer à vide aux fréquences présentées dans le Tableau 2.2.

Notation Européenne	Mi ₂	La ₂	Ré ₃	Sol ₃	Si ₃	Mi ₄
Notation Américaine	E2	A2	D3	G3	B3	E4
Fréquence (Hz)	82.41	110.00	146.83	196.00	247.00	329.63

TABLEAU 2.2 – Table associant la fréquence de vibration à chaque corde d'une guitare

La corde seule ne produit pas de son car sa surface est trop petite pour provoquer une vibration de l'air, il est donc nécessaire de passer par un résonateur pour qu'un son soit produit. Ce résonateur ne se contente pas d'amplifier cette vibration, il va aussi jouer un rôle de filtre et d'affineur de timbre.

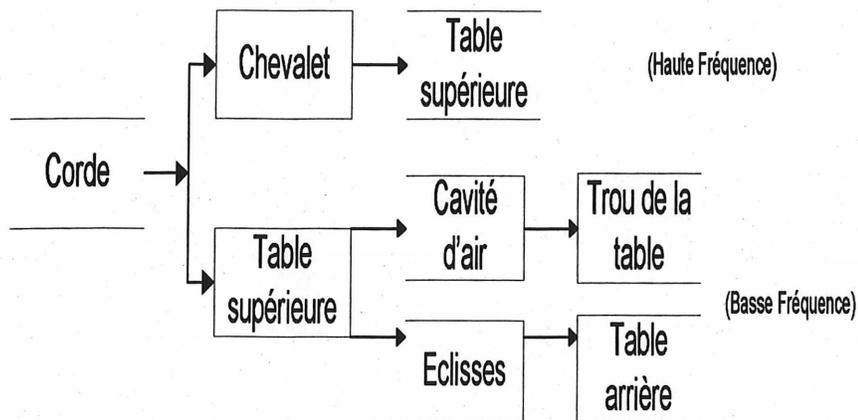


Figure 2.4 – Schématisation simple d'une guitare. Aux basses fréquences le son est émis par la table arrière et la table supérieure. Aux hautes fréquences, la plupart du son est émis par la table supérieure

On distingue deux modes de propagation du son (schématisé par la Figure 2.4) :

- Dans les basses fréquences (82 - 200Hz) : On est dans la zone de résonance de la table supérieure du corps qui transmet, par les éclisses et la cavité, ses vibrations à la table inférieure. Le son prend de l'amplitude au sein de la cavité et ressort par le trou de la table supérieure.
- Dans les hautes fréquences (200 - 1000Hz) : Le son est transmis essentiellement par la vibration de la table supérieure. C'est ce qui se passe par exemple dans le cas d'un diapason, où le son devient audible quand on le pose sur une surface quelconque.

2.2.2 Caractéristiques

Une corde peut être caractérisée par :

- L : longueur de la corde vibrante (entre les deux sillets)
- mv : masse volumique du matériau de fabrication des cordes
- D : diamètre de la corde (non filée)

- S : section de la corde (calculée à partir de D)
- μ : masse par unité de longueur
- T : tension de la corde montée au repos

Pour une corde donnée dont on connaît les caractéristiques, on peut calculer :

- La masse par unité de longueur

$$\mu = mv * S$$

- La célérité (vitesse à laquelle se déplace l'onde sur la corde) :

$$C = \left(\frac{T}{\mu}\right)^{\frac{1}{2}}$$

- Pour parcourir une période complète P , l'onde devra parcourir un aller retour sur la corde, donc 2 longueurs :

$$P = \frac{2L}{C}$$

- Donc, la fréquence de vibration de la corde est :

$$F = \frac{1}{P}$$

La longueur : La longueur de la corde (en mètre) intervient directement sur la période et inversement sur la fréquence. Pour doubler la fréquence, il faut diviser par deux la longueur de la corde. C'est bien sûr le principe de la division d'un manche par le fretage, les longueurs étant calculées par des divisions par "racine douzième de deux". Une corde courte ne vibre pas longtemps, c'est pourquoi plus un instrument est grave, plus sa longueur de corde vibrante (diapason) est grande.

Le diamètre - La surface : Ces deux caractéristiques sont bien sûr liées et permettent de jouer sur le facteur μ et donc également sur la fréquence. Il faut diviser la masse par quatre pour doubler la fréquence et il en est de même pour la surface.

La tension de la corde : La tension (usuellement entre 5 et 10 kg) intervient proportionnellement sur la vitesse de propagation de l'onde sur la corde. Il faut la multiplier par quatre pour doubler la fréquence.

2.2.3 Cas de la corde pincée

Lorsqu'on agit sur une corde à l'aide d'un médiator, une onde parcourt la corde en aller-retour, produisant une vibration de la corde à une fréquence précise, la fondamentale, mais la corde est le siège d'autres vibrations secondaires :

- Elle n'est pas parfaite, et génère elle-même des harmoniques.
- Elle est fixée entre deux points qui réagissent à la vibration.
- Elle connaît des variations de tension dues à son allongement lors des vibrations.

La corde pincée comporte un spectre quasi harmonique dont les raies s'écartent très légèrement au fur et à mesure qu'elles sont placées plus haut. Il est possible d'expliquer cela en modélisant une partie de la corde vibrante comme sur la Figure 2.5.

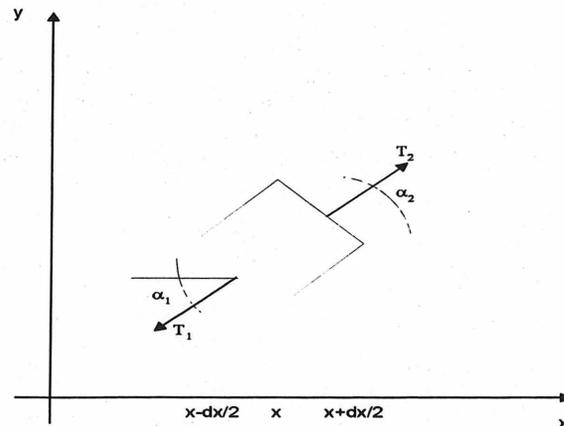


Figure 2.5 – Modélisation d'une infime partie de la corde

Si μ est la masse linéique de la corde, alors le Principe Fondamentale de la Dynamique (P.F.D) appliqué au système ci-dessus donne :

$$(\mu \cdot dx) \cdot \frac{\partial^2 y(x, t)}{\partial t^2} = T_2 \sin \alpha_2 - T_1 \sin \alpha_1 \quad (2.1)$$

On considère ici que les déplacements sont faibles et que $T_1 = T_2 = T_0$ tension à l'équilibre. De plus, $\sin(\alpha) = \text{tg}(\alpha)$ pour α petit, et ainsi on peut considérer $\text{tg}(\alpha)$ comme la pente au point considéré, ici en $x-dx/2$ et $x+dx/2$.

Alors, avec un développement limité d'ordre 2, le P.F.D devient :

$$\frac{\partial^2 y}{\partial t^2} - \frac{T_0}{\mu} \frac{\partial^2 y}{\partial x^2}(x, t) = 0 \quad (2.2)$$

L'effort transversal vaut :

$$F_T = -T_0 \cdot \frac{\partial y}{\partial x} \quad (2.3)$$

Dans un premier temps, on considère les conditions aux limites :

- en $x = 0$, $y = 0$
- en $x = L$, $y = 0$

Alors, à partir de la solution générale de la forme :

$$y(x, t) = A \sin(kx) \cos(\omega t + \phi) + B \cos(kx) \cos(\omega t + \phi) \quad (2.4)$$

On trouve :

$$y(t) = A \sin\left(\frac{n\pi}{L}t + \phi\right) \quad (2.5)$$

avec n entier.

On n'a que certaines valeurs possibles pour les modes de vibration qui sont les harmoniques. C'est-à-dire que la corde ne peut vibrer qu'autour des noeuds en L/n comme sur la Figure 2.6.

Cependant, dans le cas d'une corde de guitare, la condition en L au niveau du chevalet n'est pas nulle, car la table supérieure oscille très légèrement. Ceci se traduit par le fait que

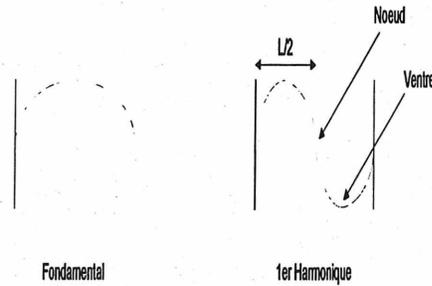


Figure 2.6 – Valeurs possibles des modes de vibrations

l'admittance mécanique n'est plus nulle, mais très petite. On a donc en $x = L$:

$$Y_{meca} = \frac{\dot{y}}{F_{tranchant}} \quad (2.6)$$

Alors, avec les formes d'ondes trouvées précédemment, on a :

$$y = A \sin(kx) e^{j\omega t} \quad (2.7)$$

d'où

$$\dot{y} = A \cdot j\omega t \sin(kx) e^{j\omega t} \quad (2.8)$$

D'après l'équation (2.3), l'effort transversal vaut :

$$F_T = -T_0 A \cdot k \cos(kx) e^{j\omega t} \quad (2.9)$$

On en déduit donc, l'admittance mécanique en $x=L$:

$$\underline{Y}(L, \omega) = -j \frac{\omega}{T_0 \cdot k} \tan(kL) \quad (2.10)$$

Avec l'hypothèse Y petit, on a :

$$\omega = \frac{n\pi}{L} \sqrt{\frac{T_0}{\mu}} + j \frac{T_0}{L} \underline{Y} \quad (2.11)$$

On trouve les harmoniques en $n/(2L)$ avec en plus la partie due à Y qui traduit deux résultats :

- La partie réelle de Y qui traduit les pertes ou facteurs d'amortissement
- La partie imaginaire de Y qui fausse l'harmonicité. Les harmoniques ne sont plus situées exactement au niveau des multiples entiers de la fréquence fondamentale.

Donc, lors de l'analyse, il va falloir tenir compte du fait que les harmoniques ne se situeront pas à des fractions entières de la fréquence fondamentale mais qu'un Δf peut entrer en jeu.

2.3 Caractéristiques des signaux périodiques

Un son pur est un signal de forme sinusoïdale, de fréquence et d'amplitude constante. Par contre, ces sons n'existent pas à l'état naturel, on ne peut les obtenir qu'avec des générateurs électroniques. En réalité, les sons musicaux qui arrivent à nos oreilles sont composés de plusieurs sinusoïdes de fréquences et d'amplitudes différentes. Ces signaux sont dit périodiques et ils ont la particularité d'être harmoniques.

2.3.1 Représentation

Il est possible de modéliser de façon assez simple le son sortant d'une guitare. En ne se concentrant que sur la partie du son que l'on retrouve après l'attaque et avant l'atténuation (voir figure 2.7), on peut voir cela comme une somme de sinusoïdes avec ses harmoniques. On peut donc représenter des sons monophoniques ou polyphoniques.

Les sons monophoniques : L'apparence physique d'un signal monophonique est une série de fréquences variables appelées harmoniques. Ils ont la forme de pics dans le domaine des fréquences avec une fréquence f_0 constante. Dans le domaine du temps, c'est $1/f_0$ qui est la période fondamentale. Un tel signal peut être modélisé de la façon suivante :

$$x_t = \sum_{k=0}^{M_j} \rho_{k,j} \cos(2\pi k F_j t + \phi_{k,j}) \quad (2.12)$$

où M_j est le nombre d'harmoniques jouées dans la note, et $\rho_{k,j}$, kF_j and $\phi_{k,j}$ sont respectivement l'amplitude, la fréquence et la phase de chaque harmonique dans la note.

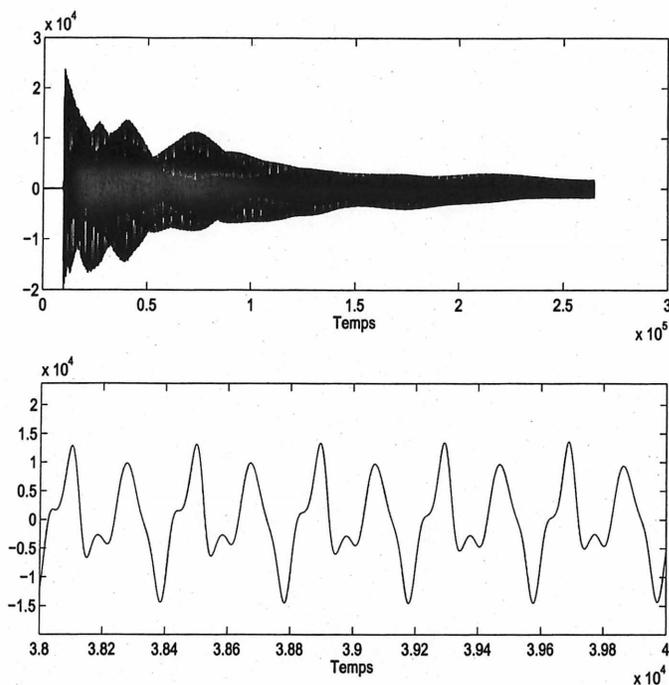


Figure 2.7 – Signal monophonique de guitare avec zoom sur une partie stationnaire

Ce genre de signaux (que l'on a obtenu grâce au logiciel *MATLAB*) peut correspondre à un joueur de guitare qui aurait *gratté* une seule corde de sa guitare. La figure 2.8 modélise la note E2 qui résonne à une fréquence de 82.41 Hz et composée de 5 harmoniques. On va donc rencontrer ce genre de signaux tout au long de ce projet et dans ces cas, M_j peut être très grand. Il va dépendre de la richesse harmonique du son joué.

Les sons polyphoniques : Ceux-ci sont les signaux où plusieurs sons sont mélangés, et de ce fait on y rencontre plusieurs fréquences fondamentales. On peut les représenter de la façon suivante :

$$x_t = \sum_{j=1}^N \sum_{k=0}^{M_j} \rho_{k,j} \cos(2\pi k F_j t + \phi_{k,j}) \quad (2.13)$$

où N est le nombre de notes jouées, M_j est le nombre d'harmoniques jouées dans la note, et $\rho_{k,j}$, kF_j and $\phi_{k,j}$ sont respectivement l'amplitude, la fréquence et la phase de chaque

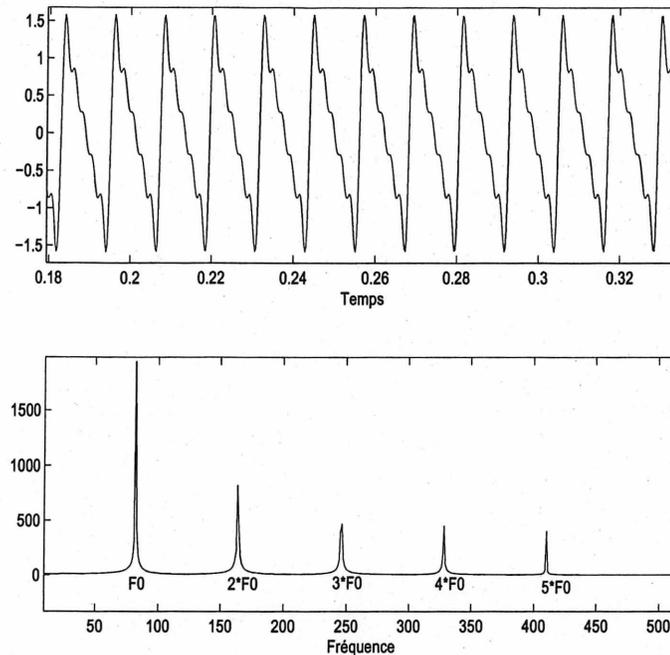


Figure 2.8 – Signal monophonique dans le domaine du temps et des fréquences modélisant la note E2

harmonique dans la note.

Ce genre de signaux peut correspondre à un joueur de guitare qui aurait *gratté* plusieurs cordes sur sa guitare. La figure 2.9 modélise les notes A2, D3 et G3 qui résonnent respectivement à une fréquence de 110 Hz, 146.83 Hz et 196 Hz qui sont composées de 3 harmoniques chacune. En analysant le spectre du signal, on retrouve bien les 3 fréquences fondamentales avec chacune leurs harmoniques ($2f_0$, $2f_1$, $2f_2$, ...). C'est avec ce genre de signaux que l'on va travailler le plus souvent, et dans notre cas N peut aller de 1 à 6 et M_j va toujours dépendre de la richesse harmonique du son joué.

2.3.2 Relations harmoniques

Comme on l'a vu précédemment, les sons musicaux que l'on va étudier sont composés de plusieurs sinusoides de fréquences et d'amplitudes différentes. Ces fréquences sont appelées

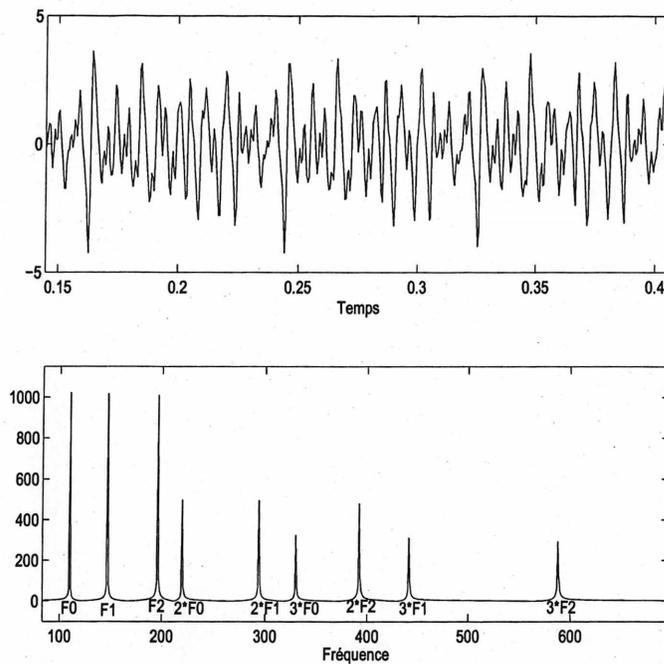


Figure 2.9 – Son polyphonique dans le domaine du temps et des fréquences

harmoniques si elles sont multiples de celle que l'on appelle la *fondamentale*. Les autres sont les partiels que l'on retrouve principalement au moment de l'attaque (moment où la corde est frappée) qui représente la partie inharmonique du signal. Il est donc important de spécifier les relations qui existent entre la fréquence fondamentale et ses harmoniques, surtout dans le cas de signaux polyphoniques où il peut y avoir des cas de recouvrements d'harmoniques.

On vient de voir d'après la figure 2.8 que si un signal est harmonique avec une fréquence fondamentale f_0 , on va retrouver des harmoniques aux fréquences $2f_0$, $3f_0$ et ainsi de suite. On peut donc imaginer que si les fréquences que l'on souhaite déterminer sont rapprochées et bien il risque d'y avoir des superpositions d'harmoniques. Cela va donc poser des problèmes :

- il ne va pas être facile de classer les différents harmoniques en fonction de leur fondamentale, car les différentes séries d'harmoniques vont se retrouver dans la même bande de fréquence.
- l'amplitude et la phase de l'enveloppe de deux sinusoides ne peuvent plus être déduites

de leur somme, car elles se superposent (elles partagent la même fréquence).

Posons F_S comme étant la fréquence fondamentale d'un son périodique S et F_R comme étant la fréquence fondamentale d'un son périodique R . Ainsi, la condition pour qu'un harmonique à la fréquence $k \cdot F_S$ d'un son S coïncide avec un harmonique à la fréquence $l \cdot F_R$ d'un son R peut être écrit comme suit :

$$k \cdot F_S = l \cdot F_R \quad (2.14)$$

Quand on généralise cette relation, on trouve que :

$$f_R = \frac{m}{n} f_S \quad (2.15)$$

avec $(m, n) \geq 1$ étant des nombres entiers. D'après l'équation (2.15), on peut dire que les harmoniques de deux sons vont concorder si le rapport de leur fréquence fondamentale est un nombre rationnel. De plus, il est facile de dire qu'en vérifiant l'équation ci-dessus, chaque m^{th} harmonique à la fréquence $mW \cdot F_S$ du son S coïncide avec chaque n^{th} harmonique à la fréquence $nW \cdot F_R$ du son R , où $W \geq 1$ est un entier.

Étant donné l'importance des relations entre harmoniques dans les signaux musicaux, il va toujours falloir faire attention à ces conditions lors du traitement. Cela explique aussi le fait que la détection multipitch est difficile à résoudre dans le cas de signaux musicaux.

2.4 Spécifications

2.4.1 Evolution de la fréquence fondamentale

Un signal acoustique évolue dans le temps, il n'est pas constant. Il nous faut donc tenir compte de ses différentes phases dans l'analyse d'un son.

Quand une note est jouée sur un instrument à cordes, on retrouve principalement 4 phases qui sont présentées sur la Figure 2.10. Sur cette figure, les échelles sont à titre indicatif uniquement.

On peut aussi nommer et décrire ces différentes phases :

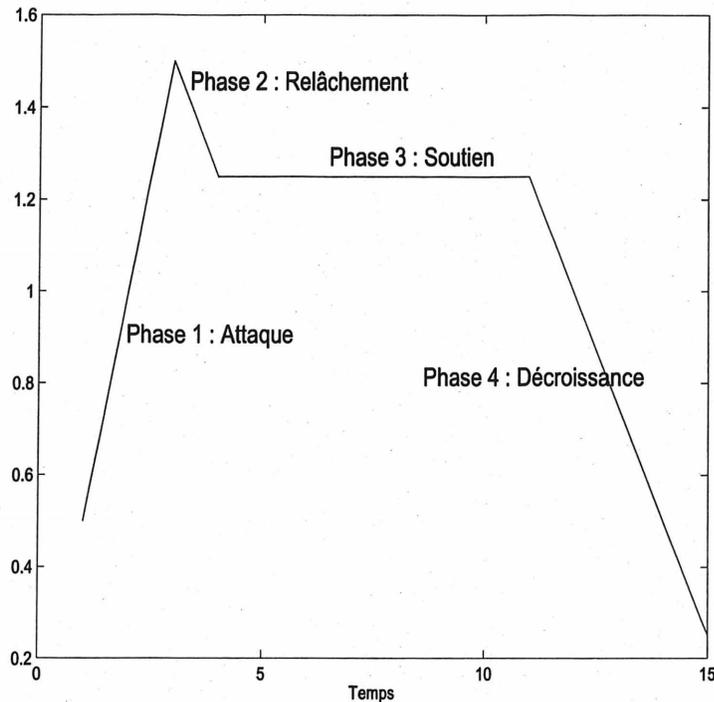


Figure 2.10 – Evolution des phases au cours du temps

1. L'attaque : C'est la période pendant laquelle la corde se met en vibration sous l'action du musicien. Les harmoniques et les partiels apparaissent.
2. Le relâchement : C'est le moment pendant lequel la corde est libérée de son moteur. Les partiels perdent de l'énergie.
3. Le soutien : C'est la partie la plus stable du signal (des harmoniques peuvent s'éteindre durant cette période). Les partiels ont disparus.
4. La décroissance : C'est la partie où le signal et notamment les fondamentales s'éteignent.

On a vu que, au moment de l'attaque, les signaux musicaux sont composés d'harmoniques mais aussi de partiels. Cela risque donc d'influencer notre analyse, car au moment de cette attaque, notre signal va être inharmonique et aussi être aléatoire à cause des bruits de fonds et autres parasites.

Il est donc important de voir, dans ces différentes phases, de quelle façon la fréquence fondamentale d'une corde de guitare évolue. Cela va nous aider à mieux situer la position de notre

analyse dans le signal pour qu'il soit le plus stationnaire possible et donc faire l'analyse la plus précise possible.

La figure 2.11 montre l'évolution de la fréquence fondamentale pour une note au cours du temps. C'est la méthode de l'autocorrélation qui a été utilisée pour calculer la valeur de la fréquence fondamentale dans chacune des trames. Cette méthode est bien connue et donne des résultats très convaincants quant à la recherche de pitch dans un signal monophonique (cf section 3.3).

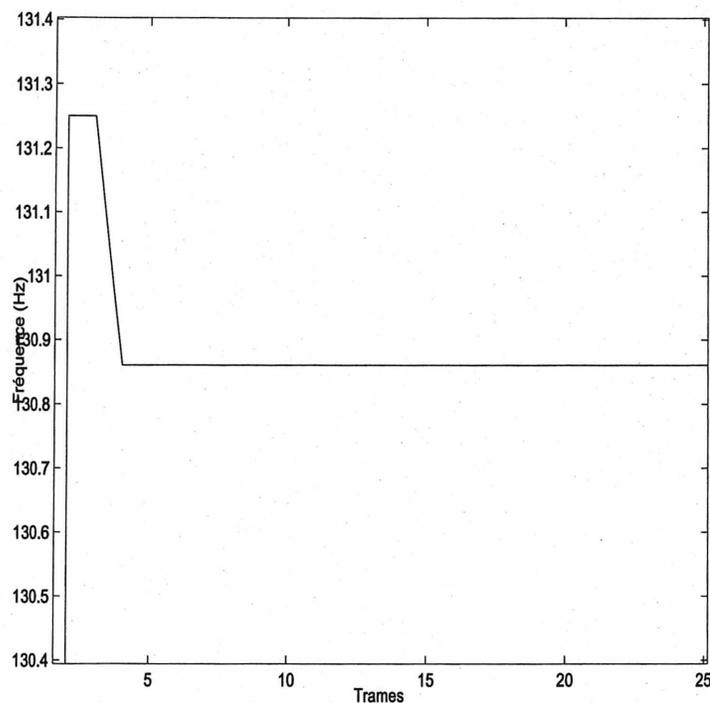


Figure 2.11 – Evolution de la fréquence fondamentale dans le temps pour la note C3 (130.81 Hz) en fonction des trames (1024 échantillons chaque).

On travaille sur un signal échantillonné à 44.1 kHz. En prenant des trames de 1024 échantillons, cela veut dire qu'on fait une analyse de la fréquence fondamentale environs toutes les 25 ms. En ayant fait le test sur la plupart des notes de la guitare, on s'est aperçu qu'il fallait attendre environ 100 ms pour que la fréquence fondamentale se stabilise.

Cela va servir à savoir à quel moment après l'attaque il faut commencer l'analyse pour avoir

les résultats les plus fiables possibles.

2.4.2 Succession d'accords

Un autre problème qui peut se poser quant au moment auquel on va faire notre analyse, est le cas des sons qui sont constitués de plusieurs accords joués successivement. En effet, les résultats seront tronqués si l'analyse se fait sur deux accords consécutifs. Pour cela, il faut faire de la segmentation temporelle afin de savoir à quel moment les attaques (changement d'accord) ont eu lieu. Il est aussi utile de noter le moment où le guitariste gratte les cordes vers le bas ou vers le haut car cela peut entraîner des non-stationnarités.

Les méthodes de segmentation temporelle vont être présentées dans la section 4.3. La Figure 2.12 présente le signal temporel avec en superposition les marqueurs (changement d'accord) et les sous-marqueurs (accord joué vers le haut ou vers la bas). Cette information donnée par les marqueurs permet donc de séparer les accords joués dans une pièce musicale. Ce pré-traitement est nécessaire pour être capable d'analyser, estimer et reconnaître les accords joués.

2.4.3 Construction des accords

Un accord est obtenu lorsqu'on joue 2 notes ou plus ensembles. Comme pour les gammes, on peut construire un accord à partir de n'importe quelle note. Il existe donc un très grand nombre d'accords que l'on peut ensuite classer selon leur type : majeur, mineur, 7ème de dominante, 9ème de dominante, etc...

Les accords dits "simples" obéissent aux mêmes règles que les accords "complexes". A savoir, la même règle de construction ou plutôt de séquence de notes. Ils sont dénommés "simples", car ils ne comportent pas de difficultés techniques dans leur pratique, comme le fameux "barré".

Les accords appartiennent à deux familles distinguées par le nombre de sons qui les composent :

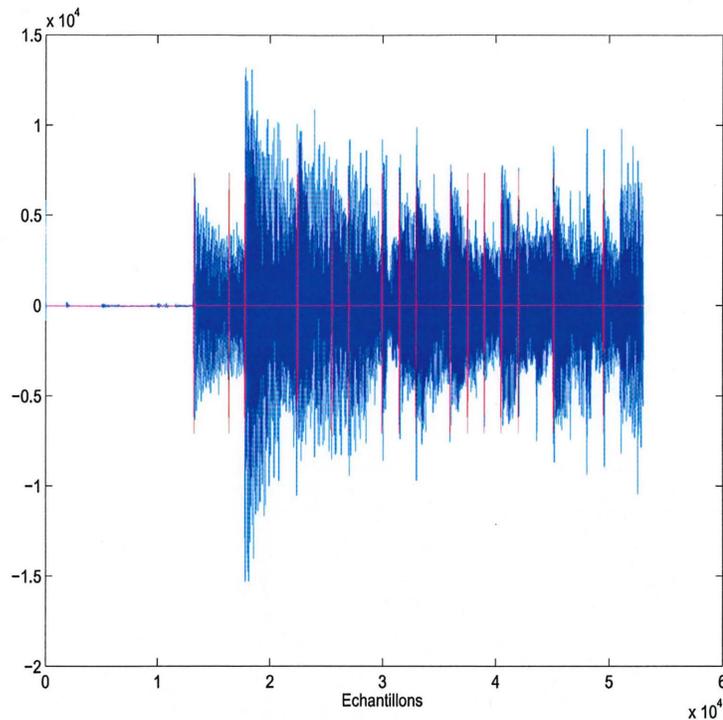


Figure 2.12 – Segmentation par accord d'une partie d'une pièce musicale

- les accords de trois sons
- les accords de quatre sons et plus.

Une règle de la composition des accords réside dans le fait que :

- l'accord est appelé *accord majeur* si la tierce jouée est majeure.
- l'accord est appelé *accord mineur* si la tierce jouée est mineure.

Les accords majeurs : Un accord majeur est composé de la fondamentale, de la tierce majeure, et d'une quinte juste. Ces 3 notes forment une triade majeure. Sur la Figure 2.13, on retrouve pour chaque accord quelles cordes doivent être jouées et dans quel fret.

Si par exemple on se concentre sur l'accord C (Do), il est composé de la façon suivante :

- la tonique est donc la note *Do* (C3).

