ÉVALUATION DES ALGORITHMES D'ESTIMATION DE LA FRÉQUENCE FONDAMENTALE DANS LE CADRE DE SIGNAUX MUSICAUX MONOPHONIQUES

Rapport de stage en vue de l'obtention du grade
de Master MIS SDI 1ère année
à l'Université Pierre et Marie Curie

Nicolas OBIN

Sous la direction de
Adrian FREED, responsable de la recherche scientifique au CNMAT, BERKELEY
Xavier RODET, responsable de l'équipe Analyse/Synthèse, IRCAM

IRCAM, 1, place Stravinsky 75004 Paris

Juillet 2005

Etudiant

Nicolas OBIN
nobin@ircam.fr

Lieu

CNMAT
Berkeley - Université de Californie

Responsables

CNMAT - Adrian Freed, adrian@cnmat.berkeley.edu
IRCAM - Xavier Rodet, rod@ircam.fr

Mes remerciements vont à :

CNMAT : Adrian Freed, David Wessel, Sylvain Le Groux, Matt Wrigth, Richard Andrews et Gillian Edgelow.
IRCAM : Xavier Rodet, Axel Roebel, Chungsin Yeh
UPF : Emilia Gomez et Xavier Serra
et Alexander Galembo, Saurabh Sood et Boris Doval.

pour leur participation, leur aide, et leur soutien durant ce stage.

David Wessel, directeur du CNMAT, professeur de musique, compositeur

A Katia...

Table des matières

Table des figures

Liste des tableaux

Résumé

Notre rapport traite du sujet de l'estimation de la fréquence fondamentale dans le cas de signaux musicaux monophoniques. Nous y proposons une nouvelle méthode d'estimation pour des signaux pseudo harmoniques basée sur le modèle sous harmonique de Terhardt ainsi que sur une méthode d'optimisation des moindres carrés avec convergence de Newton. Nous avons testé notre algorithme sur une large base de données de piano, et avons obtenu des résultats prometteurs. Par ailleurs nous sommes à l'origine de l'évaluation de plusieurs algorithmes d'estimation de la fréquence fondamentale sur une large base de données monophonique, évaluation qui a mis en valeur les performances de l'algorithme YIN.

Présentation du stage

Le CNMAT - Université de Berkeley, Californie

Le CNMAT, *Center for New Music and Audio Technology*, est un laboratoire faisant partie de l'université de Berkeley en Californie. Ses champs de recherches se trouvent dans le domaines des technologies audio, ainsi que le développement d'outils pour la création musicale contemporaine. Son directeur en est David Wessel, et son directeur de la recherche scientifique Adrian Freed. Les champs de recherche du CNMAT sont multiples, et nous pouvons y trouver autant des axes de recherche scientifiques tels que l'estimation de la f_0, la spatialisation du son, le rythme, le format SDIF, que de l'application industrielle : collaboration avec le fabriquant de guitare Gibson USA, le développement d'outils pour la création musicale contemporaine avec Max/Msp et la création musicale avec des travaux avec Gyorgy Lygeti, Steeve Coleman et des étudiants de la filière musicale de l'université de Berkeley. L'estimation de la f_0 est un sujet de recherche continu au CNMAT, qui a déjà développé par le passé des algorithmes d'estimation tels que *Additive*, les travaux de Tristan Jehan, et les recherches menées actuellement sur l'estimation de f_0 multiples.

L'IRCAM

Fondé en 1974 à l'initiative du président de la République Française d'alors Georges Pompidou, et présidé (aujourd'hui présidence d'honneur) par le compositeur Pierre Boulez, l'IRCAM (*Institut de Recherche et de Coordination Acoustique Musique*) a pour objectif de faire collaborer compositeurs et scientifiques dans le développement de nouvelles techniques de d'analyse, de modification et de synthèse, de production, de transmission et de perception du son. Il regroupe de nombreux pôles de recherche dans des domaines regroupés autour de la musique comme la psycho- acoustique, l'acoustique des instruments, l'acoustique des salles, l'analyse, transformation et la synthèse du son et des applications temps réels, jusqu'à la production musicale. L'IRCAM constitue en soi une «exception culturelle», unique à travers le monde, que ce soit dans la production musicale, ou sur l'ampleur des moyens mis en oeuvre pour la recherche scientifique.

Sujet du Stage

Notre stage se présente comme une collaboration entre l'équipe analyse-synthèse de l'IRCAM sous la direction de Xavier Rodet et le CNMAT de l'Université de Berkeley sous la direction d'Adrian Freed, directeur de la recherche scientifique. Le sujet en était à l'origine l'évaluation des algorithmes d'estimation de la fréquence fondamentale dans le cas de signaux audio musicaux monophoniques. Cependant nous avons été amené au cours de notre stage à une redéfinition des objectifs premiers, puisqu'il nous a été demandé de développer également un nouvel algorithme d'estimation de la f_0 dans le cas des instruments pseudo-harmoniques tels que le piano ou les instruments à cordes pincées. Ce stage constitue un premier pas dans le développement à plus long terme d'une estimation générale des algorithmes, ainsi que dans le développement d'une série d'algorithmes correspondant aux principaux comportements physiques rencontrés dans l'application musicale, afin d'avoir un

ensemble d'algorithmes permettant une meilleure analyse de chaque type instrumental. La répartition du stage s'est répartie comme suit :
- Premier mois : mise à jour des connaissances scientifiques sur le problème de l'estimation de la f_0.
- Deuxième mois : développement d'un nouvel algorithme dans le cas pseudo-harmonique.
- Troisième mois : constitution d'une base de données audio de sons monodiques, et évaluation des algorithmes d'estimation de la f_0 .

Introduction

Depuis maintenant plus de 50 ans, l'estimation de la fréquence fondamentale du signal audio a été l'objet de multiples recherches et d'un constant défi pour le chercheur. Les enjeux de cette estimation sont d'importance, car ils regroupent des domaine aussi variés que :
- le domaine industriel : application à la reconnaissance vocale, à la synthèse vocale, et aux télécommunications. Il a notamment été montré que la fréquence fondamentale ainsi que ses variations au cours du temps sont des caractéristiques importantes permettant de transmettre les émotions véhiculées par la voix. Ceci est évidemment important pour faire de la synthèse et de la transmission plus réaliste de la voix.
- le domaine médical : de même la fréquence fondamentale est un critère permettant de déterminer certaines pathologies comme les tendances suicidaires par exemple.
- le domaine musical : la détermination de la fréquence fondamentale serait très utile notamment à la transcription automatique de partition dans le champ de la musicologie et de l'ethno-musicologie, le suivi de partition en temps réel ainsi que pour la synthèse de sons en temps réel à partir de l'analyse de la fréquence fondamentale.

L'histoire de la détermination de la fréquence fondamentale montre l'importance du sujet, puisque plusieurs centaines d'algorithmes et autant de méthodes ont vus le jour, depuis la simple fonction d'autocorrélation jusqu'aux chaînes de Markov. Pour faire simple dans le domaine qui nous concerne, c'est-à-dire celui de l'application musicale, on décompose généralement le signal en 3 composantes fondamentales, à savoir le timbre (qui se rapporte directement à l'instrument), la durée d'une note, et enfin la hauteur de la note. C'est cette dernière caractéristique du signal que nous proposons d'étudier dans notre présent rapport. Notre rapport s'articulera en trois partie de la manière suivante : Dans une première partie nous nous intéresserons à définir la fréquence fondamentale dans le cadre d'une application musicale, et étudierons ses spécifités à travers l'étude modale des instruments de musiques. Dans une seconde partie, nous expliquerons les raisons qui nous ont mené au développement d'un nouvel algorithme, et présenterons celui-ci. Nous montrerons notamment quelques résultats très encourageants. Enfin, dans une dernière partie, nous ferons la présentation des algorithmes que nous avons retenu pour notre évaluation, présenterons notre base de données, et expliciterons notre méthode d'évaluation, et enfin de présenter nos résultats. Nous conclurons enfin sur l'ensemble de notre travail, et ouvrirons des perspectives en vue de travaux futurs.

Chapitre 1

Introduction au problème de la détermination de la fréquence fondamentale

Sommaire

1.1 «Pitch» et fréquence fondamentale

Avant toute chose, il est important de résoudre un point important et de lever une ambiguïté que nous avons souvent rencontré au cours de notre stage : pitch et f_0, est-ce la même chose, et si non, quelles sont les différences ? L'ambiguïté vient essentiellement du fait qu'en anglais, on utilise le même vocable pour deux choses différentes : pitch signifie à la fois la hauteur de note perçue et le fréquence fondamentale. Cependant ces deux choses ne sont pas du tout équivalentes dans la plupart des cas : Terhardt [61] et à sa suite Daï [12] ont développé chacun un modèle perceptif de la hauteur de note, le premier soutenant que la hauteur de note perçue est déterminée par une somme pondérée des premiers partiels , le second privilégiant un seul partiel dit dominant dans une zone de référence. Quoi qu'il en soit, ces modèles montrent que la hauteur de note, ou modèle perceptif, ne correspond généralement pas à la fréquence fondamentale du signal. En l'occurrence et pour donner un exemple, une fréquence fondamentale peut posséder de nombreux pitch différents en agissant sur l'écart entre les partiels du signal, comme l'a montré Jarvelainen [30][31].

1.2 Définition de la fréquence fondamentale dans le cadre d'une application musicale : Le modèle harmoniques plus bruit

Dans un premier temps, nous nous proposons de recadrer le problème de l'évaluation de la fréquence fondamentale, en définissant la fréquence fondamentale (où f_0).