	C (Do)	D (Ré)	E (Mi)	F (Fa)	G (Sol)	A (La)	B (Si)
T		2		1	3		
A	1	3		1	3	2	3
B	0	2	1	2		2	3
	2	0	2	3		2	3
	3		2	3	2	0	1
			0	1	3		

Figure 2.13 – Accords majeurs avec leurs compositions

- la tierce majeure est le *Mi* (E3).
- la quinte est le *Sol* (G3).

La composition d'un accord est la combinaison de ces notes sur le manche. Cependant, ces notes peuvent se répéter sur l'ensemble des cordes, afin de rendre la sonorité plus "riche" et plus "puissante" en terme de complémentarité sonore.

On peut donc présenter l'accord C par le diagramme présenté à la figure 2.14.

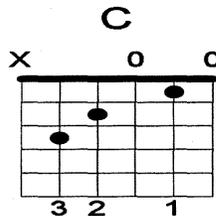


Figure 2.14 – Accord C avec sa composition

On observe que l'accord est bien "ouvert" car il est joué "en haut du manche", près du sillet. Notons que l'expression "en haut du manche" évoque les sons les plus graves, donc bas en fréquence, pouvant être obtenus sur une guitare.

Ce diagramme est en fait une représentation graphique des notes jouées, corde par corde, dans l'accord (de la première à la sixième case, soit du *Mi aigu* au *Mi grave*) :

- la 1^{ière} corde est jouée "à vide (ou case 0)" et la notation du diagramme le symbolise par le petit signe \circ sur un *E* qui est la tierce majeure de *C*.

- la 2^{ième} corde est jouée à la 1^{ième} case (représentée par le signe • sur le diagramme), ce qui donne un *C* et constitue la tonique.
- la 3^{ième} corde est jouée "à vide (ou case 0)" et la notation du diagramme le symbolise par le petit signe o sur un *G* qui est la quinte de *C*.
- la 4^{ième} corde est jouée à la 2^{ième} case (représentée par le signe • sur le diagramme) ce qui donne *E* et constitue la tierce majeure.
- la 5^{ième} corde est jouée à la 3^{ième} case (représentée par le signe • sur le diagramme) ce qui donne *C* et constitue la tonique.
- la 6^{ième} corde n'est pas jouée (représentation par le signe x sur le diagramme).

Si la dernière corde n'est pas jouée, c'est pour "attaquer" l'accord en descendant lors du frotter, avec le doigt ou le médiator, et de commencer cette attaque par la tonique.

Relations fréquentielles : La technique de construction des accords nous montre que selon l'accord que l'on va jouer, que des notes différentes vont entrer en jeu. On peut associer une fréquence fondamentale à chacune de ces notes. Dans le cas de l'accord C elles sont :

- la 6^{ième} corde n'est pas jouée donc pas de fréquence.
- la 5^{ième} corde est jouée en C3 de fréquence fondamentale 130.81Hz.
- la 4^{ième} corde est jouée en E3 de fréquence fondamentale 164.81Hz.
- la 3^{ième} corde est jouée en G3 de fréquence fondamentale 196.00Hz.
- la 2^{ième} corde est jouée en C4 de fréquence fondamentale 261.63Hz.
- la 1^{ière} corde est jouée en E4 de fréquence fondamentale 329.63Hz.

Comme on l'a vu précédemment, pour que deux signaux aient des harmoniques qui se recouvrent, il faut que $k \cdot F_s = l \cdot F_R$ avec F_s et F_R , respectivement les fréquences fondamentales des sons *S* et *R*, et les produits $k \cdot f_s$ et $l \cdot f_s$ étant les valeurs des fréquences de ces harmoniques.

Or, en prenant le simple fait qu'il y a un rapport de $3/2$ entre la tonique et sa quinte on va retrouver cette ambiguïté pour certains harmoniques. En effet, quand R est à la quinte de S , la moitié des harmoniques de S sont aussi des harmoniques de R .

En prenant le cas concret de notre accord C, on a G3 qui est la quinte de C3. Donc, en terme de fréquence, on va avoir $G3 = 196.00 \text{ Hz}$ et $C3 = 130.81 \text{ Hz}$, est on retrouve bien $196.00 \simeq 3/2 * 130.81 = 196.20$. Cela veut donc dire qu'on va avoir des recouvrements d'harmoniques entre la tonique et sa quinte.

De plus, on remarque qu'on va retrouver la note C3 et la note C4 dans cet accord. Ces deux notes sont espacées d'une octave, c'est à dire que tous les harmoniques de C4 vont être des harmoniques de C3. Il en va de même entre le E3 et le E4.

Ceci est valable dans tous les accords majeurs, car ils sont tous construits de la même façon. L'analyse d'un accord va donc être rendue difficile à cause des relations fréquentielles qui existent entre les différentes notes composant un accord.

2.5 Conclusion

Dans cette première partie, on a vu les généralités sur l'acoustique des instruments à cordes, en particulier les facteurs qui caractérisent un son émis par un de ces instruments. Ils sont au nombre de 4 et sont respectivement sa hauteur, son intensité, sa durée et son timbre.

Étant donné que tous les signaux que l'on va analyser sont émis par une guitare, un bref descriptif de celle-ci était nécessaire. On sait qu'à vide, les 6 cordes de notre guitare oscillent à des fréquences qui leur sont propres. On a aussi vu qu'en modélisant une infime partie de la corde, on prouvait que les harmoniques de notre fréquence fondamentale ne sont pas situés à des facteurs entiers de celle-ci.

Par la suite, on a vu que les sons émis par une guitare peuvent être modélisés par des signaux sinusoïdaux composés de plusieurs fréquences ayant chacune des harmoniques (signaux polyphoniques ou multipitch). Dans le cas des sons polyphoniques (un accord de guitare par exemple), il va falloir faire attention aux recouvrements entre harmoniques appartenant à

différentes fréquences fondamentales. Ceci a été prouvé dans la partie concernant la construction des accords, et principalement les accords majeurs.

Après avoir présenté l'instrument que l'on veut analyser, une revue de littérature sur l'analyse de pitch ainsi que sur l'analyse multipitch est réalisée.

CHAPITRE 3

ANALYSE MULTIPITCH : ETAT DE L'ART

Une revue de littérature a été réalisée sur les sujets en rapport avec l'estimation de la fréquence fondamentale dans les signaux audio. Notamment sur la transcription automatique de la musique, les différentes techniques d'analyse multipitch que ce soit avec des méthodes de traitement de signal ou avec des méthodes plus novatrices et l'analyse des instruments musicaux.

Dans ce chapitre, on présente tout d'abord les méthodes qui ont permis de récolter tous les articles, thèses et autres documents. Ensuite on va présenter les différents types d'approche de résolution du problème puis les différents algorithmes existant aussi bien pour l'analyse de pitch que pour l'analyse multipitch.

3.1 Méthodes de recherche

Les recherches qui sont faites dans le Laboratoire de Recherche sur la Parole et l'Audio de l'Université de Sherbrooke portent sur plusieurs domaines, mais notamment le codage de parole et la voix IP. Cependant, l'analyse de signaux musicaux n'est pas un sujet qui avait été traité auparavant. De ce fait, on ne partait d'aucune base lorsque la recherche de littérature a commencé. On s'est donc renseigné sur les laboratoires universitaires ou privés qui travaillaient de la domaine pour faire une première sélection sur les travaux pertinents qui avaient déjà été réalisés ou qui étaient en cours.

Après s'être un peu plus familiarisé avec le sujet, on a commencé à faire des recherches de publications sur celui-ci. Cette recherche a bien entendu été facilitée par le grand nombre d'information que l'on peut trouver sur le *World Wide Web*, notamment les pages web des différents laboratoires universitaires sur lesquels les publications sont en ligne, ou même les pages web personnelles des différents auteurs. De plus, un grande quantité de journaux et

de publications dans les conférences internationales traitant du traitement de signal étaient à notre disposition dans le laboratoire et les bibliothèques. En ce qui concerne les journaux, la recherche s'est principalement faite dans le Journal of Acoustical Society of America, IEEE Transactions on Acoustic, Speech and Signal Processing et le Journal of the Audio Engineering Society. Pour les conférences, les deux principales dans lesquels les recherches ont été effectuées sont l'ICASSP (International Conference on Acoustic, Speech and Signal Processing) et l'AES (Audio Engineering Society). A travers les premiers articles trouvés, nous avons vu que les recherches significatives dans le domaine se faisaient principalement dans certains laboratoire tel le Machine Listening Group du Massachusetts Institute of Technology, le Center for Computer Research in Music and Acoustics de l'University of Stanford, l'Institut de Recherche et Coordination Acoustique/Musique de Paris, le Signal Processing Laboratory à Tampere University of Technology et l'Acoustic Laboratory de Helsinki University of Technology.

Avec toutes les principales publications trouvées, on a pu étendre notre liste grâce aux références présentes dans chacune de ces publications. Le résultat de ces recherches est donc une importante collection de publications sur le sujet traité et ses dérivés.

3.2 Estimation de la fréquence fondamentale

3.2.1 Problématique

Un des problèmes dans l'analyse des signaux musicaux est la séparation et l'identification des signaux composant une source musicale complexe. Par complexe on entend qui contient des fréquences différentes. Cette analyse passe par la détermination de la fréquence fondamentale de chacune des composantes de ce signal. Cela représente une part importante de tous les systèmes de traitement audio, quelques propositions ont déjà été faites, et ce problème est considéré comme très intéressant car difficile à résoudre.

La détection de la fréquence fondamentale d'un signal monophonique est un problème que l'on peut considérer comme résolu étant donné que de nombreux algorithmes ont été proposés

et qu'ils sont fiables (cf section 3.3).

En ce qui concerne la détection multipitch, les premiers travaux remontent aux années 1970 avec Moorer [Moorer, 1977]. Il a développé un système permettant de séparer le son de deux instruments joués simultanément. Cependant, afin d'être robuste, il fallait que chaque son soit monophonique, périodique et qu'il n'y ait pas de recouvrement d'harmonique entre les deux sons. Il s'agissait donc du premier essai de détection multipitch qui en a engendré de nombreux autres (cf section 3.4).

3.2.2 Schéma général d'un algorithme

Il est possible de classer les algorithmes selon le type de technique utilisé dans le pré-traitement de l'algorithme. C'est à dire que cela revient à classer selon le choix de représentation du signal. Il existe principalement 3 types de classes qui sont :

- Classe des algorithmes temporels
- Classe des algorithmes temporels à court terme
- Classe des algorithmes fréquentiels à court terme

Indépendamment de la classe d'algorithmes, la chaîne de traitement d'un algorithme donné comporte en général 3 parties qui sont représentées Figure 3.1.

Comme on vient de le voir précédemment, le pré-traitement consiste à mettre en forme le signal et à le représenter sous une forme utilisable par le traitement fondamental. La mise en forme du signal analogique qui sort de la chaîne d'acquisition du signal acoustique, consiste à numériser ce signal et à appliquer certaines transformations comme du filtrage (passe-haut, passe-bas, non-linéaire), sous-échantillonnage, seuillage, Puis, selon le traitement fondamental, une représentation de ce signal est obtenue par application d'une transformation : représentation temporelle (aucune transformation), représentation à court terme temporelle (extraction de trames successives avec recouvrement et fenêtrage éventuelle), représentation à court terme fréquentielle (transformation temps-fréquence à court terme).

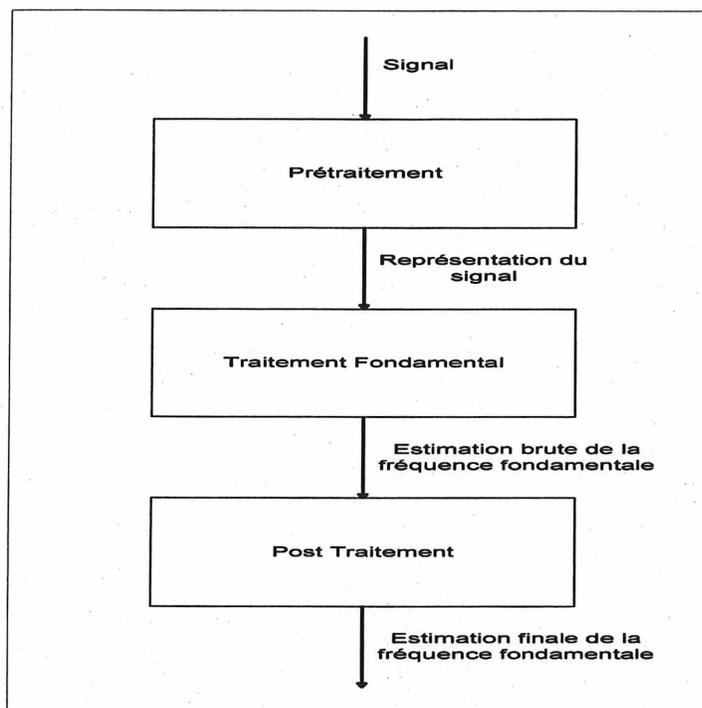


Figure 3.1 – Schéma général d'un algorithme de détermination de la fréquence fondamentale

Le but de ce pré-traitement est donc de mettre en évidence la structure périodique du signal en rehaussant l'information utile.

Le traitement fondamental dépend de l'algorithme choisi. C'est cette partie qui est la plus importante, en ce sens que c'est là qu'est effectuée l'estimation proprement dite de la fréquence fondamentale. Il n'y a donc pas de critères communs que l'on va retrouver entre chaque algorithme.

Le but du post-traitement est de corriger certaines erreurs comises par le traitement fondamental, d'affiner les résultats obtenus, et d'extraire des caractéristiques quand aux résultats trouvés. La correction des erreurs peut se faire par une vérification de la concordance des résultats par rapport aux résultats des trames précédentes. L'affinement des résultats peut se faire par un calcul plus précis de la valeur des fréquences fondamentales avec des méthodes d'interpolation fréquentielles. Quant à l'extraction de caractéristiques, cela peut être vu comme la détermination de la note ou l'accord de musique qui a été joué.

3.3 Algorithme de résolution de pitch

Dans un premier temps, une recherche sur les algorithmes de résolution de pitch (une seule fréquence fondamentale) a été réalisée pour comprendre le fond du problème à résoudre. La détection de la fréquence fondamentale d'un signal mono-fréquence est un sujet bien étudié et maîtrisé dans la littérature spécialisée.

Nous allons faire un compte rendu de ces algorithmes selon le domaine dans lequel l'analyse est faite. Ici, on s'intéresse uniquement aux algorithmes temporels ou fréquentiels à court terme, c'est à dire faisant des analyses trame par trame.

3.3.1 Algorithmes temporels à court terme

Comme on l'a vu dans la section 2.3, les signaux que nous traitons sont périodiques. Or, un signal périodique voit sa forme d'onde se répéter identique à elle-même. Cependant, on a aussi vu que l'évolution de la fréquence fondamentale se fait de façon rapide. Cela impose donc de considérer la notion de périodicité de façon locale, et ceci peut être effectué en analysant le signal grâce aux techniques d'analyse temps-fréquence. En pratique, nous avons vu que cela revient à extraire du signal des trames qui se recouvrent, et auxquelles il est possible d'appliquer une transformée. La donnée de l'ensemble des trames constitue, sous certaines conditions, une représentation à court terme du signal.

La représentation à court terme dans le domaine temporel du signal consiste à extraire une représentation trame par trame du signal sans effectuer de transformation sur ces trames. L'idée est alors de travailler sur chaque trame et de fournir pour chacune une valeur de période. Dans certains cas, le signal est supposé stationnaire sur la trame donnée (par exemple pour l'autocorrélation). La taille de la trame est un paramètre important, et dans le cas où cette taille est fixe, elle est choisie pour contenir dans tous les cas au moins 3 périodes du signal. Il faut donc connaître un majorant des périodes attendues du signal. Cela va dépendre de l'instrument analysé, mais le calcul sera effectué de la même façon.

On verra que pour la plupart des algorithmes, il s'agit de trouver les extremum d'une fonction

de la période test. Si τ est la période test, nous noterons cette fonction de périodicité $FP(\tau)$.

Algorithme d'autocorrélation : Le principe des méthodes de corrélation est d'utiliser le fait que le signal doit se répéter à peu près identiquement à lui même sur des périodes successives. Il est donc possible de corréler le signal de départ et une version décalée de ce signal d'une valeur correspondant à la période cherchée. Le résultat est que la fonction de corrélation présentera un pic à la période attendue du signal observé.

Le cas particulier de la méthode d'autocorrélation travaille sur une version discrète de la fonction d'autocorrélation d'un signal (déjà multipliée par une fenêtre). Cette méthode consiste donc à effectuer une corrélation entre deux parties de la trame de même taille et décalée du décalage τ . La fonction de périodicité de l'autocorrélation est décrite dans l'équation 3.1.

$$FP_{AutoC}(\tau) = \frac{1}{n} \sum_{k=1}^{n-\tau} x_k * x_{k+\tau} \quad (3.1)$$

Le résultat de la fonction d'autocorrélation est que pour un signal périodique de période T_0 , chaque multiple de la période T_0 montre un maximum d'égale amplitude. Le procédé pour extraire la période est donc de choisir le premier pic (celui qui correspond au plus petit décalage).

De nombreux algorithmes basés sur la méthode de l'autocorrélation ont été proposés. En 1990 Brown a publié des résultats où la détermination du pitch se faisait par cette méthode d'autocorrélation. Le but de cette étude était de voir, si oui ou non la méthode d'autocorrélation pouvait s'adapter à la détermination de la fréquence fondamentale en prenant comme entrée des signaux musicaux. De plus un test était fait pour voir si une autocorrélation plus restreinte donnait de meilleurs résultats, et dans quelles conditions [Judith.C.Brown, 1992].

Les résultats sont importants car Brown a montré que les deux techniques donnaient des résultats cohérents pour la détermination de la fréquence fondamentale. Cependant, et il est important de noter que dans notre cas, cette méthode ne donne pas de résultats convaincants avec des signaux multipitch. De plus, cette méthode n'est pas exploitable lorsqu'on a à faire

à une large échelle de pitch et aussi lorsque le spectre rencontré est varié. Comme ce sont les caractéristiques premières des signaux musicaux multipitch, cette technique ne peut être optimisée pour ce type de signaux.

Algorithme basé sur la forme d'onde : Certains détecteurs de pitch, sont basés sur la mesure du minimum et du maximum dans le domaine du temps. Par exemple, l'algorithme de Miller [Miller, 1975] détecte tout d'abord l'excursion de cycles dans la forme d'onde en observant l'intervalle entre les principaux croisements par zéro. Le reste de l'algorithme essaye d'identifier le principal cycle d'excursion, par exemple ceux qui correspondent à la vraie période du pitch. L'algorithme a été rapporté comme donnant de bons résultats, mais pas aussi précis que d'autres algorithmes.

Conclusion Les algorithmes temporels à court terme (utilisant une représentation à court terme dans le domaine temporel du signal) présentent l'avantage par rapport aux algorithmes temporels d'être plus fiables en ce sens qu'ils sont moins sensibles aux distorsions de phases, au bruit, et aux signaux parasites. Ces algorithmes ont tous comme principe de minimiser un critère d'erreur fonction de la période. Ce critère peut être une distance dans un espace structuré, ou correspondre à un corrélation, ou bien une simple fonction sans justification théorique.

3.3.2 Algorithmes fréquentiels à court terme

D'un point de vue théorique, les séries de Fourier indiquent qu'un signal parfaitement périodique peut être décomposé en une somme d'exponentielles dont les fréquences sont toutes multiples d'une fréquence particulière que l'on appelle fréquence fondamentale (elle correspond à l'inverse de la plus petite période du signal). Les différentes exponentielles sont appelées harmoniques du signal, et chacune d'entre elle est associée à une amplitude.

D'un point de vue pratique, cette décomposition en série de Fourier est remplacé par une transformée de Fourier (ou une autre transformée) qui elle aussi fournit les harmoniques de notre signal. De plus, il faut préciser que la seule analyse réalisable est une analyse à court

terme (et à support temporel borné) dans le domaine fréquentiel. Cette analyse nécessite l'utilisation d'une fenêtre sur le signal ce qui modifie la propriété de décomposition (de séparation) en harmoniques. Il faut donc faire attention quant au choix de la taille de la fenêtre car c'est elle qui va nous donner notre résolution fréquentielle en fonction de la fréquence d'échantillonnage. En effet, si notre fréquence d'échantillonnage est de 4.41 kHz et que notre taille d'analyse est de 4096 échantillons, alors la résolution fréquentielle sera de l'ordre du Hertz. Cela se rapproche donc du choix de la taille de la trame dans le cas des algorithmes temporels à court terme.

L'idée de fond des algorithmes utilisant cette représentation est d'exploiter la structure harmonique des signaux périodiques. Cela peut se faire de plusieurs façons, en utilisant l'une ou l'autre des caractéristiques de cette structure, en tentant d'identifier directement les harmoniques par appariement, ou bien en cumulant les amplitudes de toutes les harmoniques par compression spectrale. Sans oublier les algorithmes fonctionnant dans le domaine cepstral, qui utilisent la régularité de la répartition des harmoniques.

Appariement d'harmoniques : Cette approche a pour but d'apparier l'ensemble des partiels trouvés dans la représentation du signal aux harmoniques d'une valeur de référence. Ceci nécessite donc d'extraire de la représentation un ensemble de partiels, chaque partiel ayant une fréquence et une amplitude. Ensuite, pour chaque valeur test de fréquence fondamentale f_0 , il s'agit d'estimer à quel point les partiels sont représentés par les harmoniques de f_0 . Donc on recherche la fréquence fondamentale dont les harmoniques représentent le mieux les variations du signal.

Une méthode proposée par Doval et Rodet fait partie de cette classe [Doval et Rodet, 91]. Elle se base sur un modèle probabiliste afin de prendre en compte le caractère bruité, la constante évolution et l'imprécision des paramètres des signaux traités. Ils se servent de la structure harmonique du signal pour déterminer le meilleur candidat à la fréquence fondamentale grâce à la technique du maximum de vraisemblance. Les tests ont été faits sur une large bande de f_0 allant de 50 à 4000 Hz avec un délai de réponse très court de 20 ms. De plus, les tests ont été faits sur des signaux musicaux monophoniques et des signaux vocaux et donnent de très

bon résultats.

Plus tard, Doval a présenté une structure d'un niveau supérieur (basé sur les modèles cachés de Markov et l'algorithme de Viterbi) qui peut être utilisée pour supprimer les plus grosses erreurs faites lors de la détection de pitch en recherchant le contour du pitch au travers des trames successives [Doval, 1994].

Autocorrelation : Il y a d'autres méthodes qui sont plus directes que l'approche probabiliste de Doval. La structure du spectre du signal harmonique est appropriée pour mettre en évidence la fréquence fondamentale. On calcule l'intervalle équidistant entre les maximums en appliquant l'autocorrelation à l'amplitude du spectre. Pour cela, nous allons prendre deux exemples.

Tout d'abord la méthode de Lahat [Lahat, 1987] : Dans ce cas le spectre du signal est aplati par un groupe de filtres passe-bande et le pitch est extrait de la fonction d'autocorrelation appliquée à la sortie des filtres.

On peut aussi parler de la technique de Kunieda [Kunieda, 1996] qui applique la fonction d'autocorrelation sur le spectre du signal. Sa méthode est appelé ACLOS (Autocorrelation of LOg Spectrum) car il prend le logarithme de l'amplitude du spectre, aplatit le résultat et calcule la fonction d'autocorrelation pour trouver l'information sur le pitch.

Ces deux méthodes sont dans un sens plus sophistiquées que la méthode d'autocorrelation directe mais le principe reste le même.

Compression spectrale : L'idée de la compression spectrale est de cumuler les amplitudes de toutes les harmoniques. Ceci est effectué de façon globale, c'est à dire en utilisant le seul fait que les harmoniques ont des fréquences multiples de la fréquence cherchée. La compression consiste donc à appliquer une homothétie de rapport entier à l'échelle des fréquences puis à sommer les spectres comprimés [Hermes, 1988]. Le résultat attendu est que les amplitudes des harmoniques s'ajoutent à la fréquence fondamentale cherchée et pas aux autres fréquences. Si on nomme \tilde{s} le spectre, le spectre comprimé d'un facteur i , \tilde{s}_i , s'obtient par la

formule :

$$\tilde{s}_i(f_0) = \tilde{s}(i * f_0) \quad (3.2)$$

et la fonction de périodicité de la compression spectrale s'écrit alors :

$$FP_{CompSpec} = \sum_{i=1}^{n(f_0)} |\tilde{s}_i| \quad (3.3)$$

La technique utilisée est globale (aucune harmonique n'est identifiée en position, amplitude ou numéro), et ne traite pas le problème des erreurs d'octaves. Tout sous-multiple de la fréquence fondamentale cumule aussi les harmoniques, ce qui se traduit par un pic à tous les sous-multiples de la fréquence fondamentale. Il faut cependant noter que le nombre d'harmoniques cumulées à f_0/k est d'autant plus faible que le facteur k est grand. Le pic à la fréquence f_0/k sera donc en général de plus faible amplitude que celui à f_0 .

Les avantages des algorithmes de compression sont multiples : ils sont assez résistants au bruit, les signaux étudiés n'ont pas besoin d'avoir de partiels à la fréquence fondamentale, ni de partiels à chaque harmonique.

Cepstre : Le cepstre est la transformée de Fourier inverse du logarithme du spectre du signal étudié. Pour un spectre composé de nombreuses harmoniques, la transformation cepstrale fournit directement une estimation de la période grâce au principe de décomposition homomorphique. Il suffit de prendre le maximum dans la zone des *quéfrences* possibles. En effet, dans ce cas, le logarithme du spectre est considéré comme périodique et quasiment sinusoïdal. Sa transformée présente donc un pic bien marqué à la période correspondante.

Cette technique a été très souvent utilisée dans le cas des signaux vocaux, car la condition citée précédemment est très souvent vérifiée pour les signaux de parole. Cependant, dans le cas où il y a peu d'harmoniques ou bien si le signal comporte des harmoniques paires faibles, les conditions ne sont plus vérifiées. De plus, cette méthode ne permet pas de prendre en compte une large échelle de pitch, donc elle n'est pas applicable aux signaux musicaux [Rabiner et coll., 1976].

Conclusion Les algorithmes fréquentiels à court terme (utilisant une représentation à court terme dans le domaine fréquentiel du signal) héritent des caractéristiques déjà signalées plus haut concernant les algorithmes temporels à court terme : ils sont plus résistants au bruit, et aux signaux parasites et fonctionnent en général sur la minimisation d'un critère d'erreur.

Mais ils présentent en plus la capacité d'exploiter la structure harmonique du signal. Certains le font de façon globale (compression spectrale), d'autres de façon plus analytique (appariement d'harmoniques). On approfondira cette capacité dans l'algorithme que nous présenterons au Chapitre 4, et qui se rattache aux algorithmes utilisant l'appariement d'harmoniques.

3.4 Algorithme de résolution multipitch

Un humain est capable de séparer un son parmi un mélange de sons. Jusqu'à quel point est-il possible d'automatiser cela ? On a vu précédemment que pour des signaux monophoniques, de nombreux algorithmes ont été développés et sont très robustes. Que va-t-il se passer quand un son va être perturbé par un autre ou plusieurs autres ? Les composantes fréquentielles de chacun vont se mélanger et donc le spectre va devenir plus complexe. La plupart des méthodes utilisées font recourt à une analyse fréquentielle. Faut-il penser à une analyse temporelle ou alors une analyse tempo-fréquentielle ?

Faire de la détection multipitch peut revenir à faire de la séparation de sources et donc mener à faire de la transcription automatique de musique. C'est dans cet esprit là que la plupart des recherches sur l'analyse multipitch ont été réalisées. Il est cependant important de noter que de nombreux travaux dans le domaine de la détection multipitch ont été faits dans le domaine de la parole mais très peu dans le domaine de la musique. En effet, on prend deux sons monophoniques que l'on mélange ensemble et le but de l'analyse est de les séparer, c'est à dire de retrouver identiques leur original. Les premiers travaux sur l'analyse multipitch remontent à 1977 avec Moorer [Moorer, 1977]. En effet, le système pouvait séparer deux sons provenant d'instruments différents. Il y avait cependant de fortes contraintes qui limitaient sérieusement la possibilité de faire de la transcription de musique. Les sons devaient être

monophoniques, périodiques et choisis de telle sorte que les harmoniques de chacun d'entre eux ne se superposent pas. Cependant, cela a initié de nombreux travaux dans le domaine, et c'est au cours des dernières années que l'on a vu les meilleures améliorations avec certaines techniques que l'on va présenter ci-dessous.

3.4.1 Modèle basé sur la note

Si on prend comme exemple de séparer deux sons espacés dans leur pitch d'un octave, et bien, celui-ci est théoriquement ambigu sans avoir plus de connaissance sur le signal. Les modèles du son sont des structures de données, qui représentent les caractéristiques du spectre dans un certain format.

Kashino et Tanaka [Kashino et Tanaka, 1993] sont les deux premiers à avoir décrit un système complet de détection basé sur le modèle de la note. Les caractères indicateurs dont ils ont tenu compte sont les suivants :

- Proximité spectrale (compacité en temps ou en fréquence)
- Concordance d'harmonique vs Non accord d'harmonique
- Changement synchrone du comportement :
 - Déclenchement commun
 - Décalage commun
 - Modulation d'amplitude/de fréquence commune
 - Mouvement équidirectionnel dans le spectre
- Proximité spatiale

Ils ont aussi proposé un algorithme pour modéliser automatiquement le son, ce qui veut dire extraire le modèle de la note à partir de l'analyse faite sur celle-ci. Cependant, ce système était très limité, mais tout de même assez générique par rapport au type d'instruments utilisés. En simulation, le premier modèle a été capable de transcrire trois notes différentes, alors que le deuxième n'a pu en transcrire que deux.

3.4.2 Modèle basé sur l'audition

Ici, nous allons introduire des représentations multidimensionnelles appelées *trame* et *correlogramme*. L'intérêt principal de ces structures est que non seulement elles expliquent un éventail des phénomènes psychoacoustiques dans l'audition, mais elles essaient aussi d'organiser le son à partir de sa source, ce qui dans le cas de la musique veut dire qu'elles sont capables de distinguer des sons.

Il n'y a que Martin [Martin, 1996] qui a utilisé la représentation par correlogramme, et il suggère que celle-ci facilite la détection de deux sons qui sont séparés entre eux d'un octave dans leur pitch, même sans introduire le modèle de la note. Cependant seule une petite simulation est présentée pour vérifier cela.

On souligne cela quand deux sons ont une relation directe entre leurs harmoniques. Par exemple, la fréquence fondamentale du plus haut est un multiple entier de celle du plus bas, alors les deux sons ne peuvent être, sans ambiguïté, séparés en deux sons, et ceci par aucune technique existante. Pour que cela devienne possible, il faut que des caractères indicateurs, tels que le temps de montée (onset) ou des modulations d'amplitudes ou de fréquences spécifiques qui associent les harmoniques à leur source, soient perceptibles.

La représentation par *correlogramme* ne fait pas appel à ces caractères indicateurs, mais par contre la représentation par *trame* utilise explicitement l'indicateur de modulation d'amplitude. Il faut tout de même noter qu'un tel indicateur n'est pas forcément présent dans tous les signaux musicaux que l'on va rencontrer.

3.4.3 Lissage spectral

Une des méthodes qui semblent les plus pertinentes pour résoudre ce problème est celle qui utilise le lissage spectral [Klapuri, 2001]. En fait ce principe a été choisi car on le retrouve dans le système auditif humain. Il a pour but de séparer les sons en supposant que l'enveloppe spectrale d'un son réel est continue.

Avant d'adapter ce principe, le son doit être traité. En effet, l'algorithme de détection de

pitch préliminaire est composé de deux parties qui sont appliquées de façon itérative sur notre signal. Dans un premier temps, l'algorithme va déterminer le pitch du son principal dans notre signal qui est bruité et composé d'autres sons harmoniques. Pour résoudre cela, on travaille dans le domaine des fréquences, et seule l'utilisation d'une *FFT* est nécessaire. Une fois que ce traitement a été fait sur tout le signal, le résidu obtenu va être traité par sous-bande dans le domaine des fréquences. A chacune de ces sous-bandes, une fréquence fondamentale plausible va être extraite et à la fin de ce traitement, ces fréquences fondamentales vont être mises en commun pour donner les fréquences fondamentales du son traité. Cela fut l'objet d'un premier travail de Klapuri [Klapuri et coll., 2000].

Par la suite, la méthode du lissage spectral a été rajoutée. En effet, lorsqu'on soustrayait le pitch principal (et ses harmoniques), il pouvait arriver que d'autres harmoniques aient été supprimés. Cela a été vu dans les relations existantes pour les signaux harmoniques dans la section 2.3.2. En effet, au lieu de prendre la valeur maximale (a_h) de l'amplitude du pitch, celui-ci va prendre le minimum entre cette valeur et celle de son voisin :

$$a_h = \min(a_h, a_{h+1})$$

Un deuxième algorithme a été implémenté. Dans un premier temps on va faire une moyenne (m_h) des amplitudes du spectre à l'intérieur d'une fenêtre de Hamming centré sur le pitch trouvé (a_h). Comme dans l'algorithme précédent, l'amplitude a_h de notre partiel va être remplacé par le minimum de a_h et de la moyenne m_h trouvée :

$$a_h = \min(a_h, m_h)$$

Les résultats trouvés par l'utilisation de cette méthode sont assez significatifs, car sur leur banc de test, ils ont réduits leur taux d'erreur de 25% à 13% dans un signal musical composé de 4 sons. Le 25% était trouvé pour l'algorithme n'utilisant aucune des deux méthodes de lissage et le 13% pour le deuxième algorithme.

3.4.4 Estimation de la fréquence fondamentale et séparation de source

On a dit précédemment que faire de la détection multipitch pouvait revenir à faire soit de l'estimation de fréquence fondamentale soit de la séparation de source. C'est donc pour cela que l'on va retrouver les mêmes étapes dans les méthodes de séparation et d'estimation des F_0 :

1. Estimer la période du son A
2. Se servir de la période du son A pour le supprimer
3. Estimer la période du son B
4. Se servir de la période du son B pour le supprimer
5. Répéter les étapes 1 à 4 jusqu'à ce que tous les sons soient estimés.

La différence principale qui va se faire entre les deux méthodes est dans la façon d'estimer les différents sons composants notre mélange de son (signal polyphonique harmonique).

En effet, il existe de nombreuses méthodes pour faire de la séparation des sons provenant d'un mélange de son. Cependant, cela se base principalement autour du réhaussement et de la suppression (ici on ne parle pas des méthodes multirésolution telles que l'analyse en composantes indépendantes) pour estimer ou reconstruire le spectre d'origine.

Réhaussement d'harmonique Cette méthode utilise la période du spectre du son que l'on veut entendre (son cible) pour le réhausser par rapport au reste [de Cheveigné, 1993]. Dans le domaine spectral, cela revient à sélectionner les harmoniques du son cible et de les réhausser par rapport aux autres. Dans le domaine du temps, le traitement peut être réalisé par un filtre ayant la réponse impulsionnelle suivante :

$$h(t) = \frac{1}{K} \sum_{n=0}^{K-1} \delta(t - nT) \quad (3.4)$$

où t est le temps, T est la période du son cible, δ est la fonction Dirac et K est le nombre de pics dans la courbe de réponse impulsionnelle. Le filtre ajoute des copies du signal qui vont

être retardées dans le temps par des multiples de la période du son cible. Le réhaussement fonctionne avec n'importe quelles sortes de perturbations, mais donne pour le meilleur une modeste amélioration quand au rapport signal à bruit à moins que la réponse impulsionnelle ne soit longue. Cependant, cette technique peut donc s'appliquer aux signaux de musique qui, on l'a vu dans la section 2.3, sont très stationnaires sur certaines périodes du signal.

Suppression d'harmoniques Dans cette technique, c'est la nature harmonique du son perturbé qui est utilisée pour la supprimer [de Cheveigné, 1993]. Dans le domaine des fréquences, cela correspond à mettre à zéro toutes les harmoniques composant ce son. Dans le domaine du temps, le même résultat peut être obtenu par un filtre en peigne qui donne une réjection infinie avec une réponse impulsionnelle courte telle que :

$$h(t) = \frac{1}{2}[\delta(t) - \delta(t - T)] \quad (3.5)$$

La suppression ne va être efficace que dans le cas où la perturbation est de type harmonique. Par exemple, cela ne va pas être utile dans le cas d'un signal perturbé par du vent ou de la pluie dans le cas d'un signal de parole. Dans le cas de signaux de musique, cela pourrait être comparé à une perturbation de batterie (instrument non harmonique).

On voit aussi souvent l'utilisation des deux méthodes en simultanée. Souvent, le réhaussement va être utilisé pour extraire le son le plus fort et la suppression pour le son le plus faible.

Méthode fréquentielle, temporelle ou tempo-fréquence Les méthodes fréquentielles et tempo-fréquence utilisent toutes les deux une analyse fréquentielle mais diffèrent dans la façon de séparer les bandes et dans la résolution fréquentielle dont elles ont besoin. Les méthodes tempo-fréquentielles séparent les bandes en fonction de la période temporelle dominante. Cela fonctionne d'autant plus si des bandes sont dominées par un son et d'autres bandes par d'autres sons. D'un autre côté, les méthodes spectrales séparent leur bandes en fonction de la position qu'elles occupent sur une échelle fréquentielle. Cependant, cela nécessite une fine résolution fréquentielle pour isoler les partiels individuellement dans le spectre.

Cependant, il y a une alternative à l'analyse fréquentielle. On a vu que le réhaussement et la suppression d'harmoniques peuvent être réalisés dans le domaine du temps à l'aide de filtres ayant des réponses impulsionnelles très simples. Le fait de faire une analyse dans le domaine du temps évite d'avoir besoin de choisir une longueur et une forme de fenêtre pour l'analyse fréquentielle ainsi qu'une méthode de transformation à utiliser (FFT, DCT, Q-Transform, ...). De plus, la performance n'est plus seulement déterminée par la résolution fréquentielle.

3.4.5 Autocorrelation

Tolonen [Tolonen, 2000a] a proposé un algorithme d'analyse multipitch efficace qui est une amélioration du modèle de Meddis et O'Mard [R.Meddis et M.Hewitt, 1991]. De nouvelles caractéristiques qui permettent l'analyse de signaux multipitch (signaux de musique ou de parole) sont présentées.

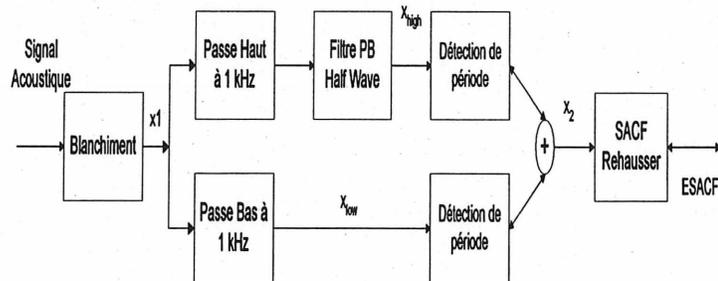


Figure 3.2 – Schéma bloc du modèle de Tolonen

Le schéma bloc du modèle est présenté à la Figure 3.2. Le premier bloc est un filtre de blanchiment qui est utilisé pour supprimer les faibles corrélations temporelles du signal. Le filtre de blanchiment est implémenté en utilisant un prédicteur linéaire modifié tel que décrit en [U.K.Laine et T.Altosaar, 1994]. La technique du *WLP* (Warped Linear Prediction) est la même qu'un prédicteur linéaire classique sauf qu'elle implémente une résolution fréquentielle bandes critiques au lieu d'une résolution fréquentielle uniforme et peut être utilisée pour réduire l'ordre du filtre de façon considérable. Ici, un filtre d'ordre 12 avec une fréquence d'échantillonnage de 22 kHz, fenêtrage de Hamming, longueur de trame de 23.2 ms est utilisé.

Ensuite le signal est séparé en deux bandes, en-dessous et au-dessus de 1 kHz. La séparation

des bandes est effectuée avec des filtres qui ont 12dB/octave d'atténuation dans la bande atténuée. Le bloc *passé-bas* inclus également un *passé-haut* d'affaiblissement ayant 12dB/octave d'atténuation en-dessous de 70 Hz. Le signal en bande haute est redressé en simple alternance et filtré *passé-bas* (filtre ayant les mêmes caractéristiques que celui utilisé pour séparer les bandes). La détection de périodicité est basée sur la méthode de l'autocorrélation. Elle est réalisée en utilisant une *TFD* du signal, compression en amplitude de la représentation spectrale puis une *TFD* inverse. Le signal x_2 de la Figure 3.2 correspondant à la fonction de *SACF* (Summary Autocorrelation function) est présenté à l'Equation 3.7, dans laquelle l'hypothèse que le vecteur d'autocorrélation d'un signal à hauteurs tonales multiples (accord) correspond à une somme d'autocorrélation dilatées des autocorrélations des hauteurs tonales individuelles, plus un terme d'erreur, doit être validée.

$$x_2 = IDFT(|DFT(x_{low})|^k + IDFT(|DFT(x_{high})|^k) \quad (3.6)$$

$$x_2 = IDFT(|DFT(x_{low})|^k + |DFT(x_{high})|^k) \quad (3.7)$$

où x_{low} et x_{high} sont les bandes hautes et basses du signal avant le bloc *Détection de période*.

La détection de périodicité utilisant la *TDF* permet de contrôler le paramètre k ou l'utilisation d'autre traitement non linéaire de la transformée fréquentielle. Ceci n'est pas directement réalisable avec les algorithmes de détection de période évoluant dans le domaine du temps. La *TDF* et *ITFD* sont utilisées pour diminuer le temps de calcul.

Enfin, le dernier bloc est celui qui permet de traiter les résultats de la fonction *SACF*. En effet, les pics dans la courbe de résultats sont relativement de bons indicateurs des *pitchs* potentiels du signal. Cependant, une telle fonction qui récapitule les périodicités contient de nombreuses redondances et informations parasites qui rendent difficile à estimer quels pics sont vrais et quels ne le sont pas. Par exemple, la fonction d'autocorrélation génère des pics à tous les multiples entiers de la fréquence fondamentale.

La courbe originale de la fonction *SACF* est tout d'abord redressée aux valeurs positives et ensuite dilatée d'un facteur 2 puis soustraite à la courbe de la fonction *SACF* redressée, et le résultat est de nouveau redressé pour n'avoir que des valeurs positives. Cela permet de

supprimer la répétition de pic en doublant le temps de retard où le pic d'origine à une plus grande amplitude que le dupliqué. Cela permet aussi de supprimer les parties proches de zéro dans la courbe retardée de la fonction SACF. La fonction résultante de ce calcul est appelé *ESACF* (Enhanced Summary Autocorrelation function).

Des tests ont été réalisés avec des accords de musique et des sources composées de sons d'instruments. Dans la plupart des cas, la détection de son combinant 2 ou 3 notes se résout facilement à condition que le niveau d'amplitude ne soit pas trop différent. Pour les combinaisons de 4 notes et plus, étant donné que les sons se masquent entre eux, la détection est moins fiable.

3.4.6 Conclusion

La partie la plus importante de ce chapitre est une revue de littérature en ce qui concerne les méthodes de résolutions de signaux monophoniques et polyphoniques. Cependant, on a aussi vu quel était le schéma général des algorithmes de détection de fréquence fondamentale. Ceux-ci sont en général composés de 3 étages que l'on nomme : Pré-traitement (filtrage, sous-échantillonnage, seuillage), Traitement fondamental (corps de l'algorithme), Post-traitement (corriger les erreurs, affiner les résultats).

Dans le domaine de la détection monophonique, on a vu que les méthodes évoluent soit dans un domaine temporel à court terme soit dans un domaine fréquentiel à court terme. Toutes les techniques présentées sont fiables, robustes et implémentables en temps réel. Cependant, aucune de ces techniques n'est directement utilisable pour faire de la détection multipitch, mais on peut tout de même se servir des leurs bases.

Du point de vue de la détection polyphonique, il est tout d'abord important de noter qu'aucune technique développée jusqu'à présent n'est totalement fiable et robuste. Il y a donc encore de nombreuses évolutions à faire dans ce domaine. La détection polyphonique à des applications dans le domaine de la séparation de source, de la représentation structurelle de signaux audio, transcription automatique de la musique et la synthèse de son.

On a vu dans ce recensement des différentes techniques de détection multipitch qu'elles uti-

lisent aussi bien des méthodes classiques de traitement de signal, des connaissances acoustiques que le fonctionnement de l'oreille humaine.

Dans le prochain Chapitre, on présente la méthode de détection de fréquences fondamentales comprises dans un signal de guitare que l'on a développée. Celle-ci utilise des techniques classiques de traitement de signal et se compose de 2 étages. Un premier qui sert de pré-traitement et qui évolue dans le domaine du temps, le deuxième qui est le corps de l'algorithme et qui évolue dans le domaine fréquentiel.

CHAPITRE 4

ALGORITHMES DÉVELOPPÉS

Les deux prochains chapitres constituent la partie la plus importante de ce mémoire de Maîtrise, c'est à dire la description des algorithmes développés et les résultats obtenus. Comme on a pu le voir dans le chapitre précédent, de nombreuses publications ont été faites sur la détection monophonique et les résultats obtenus sont robustes, applicables de façon commercial et fonctionnant en temps réel.

Il en est tout autrement pour la détection multipitch. En effet, les publications dont on a parlé précédemment ont toutes été réalisées pour faire de la transcription automatique de musique pour des signaux polyphoniques. De plus, ces systèmes sont assez limités car ils ne fonctionnent que pour des ordres de polyphonies de 2 ou 3 et pour une petite quantité d'instruments. Ceci est surtout dû au fait qu'aucune nouvelle fonction mathématique n'a été développée pour l'analyse polyphonique et donc les systèmes essaient d'utiliser les mêmes principes de détection de pitch que ceux qui sont utilisés dans le cas des signaux monophoniques.

Dans ce chapitre, on présente une méthode d'analyse multipitch à haute résolution pour détecter des accords de guitare. Cela ne se limite pas seulement à détecter les fréquences fondamentales présentes dans un mélange de son, mais aussi d'analyser les caractéristiques d'un signal acoustique provenant d'une guitare. De plus, lorsque les résultats de l'analyse le permettent, les accords qui ont été joués sont retrouvés. Les différentes parties de l'algorithme, ainsi que les différentes techniques sont présentées dans ce chapitre.

4.1 Description

Comme on l'a vu dans la section 3.2.2, un algorithme de détection de fréquence fondamentale peut être séparé en 3 blocs principaux :

- Pré-traitement (filtrage, sous-échantillonnage, seuillage)
- Traitement fondamental (corps de l'algorithme)
- Post-traitement (corriger les erreurs, affiner les résultats)

L'algorithme que l'on présente va respecter les 3 étapes précédentes qui vont être détaillées dans les sections 4.2 et 4.3.

4.1.1 Caractéristiques tirées des autres algorithmes

On a vu que, pour avoir une détection multipitch robuste, il était impossible de se servir des méthodes développées pour la détection monophonique. Cependant, il y a des choses à apprendre des méthodes qui ont été développées pour la transcription automatique de signaux polyphoniques, notamment par rapport à la détection de transitoire et de début de note pour segmenter de façon temporelle le son traité. Étant donné que dans notre cas on se concentre exclusivement sur un instrument (la guitare), il était essentiel de bien comprendre son comportement et ses caractéristiques. Deux aspects qui ont bien été étudiés et présentés dans la section 2.

Il y a une chose très importante à tirer des méthodes déjà développées, c'est que le problème de détection de 2 fréquences fondamentales séparées chacune d'un octave ($f_1 = 110 \text{ Hz}$ et $f_2 = 220 \text{ Hz}$ par exemple) est loin d'être résolu. En effet, aucune technique, qu'elle soit mathématique, psychoacoustique, ou d'un autre domaine, n'a réussi à résoudre ce problème.

4.1.2 Fondement de notre méthode

Il est important de préciser ici que le but de notre algorithme est de détecter le nombre et la valeur des fréquences fondamentales présentes dans un mélange de son formé par un accord de guitare. Les solutions utilisées en pratique pour déterminer les différentes notes des sons multipitch consistent à séparer le plus tôt possible dans la chaîne d'acquisition les différents sons monophoniques composant le signal polyphonique acoustique. On sait aussi que le spectre d'un tel son se prête bien à la décomposition harmonique de notre signal

d'entrée. Dans les chapitres précédents, on a aussi noté qu'il était important de distinguer le moment le plus stationnaire de notre son pour avoir l'analyse la plus précise possible.

4.2 Différents algorithmes

4.2.1 Présentation

La Figure 4.1 représente le schéma bloc de l'algorithme final. Comme on peut le voir, celui-ci est composé de 2 étages. Dans un premier temps, une analyse temporelle est faite sur le signal d'entrée. Cette analyse a pour but de détecter l'attaque de la note ou l'accord joué afin d'avoir un signal stationnaire pour l'analyse fréquentielle. L'entrée de l'étage d'analyse fréquentielle est une trame du signal $y'[n]$. Le choix de la longueur de la trame d'analyse va être expliqué dans la section 5.2. Le rôle de l'étage d'analyse fréquentielle est de détecter le nombre de fréquences et la valeur des fréquences fondamentales comprises dans notre signal d'entrée. De plus, si l'analyse le permet, l'accord qui a été joué peut être retrouvé grâce à ces données (représentant le post-traitement de la Figure 3.1).

Toutes les étapes de cet algorithme vont être présentées et expliquées dans les sections 4.3 et 4.4. Différents cas de figure vont être présentés selon le type de signal d'entrée.

4.2.2 Premières itérations et évolutions

Il est évident que le schéma bloc de l'algorithme présenté à la section 4.1 n'est pas le premier qui ait été trouvé. De nombreuses itérations, comprenant de nombreuses recherches et de nombreux tests, ont été réalisées pour arriver à mettre en oeuvre cet algorithme.

Dans un premier temps, il était essentiel que l'algorithme soit totalement robuste aux signaux synthétiques. Par signaux synthétiques, on entend des signaux qui reprennent certaines des caractéristiques des signaux de guitare, mais qui ne sont tout de même pas comparables aux signaux de guitare réels. De ce fait, aucun des tests présentés par rapport à la détection de transitoire et recherche d'attaque n'avait été effectués, ce sont des améliorations qui ont été apportées aux premières itérations, au même titre que l'utilisation d'un calcul d'interpolation

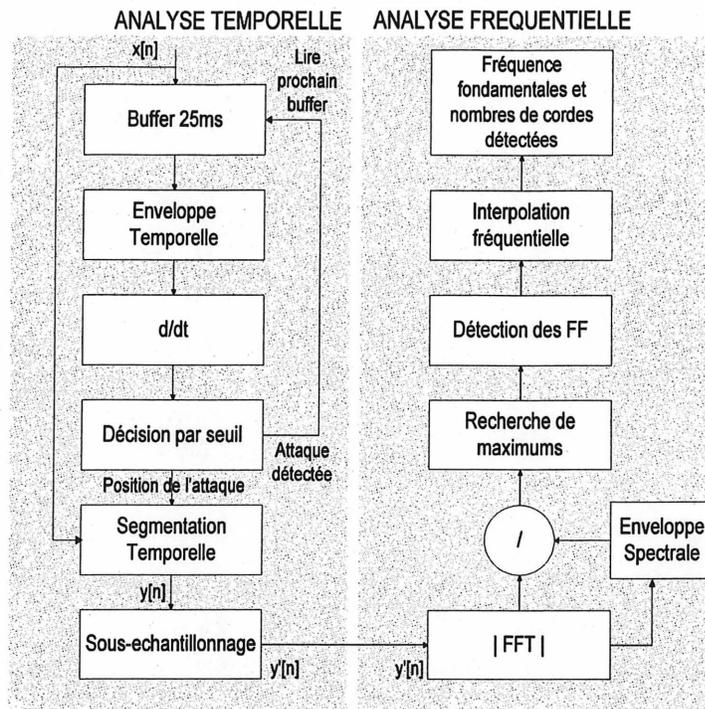


Figure 4.1 – Schéma bloc de l'algorithme

pour affiner la valeur des fréquences fondamentales détectées.

De plus, on n'utilisait qu'une seule guitare au début. On pouvait donc se servir de sa couleur (sonorité) pour développer les algorithmes. Cependant, pour des tests réalisés avec d'autres guitares, les résultats obtenus n'étaient plus aussi fiables. C'est donc pour cela qu'on a introduit l'enveloppe spectrale pour à la fois rehausser les raies spectrales de faibles amplitudes mais aussi être générique quel que soit l'instrument utilisé.

Ensuite, on a beaucoup joué sur la longueur de la taille d'analyse pour avoir les résultats optimaux tout en analysant une longueur temporelle de signal adéquate (ces résultats vont être présentés dans la section 5.2).

C'est donc étape par étape que l'algorithme final est présenté dans les sections suivantes.

4.3 Segmentation temporelle

Comme on l'a déjà expliqué à plusieurs reprises dans ce mémoire, il est impératif de faire notre analyse fréquentielle sur une portion stationnaire de notre signal d'entrée. C'est donc pour cela que l'on a besoin d'un étage d'analyse temporelle avant de commencer l'analyse fréquentielle. Cet étage avait pour but de détecter l'attaque de la note ou l'accord joué.

La détection d'attaque ou de transitoire peut se réaliser par plusieurs méthodes. Il y a eu de nombreuses publications à ce sujet allant d'une détection simple à une détection complexe selon la précision temporelle souhaitée. Dans notre cas, étant donné qu'il n'est pas nécessaire d'avoir une précision à l'échantillon près, seulement les méthodes de base ont été retenues pour ce pré-traitement.

Méthodes qui sont aussi simples qu'un calcul d'énergie par sous-bande, qu'un calcul de la dérivée de l'enveloppe temporelle de nouveau par sous-bande ou alors par une technique plus complexe utilisant l'opérateur d'énergie de Teager. Ces trois techniques ainsi que les résultats obtenus vont être présentés ci-dessous.

4.3.1 Par un calcul d'énergie

L'énergie pour un signal discret $x[n]$ sur l'intervalle $[N_1, N_2]$ est la somme des amplitudes s au carrée pour chacun de ces échantillons, et s'écrit de la façon suivante :

$$W_s(N_1, N_2) = \sum_{n=N_1}^{N_2} s^2(n) \quad (4.1)$$

Dans l'algorithme, on utilise une fenêtre de Hamming que l'on fait glisser d'une demie fois sa longueur sur tout le signal, résultant donc à un recouvrement de 50% entre chaque trame. Ceci, de manière à ce que la valeur maximale correspondant à une transitoire soit au milieu d'une des trames.

Il s'agit d'un calcul assez simple mais long car une opération de multiplication est effectuée pour chaque échantillon composant notre signal. Même si les résultats sont convaincants,

probablement qu'en termes de rapidité, cette méthode de détection d'attaque par l'énergie n'est pas la meilleure.

La comparaison des résultats entre les différentes méthodes va se faire en fonction de la précision de détection de l'attaque. En d'autres termes, quel est l'écart entre l'échantillon de détection et le véritable échantillon d'occurrence de l'événement. Il est vrai qu'on a dit qu'on ne souhaitait pas être à l'échantillon près, mais on ne veut pas non plus être à une trop grande distance.

La taille de la fenêtre de Hamming utilisée est de 1024 échantillons. Étant donné que le signal est échantillonné à 44.1 kHz, cela représente environ 25 ms du signal. L'énergie du signal est donc calculée pour chacune de ces trames. La prise de décision, à savoir si une attaque a eu lieu ou non est basée sur une décision de type seuillage. En effet, il est possible de savoir qu'elle est l'amplitude maximale du signal d'entrée, de ce fait on peut connaître l'énergie maximale. Le seuil va donc être fixé légèrement en deçà de cette valeur. Plusieurs résultats de tests vont être présentés ci-dessous. Les cas de signaux représentant un seul accord mais aussi une succession d'accord vont être présentés.

Comme on peut le voir sur les Figures 4.2, 4.3 et 4.4, la détection de transitoire se fait de façon plus ou moins précise selon le type de signaux traités. Dans le cas d'un accord simple, la détection de transitoire se fait de façon précise. On peut voir sur la courbe (b) de la Figure 4.2 que l'attaque a lieu à l'échantillon 24080. D'après la courbe (d), la détection d'attaque s'est faite dans la trame 47 soit correspondant à l'échantillon : $47 \times 512 = 24064$. Étant donné que l'information temporelle présente lors d'une attaque est plus importante qu'avant celle-ci, cela va engendrer un pic au niveau du calcul de l'énergie dans les trames. Cependant, la détection d'attaque est plus complexe à faire lorsque plusieurs accords se succèdent. Il peut y avoir deux cas : succession d'accord rapprochés ou éloignés. Le cas d'une succession d'accords rapprochés est présenté à la Figure 4.3. On peut nettement voir sur le signal (a) que 4 transitoires sont présentes. Cependant, si celles-ci sont trop proches les unes des autres, alors l'algorithme de détection par le calcul d'énergie ne pourra pas toutes les détecter parfaitement avec l'application d'un simple seuil comme décrit précédemment. C'est en effet ce que l'on constate sur le courbe (b), c'est à dire que la 2^{ième} transitoire n'est pas

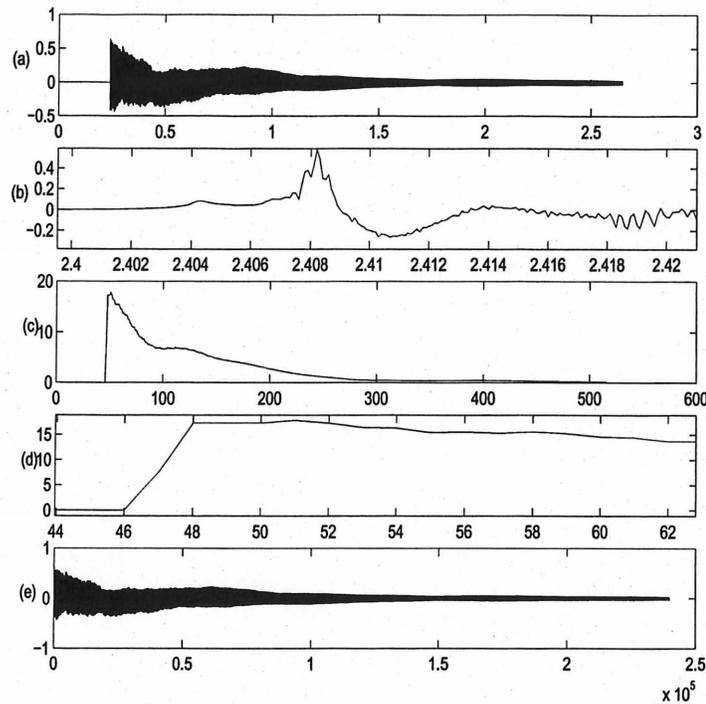


Figure 4.2 – (a) est le signal d'entrée, (b) est un zoom sur le signal d'entrée pour localiser l'échantillon d'attaque, (c) est la courbe d'énergie par trame, (d) est un zoom sur l'énergie pour déterminer la trame d'occurrence de l'attaque, (e) le signal à partir de l'attaque.

détectée. Il y a trop de fluctuations pour qu'une simple prise de décision par seuil soit assez robuste pour être fiable. Cependant, dans le cas où les accords sont suffisamment espacés, la détection est très précise. Ceci est facilement constatable sur la Figure 4.4, et simplement dû au fait que les accords sont éloignés et donc que l'information temporelle d'un accord à l'autre ne se chevauche pas.

On peut donc dire que la détection de transitoire par calcul d'énergie est fiable dans le cas des accords isolés (une succession d'accords éloignés peut être vue comme une suite d'accords isolés), mais peu fiable dans le cas d'accords successifs. Il est donc impératif de trouver une autre méthode de détection de transitoire qui ait la même fiabilité dans le cas d'accords isolés et rapprochés. Pour ce faire, on présente le cas de la méthode de la dérivée de l'enveloppe temporelle.