1.2.1 Modèle à une seule sinusoïde

Commençons par un exemple trivial, en considérant la fonction :

$$s(t_k) = A.\sin(\omega_0 t_k) \tag{1.1}$$

Il est ici évident que cette fonction est T_0 - périodique et donc de fréquence fondamentale (définie par l'inverse de la période fondamentale)

$$f_0 = \frac{1}{T_0} \tag{1.2}$$

1.2.2 Modèle stationnaire harmonique

Dans le cas d'un signal stationnaire harmonique, nous savons que le signal temporel peut s'écrire comme une somme de sinusoïde : c'est le modèle sinusoïdal.

$$s(t_k) = \sum_{n=0}^{N} A_n \sin(n\omega_0 t_k + \phi_n) + b(t_k) \tag{1.3}$$

$$\text{où} \begin{cases} \omega_0 \text{ est la pulsation du signal à l'instant } t_k \\ A_n \text{ est l'amplitude du partiel n} \\ \phi_n \text{ est la phase du partiel n} \end{cases}$$

Dans ce modèle nous avons encore la fréquence fondamentale f_0 définie comme l'inverse de la période fondamentale T_0 du signal, si l'on fait abstraction du bruit.

1.2.3 Modèle localement stationnaire harmonique

Généralement c'est ce modèle qui est utilisé pour modéliser le comportement des signaux musicaux :

Les propriétés du signal musical sont en effet les suivantes :
- localement stationnaire
- pseudo périodique

$$s(t_k) = \sum_{n=0}^{N} A_n(t_k) \sin(n\omega_0(t_k)t_k + \phi_n(t_k)) + b(t_k) \tag{1.4}$$

1.2.4 Modèle localement stationnaire non périodique

Dès lors que les «modes» du signal ne sont plus dans le rapport simple $\omega_k = k\omega_0$, c'est-à-dire lorsque les modes ne sont pas régulièrement espacés, le signal n'est alors plus périodique (nous renvoyons à l'annexe pour une démonstration complète). Quel sens prend alors le terme de période fondamentale, puisque le motif ne répète plus à kT_0 ? Il n'existe alors pas de definition unitaire, mais on peut constater deux tendances :
- une tendance mathématique, qui consiste à minimiser l' «apériodicité» du signal, c'est-à-dire la différence entre une série harmonique et les partiels présent dans le signal. C'est le modèle utilisé dans la majorité des cas non périodique, comme les signaux théoriquement harmonique, mais non réelement, ainsi que les signaux inharmoniques.
- une tendance physique, qui consiste à considérer comme fréquence fondamentale la fréquence du mode dit fondamental. C'est le cas des signaux pseudo - harmoniques, comme le piano par exemple. Dans ce cas, l'application d'un modèle mathématique conduirait nécessairement à une erreur trop haute.

En un mot, il n'existe pas aujourd'hui de définition unitaire de la fréquence fondamentale, ce qui constitue un problème majeur ; problème complexe de considérations à la fois psycho - acoustique (il faut que la fréquence fondamentale corresponde à une note effectivement entendue : une fréquence fondamentale à 330 Hz pour un son entendu à 500 n'aurait pas de sens), mathématique et physique.

1.3 Problèmes posés par la réalité du signal musical : étude de la physique des instruments

Avant de nous attacher à la constitution d'un nouvel algorithme ainsi que l'étude des algorithmes existants, nous voudrions faire un détour par la réalité du signal musical et dresser une étude de la physique des instruments de musiques, montrer les spécificités qui leur sont liés, ainsi que les problématiques qu'ils posent pour l'estimation de la fréquence fondamentale. Dans une première partie nous allons développer une étude modale détaillée de la physique des instruments, avant de résumer les propriétés dans un second temps. Ensuite nous nous proposons de faire un détour par la considération de l'ensemble des problèmes rencontrés par l'estimation de la fréquence fondamentale.

1.3.1 Étude modale de la physique des instruments

1.3.1.1 Vibrations transverses d'une corde de raideur K fixée à ses deux extrémités

On considère une corde de longueur L, de densité linéïque, de section S, de module d'Young E, de raideur K et de tension appliquée T.

L'équation du déplacement transverse de la corde s'écrit alors :

$$\rho\frac{\partial^2 y}{\partial t^2} = T\frac{\partial^2 y}{\partial x^2} - ESK\frac{\partial^4 y}{\partial x^4} \tag{1.5}$$

En considérant comme condition limite que la corde est fixée à ses deux extrémités, on obtient alors les modes propres :

$$f_n = nf_0^1[1+\beta+\beta^2+\frac{n^2\pi^2}{8}\beta^2] \tag{1.6}$$

$$\text{où}\begin{cases} f_0^1 \text{ est la fréquence fondamentale de la corde pour } K=0 \\ \beta = 2\frac{K}{L}\sqrt{\frac{\pi E}{T}} \text{ est le facteur d'inharmonicité} \end{cases}$$

Comme nous pouvons le voir dans la relation précédente, l'introduction de la raideur de la corde dans le modèle est à l'origine de l'inharmonicité des modes.

1.3.1.2 Les pavillons

De manière générale on peut écrire l'équation de propagation dans un pavillon de la manière suivante :

$$\bigtriangledown^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0 \tag{1.7}$$

Pavillons coniques et cylindriques

L'équation de la pression acoustique dans un pavillon est de la forme :

$$p(x) = Ax^{\epsilon+\frac{1}{2}}J_{\epsilon+\frac{1}{2}}(kx) \tag{1.8}$$

pour laquelle $J_\nu(z)$ est la fonction de Bessel d'ordre ν.

La condition aux limites en x = L (bords ouvert, pression acoustique nulle) donne alors les modes propres :

$$J_{-\epsilon+\frac{1}{2}}(\frac{\omega L}{c}) = 0 \tag{1.9}$$

Étude de l'influence de ϵsur les modes propres :
Si ϵ= 0 : pavillon cylindrique, alors les modes ne comprennent que les modes impairs 1, 3, 5, ...

$$\omega_n = (2n+1)\omega_1 \tag{1.10}$$

$$\text{où}\begin{cases} \omega_1 = \frac{\pi c}{2L'} \\ L' \text{ est la longueur acoustique du pavillon} \end{cases}$$

Si $\epsilon = 2$: pavillon conique, alors la série des modes est entière 1, 2, 3, ...

$$\omega_n = n\omega_1 \tag{1.11}$$

$$\text{où}\begin{cases} \omega_1 = \frac{\pi c}{(L'+x_1)} \\ L' \text{ est la longueur acoustique du pavillon} \\ x_1 \text{ est la distance de l'ouverture au sommet du cône} \end{cases}$$

Pavillons coniques tronqués

La modélisation nous donne alors l'impédance d'entrée, c'est-à-dire l'impédance en x=L :

$$Z_{in} = \frac{j\rho c}{S_1} \frac{\sin(kL)\sin(k\theta_1)}{\sin(k(L+\theta_1))} \tag{1.12}$$

$$\text{où} \begin{cases} L \text{ est la longueur du pavillon} \\ \theta_1 \text{ est l'angle relatif à la distance } x_1 \text{ de l'ouverture du cône} \end{cases}$$

θ_1 est défini de la manière suivante :

$$k\theta_1 = \tan^{-1}(kx_1) \tag{1.13}$$

La condition limite en x=L, c'est-à-dire $Z_{in} = 0$ donne alors les modes propres :

$$\omega_n = n\omega_1 \tag{1.14}$$

$$\text{où} \begin{cases} \omega_1 = \frac{\pi c}{L} \end{cases}$$

Pavillons de Bessel

Dans le cas du pavillon de Bessel, la géométrie du pavillon est décrite de la façon suivante :

$$a = b(x + x_0)^{-\gamma} \tag{1.15}$$

$$\text{où} \begin{cases} a \text{ est inférieur au rayon du pavillon de l'ouverture du pavillon} \\ b \text{ est un facteur adaptatif} \\ x \text{ est la distance entre l'embouchure et l'ouverture du pavillon} \\ x_0 \text{ est la position de l'ouverture du pavillon} \end{cases}$$

On peut alors approximer les modes propres par :

$$f_n \approxeq \frac{c}{4(l+x_0)}(2n-1) + \beta[\gamma(\gamma+1)]^{\frac{1}{2}} \tag{1.16}$$

Dans le cas ou $\gamma = 1$, on peut faire une bonne approximation des modes propres par :

$$f_n \approxeq \frac{nc}{2(l+x_0)} \tag{1.17}$$

1.3.1.3 Les «pipe resonator s»

«Pipe resonator » fermé à une extrémités et ouvert à l'autre

L'impédance d'entrée s'écrit :

$$Z_{in} = -jZ_0 cot(kL) \tag{1.18}$$

$$\text{où} \begin{cases} Z_0 = \frac{\rho c}{S} \\ \rho \text{ est la densité de l'air} \\ c \text{ est la vélocité du son dans l'air} \\ S \text{ est la section du «\textit{pipe resonator} »} \end{cases}$$

Ce qui nous donne, en supposant que que le pipe est ouvert en x=0, la condition suivante :

$$Z_{in} = 0 \tag{1.19}$$

Soit les modes :

$$\omega_n = (2n-1)\omega_1 \tag{1.20}$$

$$\text{où } \omega_1 = \tfrac{\pi c}{2L}$$

« Pipe resonator » ouvert à ses deux extrémités

L'impédance d'entrée s'écrit :

$$Z_{in} = jZ_0 \tan(kL) \tag{1.21}$$

En utilisant encore la condition limite :

$$Z_{in} = 0 \tag{1.22}$$

On en déduit les modes propres :

$$\omega_n = n\omega_1 \tag{1.23}$$

$$\text{où } \omega_1 = \tfrac{\pi c}{L}$$

1.3.1.4 Membranes circulaires

L'équation des déplacements transversaux s'écrit en coordonnées cartésiennes :

$$\frac{\partial^2 z}{\partial t^2} = \frac{T}{\sigma}(\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}) = c^2 \bigtriangledown^2 z \tag{1.24}$$

$$\text{où} \begin{cases} \sigma \text{ est la densité surfacique} \\ c = \sqrt{\frac{T}{\sigma}} \end{cases}$$

Ce qui donne en coordonnées polaires :

$$\frac{\partial^2 z}{\partial t^2} = c^2(\frac{\partial^2 z}{\partial r^2} + \frac{1}{r}\frac{\partial z}{\partial r} + \frac{\partial^2 z}{\partial \phi^2}) \tag{1.25}$$

En posant :

$$z(r,\phi,t) = R(r)\Phi(\phi)e^{j\omega t} \tag{1.26}$$

On obtient l'équation découplée :

$$\begin{cases} \frac{\partial R}{\partial r} + (\frac{\omega^2}{c^2} - \frac{m^2}{r^2}R) = 0 \\ \frac{\partial^2 \Phi}{\partial \phi^2} + m^2\Phi = 0 \end{cases} \tag{1.27}$$

Soit les solutions :

$$\Phi(\phi) = Ae^{\pm jm\phi} \tag{1.28}$$

Et pour la variable r, les fonctions de Bessel d'ordre m $J_m(r)$ avec $m \in [1, \nu]$.
La fréquence $f_{m,n}$ du mode propre (m,n) est alors le n-ième zéro de la fonction de Bessel d'ordre m.