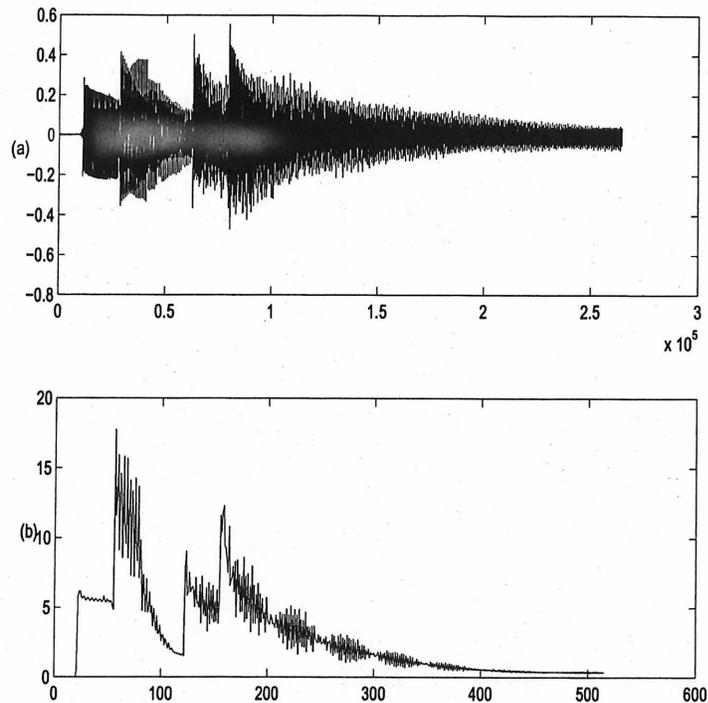


Figure 4.3 – (a) est le signal d'entrée composé d'une succession d'accords rapprochés, (b) est la courbe d'énergie par trame

4.3.2 Par la dérivée de l'enveloppe temporelle

La deuxième technique de détection de transitoire qui a été évaluée est celle de la dérivée de l'enveloppe temporelle. Lorsque l'on analyse un signal de guitare (la courbe (a) de la Figure 4.6) on remarque que l'amplitude du signal va s'accroître rapidement au moment de l'attaque.

Cependant, étant donné que ce signal est sinusoïdal, il ne représente pas la sensation perçue, il faut donc rechercher une autre représentation temporelle du signal plus proche de la sensation. Une méthode consiste à suivre l'énergie véhiculée par le signal sur une durée conventionnelle. Le résultat obtenu indique une forme générale du signal qui est liée *l'enveloppe temporelle*. Cette forme n'est pas unique, elle dépend de la méthode de calcul utilisée et de la durée d'observation. En effet, l'enveloppe temporelle va être une vue grossière de l'évolution de notre signal dans le temps.

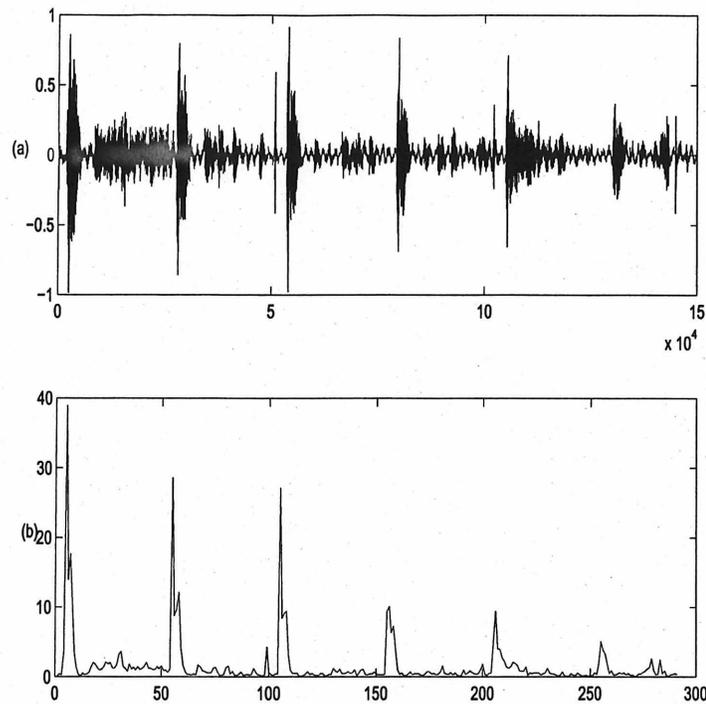


Figure 4.4 – (a) est le signal d'entrée composé d'une succession d'accords non rapprochés avec une addition de bruit blanc entre les accords, (b) est la courbe d'énergie par trame

Dans le cas de la méthode de la dérivée de l'enveloppe temporelle on va dans un premier temps calculer l'enveloppe temporelle du signal d'entrée. Pour cela on va tout d'abord en prendre sa valeur absolue, donc la phase négative du signal est redressée par inversion de phase. Le signal est ensuite filtré par un passe-bas pour obtenir un signal continu qui constitue un suivi relativement fidèle de l'enveloppe d'amplitude du signal d'origine.

Dans l'algorithme, c'est un filtre très simple de type *sinuscardinal* qui est utilisé et qui est présenté à la Figure 4.5. Pour être le plus précis possible, il va falloir varier la taille du filtre, autrement dit faire varier sa fréquence de coupure. Plus le filtre va être long, plus il va être précis, mais plus le temps de calcul va être long, donc il est très important de prendre cela en compte.

Pour ce faire, des tests ont été effectués avec une longueur de filtre de 512, 1024 et 2048 échantillons et les résultats sont présentés dans les Figures 4.6, 4.7 et 4.8.

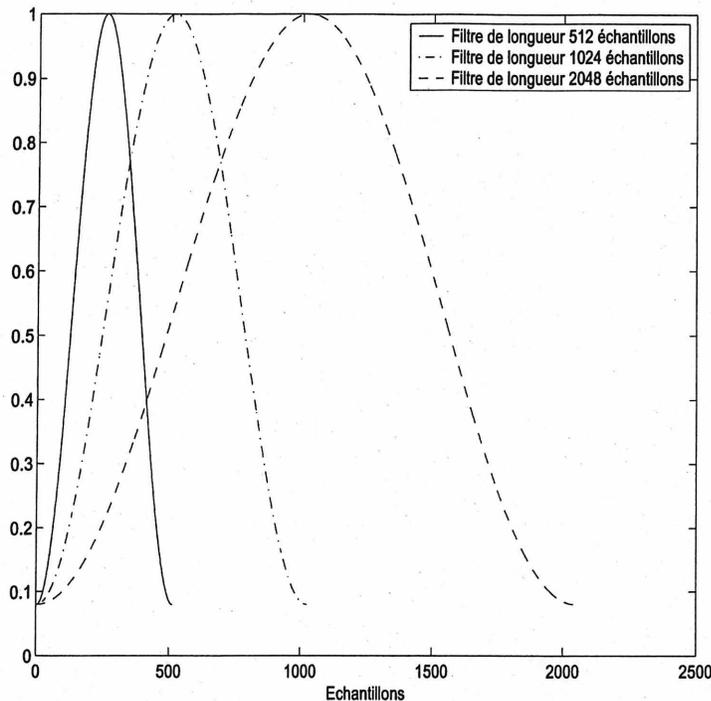


Figure 4.5 – Filtre passe bas de type sinuscardinal de longueur : 512, 1024 et 2048 échantillons

Comme on peut le voir sur les Figures 4.6, 4.7 et 4.8, plus le nombre d'échantillons constituant le filtre est élevée, plus proche du signal d'origine rectifié va être l'enveloppe temporelle. Cependant, et comme on l'a spécifié précédemment, le temps de calcul n'est pas le même selon la longueur du filtre. En effet, pour les 3 exemples précédemment illustrés, le temps de calcul double à chaque fois que l'on double le nombre d'échantillons. Il faut donc faire un compromis entre précision et rapidité de calcul. Cependant, on peut tout de suite éliminer le cas où le filtre à une longueur de 512 échantillons car d'après la Figure 4.6, la dérivée de l'enveloppe spectrale ne fait ressortir aucun événement en particulier et donc la précision n'est pas celle souhaitée. Par contre quand on analyse la Figure 4.8, il est clair que la courbe de la dérivée de l'enveloppe temporelle montre la présence d'une transitoire avec un premier pic se démarquant des autres. Il en est de même en analysant la Figure 4.7, ce premier pic séparé du reste de l'information fournie par la dérivée de l'enveloppe temporelle nous confirme la présence d'une transitoire. La précision se rapproche donc très nettement du cas où la longueur du filtre est la plus grande mais le temps de calcul est 2 fois inférieur.

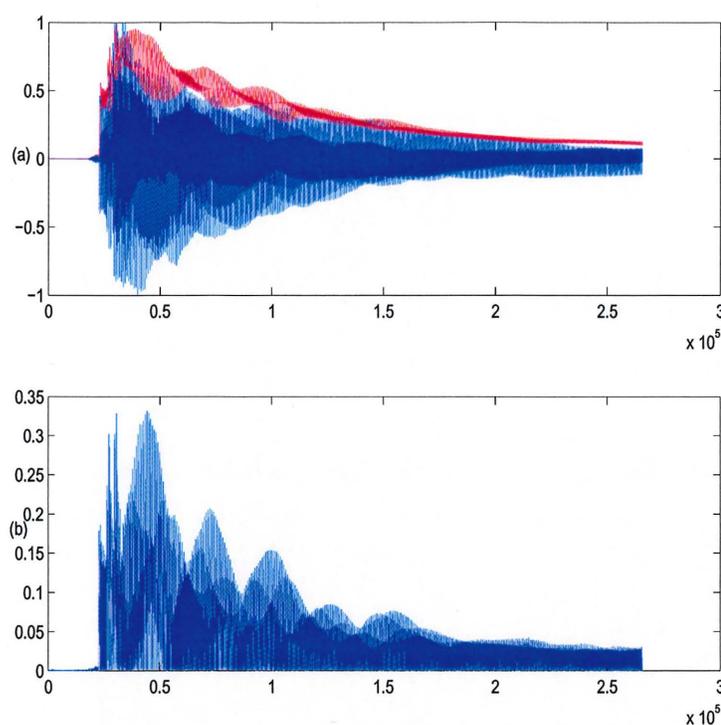


Figure 4.6 – (a) la courbe bleue est le signal d'entrée et la courbe en rouge et son enveloppe temporelle avec un filtre de longueur 512 échantillons, (b) est la courbe de la dérivée de l'enveloppe temporelle

Les algorithmes qui ont été développés par la suite utilisent un filtre passe-bas de longueur 1024 échantillons car il associe précision de détection et rapidité du temps de calcul.

Comme on l'a fait dans le cas de la méthode par le calcul de l'énergie, on présente dans les Figures 4.9 et 4.10 les résultats de la détection de transitoire dans le cas de sons ayant des accords rapprochés et éloignés.

Comme on peut le constater sur les Figures 4.9 et 4.10, la détection des transitoires se fait de façon plus précise que ce qui a été présenté dans les Figures 4.2, 4.3 et 4.4 et notamment dans le cas où les accords sont rapprochés.

On peut donc dire d'un point de vue précision (aussi bien dans le cas d'accord seul ou de succession d'accords rapprochés et éloignés) que les résultats obtenus avec la méthode de calcul de la dérivée de l'enveloppe temporelle semblent nettement plus fiables que ceux

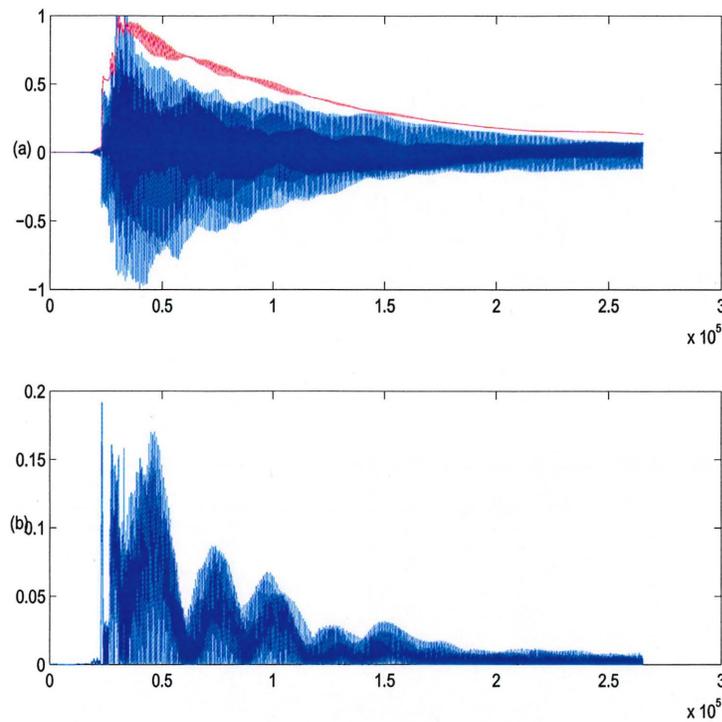


Figure 4.7 – (a) la courbe bleue est le signal d'entrée et la courbe en rouge et son enveloppe temporelle avec un filtre de longueur 1024 échantillons, (b) est la courbe de la dérivée de l'enveloppe temporelle

obtenus avec la méthode du calcul d'énergie par trame.

Il existe cependant différentes manières de calculer l'enveloppe temporelle d'un signal et c'est justement ce que l'on va présenter dans la section suivante avec l'introduction de *l'opérateur d'énergie de Teager*.

4.3.3 Par la méthode de Teager

La méthode dite de Teager a été introduite par Kaiser [J.F.Kaiser et T.F.Quateri, 1993] est permet de calculer par un nouvel outil l'énergie d'un signal, outil que l'on nomme *l'opérateur d'énergie de Teager* ou plus communément *TEO*. Comme on l'a vu précédemment, l'énergie d'un signal est définie comme la somme des amplitudes au carré de notre signal. Le TEO est une méthode alternative permettant de calculer de façon moins complexe l'énergie d'un

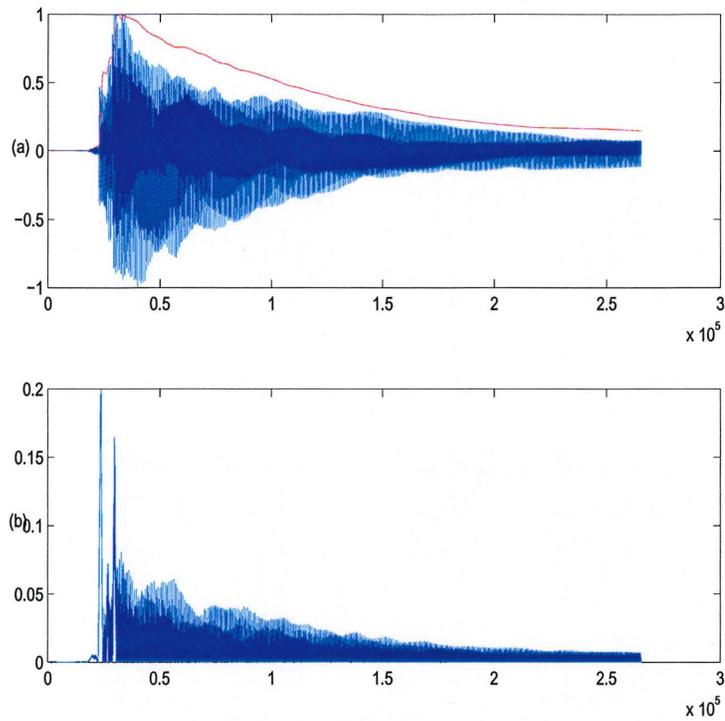


Figure 4.8 – (a) la courbe bleu est le signal d'entrée et la courbe en rouge est son enveloppe temporelle avec un filtre de longueur 2048 échantillons, (b) est la courbe de la dérivée de l'enveloppe temporelle

signal. De plus, il peut en ressortir l'enveloppe d'énergie en utilisant moins de calculs. Toutes les notions mathématiques sont données dans [J.F.Kaiser, 1993]. L'opérateur d'énergie de Teager est défini dans l'Equation 4.2.

$$TEO(x(t)) = ((x'(t))^2 - (x(t))(x''(t))) \quad (4.2)$$

où $x'(t)$ est la dérivée de $x(t)$ et $x''(t)$ est la dérivée de $x'(t)$. Sous certaines hypothèses, le TEO peut être vu comme une composition linéaire de l'amplitude A au carrée et de la fréquence instantanée Ω au carrée, exprimé par l'Equation 4.3

$$TEO(x(t)) \approx A^2\Omega^2 \quad (4.3)$$

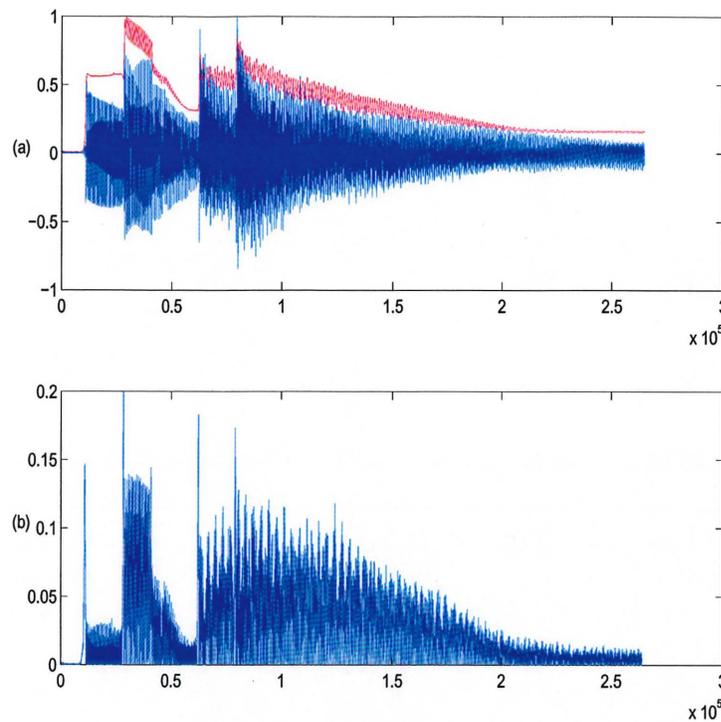


Figure 4.9 – (a) est le signal d'entrée composé d'une succession d'accords rapprochés avec son enveloppe temporelle, (b) est la courbe de la dérivée de l'enveloppe temporelle

Les caractéristiques qui ressortent de l'Equation 4.3 sont que l'opérateur permet de suivre la fréquence instantanée et détermine l'enveloppe temporelle de n'importe quel signal composé de plusieurs sinusoïdes.

C'est donc cette dernière caractéristique qui va nous être la plus utile. En effet, comme on l'a vu précédemment, il est possible de déterminer les transitoires grâce à l'enveloppe temporelle du signal. La différence majeure avec la technique utilisée précédemment est que l'on se servait d'un filtre pour déterminer notre enveloppe, alors que dans le cas de l'opérateur de Teager c'est un résultat instantané que l'on obtient, donc avec un temps de calcul nettement inférieur.

Comme on peut le constater en comparant les Figures 4.12 et 4.9, la détection des transitoires est plus précise avec la méthode de Teager que par le calcul de la dérivée de l'enveloppe temporelle. Cependant, quand on compare les résultats obtenus sur les Figures 4.11, et 4.13

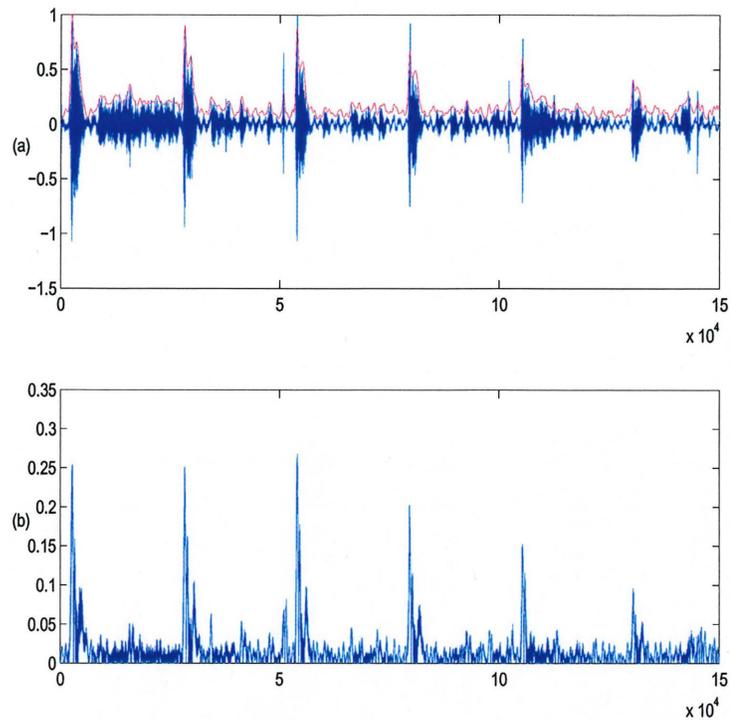


Figure 4.10 – (a) est le signal d'entrée composé d'une succession d'accords éloignés avec son enveloppe temporelle, (b) est la courbe de la dérivée de l'enveloppe temporelle

avec ceux obtenus aux Figures 4.7, et 4.10 la précision de détection est quasiment identique d'une technique à l'autre et de plus le temps de calcul est quasiment identique, donc on ne peut pas se fonder sur ce critère pour choisir l'une ou l'autre des méthodes.

4.3.4 Conclusion

Comme on peut le voir sur la Figure 4.14, la technique qui a été retenue pour faire la segmentation temporelle est celle de la dérivée de l'enveloppe temporelle. Le choix s'est restreint entre la technique de l'opérateur de Teager et celle de la dérivée de l'enveloppe temporelle. Étant donné qu'au cours de ce projet on ne travaille que sur des accords isolés et que les résultats obtenus pour faire de la détection de transitoire avec la méthode de la dérivée de l'enveloppe temporelle était très pertinents, c'est cette dernière qui a été employée.

Dans la partie *Analyse Temporelle*, la méthode de la dérivée de l'enveloppe temporelle est

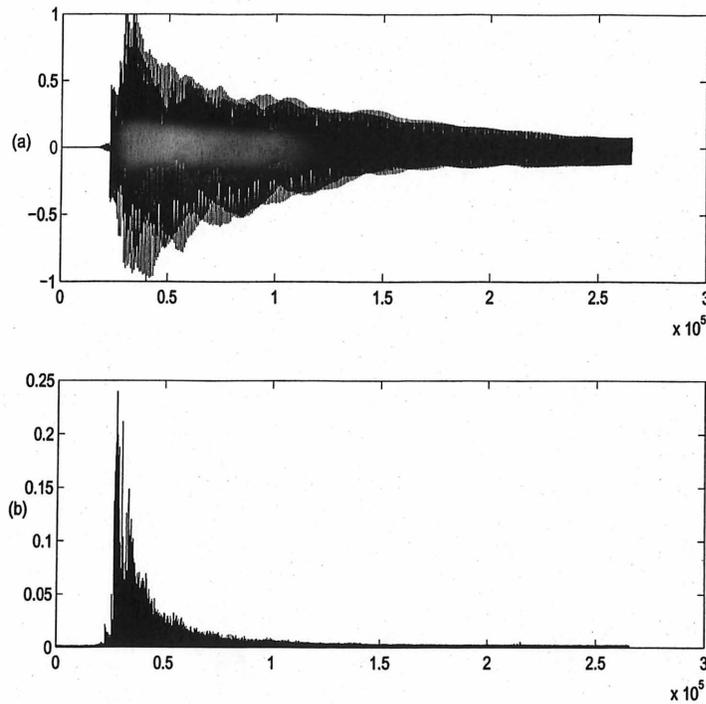


Figure 4.11 – (a) est le signal d'entrée, (b) est la courbe de l'enveloppe temporelle donnée par Teager

appliquée à des courtes trames du signal. Tout d'abord, comme on la vu dans la section 4.3.2, la longueur du filtre a été choisie à 1024 échantillons. De plus, on a vu dans le calcul de l'énergie que le fait de faire des calculs par trame n'affecte pas la précision de la détection des transitoires. De ce fait, on analyse le signal trame par trame avant de décider le moment où a eu lieu l'attaque. On sait qu'au moment de l'attaque, la dérivée de l'enveloppe temporelle croit de façon significative. On va donc déterminer qu'on est en présence d'une attaque lors que la dérivée de l'enveloppe temporelle calculée dans une trame augmente de façon significative (rapport de 1/5) par rapport à celle de la trame précédente. Cela nous permet d'obtenir le signal à partir de la trame dans laquelle a eu lieu l'attaque. En effet, tant qu'il n'y a pas une augmentation significative de la dérivée de l'enveloppe temporelle, l'algorithme continue de calculer la dérivée de l'enveloppe pour chacune des trames jusqu'à ce que celle-ci dépasse le seuil.

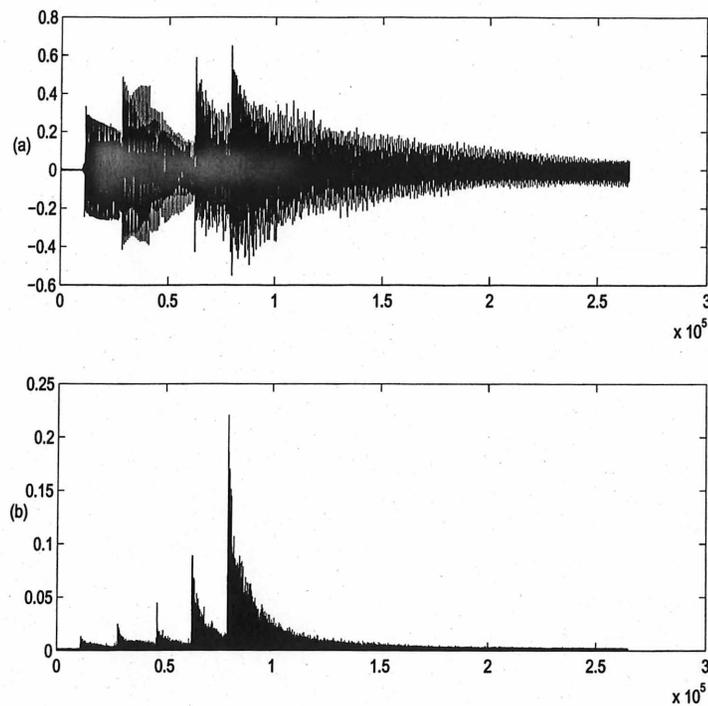


Figure 4.12 – (a) est le signal d'entrée composé d'une succession d'accords rapprochés, (b) est la courbe de l'enveloppe temporelle donnée par Teager

Prenons le cas où la dérivée de l'enveloppe temporelle se trouve au-dessus du seuil fixé pour une trame donnée. Dans ce cas là, l'échantillon où l'attaque a été détecté est mise en mémoire dans le programme. Le bloc de traitement *segmentation temporelle* prend en entrée le signal d'origine $x[n]$ ainsi que l'échantillon mis en mémoire. De ce fait, ce bloc nous permet d'avoir le signal d'origine mais à partir de l'attaque. Cependant, comme on l'a vu dans la section 2.4.1, il est essentiel d'analyser le signal seulement 100 ms après la détection de l'attaque pour avoir le signal à analyser le plus stationnaire possible. C'est donc dans ce même bloc de traitement qu'est réalisée cette opération. Par conséquent, la sortie du bloc *segmentation temporelle* est le signal d'origine à partir du moment où il est analysable (à partir de la partie la plus stationnaire du signal).

Au cours de ce projet on ne travaille que sur des accords dont les notes sont jouées sur les 5 premiers frets de la guitare. En terme de fréquences fondamentales cela veut dire que nos

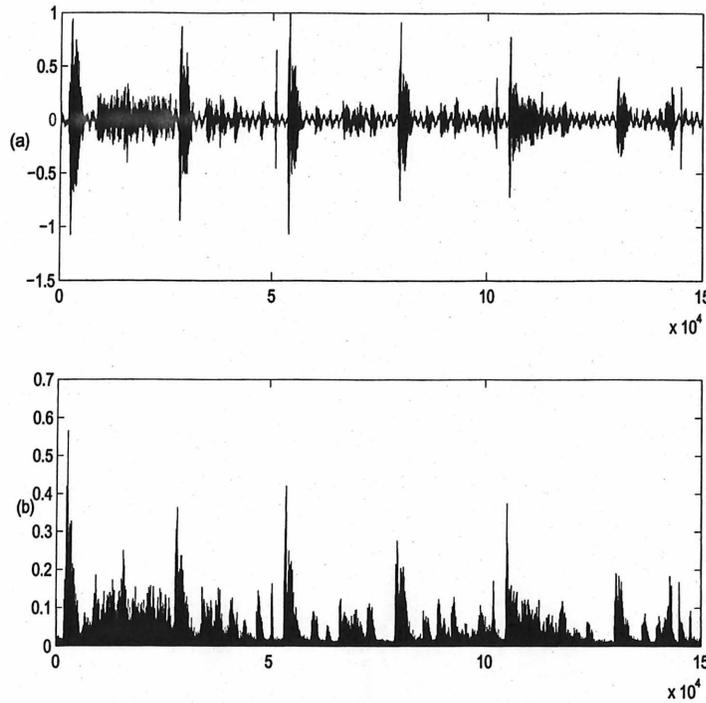


Figure 4.13 – (a) est le signal d'entrée composé d'une succession d'accords éloignés, (b) est la courbe de l'enveloppe temporelle donnée par Teager

F_0 sont comprises entre 82.41 Hz et 440 Hz, donc la bande de fréquence en tenant compte des harmoniques peut être limitée à 2000 Hz (soit jusqu'à 5 harmoniques pour la fréquence fondamentale la plus haute). Étant donné que le signal d'entrée est échantillonné à 44.1 kHz, soit une largeur de bande de 22 kHz, on va sous-échantillonner notre signal par 10 et donc réduire notre largeur de bande à 2.2 kHz permettant d'avoir toute l'information nécessaire à une analyse robuste.

Le signal que l'on va obtenir en sortie de la partie *Analyse Temporelle* est donc le signal sous-échantillonné par 10 et pris à partir de l'attaque. C'est à partir de ce signal que va se faire l'analyse fréquentielle permettant de retrouver le nombre et la valeur des fréquences fondamentales composant notre accord de guitare.

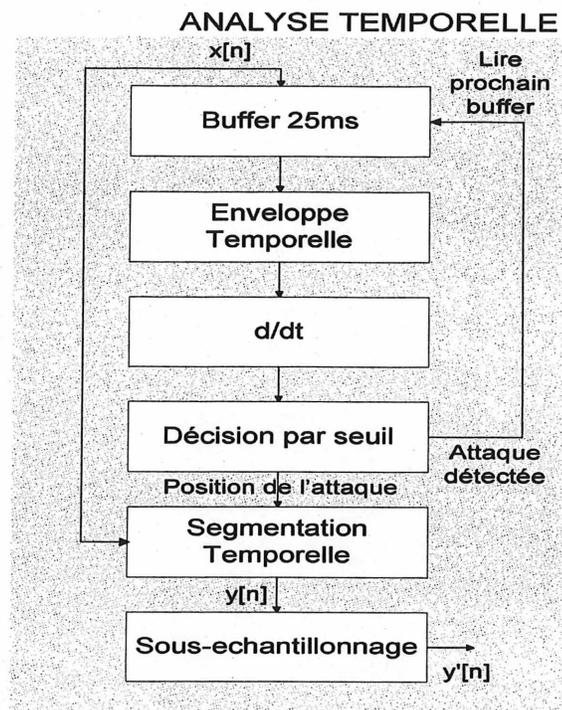


Figure 4.14 – Schéma bloc de l'algorithme de segmentation temporelle

4.4 Détermination des fréquences fondamentales

Comme on l'a vu dans la partie précédente, l'*Analyse Temporelle* ressort le signal sous-échantillonné à partir de l'attaque. C'est donc dans cette deuxième partie qu'est l'*analyse fréquentielle* qu'est réalisée la partie la plus importante de l'algorithme. En effet, cette analyse va nous permettre d'avoir en sortie de l'algorithme le nombre de fréquences fondamentales qui ont été jouées, associées avec leur valeur. De plus, si les résultats le permettent, l'accord qui a été joué en sera déduit.

Le schéma bloc de l'algorithme de détection des fréquences fondamentales est présenté à la Figure 4.15. Comme on peut le voir, celui-ci se base sur une *FFT* du signal d'entrée ($y'[n]$) sur laquelle sont appliquées des fonctions que l'on va détailler par la suite.

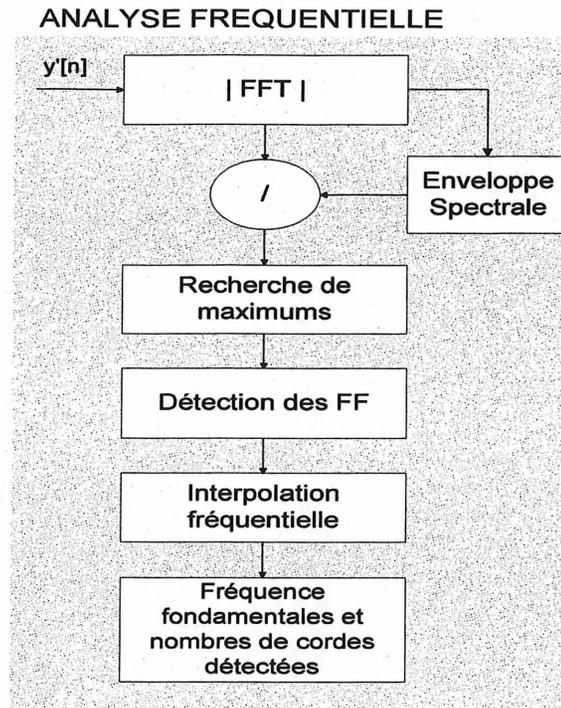


Figure 4.15 – Schéma bloc de l’algorithme de détection des fréquences fondamentales

4.4.1 Transformée de Fourier

Comme on l’a vu dans la section 4.1.2, on va se servir principalement de la composition harmonique des signaux à étudier pour les analyser. Donc pour pouvoir analyser les composantes sinusoïdales présentes dans le signal on va utiliser la *Transformée de Fourier Discrète* dont on va rappeler la définition.

Transformée de Fourier Discrète Notre approche se fonde en grande partie sur une étude détaillée des signaux possédant une composante périodique ou presque périodique. Pour l’analyse de ce type de signaux, l’outil le mieux adapté est la transformée de Fourier. Mais comme l’exploitation informatique de cette étude nous conduit à utiliser des signaux discrets et finis, alors il s’agit d’une transformée de Fourier appliquée à ce type de signaux, et appelée Transformée de Fourier Discrète. Rappelons donc la définition de la Transformée de Fourier Discrète.

Soit $x = (x_k)_{k=1,N}$ un signal discret fini composé de N points, nous appelons Transformée de Fourier Discrète (ou plus simplement TFD) de x , l'ensemble des N points définis par :

$$\tilde{x}_n = TFD(x)(n) = \sum_{k=1}^N x_k e^{-i2\pi nk/N} \quad (4.4)$$

avec $0 \leq n \leq N - 1$.

Cette transformation est inversible et la transformation inverse s'écrit :

$$TFD^{-1}(\tilde{x})(n) = \sum_{k=0}^{N-1} \tilde{x}_k e^{i2\pi nk/N} \quad (4.5)$$

avec $1 \leq k \leq N$.

La *TFD* peut être interprétée comme le résultat de l'échantillonnage de la transformée de Fourier (continue) du signal discret. Mieux, à partir de la *TFD*, il est théoriquement possible de reconstruire cette transformée de Fourier en tout point. De plus, les maxima de la transformée de Fourier jouent un rôle particulier, puisque chaque maximum local est susceptible de représenter une composante sinusoïdale du signal. Finalement, grâce à la reconstruction du spectre continu, on peut utiliser la *TFD* pour l'analyse et l'estimation des paramètres des sinusoïdes présentes dans le signal.

Représentation de la *TFD* La représentation spectrale est obtenue à partir de la transformée de Fourier de la représentation temporelle. Cette représentation s'appelle aussi le spectre. Le spectre permet de déterminer si le signal provient d'une source harmonique (signaux de musique ou de parole) ou une source non-harmonique. Concernant les spectres harmoniques, ils sont caractérisés par la présence de raies appelées *partiels* espacées de manière constante et régulière le long de l'axe des fréquences. Le partiel de fréquence minimale f_0 , est appelée la fondamentale. Tous les autres sont appelées harmoniques et, bien entendu, leurs fréquences sont successivement $2f_0$, $3f_0$, ...

La précision de la *TFD* est basée sur sa résolution fréquentielle qui diffère en fonction de la fréquence d'échantillonnage et du support temporel choisit. Par exemple si on prend

un signal d'entrée échantillonné à 44.1 kHz, et qu'on analyse 4096 échantillons du signal, alors la résolution fréquentielle est $4096/44100 \sim 10$ Hz. Plus la résolution fréquentielle est petite, plus l'écart entre deux raies spectrales sera faible, en d'autres termes pour avoir la plus grande précision possible il faut avoir la plus petite résolution fréquentielle possible. Cependant, petite résolution fréquentielle est synonyme d'un long support temporel. De ce fait, la résolution fréquentielle va dépendre du délai d'analyse dont on dispose. Ici tous les signaux analysés dans la partie *analyse fréquentielle* sont échantillonnés à 4.41 kHz (résultant du sous-échantillonnage par 10 du signal d'entrée).

Les Figures 4.16 et 4.17 représentent la Transformée de Fourier Discrète de deux portions de signal échantillonné à 4.41 kHz et simulant une note de guitare (G3 de fréquence fondamentale 196 Hz). Les supports temporels choisis sont respectivement de 4096 et 1024 échantillons, soit un support temporel d'environ 1 s et 250 ms. De ce fait, la résolution fréquentielle est respectivement de 1 Hz et 4 Hz, et comme on peut le voir sur les Figures 4.16 et 4.17 la précision va être meilleure avec un long support temporel.

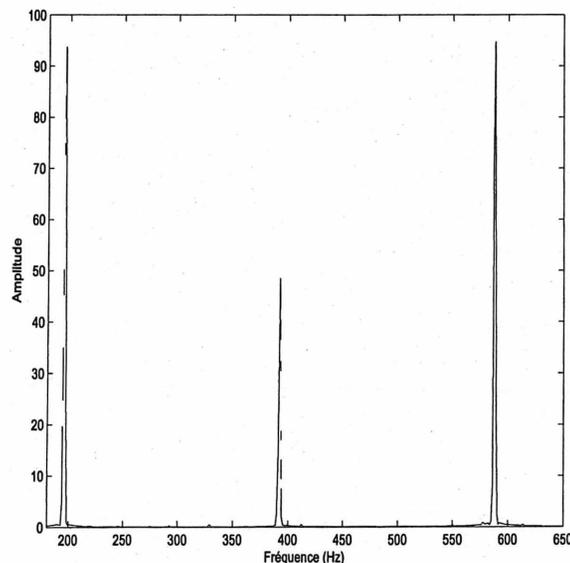


Figure 4.16 – Transformée de Fourier Discrète d'une longueur de signal de 4096 échantillons

Plus le signal analysé est riche en fréquences fondamentales (donc plus le nombre de cordes jouées est grand), plus le spectre va être riche en terme de nombre de pics. Ceci est démontré

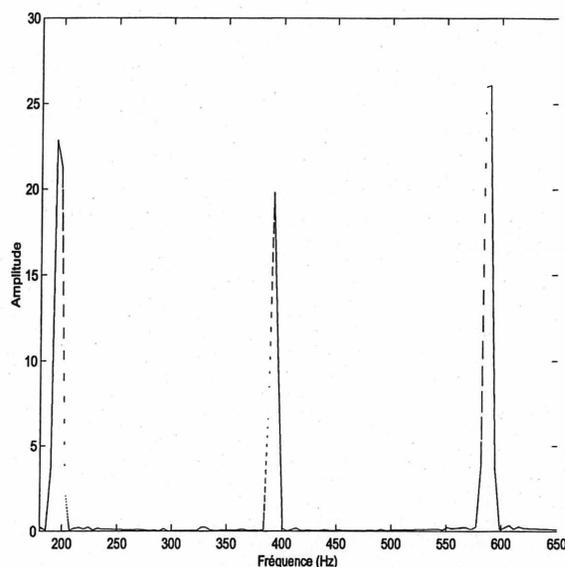


Figure 4.17 – Transformée de Fourier Discrète d'une longueur de signal de 1024 échantillons

sur la Figure 4.18, où le nombre de notes jouées varie de 2 à 6 avec pour chacun des signaux un support temporel de 2048 échantillons, soit environ $1/2$ s du signal. Les signaux sont toujours échantillonnés à 4.41 kHz, ce qui veut dire qu'on a une résolution fréquentielle d'environ 2 Hz. Donc pour pouvoir détecter deux raies spectrales, il faut qu'elles soient espacées d'au moins 2 Hz.

Comme on peut le voir sur la Figure 4.18, il y a de nombreux recouvrements d'harmoniques surtout dans les cas allant de 4 à 6 notes jouées. Dans le Tableau 4.1 sont présentées les valeurs des fréquences fondamentales avec leurs harmoniques. Il sera de ce fait plus facile de voir les recouvrements d'harmoniques que l'on va rencontrer dans les différents spectres. Les fréquences qui vont se recouvrir les unes avec les autres sont spécifiées en *italique* dans le tableau.

Comme on peut le voir, lorsque le signal analysé est composé de 2 ou 3 notes, il y a peu souvent de recouvrement d'harmonique. D'ailleurs, dans ces deux cas-ci, les notes jouées étaient prises au hasard tandis que dans les 3 autres cas, ce sont des accords spécifiques qui ont été joués et qui sont respectivement (selon le nombre de notes jouées) : C (Do), A (La), F (Fa) (dont la composition était présentée dans la section 2.4.3) et plus propice aux

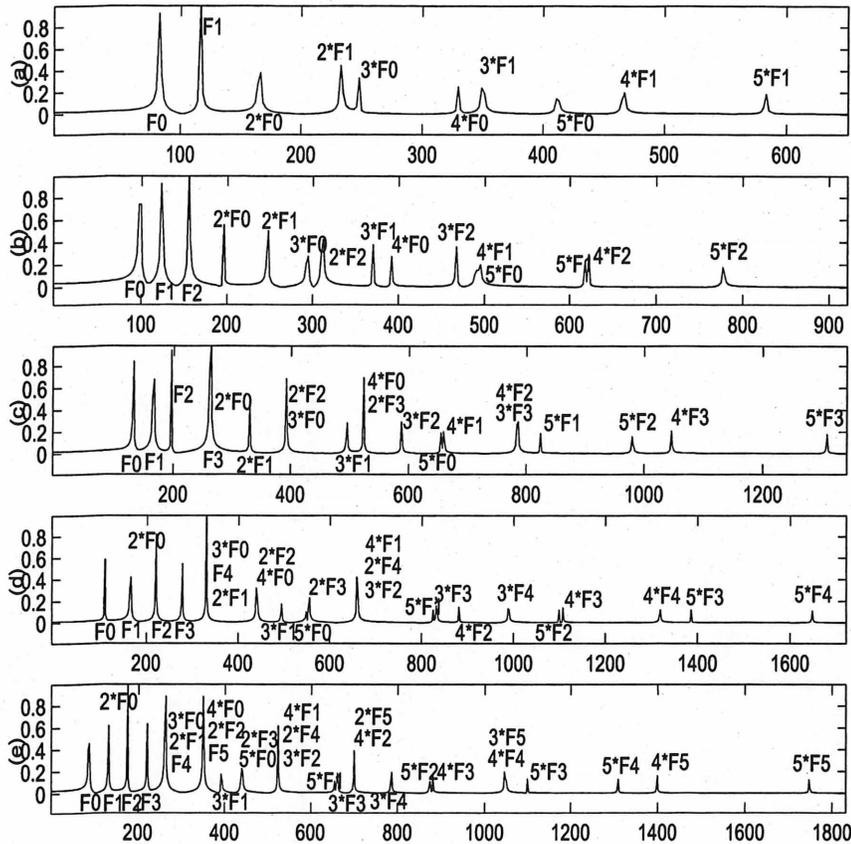


Figure 4.18 – (a) La *TFD* composée de 2 notes : E2 et A#2, (b) La *TFD* composée de 3 notes : G2, B2 et D#3, (c) La *TFD* composée de 4 notes : C3, E3, G3, C4, (d) La *TFD* composée de 5 notes : A2, E3, A3, C#4 et F4, (e) La *TFD* composée de 6 notes : F2, C3, F3, A3, C4 et F4.

recouvrements d'harmoniques.

Enveloppe spectrale L'ensemble des amplitudes du spectre matérialise l'enveloppe spectrale du son considéré, ou plus simplement on peut la définir comme étant la courbe qui rejoint les sommets de chaque raie spectrale.

L'enveloppe spectrale est calculée par lissage spectral dans le domaine de la *TFD* à l'aide d'une fenêtre de *Hamming* glissante. La longueur de la fenêtre de *Hamming* a été choisie de façon à ce quelle couvre au moins 3 harmoniques (d'une fréquence fondamentale moyenne) en

nb de notes	note	f_0	$2 \times f_0$	$3 \times f_0$	$4 \times f_0$	$5 \times f_0$
2	E2	82.41	164.82	247.23	329.64	412.50
	A#2	116.54	233.08	349.62	466.16	582.70
3	G2	98	196	294	392	490
	B2	123.47	246.94	370.41	493.88	617.35
	D#3	155.56	311.12	466.68	622.24	777.80
4	C3	130.81	261.62	392.43	523.24	654.05
	E3	164.81	329.62	494.43	659.24	824.65
	G3	196	392	588	784	980
	C4	261.83	523.26	784.89	1046.52	1308.15
5	A2	110	220	330	440	550
	E3	164.81	329.62	494.43	659.24	824.65
	A3	220	440	660	880	1100
	C#4	277.18	554.36	831.54	1108.72	1385.90
	F4	329.63	659.26	988.89	1318.52	1648.15
6	F2	87.30	174.60	261.90	349.20	436.50
	C3	130.81	261.62	392.43	523.24	654.05
	F3	174.61	349.22	523.83	698.44	873.05
	A3	220	440	660	880	1100
	C4	261.83	523.26	784.89	1046.52	1308.15
	F4	349.63	699.26	1048.89	1398.52	1748.15

TABLEAU 4.1 – Fréquences fondamentales et harmoniques composant des différents notes testées

termes d'échantillons. Évidemment, il y a des fréquences pour lesquelles la taille de la fenêtre va être optimale et d'autres moins, mais d'après les tests qui ont été réalisés, les résultats sont toujours concluants. Étant donné qu'on utilise une fenêtre glissante, le longueur de la fenêtre de *Hamming* a été choisie à 120 échantillons, correspondant donc à une fréquence fondamentale de 260 Hz. Comme on l'a vu dans la section 4.3.4 on se limite aux 5 premières frets de la guitare, c'est à dire des fréquences allant de 82.41 Hz à 440 Hz, donc avec une fréquence moyenne correspondant à environ 260 Hz.

Comme on peut le voir sur les Figures 4.19, 4.20, 4.21 et 4.22, la taille de la fenêtre de Hamming utilisée pour le lissage spectral donne des résultats aussi bon dans les plus basses fréquences (cf Figure 4.19) que dans les plus hautes (cf Figure 4.22) ainsi que pour les accords (cf Figure 4.23)

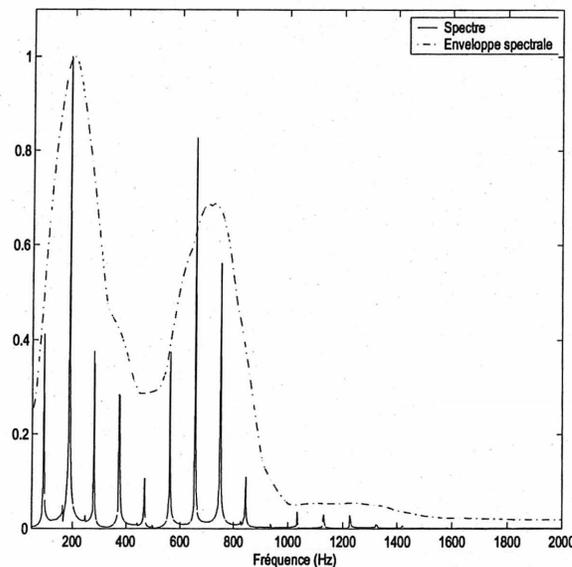


Figure 4.19 – Spectre et enveloppe spectrale d'une portion (2048 échantillons) de la note F#2 de fréquence fondamentale 92.50 Hz

Dans l'algorithme on se sert de l'enveloppe spectrale pour normaliser le spectre du signal étudié, et par la même occasion rendre l'algorithme générique pour n'importe quelle guitare utilisée. En effet, chaque guitare va avoir une couleur différente, par couleur on entend l'amplitude que va avoir le son en l'enregistrant, et donc l'amplitude du spectre va aussi

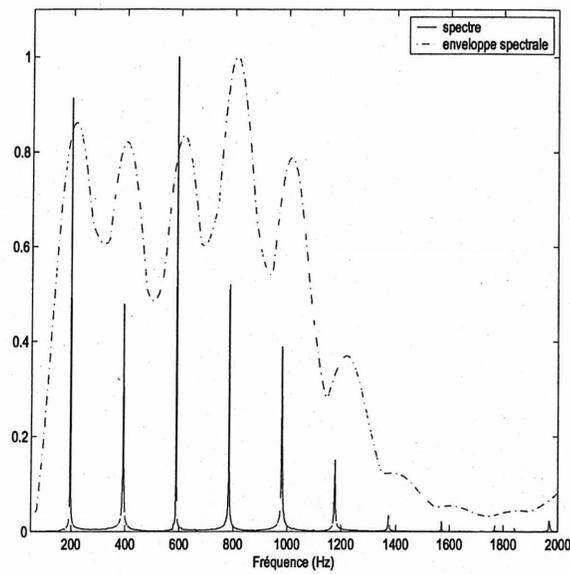


Figure 4.20 – Spectre et enveloppe spectrale d'une portion (2048 échantillons) de la note G3 de fréquence fondamentale 196 Hz

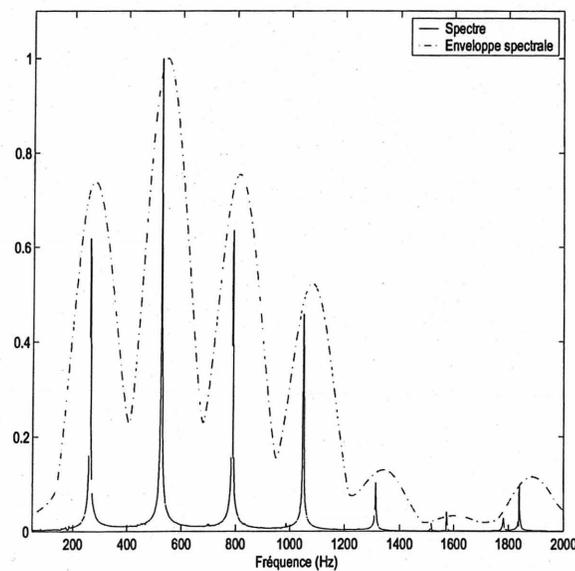


Figure 4.21 – Spectre et enveloppe spectrale d'une portion (2048 échantillons) de la note C4 de fréquence fondamentale 261.63 Hz

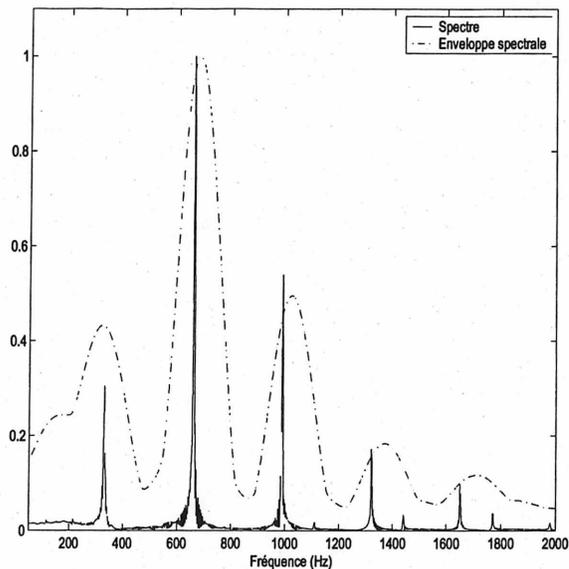


Figure 4.22 – Spectre et enveloppe spectrale d'une portion (2048 échantillons) de la note E4 de fréquence fondamentale 329.63 Hz

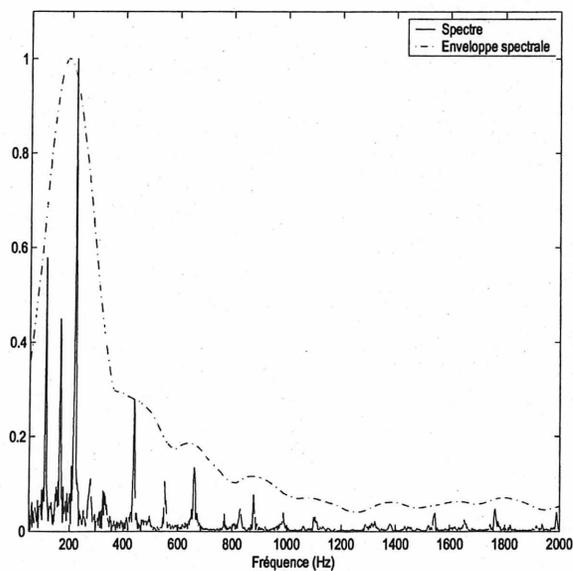


Figure 4.23 – Spectre et enveloppe spectrale d'une portion (2048 échantillons) de l'accord A (La)

dépendre de la couleur de la guitare.

De plus, certains harmoniques vont avoir une faible amplitude par rapport aux autres, car il peut théoriquement exister une infinité d'harmoniques, mais l'énergie des signaux est limitée, ce qui rend négligeable l'amplitude des harmoniques de haut rang. De ce fait, il est essentiel de rehausser ces harmoniques pour qu'ils ne soient pas considéré comme du bruit par l'algorithme.

C'est donc en calculant l'enveloppe spectrale (du spectre analysé) puis en divisant le spectre d'origine par celle-ci que l'on va pouvoir rehausser ces raies spectrales et donc avoir une information spectrale plus riche en terme de nombres d'harmoniques.

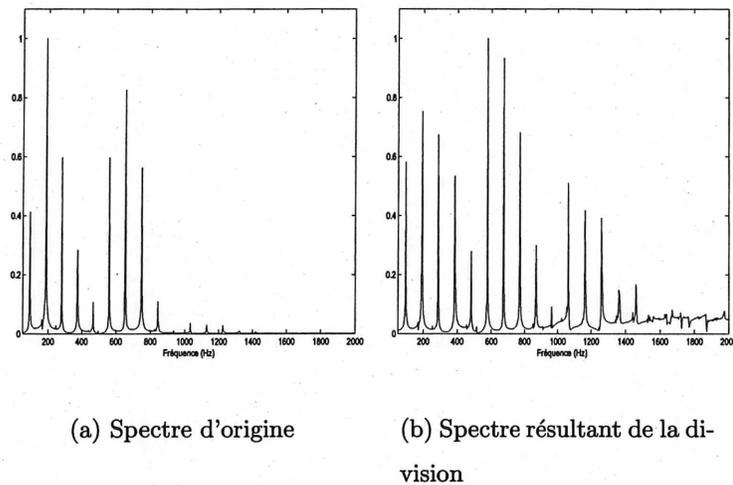


Figure 4.24 – (a) Spectre d'origine et (b) spectre résultant de la division par l'enveloppe spectrale d'une portion (2048 échantillons) de la note F#2 de fréquence fondamentale 92.50 Hz

Comme on peut le voir sur les Figures 4.24, 4.25, 4.26, 4.27, 4.28, l'amplitude de nombreuses raies spectrales, notamment dans les hautes fréquences, est rehaussée par rapport à l'amplitude qu'elles avaient dans le spectre d'origine. Il est évident que toutes ces raies spectrales ne sont pas des partiels (fréquences fondamentales ou harmoniques), mais il est tout de même primordial que les partiels soient rehaussés et dans une étape future, un traitement sera effectué afin de ne garder que les composantes sinusoïdales du signal (les partiels).

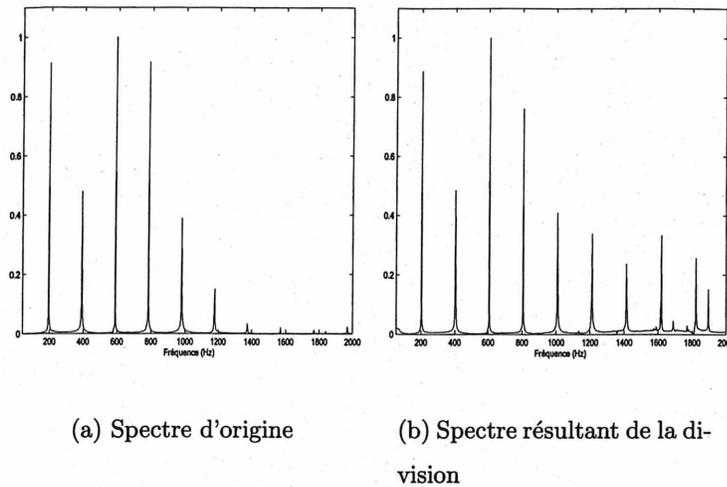


Figure 4.25 – (a) Spectre d'origine et (b) spectre résultant de la division par l'enveloppe spectrale d'une portion (2048 échantillons) de la note G3 de fréquence fondamentale 196 Hz

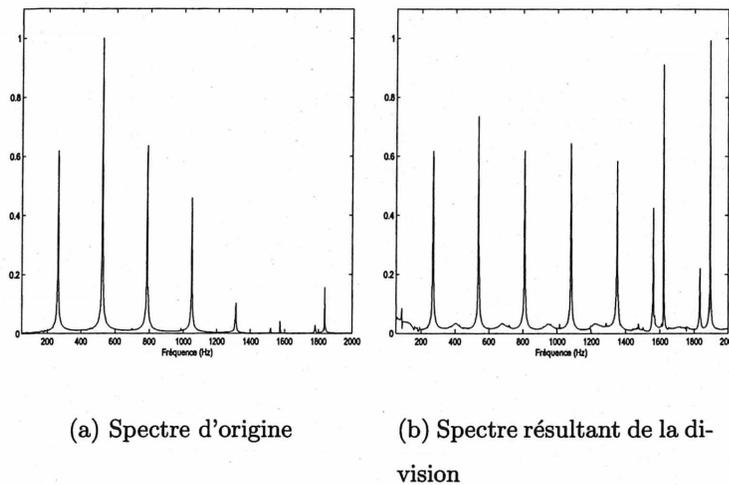


Figure 4.26 – (a) Spectre d'origine et (b) spectre résultant de la division par l'enveloppe spectrale d'une portion (2048 échantillons) de la note C4 de fréquence fondamentale 261.63 Hz

Le spectre résultant est donc composé de raies spectrales qui sont soit les partiels soit du bruit. Bruit qu'il va falloir éliminer si on ne veut pas qu'il perturbe notre analyse au moment de la détection des fréquences fondamentales. Pour cela, on applique un seuil au spectre, seuil

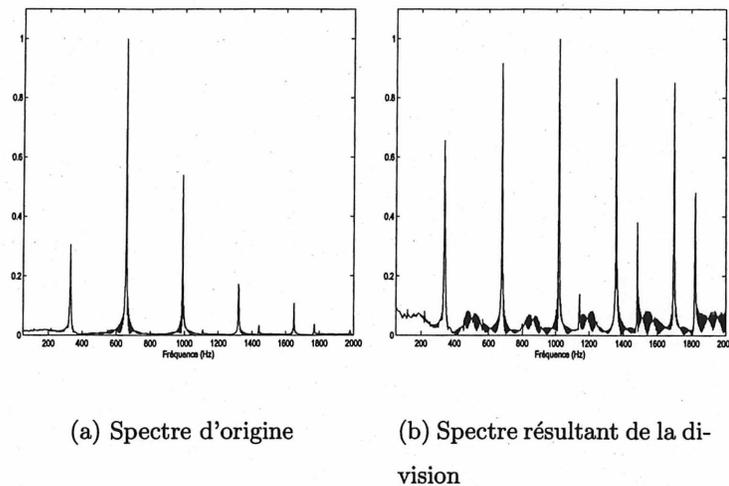


Figure 4.27 – (a) Spectre d'origine et (b) spectre résultant de la division par l'enveloppe spectrale d'une portion (2048 échantillons) de la note E4 de fréquence fondamentale 329.63 Hz

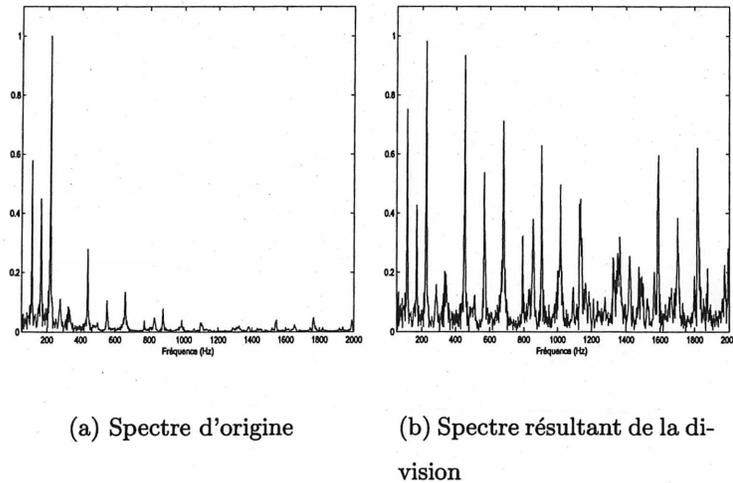


Figure 4.28 – (a) Spectre d'origine et (b) spectre résultant de la division par l'enveloppe spectrale d'une portion (2048 échantillons) de l'accord A

en dessous duquel toutes les amplitudes sont mise à zéro, et au-dessus duquel les amplitudes resteront les mêmes. Étant donné que toutes les amplitudes sont normalisées et que les raies spectrales dont on ne voulait pas qu'elles soient considérées comme du bruit sont rehaussées,

le seuil appliqué est fixé de façon à ce que toutes les amplitudes inférieures à 0.3 soient mise à zéro.