1.3.1.5 Théorie des poutres

Théoriquement les modes propres de barres fines à une dimension peuvent être résolue graphiquement à partir de l'équation.

Ce qui nous donne les modes :

$$f_n = \frac{\pi K}{8L^2}\sqrt{\frac{E}{\rho}}[3.011^2, 5^2, 7^2, ..., (2n+1)^2] \tag{1.29}$$

1.3.2 Application à l'étude modale des instruments de musique

1.3.2.1 Les instruments à cordes

Les instruments à cordes peuvent être classés en trois catégories physiques différentes, entraînant chacune des processus physiques très différenciés :
- Les cordes frottées : violon, alto, violoncelle et contrebasse
- Les cordes pincées : guitare, harpe, clavicorde, koto
- Les cordes frappées : piano

Cependant une seconde classification est possible pour caractériser la physique des instruments à corde, selon que ceux ci sont en régime :
- Oscillations forcées : cordes frottées.
Les modes forces sont alors en rapport harmoniques
- Oscillations libres : cordes pincées,et cordes frappées
Les modes libres sont alors en rapport inharmonique, comme montré dans la partie précédente.

Cette seconde classification nous est apparue plus pertinente dans le cadre de notre étude modale, puisqu'elle fait directement apparaître des différences dans le comportement modale de l'instrument : harmonique - pseudo harmonique.

1.3.2.2 Les cuivres

Les instruments à vent de type trompette, trombone, tuba, cor, etcÉ peuvent être modélisés par l'approximation des Bessel horns qui pour ce qui nous intéresse nous donne les modes propres suivant :

$$f_n = \frac{c}{4(l+x_0)}(2n-1) + \beta[\gamma(\gamma+1)]^{\frac{1}{2}} \tag{1.30}$$

Dans le cas où γ= 1, on peut faire une bonne approximation des modes propres par :

$$f_n = \frac{nc}{2(l+x_0)} \tag{1.31}$$

Ce qui nous donnerait une série harmonique entière.

Pour les instruments tels que la trompette et le trombone, le facteur γ est typiquement de l'ordre de 0.7 et il n'y a alors pas de relation simple entre les modes.

Cependant des ajustements sont réalisés lors de la faction pour accorder les harmoniques. La fréquence fondamentale est basse par rapport aux harmoniques supérieures, donnant une série harmonique approximant typiquement (0.7, 2, 3, ...)

1.3.2.3 Les bois - Les instruments à anche et à embouchure

Nous proposons ci-après une revue des instruments à vent et leur appartenance respective :
- Instruments sans anche : cornet, serpent : pavillon conique. Toutes les harmoniques sont présentes.
- Instruments avec embouchure : on trouve à la fois des pavillon cylindriques et coniques
- La famille des clarinettes : clarinette, chalumeau : pavillon cylindrique. La série des harmoniques est impaires.
- La famille des hautbois, bassons ainsi que la famille des saxophones : pavillon conique. Toutes les harmoniques sont présentes.

La clarinette
Caractérisation physique : pavillon cylindrique, simple anche, trous.
Caractérisation harmonique :
- Faiblesse des harmoniques impaires en général
- Seconde harmonique absente en basses fréquences (D3, D4)
- Seconde harmonique relativement forte en hautes fréquences (G5)
- Richesse harmonique en basses fréquences (25 pour D3)
- Pauvreté harmonique en hautes fréquences : 3 pour G5

Le hautbois
Caractérisation physique : pavillon conique, double hanche, trous.
Caractérisation harmonique :
- Harmoniques complètes
- Faiblesse relative de la fondamentale
- Richesse harmonique

Le basson
Caractérisation physique : pavillon conique, double anches, instrument à clefs.
Caractérisation harmonique :
- Harmoniques complètes
- Fondamentale faible en basses fréquences
- Pauvreté harmonique

Le saxophone
Caractérisation physique : pavillon conique, anche simple, instrument à clefs
Caractérisation harmonique :
- Harmoniques complètes
- Forte fondamentale
- Fréquence de coupure relativement élevée

1.3.2.4 Les flûtes

On discerne dans la famille des flûtes trois types ayant des caractéristiques différentes :
- un corps cylindrique fermé à une extrémité et ouverte à l'autre : la flûte de pan, la flûte traversière et le piccolo
- un corps cylindrique ouvert aux deux extrémités : le shakuhachi

- un corps conique tronqué (non réductible aux simples corps cylindriques) : la flûte à bec

La flûte de pan- La flûte traversière - Le piccolo
La flûte de pan peut être modélisée comme un«*pipe resonator* » ouvert d'un côté et fermé de l'autre, et ne présente donc que des harmoniques d'ordre impaire.

Le shakuhachi
Le shakuhachi peut être modélisé comme un «*pipe resonator* » ouvert à ses deux extremités, et présente donc une structure harmonique complète.

Remarque : Il doit être noté par ailleurs que le shakuachi a un mode de jeu qui lui est caractéristique, à savoir que le soufflé du joueur est un facteur déterminant, introduisant un bruit intrinsèque dans la considération de l'instrument.

La flûte à bec
La flûte à bec est modélisée comme un cône tronqué, et présente donc une structure harmonique complète.

Ces différences physiques conduisent comme nous venons de le voir à des propriétés harmoniques différentes : la plus importante étant l'absence ou, pour se placer dans un contexte plus réaliste, la faiblesse des harmoniques paires en regard des harmoniques impaires, chaque flûte étant approximativement harmonique.

1.3.2.5 Les instruments à percussion

Comme nous allons le voir, les instruments à percussion sont essentiellement inharmoniques. Cependant on peut diviser ces instruments en deux parties distinctes :
- les instruments qui procurent une hauteur même approximative
- ceux qui ne le procurent pas

Nous ne considérerons dans cette partie uniquement les instruments à percussion qui sont à l'origine d'une hauteur même indéterminée, et laisserons de côté ceux qui ne la procurent pas.

D'un point de vue physique on peut dors et déjà constituer des familles :
- les membranes
- simple membrane
- deux membranes couplées à une cavité d'air
- les percussions métalliques (à une ou deux dimensions)

Les membranes

La fréquence $f_{m,n}$ du mode propre (m,n) est alors le n-ième zéro de la fonction de Bessel d'ordre m. Le résultat nous donne des partiels inharmoniques.

Cependant dans le cas d'une membrane réelle, d'autres facteurs doivent être pris en compte, et qui donnent des modes propres éloignés de la solution idéale :
- le chargement de l'air qui a pour effet de diminuer la fréquence propre
- la raideur de la membrane
- la «*stiffness to shear* »
Les deux derniers ayant pour effet de l'accroître. Cependant dans une membrane fine, le facteur dominant est la charge en air.

Timpani, Tabla, Kendang

Caractérisation physique :
- Simple membrane
- Corps hémisphérique

Il est à remarqué que le rang des modes propres calculés donne pour les modes (1,1), (2,1), (3,1), (4,1) et (5,1) des rapports très proches de la série harmonique 2 :3 :4 :5 :6, ce qui confère au timpani une sensation de quasi harmonicité.

Remarque : le pitch du timpani est caractérisé par son mode (1,1), et non pas sa fondamentale manquante.

Dans cette même catégorie, on retrouve :
- timpani (orchestre classique occidental)
- tabla (inde)
- kendang (gamelan indonésien)

Les instruments à percussion métalliques

Théoriquement les modes propres de barres fines à une dimension peuvent être résolue graphiquement et nous donne les modes :

$$f_n = \frac{\pi K}{8L^2}\sqrt{\frac{E}{\rho}}[3.011^2, 5^2, 7^2, ..., (2n + 1^2)] \tag{1.32}$$

Ces modes sont très éloignés d'une série harmonique.

Le glockenspiel

Cependant pour le glockenspiel, la hauteur déjà haute de la fondamentale, soit 2160Hz pour la note la plus basse, fait que l'inharmonicité des partiels importe peu, puisqu'on sait perceptuellement que l'oreille discrimine mal les hautes fréquences.

Marimba, xylophone, gamelan

Pour d'autres instruments pour lesquels la propriété énoncée ci-dessus n'est plus valable, les instruments sont généralement «tunés» pour approximer une série harmonique par exemple par l'ajout de résonateurs à la fréquence fondamentale de la barre, ou alors par le creusement stratégique de la barre en certains de ses points afin de «tuner» les harmoniques vers un rapport approximativement harmonique.

Tuner une poutre

En creusant la poutre à un endroit ou le moment de torsion pour un mode donne est élevé diminue considérablement la fréquence de ce mode.