Par rapport aux résultats qui ont été présentés précédemment, ce sont ceux de la Figure 4.28 qui vont illustrer le mieux ce phénomène. En effet, dans la Figure 4.29, on compare le spectre résultant de la division par l'enveloppe spectrale du spectre d'origine avec le spectre résultant de l'application du seuil. Comme on peut facilement le voir, les amplitudes en-dessous du seuil ont été mise à zéro tandis que les autres ont garder leur valeurs initiales.

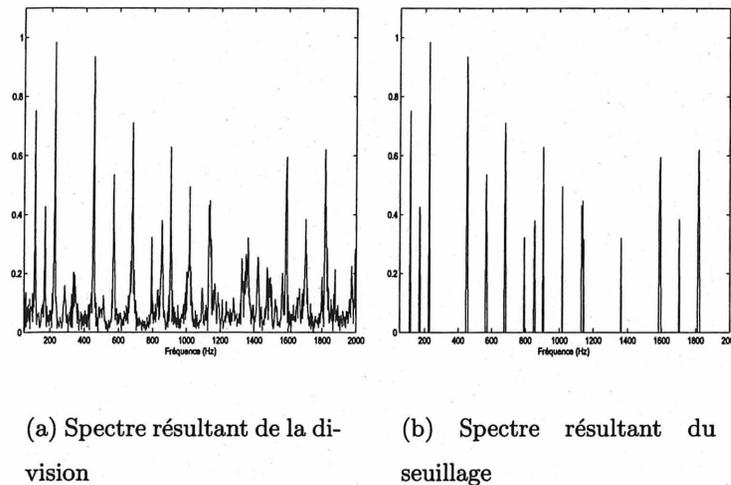


Figure 4.29 – Spectre résultant de la division de l'enveloppe spectrale et spectre résultant du seuillage d'une portion (2048 échantillons) de l'accord A

Il n'était pas aussi pertinent de montrer les résultats obtenus lorsqu'une seule note était jouée car l'enveloppe spectrale étant tellement proche du spectre d'origine que cela n'entraînait pas autant de bruit que dans le cas des accords. C'est d'ailleurs à cause du fait que l'on travaille principalement sur les accords qu'on utilise l'enveloppe spectrale pour rehausser les partiels de faible amplitude et qu'un seuil est appliqué au spectre résultant pour éliminer le bruit.

4.4.2 Détection des maximums

C'est à partir du spectre résultant du seuillage (cf graphique (b) de la Figure 4.29) que l'on va détecter les maximums du spectre.

La détection de ces maximums se fait de façon très simple. En effet, dans le spectre, chaque raie spectrale (canal fréquentiel) dont l'amplitude est supérieure à celle de ses voisines (de gauche et de droite) est considérée comme un maximum du spectre.

En terme de code *MATLAB*, la détection de maximum s'écrit de la manière suivante :

```

b = zeros(length(x));
for i = 2 :length(signal)-1
    if ((x(i) ≥ x(i-1)) & (x(i) > x(i+1)) & (x(i) ≈ x(i+1)));
        b(i) = 1;
    end
pff = find(b==1);
max = x(pff);

```

On applique cette routine à tout le spectre et on en ressort la position de chaque maximum en terme de position sur non pas une échelle des fréquences mais des canaux de FFT (bins).

Comme on peut le voir sur la Figure 4.30, chacune des composantes sinusoïdale représentant le signal occupent plusieurs canaux fréquentiels. On va donc, à l'aide de la routine de détection des maximums ne garder que le canal spectral qui représente le plus la composante sinusoïdale, c'est à dire la raie d'amplitude maximale. Maximums dont on va retrouver les valeurs en échantillons dans le Tableau 4.2.

maximums	52	77	103	205	257	259	307	359	386	409	460	511	514
----------	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

TABLEAU 4.2 – Maximum (en terme d'échantillon) du spectre de la Figure 4.30

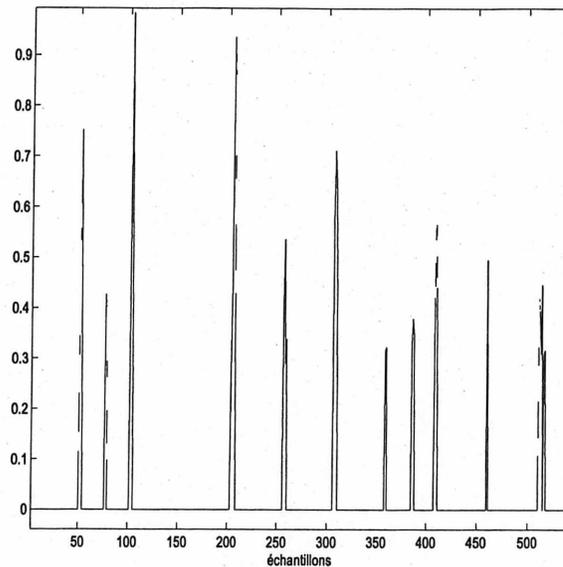


Figure 4.30 – Zoom sur le début du spectre (b) de la Figure 4.29

4.4.3 Approche logique pour la détection des fréquences fondamentales

C'est donc dans le bloc *Détection des FF* de la Figure 4.15 que va se faire la détection des fréquences fondamentales. On évolue toujours dans le domaine spectral, mais cette fois, on va se limiter uniquement aux maximums que l'on a détectés. Maximums qui, comme on l'a vu précédemment, représentent les composantes sinusoïdales du signal analysé. En d'autres termes, les maximums représentent les fréquences fondamentales et les harmoniques de toutes les notes présentes dans l'accord de guitare qui a été joué.

Étant donné que le but final est d'obtenir la valeur des fréquences fondamentales présentes afin de pouvoir estimer l'accord qui a été joué, il faut mettre en valeur les fréquences fondamentales par rapport à leurs harmoniques. Comme on l'a déjà vu à plusieurs reprises dans ce mémoire, tout signal harmonique est caractérisé par sa fondamentale f_0 et ses harmoniques dont les fréquences sont respectivement $2f_0$, $3f_0$, ...

On va donc se baser sur cette caractéristique des signaux harmoniques pour mettre en valeur la fréquence fondamentale par rapport aux harmoniques. Cela est effectué à l'aide d'une

procédure itérative que l'on applique aux tableaux des maximums détectés (positions de maximums que l'on a transposées en fréquence au lieu des canaux de FFT).

Pour cela, on vérifie (en partant de la première fréquence détectée comme maximale que l'on nomme ici *fréquence testée*) si l'on retrouve un nombre de fois entier (2, 3, 4, ... fois) la *fréquence testée*. Cependant, on s'est aperçu qu'étant donné la résolution fréquentielle qu'on obtient, que les valeurs des fréquences ne seront pas exactes et dans le cas des signaux de guitare (réels), on avait vu que les harmoniques ne sont pas situées exactement au niveau des fractions entières de la fréquence fondamentale. C'est donc pour cela qu'on ajoute un Δf au calcul précédent pour prendre en compte les caractéristiques énoncées précédemment.

En terme de code *MATLAB*, l'estimation des fréquences fondamentales s'écrit de la manière suivante

```

for s = 1 :length(ffharm)
    for k = 2 :5
        for w = 2 :length(ffharm)
            if( (ffharm(w) ≥ (k*ffharm(s)-(N*2/FSe))) & (ffharm(w)
                ≤ (k*ffharm(s)+(N*2/FSe))))
                freqfond(s)=ffharm(s);
                ffharm(w)=0;
            end
        end
    end
end
end
end

```

où *ffharm* représente le tableau des valeurs des maximums transposées en fréquence, et où Δf vaut $N*2/FSe$ avec N le nombre d'échantillon analysé (soit 2048) et FSe la fréquence d'échantillonnage (soit 4410 Hz).

Comme on peut le voir dans le code précédent, avec le nombre k on peut détecter jusqu'à 5 harmoniques appartenant à une fréquence fondamentale. Si certaines des fréquences du

tableaux représentent la *fréquence testée* de manière à ce qu'elles soient des harmoniques de celle-ci alors on les soustrait du spectre et la *fréquence testée* est mise en mémoire dans le tableau des fréquences fondamentales (*fregfond*) dans le code illustré précédemment.

4.4.4 Interpolation

Grâce à la procédure décrite ci-dessus, on a maintenant comme données un tableau, de longueur maximale 6, dans lequel sont sauvegardées les valeurs des fréquences fondamentales que l'on vient de détecter.

Afin de gagner en précision, on va utiliser un calcul d'interpolation autour des raies spectrales maximales détectées dans la section 4.4.2, pour affiner la valeur des fréquences fondamentales détectées. Ce calcul d'interpolation est fait sur le spectre résultant de la division du spectre d'origine par son enveloppe spectrale.

$$k' = \frac{\sum_{j=k-3}^{j=k+3} j \cdot A_j}{\sum_{j=k-3}^{j=k+3} A_j} \quad (4.6)$$

où k est l'indice d'une raie spectrale maximale (valeur des maximums présents dans le Tableau 4.2) de la *TFD* et A_k son amplitude, k' est le nouvel indice résultant du calcul de l'interpolation (n'étant pas forcément un entier).

Comme on peut le voir dans le Tableau 4.3, on peut gagner jusqu'à 0.50 Hz de précision avec le calcul d'interpolation appliquée aux fréquences fondamentales détectées.

Il existe des calculs d'interpolation plus complexe et encore plus précis [Quinn, 1997], mais des tests ont été réalisés et les résultats ne sont pas plus convaincants qu'avec le calcul qu'on utilise.

notes et FF	FF détectées	FF calculées après interpolation
F#2 = 92.50 Hz	93.67 Hz	93.17 Hz
G3 = 196 Hz	196.88 Hz	196.43 Hz
C4 = 261.63 Hz	262.50 Hz	262.12 Hz
E4 = 329.63 Hz	330.28 Hz	329.80 Hz

TABLEAU 4.3 – Tableau des fréquences fondamentales détectées avant et après interpolation

4.5 Conclusion

Dans ce chapitre, on a présenté étape par étape l'algorithme de détection multipitch appliqué à des signaux de guitare. On a tout d'abord vu quels ont été les fondements de l'algorithme et de quelles bases on est parti pour le concevoir. Il est évident que l'algorithme présenté a connu de nombreuses évolutions par rapport aux premiers développements qui ont été réalisés.

On a vu que l'algorithme présenté est décomposé en 2 étapes. Un premier étage appelé *Analyse Temporelle* permettant de ne se concentrer que sur le signal après que l'attaque (moment où une corde est grattée) ait eu lieu. Cette analyse est basée sur le calcul de la dérivée de l'enveloppe temporelle qui augmente de façon significative au moment de l'attaque. De plus, étant donné que les fréquences qu'on analyse sont comprises entre 80 Hz et 2000 Hz, on peut réduire la bande de fréquence du signal d'entrée qui est elle de 22 kHz. Donc après avoir détecté le signal à partir de l'attaque, celui-ci est sous-échantillonné par un facteur 10. Dans le deuxième étage appelé *Analyse Fréquentielle*, c'est une portion (2048 échantillons, soit 1/2 s) du signal résultant de l'*Analyse Temporelle* qui est analysée à l'aide d'une Transformée de Fourier Discrète (*TFD*). De nombreuses modifications sont appliquées au spectre de manière à ce qu'on ne retrouve que les composantes sinusoidales du signal d'entrée. Une fois celles-ci détectées, une approche itérative basée sur les relations harmoniques entre les fréquences fondamentales et les harmoniques permet de mettre en valeur les fréquences fondamentales recherchées par rapport aux partiels composant le spectre. Finalement, afin d'être le plus précis possible en terme de valeur des fréquences fondamentales, un calcul d'interpolation

appliqué autour des fondamentales détectées est utilisé.

Le chapitre suivant va donc présenter les résultats de l'application de cet algorithme à des signaux multipitch synthétisés avec un degré de polyphonie allant de 1 à 6, ainsi qu'à des notes et accords de guitare enregistrés.

CHAPITRE 5

SIMULATIONS ET RÉSULTATS

Dans ce chapitre, nous présentons les résultats de simulation de l'algorithme présenté dans le chapitre précédent. Comme on l'a vu à plusieurs reprises dans ce mémoire, l'algorithme de détermination multipitch a été développé par rapport à un instrument précis : la guitare.

Les signaux tests utilisés sont donc soit des signaux multipitch synthétisant la partie dite stationnaire d'une note (1 fréquence fondamentale) ou d'un accord (2 à 6 fréquences fondamentales) de guitare, soit des signaux de guitares enregistrés à l'aide d'un micro et allant d'un degré de polyphonie de 1 à 6. Dans le suite de ce chapitre, on appelle *accord* tout sons simulés ou réels qui combine au moins 2 fréquences fondamentales.

5.1 Différents signaux traités

Comme on l'a vu précédemment, les fréquences tests utilisées sont restreintes aux 5 premiers frets de la guitare allant donc de 82.41 Hz à 440 Hz. Toutes les fréquences plausibles sont détaillées dans le Tableau 5.1 où la corde 6 (grosse corde) correspond à celle sur le haut du manche de la guitare.

Ce sont ces fréquences fondamentales mélangées dans des signaux artificiels et des signaux de guitare que l'algorithme va estimer. On va, dans les sous-sections suivantes présenter les signaux artificiels ainsi que les signaux de guitare en séparant les signaux correspondant aux mélodies de ceux correspondant aux accords.

5.1.1 Signaux artificiels

Pour simuler les signaux artificiels dont on avait besoin pour les tests, un script *MATLAB* permettant de choisir le nombre et la valeur des fréquences fondamentales composant le signal

Corde	A vide	Fret 1	Fret 2	Fret 3	Fret 4	Fret 5
1	329.63	349.63	370.00	392.00	415.30	440.00
2	247.00	261.63	277.18	293.66	311.13	329.63
3	196	207.65	220.00	233.08	246.94	261.63
4	146.83	155.56	164.81	174.61	185.00	196.00
5	110.00	116.54	123.47	130.81	138.59	146.83
6	82.41	87.30	92.50	98.00	103.83	110.00

TABLEAU 5.1 – Tableau des fréquences fondamentales (Hz) à vide et des 5 premiers frets d'une guitare

a été écrit. Il est possible de simuler des signaux de 1 à 6 fréquences fondamentales (avec les valeurs présentées dans le Tableau 5.1) et aussi le nombre d'harmoniques pour chaque fréquences fondamentales.

Le script se présente de la manière suivante :

Voici les fréquences fondamentales (Hz) des différentes cordes :

1^{ère} corde = 329.63 349.63 370.00 392.00 415.30 440.00

2^{ème} = 247.00 261.63 277.18 293.66 311.13 329.63

3^{ème} = 196 207.65 220.00 233.08 246.94 261.63

4^{ème} = 146.83 155.56 164.81 174.61 185.00 196.00

5^{ème} = 110.00 116.54 123.47 130.81 138.59 146.83

6^{ème} = 82.41 87.30 92.50 98.00 103.83 110.00

Combien d'harmoniques voulez-vous générer pour chaque fréquence (1-8) :5

Voulez vous créer un signal mono(0) ou multi(1) fréquence :

La première des possibilités et de créer un signal mono ou polyphonique. Selon le choix on va ensuite choisir la valeur des fréquences fondamentales qui vont composer le signal à traiter.

Voulez vous créer un signal mono(0) ou multi(1) fréquence :1

Choisir le nombre de corde que vous voulez simuler (2-6) :3

Choisir la fréquence de la 1^{ière} corde que vous voulez simuler :92.50

Choisir la fréquence de la 2^{ième} corde que vous voulez simuler :155.56

Choisir la fréquence de la 3^{ième} corde que vous voulez simuler :293.66

Ce sont donc les fréquences que l'on vient de sélectionner qui vont composer le signal à analyser. Celui-ci est échantillonné à 44.1 kHz et sa longueur est de 2048 échantillons. Le signal temporel résultant est présenté à la Figure 5.1.

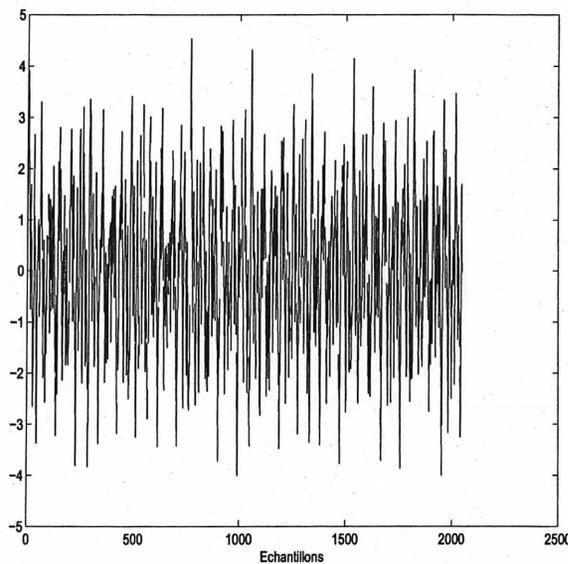


Figure 5.1 – Signal temporel synthétisé composé de 3 fréquences fondamentales

Les signaux simulés comme ceux présentés à la Figure 5.1 peuvent être comparés à la partie stationnaire d'un signal de guitare. En effet, une fois les fréquences fondamentales choisies, le signal à analyser est une somme de sinusôides (autant que le nombre de fréquences fondamentales choisies) ayant comme valeur de fréquences fondamentale celles choisies. L'amplitude des harmoniques est égale à l'amplitude de la fondamentale divisé par le rang de l'harmonique. Pour générer ce signal, on s'est servi de l'équation 2.13 présentée dans la section 2.3. Pour traiter ces signaux il n'est donc pas nécessaire de procéder à la détection de transitoire où recherche d'attaque. La partie *analyse temporelle* présentée dans la section 4.3 n'est pas utilisée dans ce cas de figure. C'est pour cela qu'uniquement 2048 échantillons du signal sont générés. On sait que pendant ce laps de temps, la fréquence fondamentale des notes

de guitare reste constante (cf Figure 2.11). Le signal synthétisé sera donc très proche d'une véritable portion d'un signal stationnaire de guitare et donc idéal pour faire les premiers tests et valider les algorithmes avant les tests avec les signaux de guitare enregistrés.

De plus, étant donné que grâce au script généré sous *MATLAB* toutes les combinaisons de fréquences possibles sont créées, cela résulte en un grand nombre de signaux ayant chacun des spectres différents grâce auxquels tous les tests voulus seront réalisés.

5.1.2 Mélodie : 1 note

Il est vrai que l'algorithme que l'on a développé a été réalisé dans le but d'analyser les signaux multipitch, mais il faut tout de même le tester dans le cas de mélodie, autrement dit quand une seule note est jouée. En effet, si l'algorithme n'est pas robuste avec une seule note comment peut-il l'être avec 6 ?

L'algorithme a donc été testé avec le générateur de signaux multipitch qui a été présenté dans la section précédente, sauf que cette fois, au moment de choisir entre le signal mono ou polyphonique et bien c'est le monophonique qui est choisit. Par rapport au script, le résultat serait celui-ci :

```
Voulez-vous créer un signal mono(0) ou multi(1) fréquence :0
```

```
Choisir la fréquence de la corde que vous voulez simuler :196
```

et la forme du signal temporel serait celle présentée sur la Figure 5.2.

On peut voir qu'au point de vue temporel, la courbe présentée sur la Figure 5.2 semble plus facilement analysable que celle présentée à la figure 5.1. En effet, sur la Figure 5.2 la période fondamentale (période à la quelle le signal se répète) est visible à l'oeil nu alors que rien n'est supposable sur la Figure 5.1.

On a aussi testé notre algorithme avec des vrais signaux de guitare monophonique. En effet, une seule note était jouée à chaque fois et chaque note individuelle du Tableau 5.1 a été testé. Il fallait donc que l'algorithme soit robuste à tous ces signaux avant de commencer l'analyse des signaux multipitch.

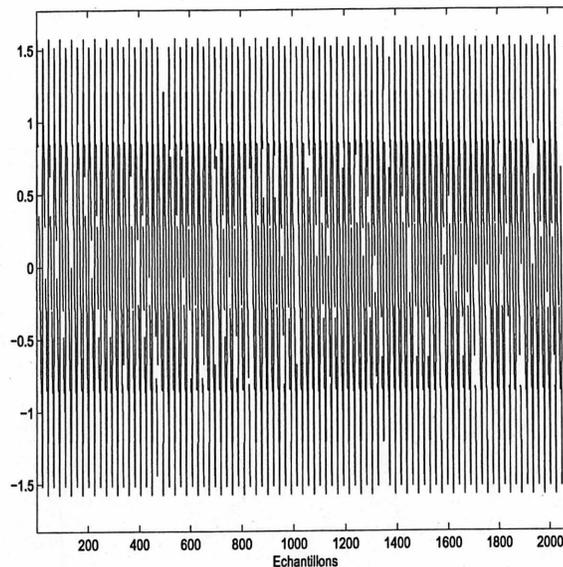


Figure 5.2 – Signal temporel synthétisé composé d'une fréquence fondamentale

Dans la Figure 5.3 sont présentées 4 notes de guitare individuelles.

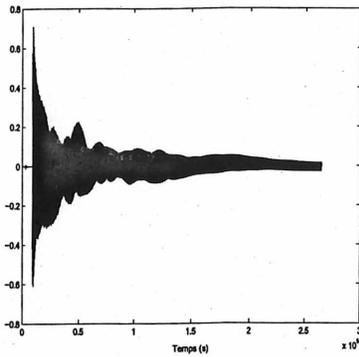
Après avoir présenté les signaux artificiels et les signaux monophoniques, on va présenter les signaux pour lesquels les algorithmes ont été développés : les accords.

5.1.3 Accord

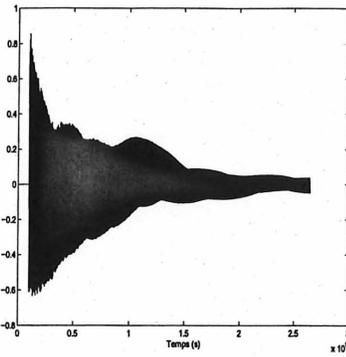
Comme on l'a présenté dans la section 2.4.3, les accords sont composés de plusieurs notes, donc de plusieurs fréquences fondamentales, qui dans notre cas sont choisies parmi celles présentées dans le Tableau 5.1.

De plus, on a trouvé dans la littérature une publication [Facoline,] (cf Annexe 2) qui regroupe tous les accords, avec leur composition, présentés de la même manière que celle de l'accord *C* sur la Figure 2.14. On note que tous les accords sont présentés avec la notation américaine soit *A*, *B*, *C*, *D*, *E*, *F*, *G* et qu'il y a plus de 1200 accords dans ce document que l'on retrouve dans l'Annexe 2. Donc un éventail d'accord largement suffisant pour les tests que l'on veut mener.

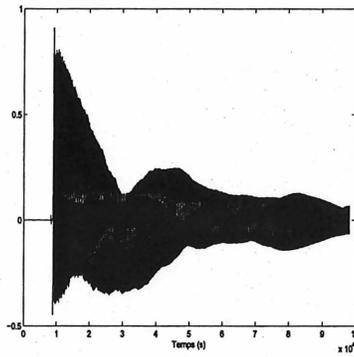
On a vu qu'avec les signaux artificiels on pouvait générer les signaux que l'on veut sans



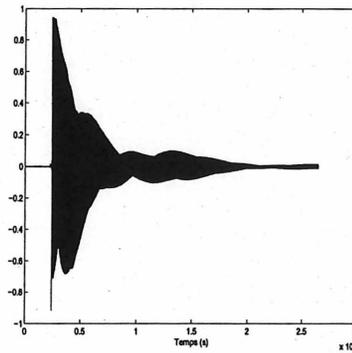
(a) note F#2 de fréquences
fondamentale 92.5 Hz



(b) note G3 de fréquences
fondamentale 196 Hz



(c) note C4 de fréquences
fondamentale 261.63 Hz



(d) note E4 de fréquences
fondamentale 329.63 Hz

Figure 5.3 – Signal temporel de 4 notes de guitare, (a) est la note F#2 de fréquence fondamentale 92.50 Hz, (b) est la note G3 de fréquence fondamentale 196 Hz, (c) est la note C4 de fréquence fondamentale 261.63 Hz, (d) est la note E4 de fréquence fondamentale 329.63 Hz

aucune difficulté. En ce qui concerne les accords de guitare, une petite partie d'entre eux ont été générés, c'est à dire 9 pour chaque polyphonie (2 à 6), résultant donc en une base de données de 45 accords de guitare.

Les accords qui ont été testés sont présentés (on peut retrouver les fréquences des différentes notes dans l'Annexe 1) dans les Tableaux 5.2 à 5.7.

Notes	Fréquences (Hz)
E2	82.41
A#2	116.54
C#3	138.59
D3	148.83
F#3	185
A3	220
B3	247
C4	261.63
F4	349.63

TABLEAU 5.2 – Tableau présentant la composition des notes testés pour un signal monophonique

Notes	Fréquences (Hz)
E2/A2	82.41/110
D3/G3	148.83/196
B3/E4	247/329.63
B2/F#3	123.47/185
E3/B3	164.81/247
A3/E4	220/329.63
A#2/E3	116.54/164.81
C#3/G4	138.59/392
C3/C4	130.81/261.63

TABLEAU 5.3 – Tableau présentant la composition des accords testés pour une polyphonie d'ordre 2

L'allure temporelle de 2 des accords décrits dans les tableaux ci-dessus sont présentées à la Figure 5.4.

Comme on peut le voir sur la Figure 5.4, il est impossible d'après l'allure temporelle de

Accord et Notes	Fréquences (Hz)
A/B. (E3/B3/C#4)	164.81/246.94/277.18
B/F# (B2/E3/A3)	123.47/164.81/220
C#m (E3/G#3/C#4)	164.81/207.65/277.18
Dadd9. (A3/D4/F#4)	220/293.66/370
Em7 (G#3/C4/F4)	207.65/261.63/349.63
Fmaj7.. (F3/A3/C4)	174.61/220/261.63
Gb/E.. (F#3/A#3/C#4)	185/233.08/277.18
D3/C4/D4	146.83/261.63/293.66
G3/D4/G4	196/293.66/392

TABLEAU 5.4 – Tableau présentant la composition des accords testés pour une polyphonie d'ordre 3

Accord et Notes	Fréquences (Hz)
Absus4 (D#3/G#3/C#4/G#4)	155.56/207.65/277.18/415.30
Bb7. (F3/A#3/D4/G#4)	174.61/233.08/293.66/415.30
C4 (F3/G3/C4/G4)	174.61/196/261.63/392
Dadd9. (E3/A3/D4/F#4)	164.81/220/293.66/370
E6 (F3/A#3/D4/G4)	174.61/233.08/293.66/392
Fsus (F3/A#3/C4/F4)	174.61/233.08/261.63/349.63
Gdmin (E3/A#3/C#4/F4)	164.81/233.08/277.18/392
G3/E3/C3/G2)	98/130.81/164.81/196
A3/E3/B2/F#2)	92.50/123.47/164.81/220

TABLEAU 5.5 – Tableau présentant la composition des accords testés pour une polyphonie d'ordre 4

spécifier le nombre de fréquences fondamentales présentes dans le signal d'origine. C'est donc pour cela que l'analyse se fait dans le domaine fréquentiel tout comme on l'a présenté à la section 4.4, car les caractéristiques du signal que l'on veut estimer y sont plus mises en

Accord et Notes	Fréquences (Hz)
Absus2/F. (A#2/F3/G#3/D#4/F4)	116.54/174.61/207.63/311.13/349.63
B7.... (B2/F#3/A3/D#4/F#4)	123.47/185/220/311.13/370
C6.. (A2/E3/G3/C4/E4)	110/164.81/196/261.63/329.63
D7#9 (B2/D#3/A3/D4/G4)	123.47/155.56/220/293.66/392
A11 (C#3/E3/B3/D4/G4)	138.59/164.81/246.54/293.66/392
G9. (B2/F3/A3/D4/G4)	123.47/174.61/220/293.66/392
Ebadd9 (A#2/D#3/A#3/D#4/F4)	116.54/155.56/233.08/311.13/349.63
Fsus4/G. (A#2/F3/G#3/D#4/F4)	116.54/174.61/207.65/311.13/349.63
Gsus2/E, (A2/E3/A3/D4/G4)	110/164.81/220/293.66/392

TABLEAU 5.6 – Tableau présentant la composition des accords testés pour une polyphonie d'ordre 5

Accord et Notes	Fréquences (Hz)
Abdim/E. (E2/B2/D3/G#3/B3/E4)	82.41/123.47/146.83/207.65/247/329.63
Gb/E. (F#2/C#3/E3/A#3/C#4/F#4)	92.50/138.59/164.81/233.08/277.18/370
A6/7. (E2/A2/E3/G3/C4/F4)	82.41/110/164.81/196/261.63/349.63
B7/add11. (E2/A2/F#3/B3/D#4/E4)	82.41/110/185/246.94/311.13/329.63
E11 (F2/A#2/D#3/G#3/C#4/F#4)	87.30/116.54/155.56/207.65/277.18/370
Dbaug.G. (F2/A2/F3/G3/C#4/F4)	87.30/110/174.61/196/277.18/349.63
E9 (F2/C3/D#3/A3/C4/F4)	87.30/130.81/155.56/220/261.63/392
F9 (F#2/C#3/E3/A#3/C#4/G#4)	92.50/138.59/164.81/233.08/277.18/415.30
A/F#, (F#2/A2/D#3/G#3/C4/E4)	92.50/110/155.56/207.65/261.63/329.63

TABLEAU 5.7 – Tableau présentant la composition des accords testés pour une polyphonie d'ordre 6

valeur.