Dans cette catégorie d'instruments tunés, on peut classer :
- le marimba
- le xylophone
- les instruments à percussion métallique provenant du gamelan

La cloche tubulaire

La cloche tubulaire quant à elle, reste dans le rapport d'inharmonicité présenté pour les vibrations transverses d'une poutre. Cependant la notion de pitch provient ici du rapport quasi-harmonique des partiels de rang 3, 4, et 5. La perception en tant que hauteur déterminée de cet instrument fait alors intervenir des notions de psycho- acoustique : nous percevons alors la fondamentale virtuelle liée à ce rapport harmonique.

Les instruments à percussion métallique utilisant des plaques bi-dimensionnelle

Dans le cas de plaques à deux dimensions comme pour le gong, les fréquences sont tellement inharmoniques que la notion de hauteur perçue disparaît : ces instruments sont dit à hauteur indéterminée, et nous ne nous y intéresserons pas dans le cadre de notre étude.

1.4 Conclusion

L'étude de la physique des instruments et notamment de la distribution des modes propres - absolument nécessaire pour cerner la particularité musicale dans le traitement du signal visant à estimer la fréquence fondamentale de la note jouée - nous amène aux conclusions suivantes :
- Aucun instrument n'est strictement harmonique : dans le meilleur des cas, il présente une bonne approximation.
- Les instruments présentant une hauteur déterminée, que l'on appelle la hauteur de la note, ne présentent pas nécessairement une série harmonique, comme nous l'avons vu pour les cordes et les instruments à percussion, membranes ou plaques métalliques. Ils peuvent alors être ou plus moins inharmoniques : cordes moins, percussions plus.
- La fondamentale peut être fréquentiellement anormalement basse par rapport à la série harmonique normale (les cuivres).
- La fondamentale n'est pas forcement l'harmonique ayant la plus grande amplitude (basson, hautbois)
- Des partiels peuvent manquer (clarinette)
- Le nombre d'harmoniques peut être plus ou moins grand selon les instruments : fréquence de coupure qui dépend de la physique de l'instrument . évidemment pour ces instruments, plus la note sera haute, moins il y aura d'harmoniques.
- Les partiels peuvent avoir des vitesses d'apparition et d'extinction différentes, comme dans le cas des cordes pincées et frappées, ce qui est dû à la dissipation d'énergie dépendante de la fréquence.

L'étude que nous venons de faire nous a conduit naturellement vers une classification des sons d'après la physique spécifique qui leur est liée, l'idée étant de développer des algorithmes pour chacune de ces spécificités, et alors de commuter ces algorithmes en fonction du son.

La classification que nous proposons est la suivante :
- harmonique
- pseudo - harmonique
- non harmonique avec une hauteur de note perceptible
- non-harmonique sans hauteur de note sans hauteur de note perceptible

Ces facteurs devront être pris en compte lors de la réalisation de notre algorithme, car un modèle uniquement harmonique échouerait bien trop souvent.

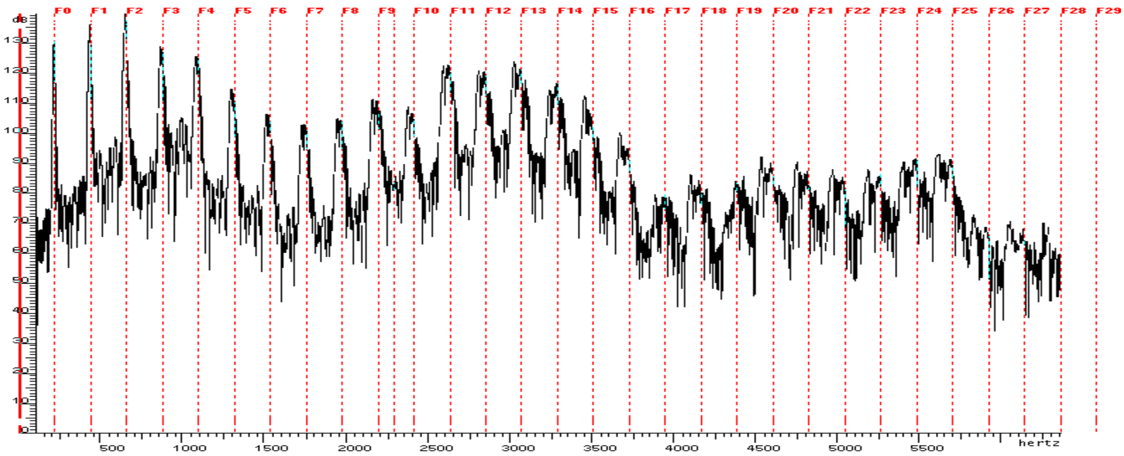

FIG. *1.1:* Spectre d'une voix chantée : les partiels sont harmoniques

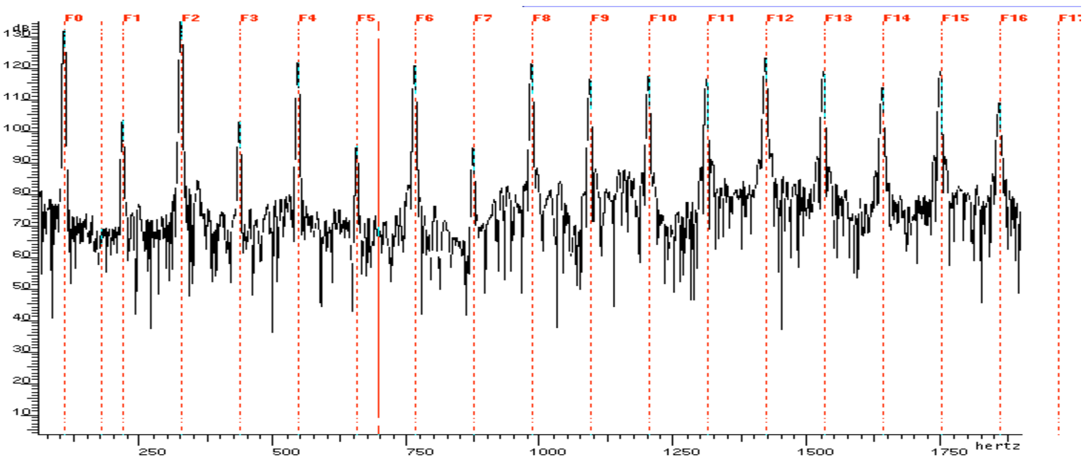

FIG. *1.2:* Spectre d'une clarinette : les partiels pairs sont plus faibles

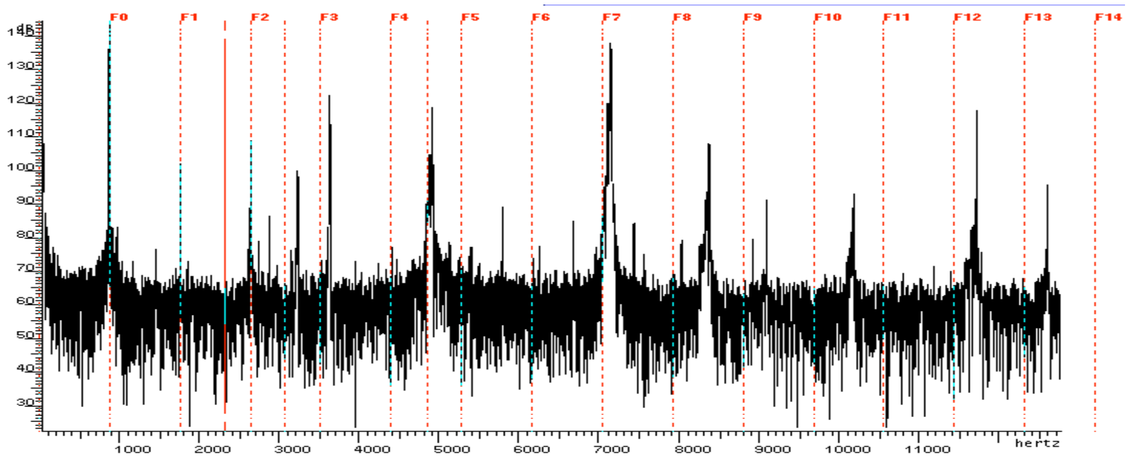

FIG. *1.3:* Spectre d'un Glockenspiel : les partiels sont inharmoniques, mais les partiels 1, 2, 3, 4 et 5 suivent une relation approximativement harmonique

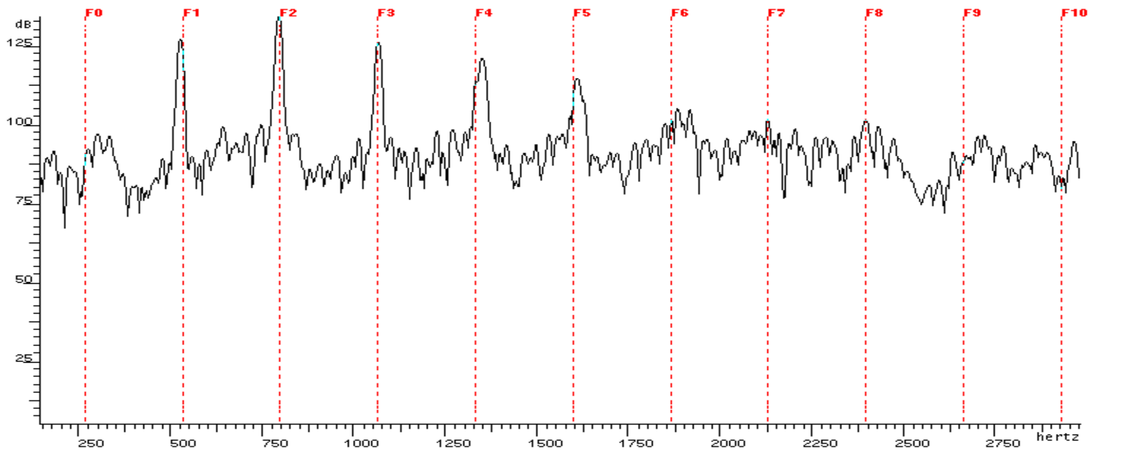

FIG. *1.4:* Spectre d'un tabla (percussion à membrane circulaire). Les partiels déterminants pour la f_0 sont les partiels 1,2,3,4 et 5