On va voir dans la section 5.2.2 quels sont les résultats obtenus par l'algorithme quand on présente les accords décrits dans les 5 tableaux précédents.

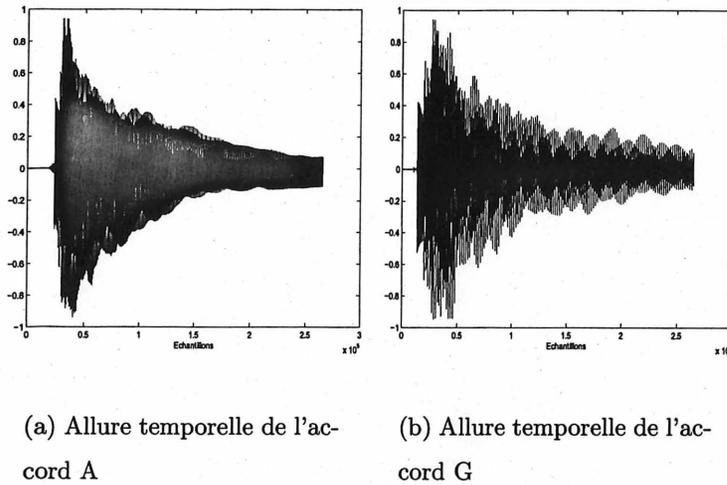


Figure 5.4 – Allure temporelle de 2 accords, (a) Accord A, (b) Accord G

5.2 Résultats

Pour vérifier la fiabilité de l'algorithme quant à l'estimation des fréquences fondamentales, on utilise un système de mesure qui s'appelle *Note Error Rate (NER)* [Klapuri, 2001]. Ce système de mesure stipule qu'une fréquence fondamentale est correctement détectée si la valeur détectée dévie de moins d'un demi-ton ($\pm 3\%$) de la véritable valeur. Le *NER* est défini comme étant la somme des fréquences fondamentales erronées divisé par le nombre de fréquences fondamentales présentes dans le signal d'origine. Ce chiffre est donc facilement interpretable en tant que pourcentage. Les erreurs sont de 3 types : les erreurs dites de substitution et de suppression qui correspondent aux fréquences fondamentales qui n'ont pu être détectées par l'algorithme et les erreurs d'insertion lorsqu'une fréquence fondamentale non présente dans le signal d'origine a été détectée.

On a vu dans la section 4.4.1 que la taille de la fenêtre d'analyse a une importance primordiale quand à la résolution en fréquence que va avoir le spectre. Plus la taille d'analyse est grande, meilleure va être la résolution. Cependant, longue fenêtre d'analyse est synonyme de long support temporel. Or on veut que notre algorithme soit le plus rapide possible afin de pouvoir possiblement analyser des accords successifs. Il ne faut donc pas analyser une séquence trop

longue du signal, mais il faut tout de même que les résultats restent fiables.

On a donc procédé à un test dans lequel on ne fait que changer la taille de la fenêtre d'analyse en allant de 4096 échantillons (soit environ 1 s du signal) à 1024 échantillons (soit environ 250 ms) par pas de 512 échantillons mais tout en gardant une *TFD* sur 4096 points (résolution fréquentielle allant de 1 Hz à 4 Hz). On vérifie donc l'évolution du *NER* et ceci dans tous les cas de polyphonie possible. Le test a été réalisé avec des signaux artificiels mais dont les fréquences ont été prises parmi celles présentées dans les Tableaux 5.3 à 5.7. Les résultats obtenus sont présentés sur la Figure 5.5.

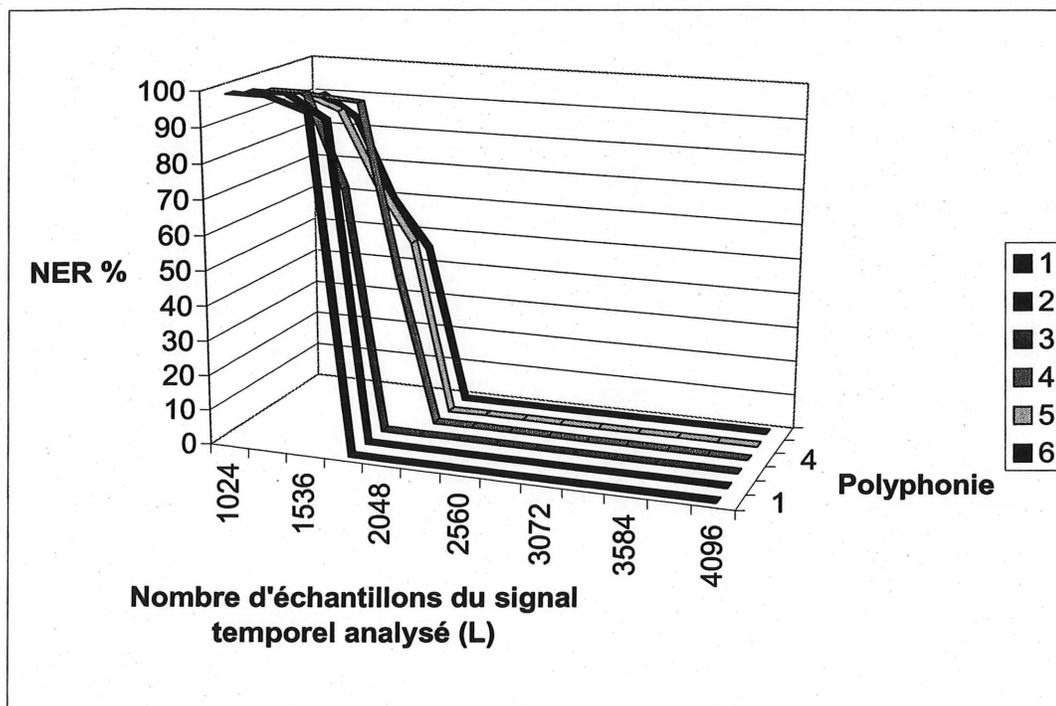


Figure 5.5 – NER par rapport à la polyphonie et aux longueurs de signal analysé

On remarque donc qu'analyser 4096 ou 2048 échantillons ne change pas le *NER*, et ceci quel que soit le degré de polyphonie. Ce test permet donc de réduire le support temporel choisit pour analyser 1/2 s du signal, ce qui pourrait permettre d'analyser plusieurs accords successifs à condition d'avoir un délai d'au moins une demi-seconde entre chaque accord.

5.2.1 Signaux artificiels

Le cas des mélodies et des accords générés par des signaux artificiels vont être présentés dans les mêmes résultats.

Afin d'être le plus proche possible de la réalité avec les signaux artificiels, on a choisit les mêmes combinaisons de notes que celles présentées précédemment. Les résultats sont présentés dans les Tableaux 5.8 à 5.13.

Fréquences testées (Hz)	Fréquences détectées (Hz)
82.41	82.84
116.54	116.19
138.59	138.78
146.83	146.32
185	185.05
220	219.47
247	246.37
261.63	261.43
349.63	349.65
9	9 ($NER = 0\%$)

TABLEAU 5.8 – Tableau présentant les fréquences testées et les fréquences détectées pour un signal monophonique artificiel

De plus, étant donné qu'il est possible de générer de façon simple (puisque informatique) toutes les combinaisons de notes possible, les résultats présentés dans la Figure 5.6 représentent un plus grand nombre d'accords que ceux présentés dans les tableaux de résultats. Il est sûr que tous les cas possibles ne sont pas tous des accords réalisables sur une guitare, donc cela permet aussi de tester l'algorithme de façon général, c'est à dire sans se préoccuper de la nature du signal (type d'instrument), mais en le voyant comme une somme de sinusoides.

Étant donné que l'on se base sur le spectre du signal pour faire l'analyse, il est important de savoir quel est l'écart minimal (Δf_{min}) qu'il y a entre 2 fréquences pour qu'elles soient

Fréquences testées (Hz)	Fréquences détectées (Hz)
82.41/110	82.84/109.74
146.83/196	146.32/195.81
247/329.63	246.37/329.21
123.47/185	123.72/185.05
164.81/247	164.60/246.37
220/329.63	219.47/329.28
116.54/164.81	116.19/164.60
138.59/392	138.78/391.60
130.81/261.63	131.25
18	17 (<i>NER</i> = 5.5%)

TABLEAU 5.9 – Tableau présentant les fréquences testées et les fréquences détectées pour une polyphonie d'ordre 2 pour les signaux artificiels

Fréquences testées (Hz)	Fréquences détectées (Hz)
164.81/246.94/277.18	164.60/246.37/277.57
123.47/164.81/220	123.72/164.60/219.47
164.81/207.65/277.18	164.60/207.64/277.57
220/293.66/370	219.47/293.71/370.10
207.65/261.63/349.63	207.64/261.43/349.65
174.61/220/261.63	174.29/219.47/261.43
185/233.08/277.18	185.05/233.46/277.57
146.83/261.63/293.66	146.31/261.43
196/293.66/392	195.81/293.71
27	25 (<i>NER</i> = 7.4%)

TABLEAU 5.10 – Tableau présentant les fréquences testées et les fréquences détectées pour une polyphonie d'ordre 3 pour les signaux artificiels

Fréquences testées (Hz)	Fréquences détectées (Hz)
155.56/207.65/277.18/415.30	156/207.64/277.57
174.61/233.08/293.66/415.30	174.29/233.46/293.71/415.27
174.61/196/261.63/392	174.29/195.81/261.43
164.81/220/293.66/370	164.60/219.47/293.71/370.10
174.61/233.08/293.66/392	174.29/233.46/293.71/391.60
174.61/233.08/261.63/349.63	174.29/233.46/261.43
164.81/233.08/277.18/392	164.60/233.46/277.57/391.60
98/130.81/164.81/196	97.90/131.25/164.60
92.50/123.47/164.81/220	92.52/123.72/164.60/219.47
36	32 (<i>NER</i> = 11.1%)

TABLEAU 5.11 – Tableau présentant les fréquences testées et les fréquences détectées pour une polyphonie d'ordre 4 pour les signaux artificiels

Fréquences testées (Hz)	Fréquences détectées (Hz)
116.54/174.61/207.63/311.13/349.63	116.19/174.29/207.64/310.92
123.47/185/220/311.13/370	123.72/185.05/219.47/310.92
110/164.81/196/261.63/329.63	109.74/164.60/195.81/261.43
123.47/155.56/220/293.66/392	123.72/156/219.47/293.71/391.60
138.59/164.81/246.54/293.66/392	138.78/164.60/246.37/293.71/391.61
123.47/174.61/220/293.66/392	123.72/174.29/219.47/293.71/391.61
123.47/155.56/220/247/370	123.72/156/219.47/370.10
116.54/174.61/207.65/311.13/349.63	116.19/174.29/207.64/310.92
110/164.81/220/293.66/392	109.74/164.60/293.71/391.60
45	39 (<i>NER</i> = 13.3%)

TABLEAU 5.12 – Tableau présentant les fréquences testées et les fréquences détectées pour une polyphonie d'ordre 5 pour les signaux artificiels

Fréquences testées (Hz)	Fréquences détectées (Hz)
82.41/123.47/146.83/207.65/247/329.63	82.84/123.72/146.32/207.64/329.21
92.50/138.59/164.81/233.08/277.18/370	92.52/138.78/164.60/233.46/370.10
82.41/110/164.81/196/261.63/349.63	82.84/109.73/195.80/261.43/349.65
82.41/110/185/246.94/311.13/329.63	82.84/109.73/185.05/246.37/310.92
87.30/116.54/155.56/207.65/277.18/370	87.14/116.19/156/207.64/277.57/370.10
87.30/110/174.61/196/277.18/349.63	87.14/109.73/195.80/277.57/349.65
87.30/130.81/155.56/220/261.63/392	87.15/131.25/156/219.47/391.60
92.50/138.59/164.81/233.08/277.18/415.30	92.52/138.79/164.60/233.46/415.27
92.50/110/155.56/207.65/261.63/329.63	92.52/109.74/156/207.64/261.43
54	46 (NER = 14.8%)

TABLEAU 5.13 – Tableau présentant les fréquences testées et les fréquences détectées pour une polyphonie d'ordre 6 pour les signaux artificiels

toutes les deux détectées. Pour cela, on a fait évoluer nos fréquences fondamentales (de 2 à 6) de 80 à 150 Hz en les incrémentant de 0.50 Hz à chaque itération. On peut donc voir sur la Figure 5.6 qu'il faut en moyenne que 2 fréquences soient espacées de 2 Hz pour qu'elles soient détectables.

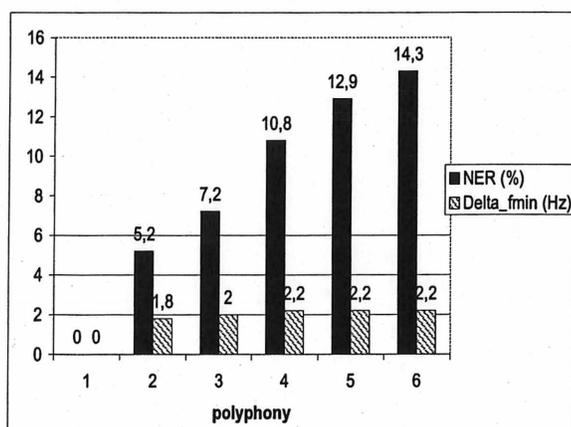


Figure 5.6 – NER pour les signaux simulés avec le Δf_{min}

Comme on peut le voir sur les tableaux précédents, les résultats obtenus avec quelques accords

et ceux obtenus avec toutes les combinaisons de notes possibles restent assez proches. De plus, comme on peut le constater dans les Tableaux 5.8, 5.9, 5.10, 5.11, 5.12, 5.13, les cas où l'on n'arrive pas à détecter la fréquence fondamentale se produisent lorsqu'il y a une relation d'octave entre 2 fréquences fondamentales. Le problème qui se pose dans ce cas de figure est que tout les harmoniques du son le plus haut (fréquence fondamentale la plus aiguë) vont être confondues avec les harmoniques pairs du son le plus bas. Il est donc impossible de les différencier dans le spectre, ce qui rend la méthode peu fiable en ce qui concerne la détection de fréquences fondamentales dans ce cas là. Cependant, comme on peut le voir dans les résultats, l'algorithme est très robuste dans les autres combinaisons de note, avec de faibles écarts de fréquences entre les notes détectées et les notes testées. En effet, les écarts évoluent entre 0.01 Hz et 0.55 Hz, donc largement dans le ($\pm 3\%$) dont il est question dans la définition du *NER*.

5.2.2 Accord

On stipule qu'un accord est détecté si toutes les fréquences fondamentales composant cet accord ont été détectées. On peut donc imaginer que si un accord est composé d'au moins deux fréquences ayant une relation d'octave entre elles, alors cet accord ne sera pas détecté.

Cependant, on a vu dans la section 2.2 que les harmoniques n'allaient pas être situées aux multiples entières de la fréquence fondamentale. De ce fait, il est possible que des fréquences qui n'ont pas pu être détectées dans le cas de signaux artificiels vont l'être dans le cas de signaux de guitare. C'est ce que l'on va voir dans les résultats présentés dans les Tableaux 5.14 à 5.19.

Les valeurs de *NER* pour chacun des cas de polyphonie, ainsi que le nombre d'accord qui ont été détectés sont présentées dans la Figure 5.7. Dans tous ces cas, la détection de transitoire est faite de façon automatique.

Comme on peut le voir dans les Tableaux 5.14 à 5.19 ainsi que dans la Figure 5.7, on détecte plus de fréquences fondamentales dans le cas de signaux de guitare que dans le cas des signaux artificiels (cf Figure 5.6). On remarque cependant que plus la polyphonie augmente,

Fréquences testées (Hz)	Fréquences détectées (Hz)
82.41	82.93
116.54	116.13
138.59	138.86
146.83	146.20
185	185.35
220	220.54
247	246.85
261.63	261.30
349.63	349.65

TABLEAU 5.14 – Tableau présentant les fréquences testées et les fréquences détectées pour des notes de guitares

Fréquences testées (Hz)	Fréquences détectées (Hz)
82.41/110	82.93/109.54
146.83/196	146.20/196.44
247/329.63	246.29/329.14
123.47/185	123.86/185.35
164.81/247	165.06/246.85
220/329.63	220.54/329.14
116.54/164.81	116.13/165.06
138.59/392	138.86/390.53
130.81/261.63	131.52

TABLEAU 5.15 – Tableau présentant les fréquences testées et les fréquences détectées pour une combinaison de 2 notes de guitare

plus faible sera le nombre d'accord détectés. Ceci s'explique par le fait qu'on a plus de chances de rencontrer un cas de relation d'octave dans un accord ayant plus de fréquences fondamentales. On sait aussi que les accords majeurs (cf section 2.4.3) sont construits d'une

Fréquences testées (Hz)	Fréquences détectées (Hz)
164.81/246.94/277.18	164.43/246.85/277.72
123.47/164.81/220	123.86/165.06/219.67
164.81/207.65/277.18	165.06/207.37/277.72
220/293.66/370	219.67/293.84/370.22
207.65/261.63/349.63	207.37/261.30/349.65
174.61/220/261.63	174.12/220.54/261.30
185/233.08/277.18	185.35/233.52/277.72
146.83/261.63/293.66	146.20/261.30
196/293.66/392	196.44/293.84

TABLEAU 5.16 – Tableau présentant les fréquences testées et les fréquences détectées pour combinaison de 3 notes de guitare

Fréquences testées (Hz)	Fréquences détectées (Hz)
155.56/207.65/277.18/415.30	156.09/207.37/277.72
174.61/233.08/293.66/415.30	174.12/233.52/293.84/415.09
174.61/196/261.63/392	174.12/196.44/261.30
164.81/220/293.66/370	165.06/220.54/293.84/370.22
174.61/233.08/293.66/392	174.12/233.52/293.84/390.53
174.61/233.08/261.63/349.63	174.12/233.52/261.30/349.65
164.81/233.08/277.18/392	165.06/233.52/277.72/390.53
98/130.81/164.81/196	97.73/131.52/165.06
92.50/123.47/164.81/220	92.17/123.86/165.06/219.67

TABLEAU 5.17 – Tableau présentant les fréquences testées et les fréquences détectées pour une combinaison de 4 notes de guitare

manière bien précise, accords dans lesquels on retrouve de nombreuses relations d'octave. On a donc, pour les tests, choisit des accords dans lesquels il y avait moins de relations d'octave de manière à voir le comportement de l'algorithme dans ces cas là. On s'est rendu

Fréquences testées (Hz)	Fréquences détectées (Hz)
116.54/174.61/207.63/311.13/349.63	116.13/174.12/207.37/310.79
123.47/185/220/311.13/370	123.88/185.35/220.54/310.79
110/164.81/196/261.63/329.63	109.54/165.06/196.44/261.30
123.47/155.56/220/293.66/392	123.88/156.09/220.54/293.84/390.53
138.59/164.81/246.94/293.66/392	138.86/165.06/246.85/293.84/390.53
123.47/174.61/220/293.66/392	123.88/174.12/220.54/293.84/390.53
123.47/155.56/220/247/370	123.88/156.09/219.67/370.22
116.54/174.61/207.65/311.13/349.63	116.13/174.12/207.37/310.79/349.65
110/164.81/220/293.66/392	109.54/165.06/293.84/390.53

TABLEAU 5.18 – Tableau présentant les fréquences testées et les fréquences détectées pour une combinaison de 5 notes de guitare

Fréquences testées (Hz)	Fréquences détectées (Hz)
82.41/123.47/146.83/207.65/247/329.63	82.93/123.88/146.20/207.37/329.14
92.50/138.59/164.81/233.08/277.18/370	92.17/138.86/165.06/233.52/370.22
82.41/110/164.81/196/261.63/349.63	82.93/109.54/196.44/261.30/349.65
82.41/110/185/246.94/311.13/329.63	82.93/109.54/185.35/246.85/310.79/329.14
87.30/116.54/155.56/207.65/277.18/370	87.03/116.13/156.09/207.37/277.72/370.22
87.30/110/174.61/196/277.18/349.63	87.03/109.54/196.44/277.72/349.65
87.30/130.81/155.56/220/261.63/392	87.03/131.52/156.09/220.54/390.53
92.50/138.59/164.81/233.08/277.18/415.30	92.17/138.86/165.06/233.52/415.09
92.50/110/155.56/207.65/261.63/329.63	92.172/109.54/156.09/207.37/261.30

TABLEAU 5.19 – Tableau présentant les fréquences testées et les fréquences détectées pour une combinaison de 6 notes de guitare

compte que l'algorithme était très robuste et que toutes les fréquences fondamentales étaient détectées quand on n'était pas dans un cas de relation d'octave.

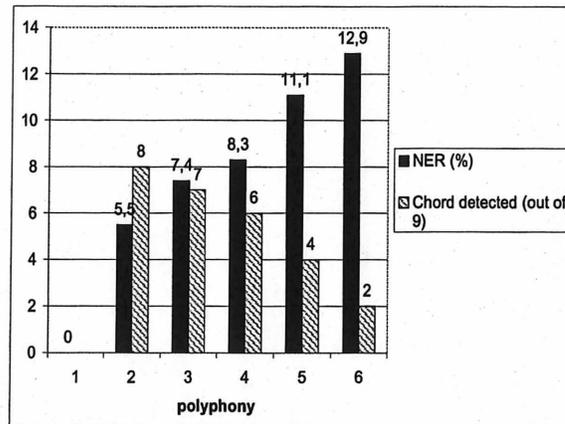


Figure 5.7 – NER pour les accords de guitare avec le nombre d'accords détectés

5.3 Conclusion

Dans ce chapitre, on a présenté les résultats de l'algorithme qui avait été décrit dans la section 4. L'algorithme a été testé à la fois avec des signaux multipitch artificiels et des signaux de guitare.

Dans un premier temps l'algorithme a été testé avec des signaux multipitch artificiels qui ont pour particularité de se rapprocher de la partie stationnaire d'un signal de guitare, partie que l'on souhaite analyser car c'est d'elle qu'on va pouvoir tirer les résultats les plus fiables étant donné que c'est là où les fréquences fondamentales présentes dans le signal sont stables. Le simulateur de signaux multipitch nous permet de choisir entre signaux monophoniques et signaux polyphoniques (dans lesquels on peut choisir l'ordre de la polyphonie). Cela permet donc de générer toutes les combinaisons de notes possibles et donc d'avoir des résultats vraiment généraux. Cependant pour tester l'algorithme par rapport aux signaux de guitare, certaines combinaisons de notes particulières ont été testées en plus. Les résultats montrent la robustesse de l'algorithme ainsi que la précision d'estimation des fréquences fondamentales (déviations de 0.55 Hz au maximum entre fréquence testée et fréquence estimée).

Les combinaisons de notes dont on s'est servi pour simuler des signaux artificiels correspondaient à de véritables accords de guitare. Ce sont donc avec ces accords (enregistrés à l'aide d'un micro) que l'on a testé l'algorithme. Il y en avait 9 pour chaque polyphonie,

résultant donc en une base de données de 54 fichiers son. Comme on a pu le constater dans les résultats, les accords qui ne sont pas détectés sont dû au fait qu'il y a une relation d'octave entre les fréquences fondamentales composant ces accords. C'est le seul problème rencontré par l'algorithme, mais problème important car d'après les caractéristiques de construction des accords, il est très fréquent de retrouver au moins deux fréquences (si ce n'est plus) ayant une relation d'octave.

La prochaine étape de l'algorithme serait donc de mettre au point une solution permettant de remédier à ce problème. Cependant, c'est un problème très complexe qui n'a pas encore trouvé de solution fiable et robuste.

CHAPITRE 6

CONCLUSION

Dans ce document, la réalisation d'un algorithme permettant d'extraire le nombre de notes jouées dans un accord de guitare et d'estimer la valeur des fréquences fondamentales de ces notes est présentée. L'algorithme a été développé sous *MATLAB* et il a été testé à la fois avec des signaux multipitch artificiels et des signaux de guitare.

Avant de présenter l'algorithme en lui-même, une description de l'acoustique des instruments à corde et en particulier les caractéristiques des sons émis par ce type d'instruments a été présenté. Étant donné que l'instrument testé ici est la guitare, une description plus avancée de son fonctionnement, notamment sur la construction des accords, l'importance des fréquences fondamentales avec les relations interharmoniques a été réalisée.

Ensuite, il était important de présenter l'état de l'art au point de vue de la détection multipitch. Un bref rappel sur les méthodes de détection monophonique a été fait. Celui-ci a permis de spécifier qu'il n'est pas ou peu possible de se baser sur ces algorithmes pour faire de la détection multipitch. Au point de vue de la détection polyphonique, il est tout d'abord important de noter qu'aucune technique développée jusqu'à présent n'est totalement fiable et robuste. Il y a donc encore de nombreuses évolutions à faire dans ce domaine. La détection polyphonique voit ses applications dans le domaine de la séparation de source, de la représentation structurelle de signaux audio, de la transcription automatique de la musique et de la synthèse de son. On a vu dans ce recensement des différentes techniques de détection polyphonique qu'elles utilisent aussi bien des méthodes classiques de traitement de signal, des connaissances acoustiques, que le fonctionnement de l'oreille humaine.

Enfin, l'algorithme développé et les résultats obtenus ont été présentés. On a vu que l'algorithme était composé de 2 étages. Un premier étage appelé *Analyse Temporelle* permet de ne se concentrer que sur le signal après que l'attaque (moment où une corde est grattée) ait eu lieu. Dans le deuxième étage appelé *Analyse Fréquentielle*, c'est une portion (2048

échantillons, soit $1/2$ s) du signal résultant de l'*Analyse Temporelle* qui est analysé et duquel on ressort le nombre et la valeur des fréquences fondamentales présentes dans le signal d'origine.

L'algorithme a été testé avec des signaux multipitch artificiels se rapprochant de la partie stationnaire d'un accord de guitare et aussi avec des accords de guitare. Les résultats montrent que l'algorithme est fiable et robuste au point de vue de l'estimation des fréquences fondamentales. Le seul problème présent est celui de la détection de 2 fréquences fondamentales espacées d'une octave l'une de l'autre. Dans le cas où une fréquence n'est pas détectée alors l'accord n'est pas détecté. Donc d'un point de vue détection d'accord, l'algorithme n'est pas aussi robuste que pour la détection des fréquences fondamentales.

Les travaux futurs à mettre en oeuvre sont donc le développement d'une nouvelle méthode permettant de tenir compte des relations d'octaves de sortes que celles-ci ne perturbent plus l'analyse des signaux. Cependant, c'est une problématique très complexe et aucune méthode de résolution n'a encore été proposée.

PREMIÈRE PARTIE

ANNEXE 1

Voici les notes du fretboard d'une guitare (les notes sont inscrites au-dessus de l'endroit où l'on pince la corde) :

Corde	E4	F4	F#4	G4	G#4	A4	A#4	B4	C5	C#5	D5	D#5	E5	F5	F#5	G5	G#5	A5	A#5	B5	C6	C#6	D6	D#6
1	B3	C4	C#4	D4	D#4	E4	F4	F#4	G4	G#4	A4	A#4	B4	C5	C#5	D5	D#5	E5	F5	F#5	G5	G#5	A5	A#5
2	G3	G#3	A3	A#3	B3	C4	C#4	D4	D#4	E4	F4	F#4	G4	G#4	A4	A#4	B4	C5	C#5	D5	D#5	E5	F5	F#5
3	D3	D#3	E3	F3	F#3	G3	G#3	A3	A#3	B3	C4	C#4	D4	D#4	E4	F4	F#4	G4	G#4	A4	A#4	B4	C5	C#5
4	A2	A#2	B2	C3	C#3	D3	D#3	E3	F3	F#3	G3	G#3	A3	A#3	B3	C4	C#4	D4	D#4	E4	F4	F#4	G4	G#4
5	E2	F2	F#2	G2	G#2	A2	A#2	B2	C3	C#3	D3	D#3	E3	F3	F#3	G3	G#3	A3	A#3	B3	C4	C#4	D4	D#4
6	A vide	Fret 1	Fret 2	Fret 3	Fret 4	Fret 5	Fret 6	Fret 7	Fret 8	Fret 9	Fret 10	Fret 11	Fret 12	Fret 13	Fret 14	Fret 15	Fret 16	Fret 17	Fret 18	Fret 19	Fret 20	Fret 21	Fret 22	(Fret 23)

Fréquences de référence pour chacune des notes ci-dessus :

1	329,63	349,63	370,00	392,00	415,30	440,00	466,16	493,88	523,25	554,37	587,33	622,25	659,26	698,46	739,99	783,99	830,61	880,00	932,33	987,77	1046,50	1108,73	1174,66	1244,51
2	247,00	261,63	277,18	293,66	311,13	329,63	349,23	369,99	392,00	415,30	440,00	466,16	493,88	523,25	554,37	587,33	622,25	659,26	698,46	739,99	783,99	830,61	880,00	932,33
3	196,00	207,65	220,00	233,08	246,94	261,63	277,18	293,66	311,13	329,63	349,23	369,99	392,00	415,30	440,00	466,16	493,88	523,25	554,37	587,33	622,25	659,26	698,46	739,99
4	146,83	155,56	164,81	174,61	185,00	196,00	207,65	220,00	233,08	246,94	261,63	277,18	293,66	311,13	329,63	349,23	369,99	392,00	415,30	440,00	466,16	493,88	523,25	554,37
5	110,00	116,54	123,47	130,81	138,59	146,83	155,56	164,81	174,61	185,00	196,00	207,65	220,00	233,08	246,94	261,63	277,18	293,66	311,13	329,63	349,23	369,99	392,00	415,30
6	82,41	87,30	92,50	98,00	103,83	110,00	116,54	123,47	130,81	138,59	146,83	155,56	164,81	174,61	185,00	196,00	207,65	220,00	233,08	246,94	261,63	277,18	293,66	311,13

Voir la guitare comme en jouant droitier, la corde 6 (grosse corde) vers vous.