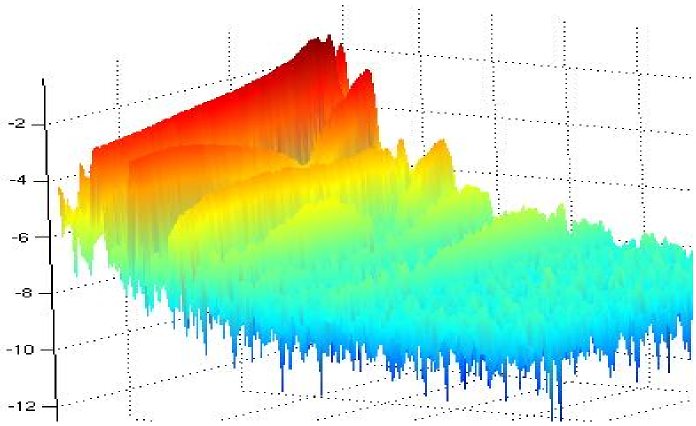

FIG. *1.5:* Spectre d'un piano au cours du temps : les harmoniques ont des temps d'amortissement différents et présentent des battements

Chapitre 2

Développement d'un nouvel algorithme d'estimation de la fréquence fondamentale

Sommaire

2.1 Motivation pour un nouvel algorithme

Comme nous l'avons dit dans le chapitre précédent, l'idée que nous voulions est de développer différents algorithmes en fonction de la physique du son spécifique à analyser. En l'occurrence, les cordes frappées comme dans le cas du piano, ainsi que des cordes pincées, comme dans le cas du clavecin, de la guitare et du violon en mode de jeu pizzicato, ainsi que de leur familles respectives, suivent un régime physique particulier qui est celui de la corde en oscillations libres.

Dans le cas d'une corde en oscillation libre, en tenant compte de la raideur de la corde, nous obtenons la relation modale suivante :

$$f_n = n f_0^1 \sqrt{1 + Bn^2} \tag{2.1}$$

$$\text{où} \begin{cases} B \text{ est le coefficient d'inharmonicité} \\ f_0^1 \text{ est la fréquence fondamentale pour une corde de raideur nulle} \end{cases}$$

Le coefficient d'inharmonicité est défini par .

$$B = 2\frac{K}{L}\sqrt{\frac{ES}{T}} \tag{2.2}$$

$$\text{où} \begin{cases} K \text{ est la raideur de la corde} \\ E \text{ est le module d'Young} \\ L \text{ est la longueur de la corde} \\ T \text{ est la tension appliquée à la corde} \end{cases}$$

Cette relation montre que les harmoniques ne sont plus également espacés, mais ont un écart croissant avec les harmoniques, écart qui dépend de la valeur du coefficient d'inharmonicité.

L'objectif du développement d'un algorithme capable à la fois de déterminer la f_0 et le coefficient d'inharmonicité est double :
- améliorer l'estimation de la f_0 dans le cas pseudo-harmonique L'amélioration elle-même peut se traduire de deux façons différentes : Dans un premier temps la prise en compte d'un modèle correspondant plus étroitement avec la distribution modale de l'instrument permet une meilleure analyse du signal. Dans un second temps, cette méthode peut nous permettre éventuellement d'éviter un certain nombre d'erreur d'octaves : en effet si nous trouvons une dont la valeur du coefficient d'inharmonicité se trouve loin de ce qu'elle devrait être, nous sommes capables alors de savoir que nous sommes dans l'erreur et alors de recommencer l'analyse avec de nouveaux paramètres.
- utilisation comme objet d'analyse : cette méthode nous permet également de déterminer avec

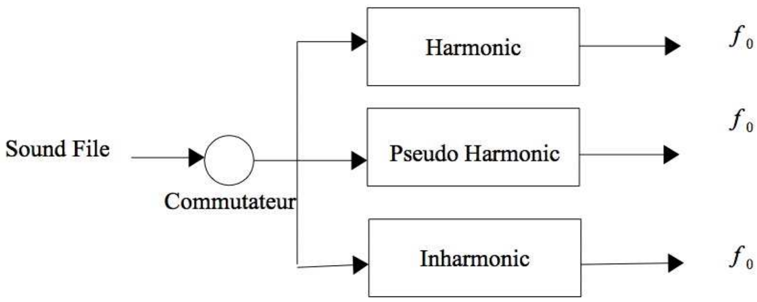

FIG. *2.1:* Proposition d'une méthode d'estimation de la fréquence fondamentale avec prise en compte des caractéristiques physiques

précision le coefficient d'inharmonicité directement à partir du signal, sans aucune autre mesure physique.

Problème de l'estimation de la frequence fondamentale dans le cas pseudo-harmonique avec un modele harmonique

Comme les partiels sont de plus en plus éloignés d'une relation harmonique dans le cas pseudo harmonique, l'application d'un modèle harmonique ou d'une régression harmonique est susceptible d'entraîner un biai systématique trop haut sur l'estimation de la fréquence fondamentale. Cependant, et comme nous le voyons dans les deux exemples suivants, le phénomène de dispersion

- décroissance des modes dependant de la fréquence - conduit à un biai qui décroît asymptotiquement vers la valeur exacte de la fréquence fondamentale quand le partiel considéré pour la régression tend vers le premier partiel.

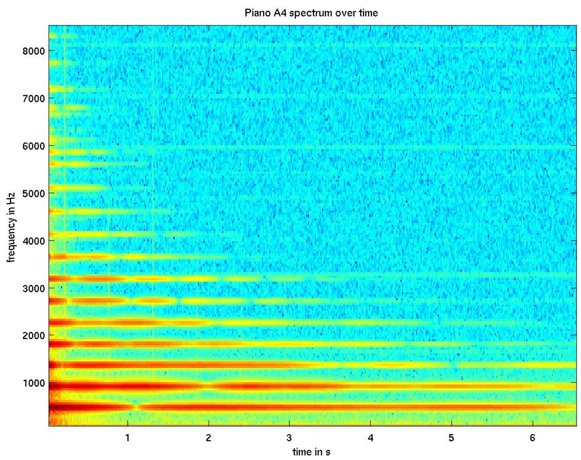

FIG. *2.2:* Spectre du piano - phénomène de dispersion comme fonction de la fréquence : les hautes fréquences décroissent plus vites que les basses fréquences

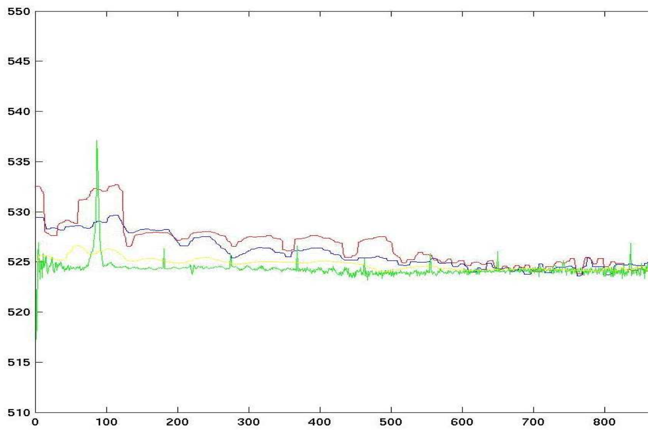

FIG. *2.3:* Estimation de la fréquence fondamentale par régression harmonique (bleu et rouge - f_0 additive) et prise en compte de l'inharmonicité

2.2 Présentation théorique de l'algorithme

Nous allons présenter dans cette partie la constitution du nouvel algorithme que nous avons développé au cours de notre stage.

Cet algorithme s'appuie sur une méthode spectrale basée sur :

- le modèle psycho-acoustique de Terhardt qui à l'origine avait été développé pour des considérations psycho acoustiques (détermination de la hauteur d'accords complexes) ainsi que pour la résolution d'accords virtuels (avec une basse fondamentale manquante), mais que nous avons détourné pour une utilisation favorable dans le cas de partiels manquants et de partiels fortement inharmoniques, ainsi que pour la possibilité de donner non pas une seule valeur de la fréquence fondamentale, mais plusieurs.
- que nous avons ensuite combiné à une méthode d'optimisation, pour l'affinement de la fréquence fondamentale conjointement à la détermination très précise et rapide du coefficient d'inharmonicité par une méthode de convergence de Newton.

Nous présenterons tout d'abord un fond théorique sous-jacent à notre algorithme, avant d'expliciter en détail notre algorithme.

2.2.1 Le modèle de Terhardt

2.2.1.1 Introduction

Le modèle de Terhardt [61] a été développé originellement pour l'extraction de la hauteur perçue dans le cas de sons complexes, comme par exemple des accords, qui comprennent une superposition d'au moins trois séries harmoniques, ou bien encore pour les instruments fortement inharmoniques, comme les cloches, mais qui présentent cependant une hauteur perçue. Cette méthode est basée sur une simple recherche des sous harmoniques des partiels présents dans le spectre, et comme nous allons le montrer, possède beaucoup d'avantages dans notre cas. Avant toute explication en détails, prenons un exemple simple de l'utilisation de ce modèle : le cas de la basse fondamentale «virtuelle» dans la musique baroque. Soit un accord forme de trois notes respectivement de fréquence fondamentale : 420, 520 et 620 Hz. En dressant un tableau des sous harmoniques de ces trois fréquences, on se rend compte qu'elles possèdent toutes une sous harmonique commune - respectant une certaine bande d'intégration - qui se trouve à 104, 103.3 et 102.9 Hz. C'est cette note pourtant non jouée dans l'accord, que nous entendrons comme basse fondamentale de l'accord. Ce que nous venons de voir dans le cas d'un son complexe, nous nous proposons de l'utiliser pour la détermination de la f_0 dans le cas de sons simples, c'est-à-dire de notes isolées. Comme nous allons le montrer, cette méthode est très avantageuse dans le cas de partiels manquants, ainsi que pour des sons fortement inharmoniques.