Voici les notes du fretboard, chacune accompagnée de sa fréquence fondamentale :

Corde	E4	F4	F#4	G4	G#4	A4	A#4	B4	C5	C#5	D5	D#5	E5	F5	F#5	G5	G#5	A5	A#5	B5	C6	C#6	D6	D#6
1	329,63	349,63	370,00	392,00	415,30	440,00	466,16	493,88	523,25	554,37	587,33	622,25	659,26	698,46	739,99	783,99	830,61	880,00	932,33	987,77	1046,50	1108,73	1174,66	1244,51
2	B3	C4	C#4	D4	D#4	E4	F4	F#4	G4	G#4	A4	A#4	B4	C5	C#5	D5	D#5	E5	F5	F#5	G5	G#5	A5	A#5
3	G3	G#3	A3	A#3	B3	C4	C#4	D4	D#4	E4	F4	F#4	G4	G#4	A4	A#4	B4	C5	C#5	D5	D#5	E5	F5	F#5
4	D3	D#3	E3	F3	F#3	G3	G#3	A3	A#3	B3	C4	C#4	D4	D#4	E4	F4	F#4	G4	G#4	A4	A#4	B4	C5	C#5
5	A2	A#2	B2	C3	C#3	D3	D#3	E3	F3	F#3	G3	G#3	A3	A#3	B3	C4	C#4	D4	D#4	E4	F4	F#4	G4	G#4
6	E2	F2	F#2	G2	G#2	A2	A#2	B2	C3	C#3	D3	D#3	E3	F3	F#3	G3	G#3	A3	A#3	B3	C4	C#4	D4	D#4
	A vide	Fret 1	Fret 2	Fret 3	Fret 4	Fret 5	Fret 6	Fret 7	Fret 8	Fret 9	Fret 10	Fret 11	Fret 12	Fret 13	Fret 14	Fret 15	Fret 16	Fret 17	Fret 18	Fret 19	Fret 20	Fret 21	Fret 22	(Fret 23)

DEUXIÈME PARTIE

ANNEXE 2

THE ULTIMATE GUITAR CHORD CHART II - by Phillip J Facoline

email- pfacol@elaine.ee.und.ac.za

The 's and ,s after the chord name means that the chord is a duplicate or alternate formation.

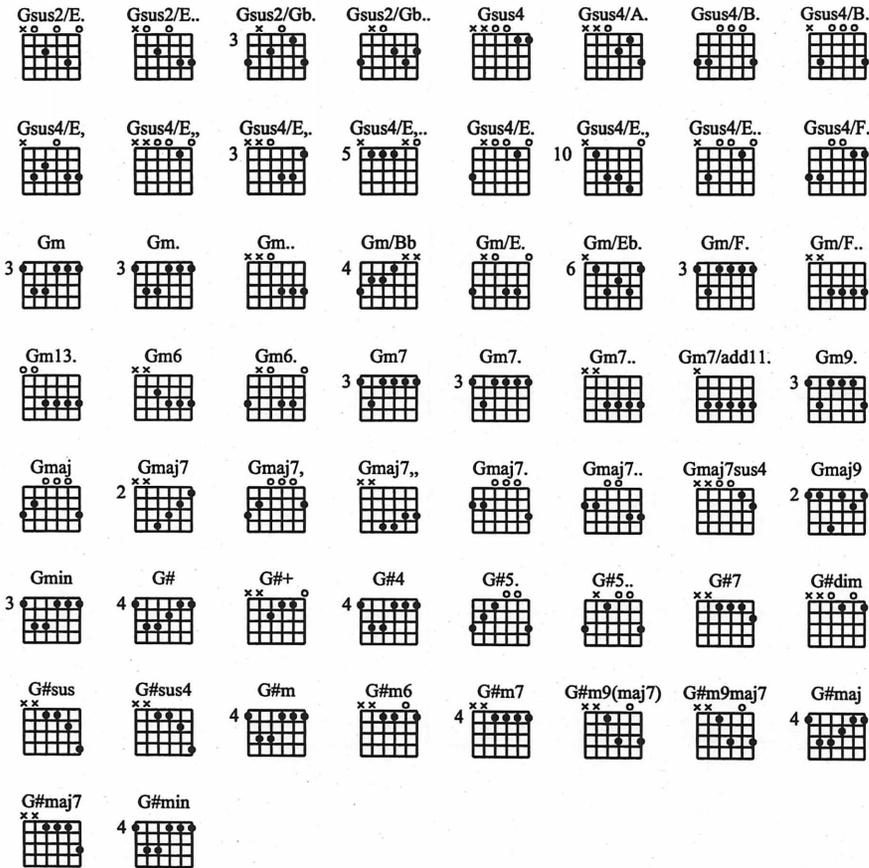
A#dim	A#sus	A#sus4	A#m	A#m7	A#maj	A#maj7	A#min	Bdim/A..	Bdim/A...	Bdim/Ab.	Bdim/Ab..	Bdim/Ab...	Bdim/G.	Bdim/G..	Bdim/G...
Bbb5.	Bb	Bb+	Bb.	Bb...	Bb/A.	Bb/Ab.	Bb/Ab..	Bdim7.	Bdim7..	Bdim7...	Bsus	Bsus.	Bsus..	Bsus2.	Bsus2..
Bb/Db.	Bb/E.	Bb/G.	Bb/G..	Bb11	Bb4	Bb5.	Bb6	Bsus2/E.	Bsus4	Bsus4/A.	Bsus4/A..	Bsus4/Ab.	Bsus4/Ab..	Bsus4/Db.	Bsus4/Eb.
Bb6.	Bb6/add9.	Bb7	Bb7.	Bb7..	Bb7sus4.	Bb9	Bbadd#11.	Bsus4/Eb..	Bsus4/G.	Bsus4/G..	Bsus4/G...	Bm	Bm(maj7)	Bm(sus9)	Bm.
Bbaug/E.	Bbdim	Bbdim/C.	Bbdim/D.	Bbdim/G.	Bbdim/G..	Bbdim/Gb.	Bbdim/Gb..	Bm...	Bm/A.	Bm/A..	Bm/A...	Bm/A....	Bm/A.....	Bm/G.	Bm/G..
Bbdim7.	Bbdim7..	Bbsus	Bbsus2.	Bbsus2/G.	Bbsus4	Bbsus4/Ab.	Bbm	Bm/G...	Bm/G....	Bm6	Bm7	Bm7(b5).	Bm7(b5)..	Bm7(b5)...	Bm7.
Bbm.	Bbm/Ab.	Bbm/D.	Bbm/Gb.	Bbm7	Bbm9	Bbmaj	Bbmaj7	Bm7..	Bm7...	Bm7....	Bm7.....	Bm7/add11.	Bm7b5	Bmaj	Bmaj7
Bbmaj7.	Bbmaj9.	Bbmin	Bb#5.	B	B(addE)	B+	B.	Bmaj7/#11.	Bmin	Bmsus9	Bmmaj7	B#5.	B#5..	Cb5.	C
B/A.	B/A..	B/A...	B/A....	B/Ab.	B/E.	B/E..	B/F#	C(add9)	C(addD)	C+	C..	C...	C....	C.....	C.....
B11	B11/13	B13	B4	B5.	B6.	B7	B7#9	C/A.	C/A..	C/A...	C/A....	C/B.	C/B..	C/B...	C/Bb.
B7(#9)	B7+	B7+5	B7..	B7....	B7/add11.	B7/add11..	B7sus4.	C/D.	C/D..	C/D...	C/D....	C/D.....	C/D.....	C/D.....	C/D.....
B7sus4..	B9	BaddE	BaddE/F#	Baug/E.	Baug/E..	Bdim	Bdim/A.	C/F.	C/F..	C11	C4	C5.	C6	C6.	C6..

C6.... 5	C6/add9. 5	C7 3	C7. 3	C7sus4. 3	C9 8	C9(11) x	C9(b5). x	Dbaug/D. x	Dbaug/G. x	Dbdim x	Dbdim/A. x	Dbdim/A.. x	Dbdim/A... x	Dbdim/B. x	Dbdim/Bb. x
Cadd2/B x	Cadd9 x	Cadd9, 3	Cadd9, 10	Cadd9,, x	Cadd9,,, 5	Cadd9. x	Cadd9.. x	Dbdim/Bb.. x	Dbdim/D. x	Dbdim/D.. x	Dbdim7. x	Dbdim7.. x	Dbsus 4	Dbsus2. 4	Dbsus4 4
Cadd9... x	Cadd9.... x	Cadd9..... x	CaddD x	Cdim x	Cdim/A. x	Cdim/Ab. x	Cdim/Ab.. 4	Dbsus4/Bb. x	Dbm x	Dbm. 4	Dbm.. 4	Dbm/A. x	Dbm/B. x	Dbm/B.. 4	Dbm7 x
Cdim/D. 2	Cdim7. x	Csus x	Csus. x	Csus.. x	Csus2 x	Csus2. 3	Csus2.. x	Dbm7(b5). x	Dbm7. x	Dbm7.. 4	Dbmaj x	Dbmaj7 x	Dbmaj7. x	Dbmaj7.. 4	Dbmin x
Csus2... 3	Csus2/A. x	Csus2/B. x	Csus2/B.. x	Csus2/E. x	Csus2/E.. 3	Csus2/E,,, 10	Csus2/E.. x	Db#5. x	Db#5.. x	D x	D(add9) x	D+ x	D. 2	D.. x	D.... x
Csus2/E.. x	Csus2/E... x	Csus2/F. x	Csus4 x	Csus4/A. x	Csus4/A.. x	Csus4/B. x	Csus4/Bb. 3	D..... 5	D/A x	D/B x	D/B. 2	D/B.. x	D/B... x	D/B.... x	D/B..... x
Csus4/D. x	Csus4/E. x	Csus4/E.. x	Csus9 7	Cm 3	Cm.. 3	Cm/A. x	Cm/Bb. 3	D/C x	D/C# x	D/C. x	D/Db. x	D/E 7	D/E. x	D/E.. x	D/E... x
Cm11 3	Cm6. x	Cm7 3	Cmaj x	Cmaj7 x	Cmaj7.. x	Cmaj7... 3	Cmaj9. x	D/E... x	D/E.... x	D/E..... 2	D/E..... 7	D/G x	D/G. 3	D/G.. x	D/G... x
Cmin 3	C# x	C#(add9) 4	C#+ x	C#4 4	C#5. x	C#7 x	C#7(b5) x	D4 x	D5. 5	D5.. 2	D5/E 7	D6 x	D6. x	D6.. x	D6... x
C#add9 4	C#dim x	C#sus 4	C#sus4 4	C#m x	C#m7 x	C#maj x	C#maj7 x	D6/add9. x	D6/add9.. x	D7 x	D7#9 4	D7(#9) 4	D7.. x	D7... 5	D7sus2 x
C#min x	Dbb5. x	Db x	Db+ x	Db. 4	Db.. x	Db... 4	Db.... 4	D7sus4 x	D7sus4. x	D9 10	D9(#5). x	D9(add6) 10	D9. x	D9.. x	D9... 5
Db/B. x	Db/Bb. x	Db/C. x	Db/C.. 4	Db5. 4	Db6. x	Db7 x	Db7.. x	D9add6 10	Dadd9 x	Dadd9, x	Dadd9.. 2	Dadd9,,, 7	Dadd9. x	Dadd9.. x	Dadd9... x

Dadd9...	Daug/E.	Ddim	Ddim/B.	Ddim/B..	Ddim/B...	Ddim/Bb.	Ddim/Bb..	Dm/B	Dm/C	Dm/C#	Dm6(5b)	Dm7	Dm9	Dmaj	Dmaj7
Ddim/C.	Ddim7.	Ddim7..	Ddim7...	Dsus	Dsus.	Dsus..	Dsus...	Dmin	Dm#5	Dm#7	D#	D#+	D#4	D#7	D#dim
Dsus2	Dsus2.	Dsus2..	Dsus2...	Dsus2/Ab.	Dsus2/B.	Dsus2/B..	D#sus	D#sus4	D#m	D#m7	D#maj	D#maj7	D#min	Eb	
Dsus2/Bb.	Dsus2/C.	Dsus2/C..	Dsus2/Db.	Dsus2/Db.	Dsus2/Db..	Dsus2/F.	Dsus2/G.	Eb(add9)	Eb+	Eb.	Eb..	Eb...	Eb/C.	Eb/D.	Eb/Db.
Dsus2/G..	Dsus2/G.	Dsus2/G..	Dsus2/G...	Dsus2/Gb.	Dsus2/Gb..	Dsus2/Gb.	Dsus2/Gb..	Eb/Db..	Eb/Db...	Eb/E.	Eb4	Eb5.	Eb6.	Eb7	Eb7.
Dsus4	Dsus4/B.	Dsus4/B..	Dsus4/C.	Dsus4/E.	Dsus4/E.	Dsus4/E..	Dsus4/E...	Eb7..	Eb7...	Ebadd9	Ebaug/E.	Ebaug/E..	Ebdim	Ebdim/B.	Ebdim/B..
Dsus4/Gb.	Dsus4/Gb..	Dm	Dm(#5)	Dm(#7)	Dm..	Dm/A	Dm/B	Ebdim/B...	Ebdim/C.	Ebdim7.	Ebsus	Ebsus2/Ab.	Ebsus4	Ebsus4/F.	Ebm
Dm/B.	Dm/B..	Dm/B...	Dm/Bb.	Dm/C	Dm/C#	Dm/C.	Dm/C..	Ebm.	Ebm/Db.	Ebm7	Ebm7.	Ebmaj	Ebmaj7	Ebmaj7.	Ebmin
Dm/C...	Dm/Db.	Dm/E.	Dm6(5b)	Dm6.	Dm6..	Dm7	Dm7(b5).	Eb#5.	Eb#5..	E	E+	E.	E..	E/A.	E/D.
Dm7.	Dm7...	Dm7/add11.	Dm9	Dmaj	Dmaj7	Dmaj7.	Dmin	E/D..	E/D...	E/D...	E/Db.	E/Db..	E/Eb.	E/Eb..	E/Eb...
Dmin/maj7.	Dm#5	Dm#7	D#	D#+	D#4	D#5.	D#7	E/Gb.	E/Gb..	E/Gb...	E11	E11/b9.	E5	E5.	E5..
D#dim	D#sus	D#sus4	D#m	D#m7	D#maj	D#maj7	D#min	E6	E6.	E6..	E7	E7#9	E7(#9)	E7(11)	E7(5b)

E7(b9)	E7.	E7..	E7...	E7....	E7/add11.	E7/b9(b5).	E7b9	Em7..	Em7...	Em7....	Em7.....	Em7/D	Em7/add11.	Em9.	Em9..
E7sus4.	E9	E9.	E9..	Eadd9.	Eadd9..	Eadd9...	Edim	Em9...	Emadd9	Emaj	Emaj7	Emaj7.	Emaj7..	Emaj7...	Emaj9.
Edim/C.	Edim/D.	Edim/Db.	Edim/Db..	Edim/Eb.	Edim7.	Edim7..	Esus	Emaj9..	Emin	Emin/maj7.	Emin/maj9.	Emsus4	E#5.	F	F(add9)
Esus.	Esus..	Esus.	Esus..	Esus...	Esus2.	Esus2..	Esus2/A.	F+	F+7+11	F.	F..	F...	F...	F/A	F/C
Esus2/A..	Esus2/Ab.	Esus2/Ab..	Esus2/Db.	Esus2/Db..	Esus2/Eb..	Esus2/G.	Esus2/G..	F/D	F/D.	F/D..	F/D...	F/E.	F/E..	F/E...	F/E....
Esus2/G...	Esus4	Esus4/Ab.	Esus4/C.	Esus4/C..	Esus4/D.	Esus4/D..	Esus4/Db.	F/E....	F/Eb.	F/Eb..	F/G	F/G.	F/G..	F11	F4
Esus4/Db..	Esus4/Eb.	Esus4/F.	Esus4/G.	Esus4/G..	Esus4/G...	Esus4/Gb.	Esus4/Gb..	F5.	F6	F6.	F6..	F6...	F6/add9.	F7	F7.
Em	Em(add9)	Em(sus4)	Em.	Em..	Em...	Em/A.	Em/A..	F7..	F7/A	F9	Fadd9	Fadd9.	Fadd9..	FaddG	Faug/D.
Em/A...	Em/B	Em/C.	Em/C..	Em/C...	Em/D	Em/D.	Em/D..	Faug/G.	Fdim	Fdim/D.	Fdim/D..	Fdim/D...	Fdim/Db.	Fdim7.	Fdim7..
Em/D...	Em/D..	Em/D...	Em/D...	Em/D.	Em/D..	Em/D...	Em/Db.	Fdim7...	Fsus	Fsus.	Fsus2.	Fsus2..	Fsus2/A.	Fsus2/A..	Fsus2/B.
Em/Eb.	Em/Eb..	Em/Gb.	Em/Gb.	Em/Gb..	Em/Gb...	Em6	Em6.	Fsus2/Bb.	Fsus2/D.	Fsus2/E.	Fsus2/E..	Fsus4	Fsus4/G.	Fm	Fm.
Em7	Em7(b5).	Em7.	Em7..	Em7...	Em7....	Em7.....	Em7.	Fm..	Fm/D.	Fm/Db.	Fm/Db..	Fm/Eb.	Fm/Eb..	Fm6	Fm6.

Fm7	Fm7.	Fm7..	Fmaj	Fmaj7	Fmaj7(+5)	Fmaj7,	Fmaj7,,	G(add9)	G+	G,	G,,	G,,,	G.	G..	G...
Fmaj7.	Fmaj7..	Fmaj7...	Fmaj7/#11.	Fmaj7/A	Fmaj7/C	Fmaj9.	Fmin	G....	G.....	G/A	G/A.	G/A..	G/B	G/C.	G/C..
Fmmaj7	F#	F#+	F#/E	F#11	F#4	F#5.	F#5..	G/D	G/E,	G/E,,	G/E,,,	G/E....	G/E..	G/E...	G/E...
F#7	F#9	F#dim	F#sus	F#sus4	F#m	F#m/C#m	F#m6	G/E....	G/E.....	G/F#	G/F.	G/F..	G/F...	G/Gb.	G/Gb..
F#m7	F#m7-5	F#maj	F#maj7	F#min	Gb	Gb+	Gb.	G/Gb...	G/Gb....	G11	G4	G5.	G5..	G6	G6(sus4)
Gb..	Gb...	Gb/Ab.	Gb/E.	Gb/E..	Gb/Eb.	Gb/F.	Gb6.	G6,	G6,,	G6,,,	G6,,,,	G6....	G6..	G6...	G6...
Gb7	Gb7(#5).	Gb7.	Gb7..	Gb7/#9.	Gb7sus4.	Gb9	Gbadd9.	G6....	G6/add9.	G6/add9..	G6/add9...	G6sus4	G7	G7#9	G7(#9)
Gbaug/E.	Gbdim	Gbdim/D.	Gbdim/D..	Gbdim/D...	Gbdim/E.	Gbdim/E..	Gbdim/Eb.	G7(b9)	G7(sus4)	G7+	G7.	G7..	G7...	G7/add11.	G7b9
Gbdim7.	Gbsus	Gbsus.	Gbsus2/Bb.	Gbsus4	Gbsus4/E.	Gbm	Gbm.	G7sus4	G7sus4.	G9	G9(11)	G9.	G9..	Gadd9	Gadd9.
Gbm..	Gbm...	Gbm/D.	Gbm/D..	Gbm/E.	Gbm/E..	Gbm/E.	Gbm/E..	Gadd9.	Gaug/E.	Gaug/E..	Gdim	Gdim/E.	Gdim/E..	Gdim/Eb.	Gdim/Eb..
Gbm/E...	Gbm7	Gbm7(b5).	Gbm7(b5)..	Gbm7.	Gbm7.	Gbm7..	Gbm7..	Gdim/Eb...	Gdim7.	Gdim7..	Gsus	Gsus.	Gsus..	Gsus...	Gsus2.
Gbm7...	Gbm7/b9.	Gbmaj	Gbmaj7	Gbmaj7.	Gbmin	Gb#5.	G	Gsus2..	Gsus2.	Gsus2..	Gsus2/B.	Gsus2/B..	Gsus2/C.	Gsus2/E.	Gsus2/E..



Chords in the key of:-

Ab or G#	Ab	Ebm	Cm	Db	Eb	Fm	Gdim
A	A	Bm	Dm	D	E	Gm	Abdim
Bb or A#	Bb	Cm	Dm	Eb	F	Gm	Adim
B	B	Dbm	Ebm	E	Gb	Abm	Bbdim
C	C	Dm	Em	F	G	Am	Bdim
Db or C#	Db	Ebm	Fm	Gb	Ab	Bbm	Cdim
D	D	Em	Gbm	G	A	Bm	Dbdim
Eb or D#	Eb	Fm	Gm	Ab	Bb	Cm	Ddim
E	E	Gbm	Abm	A	B	Dbm	Ebdim
F	F	Gm	Am	Eb	C	Dm	Edim
Gb or F#	Gb	Abm	Bbm	B	Db	Ebm	Fdim
G	G	Am	Bm	C	D	Em	Gbdim

_aug	=	_#5	(Augmented triad)
_(no 3rd)	=	_5	(Root and 5th - Power chord)
_6/7sus4	=	_6/7sus	(Sus4 triad plus 6th, minor 7th)
_dom7	=	_7	(Major triad, minor 7th)
_7/add11	=	_7/11	(Minor triad, minor 7th, plus 11th)
_2	=	_add9	(Major triad, plus 9th)
_sus	=	_sus4	(No 3rd but a 4th from a major triad)
_#7	=	_maj7	(Major triad, major 7th)
_sus2	=	_add9(no 3rd)	(No 3rd but a 2th from a major triad)
_6/9	=	_6/add9	(Major triad plus 6th and 9th)
_9(#7)	=	_maj9	(Major triad, major 9th plus 7th)
_o7	=	_m7(b5)	(Diminished triad, minor 7th : half-diminished)
_dom9	=	_9	(Major triad, minor 7th plus 9th)

C/D is an C chord with a D bass note (xx0010).

A/B is an A chord with a B bass note (x22220).

G/D is an G chord with a D bass note (xx0003). And so on.

If you are playing in a band with a bass player you don't need to play the bass note, just play it as a normal chord (eg. C/D = C and so on).

THE ULTIMATE GUITAR CHORD CHART II

Special thanks to Martin.Leclerc@Sun.com and Mario.Dorion@Sun.com for the use of their Postscript code and thanks to all those people that responded to my initial postings for guitar chords. Please check chords and notify me of any errors via email at:- Pfacol@elaine.ee.und.ac.za

Remember: Make as many copies as you want, pass it on to your friends.

Références

- BERGMAN, A. (1990). *Auditory scene analysis*. MIT Press.
- B.G.QUINN, E. (2001). *The estimation and tracking of frequency*. volume Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.
- BOUILLAUD, M., GOGNON, G. (1997). *Estimation de partiels de guitare par méthode paramétrique haute résolution : Matrix-pencil*. Mémoire de maîtrise, ATS.
- BROWN, G., COOKE, M. (1994). *Perceptual grouping of musical sounds : A computational model*. Journal of New Music Research, volume 23, pages 107–132.
- DE CHEVEIGNÉ, A. (June 1993). *Separation of concurrent harmonic sounds : Fundamental frequency estimation and a time-domain cancellation and a time-domain cancellation model of auditory processing*. Journal of Acoustical Society of America, volume 93, numéro 6, pages 3271–3289.
- DOVAL, B. (1994). *Estimation de la fréquence fondamentale de signaux sonores*. Thèse de doctorat, Université de Paris VI.
- DOVAL, B., RODET, X. (91). *Estimation of fundamental frequency of musical sound signals*. Proc IEEE-ICASSP, Toronto, pages 3657–3660.
- ERONEN, A., KLAPURI, A. (2000). *Musical instrument recognition using cepstral coefficients and temporal features*. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2000.
- FACOLINE, P. J. *The ultimate guitar chord chart ii*. Annexe 2.
- FLETECHER, N. H. (1991). *The physics of musical instruments*. Springer-Verlag.
- HERMES, D. (Jan 1988). *Measurement of pitch by subharmonic summation*. Journal of Acoustical Society of America, volume 83, numéro 1, pages 257–264.
- HESS, W. (1983). *Pitch determination of speech signals*. Springer-Verlag.
- HUA, Y., SARKAR, T. K. (Mai 1990). *Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise*. IEEE Transactions on Acoustics, Speech, And Signal Processing, volume 38, numéro 5, pages 814–824.
- JEHAN, T. (1997). *Musical signal parameter estimation*. Mémoire de maîtrise, IFSIC - CNMAT.
- J.F.KAISER (1993). *Some useful properties of teager's energy operator*. IEEE International Conference on Acoustics, Speech and Signal Processing, Minneapolis, MN, USA, volume 3, pages 149–152.
- J.F.KAISER, R., T.F.QUATERI (1993). *Detection of transient signals using the energy operator*. IEEE International Conference on Acoustics, Speech and Signal Processing, Minneapolis, MN, USA, volume 3, pages 145–148.
- JUDITH.C.BROWN (September 1992). *Musical fundamental frequency tracking using a pattern recognition method*. Journal of Acoustical Society of America, volume 92, numéro 3, pages 1394–1402.
- KARJALAINEN, M., TOLONEN, T. (March 15-19 1999). *Multi-pitch and periodicity analysis model for sound separation and auditory scene analysis*. Proc. IEEE Int. Conf.

- Acoustics, Speech, and Signal Processing (ICASSP'99), Phoenix, Arizona, volume 2, pages 929–932.
- KASHINO, K., TANAKA, H. (1993). *A sound source separation system with the ability of automatic tone modelling*. Proceedings of the International Computer Music Conference.
- KLAPURI, A. (1998). *Automatic transcription of music*. Mémoire de maîtrise, Tampere University of Technology.
- KLAPURI, A. (Oct. 17-20 1999). *Pitch estimation using multiple independent time-frequency windows*. Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, pages 115–118.
- KLAPURI, A. (2001). *Multipitch estimation and source separation by the spectral smoothness principle*. ICASSP 2001, Salt Lake City, USA, 2001.
- KLAPURI, A., VIRTANEN, T., HOLN, J.-M. (2000). *Robust multipitch estimation for the analysis and manipulation of polyphonic music signals*. In Proc. COST-G6 Conference on Digital Audio Effects, DAFx-00, Verona, Italy, 2000.
- KOOTSOKOS, P. (1997). *A review of the frequency estimation and tracking problems*.
- KUNIEDA, SHIMAMURA, S. (1996). *Robust method of measurement of fundamental frequency by a cross-autocorrelation of log spectrum*. IEEE Trans. on Acoustics, Speech and Signal Processing.
- LAHAT, NIEDERJOHN, K. (June 1987). *Spectral autocorrelation method for measurement of the fundamental frequency of noise-corrupted speech*. IEEE Trans. on Acoustics, Speech and Signal Processing.
- MARTIN, K. (1996). *Automatic transcription of simple polyphonic music : Robust front end processing*. MIT Media Laboratory Perceptual Computing Section Technical Report 399.
- MILLER (1975). *Pitch detection by data reduction*. IEEE Trans. on Acoustics, Speech and Signal Processing.
- MOORER (November 1977). *On the transcription of musical sound by computer*. Computer Music Journal.
- QUINN, B. (March 1997). *Estimation of frequency, amplitude and phase from dft of a time series*. IEEE Transactions on Signal Processing, volume 45, pages 814–817.
- QUIROS, F. J. C., ENRIQUEZ, P. F.-C. (1994). *Real-time, loose-harmonic matching fundamental frequency estimation for musical signals*. International Conference for Acoustics, Speech and Signal Processing, Adelaide, volume 2, pages 221–224.
- RABINER, L., CHENG, M., ROSENBERG, A., MCGONEGAL, C. (October 1976). *A comparative performance study of several pitch detection algorithms*. IEEE Transactions on Acoustics, Speech and Signal Processing, volume ASSP-24, numéro 5, pages 399–418.
- R.MEDDIS, M.HEWITT (June 1991). *Virtual pitch and phase sensitivity of a computer model of the auditory periphery*. Journal of Acoustical Society of America, volume 89, pages 2866–2882.
- ROSENTHAL, D., OKUNO, H. (1998). *Computational auditory scene analysis*. Lawrence Erlbaum Associates, NJ.

- SCHWARZ, D. (1998). *Spectral envelopes in sound analysis and synthesis*. Mémoire de maîtrise, University of Stuttgart - Ircam.
- SHIRADO, T., YANAGIDA, M. (Mai 1997). *Extraction of fundamental frequencies from duet sounds*. IEICE Transactions Fundamental, volume E80-A, numéro 5, pages 912–915.
- TOLONEN, T. (1998). *Model-based analysis and resynthesis of acoustic guitar tones*. Mémoire de maîtrise, Helsinki University of Technology.
- TOLONEN, T. (November 2000a). *A computationally efficient multipitch analysis model*. IEEE Transactions on Speech And Audio Processing, volume 8, numéro 6.
- TOLONEN, T. (Septembre 2000b). *Object-based sound source modeling for musical signals*. 109th Convention of the AES, California, USA.
- U.K.LAINE, M., T.ALTOAAR (1994). *Warped linear prediction (wlp) in speech and audio processing*. IEEE International Conference on Acoustics, Speech and Signal Processing, volume III, pages 349–352.
- VIRTANEN, T., KLAPURI, A. (2000). *Separation of harmonic sound sources using sinusoidal modeling*. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2000.
- VIRTANEN, T., KLAPURI, A. (2001). *Separation of harmonic sounds using multipitch analysis and iterative parameter estimation*. Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York.
- WISE, J. D., CARPIO, J. R., PARKS, T. W. (Octobre 1976). *Maximum likelihood pitch estimation*. IEEE Transactions on Acoustics, Speech and Signal Processing, volume ASSP-24, numéro 5, pages 418–423.