2.2.1.2 Théorie

Nous nous proposons de développer la théorie de la hauteur virtuelle en suivant un exemple très simple : une note dont les partiels sont en relations harmoniques complète, en négligeant les décalages de partiels.

Soit donc la série harmonique suivante : 110, 220, 330, 440, 550, ... Hz

Renversant la présentation faite par Terhardt, la question est de savoir, étant donné un partiel à la fréquence f_i, dans un premier temps si il fait partie d'une série harmonique correspondant à la fréquence fondamentale, et si oui, le combien-t-ième partiel de la série est-il ?

La première étape est de dresser la table des sous harmoniques pour chacune de ces fréquences :

Fréquences du partiel (Hz)	**110**	**220**	**330**	**440**	**550**
2	55	**110**	165	220	275
3	36,7	63,3	**110**	146,7	183,3
4	27,5	55	82,5	**110**	137,5
5	22	44	66	88	**110**
6	...	...	...	...	...

TAB. *2.1:* Tableau de sous harmoniques pour la série fréquentielle 110, 220, 330, 440, 550 Hz

Dresser la série des sous-harmoniques revient en fait à reconstituer des séries harmoniques potentielles pour la détermination de la fréquence fondamentale.

Définition des critères d'extraction de la f_0

Premièrement on définit un intervalle dit d'intégration *delta* qui sera la limite de «*pitch shift*» ou d'inharmonicité que l'on décide d'assumer pour l'extraction de la hauteur. Terhardt a montré que celui-ci devait être compris entre 0.01 et 0.05, la plus petite valeur correspondant à l'extraction d'un seul pitch, et plus l'on tend vers la seconde, plus le nombre de hauteurs secondaires (ou fortement inharmonique) croit.

D'un point de vue psycho-acoustique, l'intervalle d'intégration peut être considère comme une représentation de la largeur d'intervalle dans laquelle différent composants de la hauteur fusionnent (au lieu de donner lieu de se discriminer en deux hauteurs).

- Intervalle d'intégration faible : un seul candidat pour la hauteur.
- Intervalle d'intégration élevé : plusieurs candidats.

Cet intervalle d'intégration défini, les critères d'extraction du pitch sont définis de la façon suivante :
Est considéré comme hauteur dans notre table des sous harmoniques la fréquence
- qui a le plus grand nombre d'occurrences dans l'intervalle d'intégration
- qui correspond au plus petit numéro de sous-harmonique
- qui peut être obtenu par le plus petit intervalle d'intégration

Par ailleurs est succeptible de constituer une hauteur toute fréquence qui apparaît au moins trois fois dans le tableau.

Formalisation

Le critère de coïncidence peut se formaliser de la façon suivante :

Deux sous harmoniques appartiennent à l'intervalle d'intégration si et seulement si :

$$-\delta \leqslant \frac{f_q}{m} . \frac{n}{f_r} \leqslant +\delta \tag{2.3}$$

avec f_q et f_r les fréquences des composants dont les sous harmoniques sont examinées, et m et n respectivement les numéros de sous harmoniques de ces composants.

Cette formulation du critère nécessite de faire le calcul pour chaque valeur de m et de n. Nous avons choisi pour notre part une autre formulation complètement équivalente de ce critère, qui est moins coûteuse en temps de calcul.

$$Ent(1+\delta)m\frac{f_r}{f_q} \geqslant (1-\delta)m\frac{f_r}{f_q} \tag{2.4}$$

où Ent(x) est l'opérateur partie entière de la variable x.
Si cette deuxième condition est vérifiée, cela nous indique que la sous harmonique $\frac{f_q}{m}$ présente une sous harmonique avec une des sous harmoniques de f_r.

Quelques exemples d'application

Fondamentale manquante Soit la série harmonique : 220, 330, 440, 550, 660,...
Nous savons que la fréquence fondamentale correspondant à cette série est 110

Série harmonique incomplète Soit la série harmonique dans laquelle manquent tous les partiels pairs : 110, 330, 550, 770,...
La fréquence fondamentale est encore ici de 110 Hz.

Nous venons de montrer que le modèle de Terhardt est particulièrement bien adapté au problème de l'estimation de la f_0 dans une application musicale, étant insensible aux problèmes de décalage de partiels, de partiels manquants en plus ou moins grand nombre, ainsi que dans le cas d'une relation inharmonique entre les partiels - cas pour lequel seulement un petit nombre de partiels déterminent la fréquence fondamentale, ainsi encore pour des instruments présentant plusieurs fréquence fondamentales simultanément. Qui plus est, son utilisation simple et peu coûteuse en temps en font une méthode très efficace pour l'estimation de la fréquence fondamentale.

2.2.2 Problème de discrétisation et méthodes d'interpolation

La discrétisation intrinsèque à tout calcul numérique est un obstacle majeur à la résolution, en l'occurrence pour ce qui nous concerne dans la recherche de la position fréquentielle exacte des pics fréquentiels. En effet, le maximum d'un pic est tout à fait susceptible de se retrouver entre deux «bins» : une interpolation est alors indispensable pour nous ramener à une précision souhaitée et indispensable à notre calcul. Sinon, par exemple pour une transformée de Fourier de 1024 points sur un son échantillonné à 44100 Hz, la précision fréquentielle n'est que de 44100/1024, c'est-à-dire de 43 Hz ! !

2.2.2.1 Zero Padding

Le zero padding est une méthode bien connue en traitement du signal et qui consiste en une interpolation idéale du signal dans le domaine fréquentiel [57].

Définition

$$ZEROPAD_{M,m}(x) = \begin{cases} x(m), |m| < N/2 \\ 0, \text{sinon} \end{cases}$$

avec $m = \pm 1, \pm 2, ..., \pm M_h$
et

$$M_h = \begin{cases} \frac{M-1}{2}, \text{si M est impair} \\ \frac{M}{2} - 1, \text{si M est pair} \end{cases}$$

Si on *zero pad* d'un facteur M le signal dans le domaine temporel, c'est-à-dire que l'on ajoute au signal M fois sa longueur avec des zéros, on augmente la résolution fréquentielle d'un facteur M également.

Théorème

$$ZEROPAD_{l,n}(x) \Leftrightarrow INTERP_L(X) \tag{2.5}$$

Remarque : Le *zero padding* est une interpolation idéale dans le domaine fréquentiel, mais en aucun cas il n'ajoute d'information au signal. Par exemple, si la taille de fenêtre est insuffisamment grande pour discriminer des fréquences, le zero padding ne le permettra pas non plus.

Cependant l'inconvénient de cette méthode est que pour avoir une résolution satisfaisante, on augmente énormément le temps de calcul. En effet, si l'on veut pour une fenêtre de 1024 échantillons et pour un signal échantillonné à 44100 Hz une précision de l'ordre de 0.1 Hz, il nous faut *zero padder* le signal approximativement d'un facteur 400.

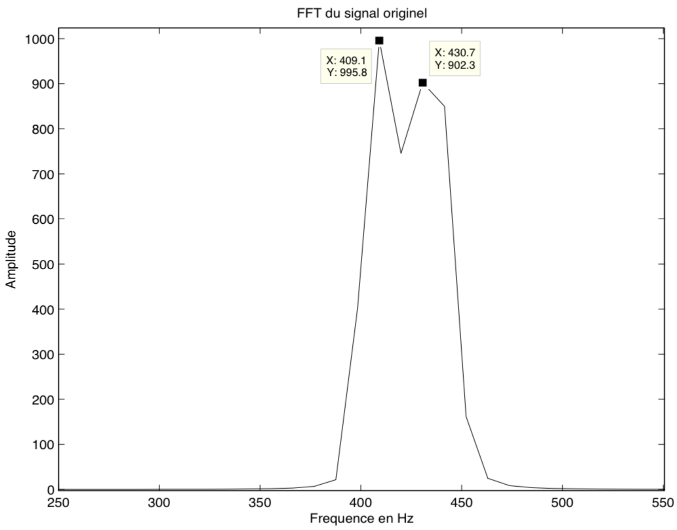

Fig. *2.4:* Détermination des pics dans la transformée de Fourier sans «zero padding»

2.2.2.2 Interpolation parabolique

Parmi le large choix d'interpolations habituellement utilisées, interpolation lagrangienne [2], régression parabolique en log-amplitude [48], nous avons choisie l'interpolation qui présente le meilleur compromis entre temps de calcul et résolution, à savoir l'interpolation parabolique [43][57].

Théorème : Soit N points (a1,...,an), alors il existe un et un seul polynôme de degré N-1 qui passent par ces N points

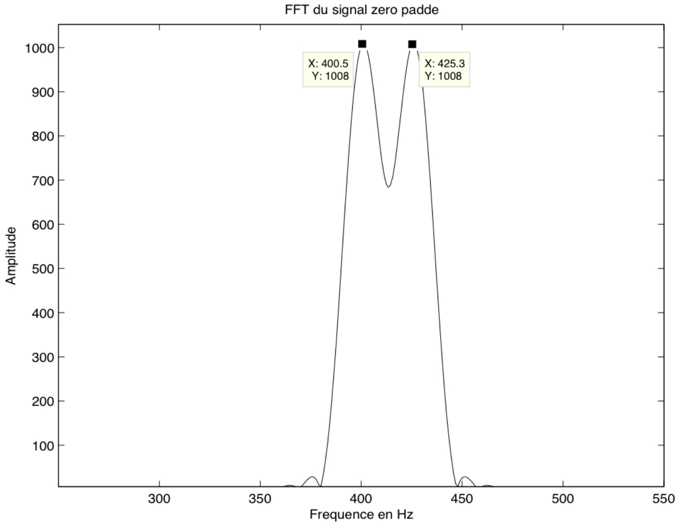

FIG. *2.5:* Détermination des pics dans la transformée de Fourier avec «zero padding»

Corollaire : Il existe un et un seul polynôme de degré 2 passant par trois points donnés x_1, x_2 et x_3. C'est cette propriété des fonctions polynomiales qui est utilisée lors de l'interpolation parabolique.

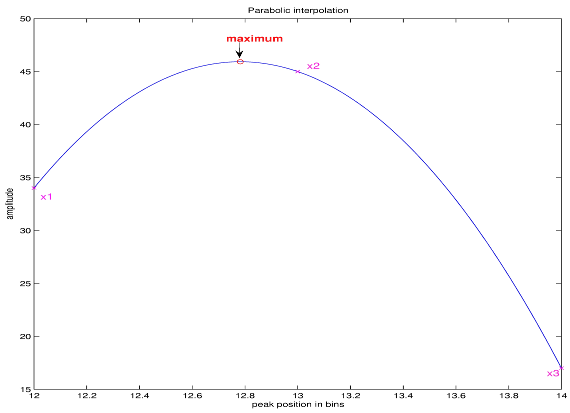

FIG. *2.6:* Interpolation parabolique : la fr ée quence du maximum d ée termin ée par cette m ée thode se rapproche plus de la solution exacte

2.2.3 Détermination du facteur d'inharmonicite : Curve Fitting - Convergence de Newton par la méthode des moindres carrés [4][13][66]

2.2.2.3 Méthode de convergence de Newton

Théorème : Soit f une fonction de classe 2 sur un intervalle [a,b] possédant une et une seule solution à l'équation f(x) = 0 sur [a,b], alors si $f'(x) \neq 0$ sur [a,b], la suite définie par :

$$p_{k+1} = p_k - \frac{f(p_k)}{f'(p_k)} \tag{2.6}$$

converge quadratiquement vers x_0.

Il existe cependant un grand inconvénient bien connu à cette méthode : si la solution recherchée n'est pas unique, il est possible que la solution diverge suivant le point d'initialisation. Cela arrive lorsque la sélection des partiels harmoniques dans le spectre en vue de la détermination du coefficient d'inharmonicité contient des partiels «bruités», c'est-à-dire qui ne correspondent pas à la série harmonique. Ces partiels bruités introduisent des minimums locaux susceptibles d'amener à de mauvais résultats. Pour lutter contre cet inconvénient, nous avons pour le moment choisi d'appliquer une méthode de convergence globale comme nous allons le voir.

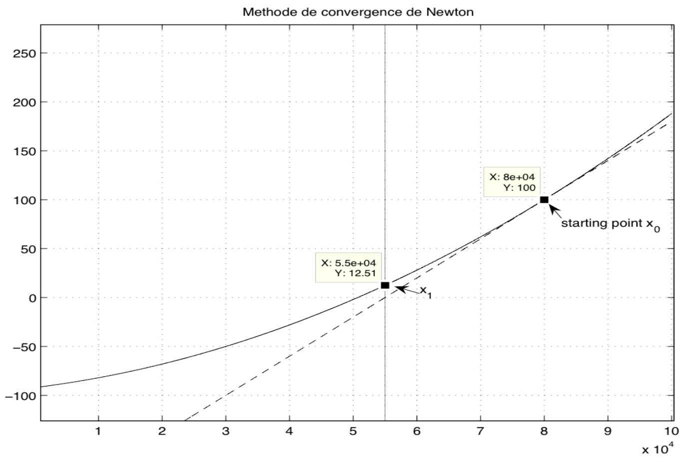

FIG. 2.7: Méthode de convergence de Newton

2.2.2.4 Curve Fitting par la méthode des moindres carrés

La fonction à minimisée dans notre cas est l'écart quadratique entre l'ensemble de partiels déterminés expérimentalement et la position théorique dans le cas de l'inharmonicité :

Théorème : Le Problème de minimisation par Newton d'une fonction sur un intervalle [a,b] est équivalent à la recherche de la racine de la dérivée de f sur ce même intervalle.

Qui plus est, et pour rendre cette méthode plus efficace, nous avons
- forcé la convergence vers des minimums, c'est-à-dire qu'à chaque pas :

$$|x_{k+1}| \leqslant |x_k| \tag{2.7}$$

- utilisé une recherche de solution globale, c'est-à-dire que l'on recherche le minimum global de la fonction et non plus seulement un minimum local dépendant du point d'initialisation.

Soit la fonction théorique :

$$f_n = nf_0^1\sqrt{1+Bn^2} \tag{2.8}$$

Et soit $(y_n)_{n\in N}$l'ensemble des fréquences des partiels déterminés dans la transformée de Fourier de notre signal.

En vue de déterminer le coefficient d'inharmonicité B et d'affiner dans un même temps le calcul de la fréquence fondamentale, nous voulons donc minimiser la fonction d'écart quadratique entre les partiels théoriques et les partiels déterminés expérimentalement dans notre spectre :

$$Err(B) = \sum_{n=0}^{N}(y_n - f_n(B))^2 \tag{2.9}$$

Cette fonction est minimisée lorsque le facteur B est celui qui explique le mieux l'ensemble de partiels déterminés dans notre spectre. La détermination se fait alors par la méthode de convergence de Newton.

2.2.2.5 Filtre médian

Définition : Soit $E = \{x_1, ..., x_n\}$ un ensemble de valeurs réelles et soit $E' = \{x'_1, ..., x'_n\}$l'ensemble de ces mêmes valeurs triées par ordre croissant.
On définit alors la médiane de cet ensemble de la manière suivante :

$$median(E') = \begin{cases} x_{p+1}, \text{ si } n = 2p+1 \\ \frac{x_p + x_{p+1}}{2}, \text{ si } n = 2p \end{cases}$$

Il a notamment été montré [53] que l'utilisation de ce filtre était avantageuse lors de post-processing pour éviter les erreurs «aberrantes», c'est-à-dire une variation trop grande de la f_0 dans un laps de temps trop court pour pouvoir être significatif d'une variation admissible de la f_0 .

2.3 Présentation de notre algorithme

Notre algorithme est un algorithme faisant partie des méthodes fréquentielles. Il a été codé en MATLAB qui est un langage de programmation particulièrement souple et rapide à implémenter.

Étape 1 : Preprocesing

Une frame de signal est fenêtrée avec une fenêtre de Blackmann-Harris, puis zero-paddée dans le domaine temporel. Ensuite on calcule la transformée de Fourier de ce signal, et on extrait les pics de la transformée de Fourier suivant une condition de ratio par rapport au partiel de plus grande amplitude. La fréquence de ces pics est alors déterminée par interpolation parabolique.

Remarque sur l'extraction des pics : L'extraction des pics dans la transformée de Fourier afin de pouvoir discriminer entre ceux qui sont des partiels et ceux qui ne sont que du bruit est un problème à part entière et aucune solution satisfaisante n'a été trouvée à ce jour. Des travaux ont été menés notamment par Axel Roebel [67] pour tenter de définir un critère de sinusoïdalité des partiels, qui semble avoir montrer des résultats prometteurs, notamment pour la discrimination partiels- bruit. Il serait avantageux de l'inclure à notre algorithme.

Étape 2 : Détermination de la f_0 - Modèle de Terhardt

Dans un second temps on applique la méthode de Terhardt pour extraire la f_0 du signal.

Étape 3 : Détermination du coefficient d'inharmonicité - Méthode de convergence de Newton

La f_0 déterminée, on sélectionne les partiels du signal sur un critère d'amplitude, puis on les interpole à nouveau. Dans une bande donnée autour de $k.f_0$, on considère comme partiel - si présent - le pic de plus grande amplitude. On utilise sur cet ensemble de pics la méthode de convergence de Newton dans un sens de moindres carrés par rapport aux harmoniques théoriques et aux harmoniques présentes. Le coefficient d'inharmonicité déterminé, on affine la valeur de la f_0.

Étape 4 : Post-Processing

Le post- processing est articulé en deux étapes :
- si la variation de la f_0 est supérieure à une valeur seuil d'une fenêtre sur l'autre, on recherche dans l'ensemble des f_0 possibles déterminées par Terhardt si il n'y a pas une valeur qui ne corresponde pas mieux.
- ensuite on applique un filtre médian pour lisser la trajectoire de la f_0, c'est-à-dire en éliminant les aberrations si il y a lieu en cas de variations trop grandes et temporellement trop courte pour constituer une véritable note.

2.4 Application à la détermination des coefficients d'inharmonicité du piano

2.4.1 Introduction

Le piano est un instrument de musique au comportement physique très complexe que nous nous proposons d'étudier dans un premier temps [35][62]. L'avantage de la détermination du coefficient d'inharmonicité dans le cas du piano est simple : de nombreuses études ont été menées jusqu'à présent, et c'est le seul instrument pseudo harmonique dont le comportement inhamonique a été étudié. Nous voudrions développer succinctement tout d'abord quelques éléments de physique du piano avant de donner les résultats de notre analyse, afin que les problèmes rencontrés puissent être mieux compris.

Pseudo harmonicité

Comme nous l'avons déjà vu, la relation harmonique entre les partiels n'est pas régulièrement espacée, mais suit la relation :

$$f_n = nf_0^1\sqrt{1+Bn^2} \tag{2.10}$$

où B est le coefficient d'inharmonicité

Fondamentale faible ou manquante

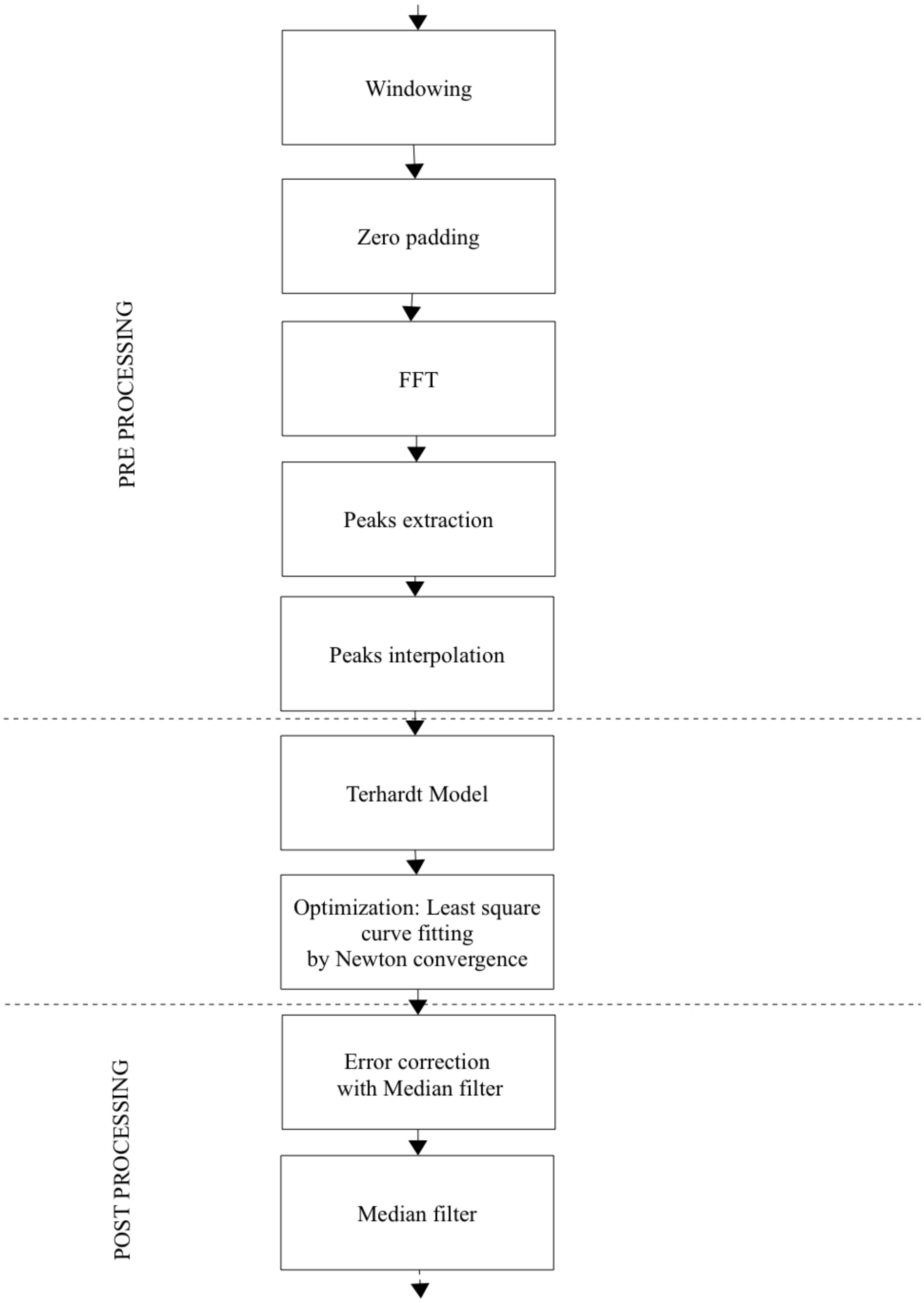

FIG. *2.8:* Représentation de la structure de notre algorithme

Dans le bas rang du piano, le partiel fondamental est manquant.

Table d'harmonie : apparition de «formants»

Le knock : le contact de la touche de piano et la table d'harmonie est à l'origine du résonance de la table d'harmonie typiquement à 90Hz

Par ailleurs, l'application du marteau sur une corde donnée produit une mise en mouvement des autres cordes, ce qui est particulièrement audible dans les hautes fréquences, toujours à cause de la table d'harmonie.

Ambitus du piano

Le piano est l'instrument de musique possédant le plus large ambitus, c'est-à-dire que les notes s'échelonnent typiquement de A0 (27.5 Hz) jusqu'à C8 (4186 Hz)

Ces quelques caractéristiques permettent de rendre compte de la complexité des processus physiques entraînée par le piano, et de la difficulté de son analyse.

2.4.2 Résultats : Détermination du coefficient d'inharmonicité du piano

Note	**A2**	**Bb2**	**B2**	**C3**	**Db3**	**D3**	**Eb3**	**E3**	**F3**	**Gb3**	**G3**	**Ab3**
B (10^{-5})	10,8	14,2	18,9	10,4	11,4	12,1	12,2	12,2	14,1	17,8	20,8	25,8
$\sigma(10^{-10})$	0,2	1,1	0,5	0,3	0,4	0,7	0,5	0,4	0,09	0,4	0,02	0,1
Note	**A3**	**Bb3**	**B3**	**C4**	**Db4**	**D4**	**Eb4**	**E4**	**F4**	**Gb4**	**G4**	**Ab4**
B (10^{-5})	27	30,5	36,2	32,2	41,1	41,3	44,8	48,2	54,1	60,6	64,6	69,4
$\sigma(10^{-10})$	0,08	0,7	0,6	6,6	8,5	5,01	1,47	1,38	1,07	13,9	24	11,2
Note	**A4**	**Bb4**	**B4**	**C5**	**Db5**	**D5**	**Eb5**	**E5**	**F5**	**Gb5**	**G5**	**Ab5**
B (10^{-5})	71,7	79,6	91,3	98,4	106	119	136	151	156	184	206	224
$\sigma(10^{-10})$	36,5	5,7	6,9	2,63	7,47	1,78	2,45	4,62	5,34	2,73	7,54	7,21
Note	**A5**	**Bb5**	**B5**	**C6**	**Db6**	**D6**	**Eb6**	**E6**	**F6**	**Gb6**	**G6**	**Ab6**
B (10^{-5})	247	252	315	325	380	343	433	462	510	544	598	694
$\sigma(10^{-10})$	5,1	1,1	7,34	9,84	1,39	9,88	47	85	17	13,1	115	78

TAB. *2.2:* Table des coefficients d'inharmonicité du piano déterminés avec notre algorithme et écart quadratique moyen par rapport à la solution

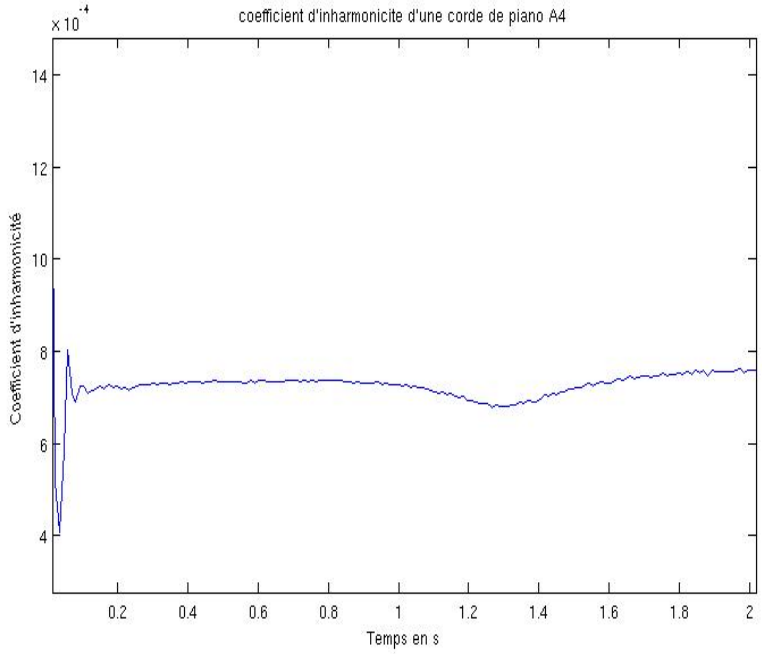

FIG. *2.9:* Coefficient d'inharmonicité du piano pour la note A4 : l'erreur de départ est due à la difficulté d'analyse lors de l'attaque où le spectre est plus confus

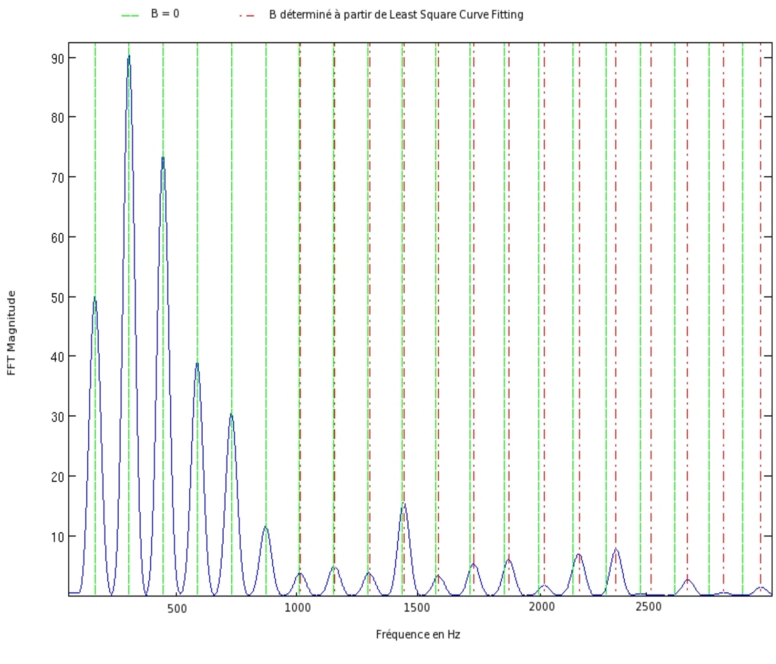

FIG. *2.10:* Comparaison des partiels dans le cas harmonique (en vert) et avec le coefficient d'inharmonicité déterminé par «least square curve fitting» (en rouge)

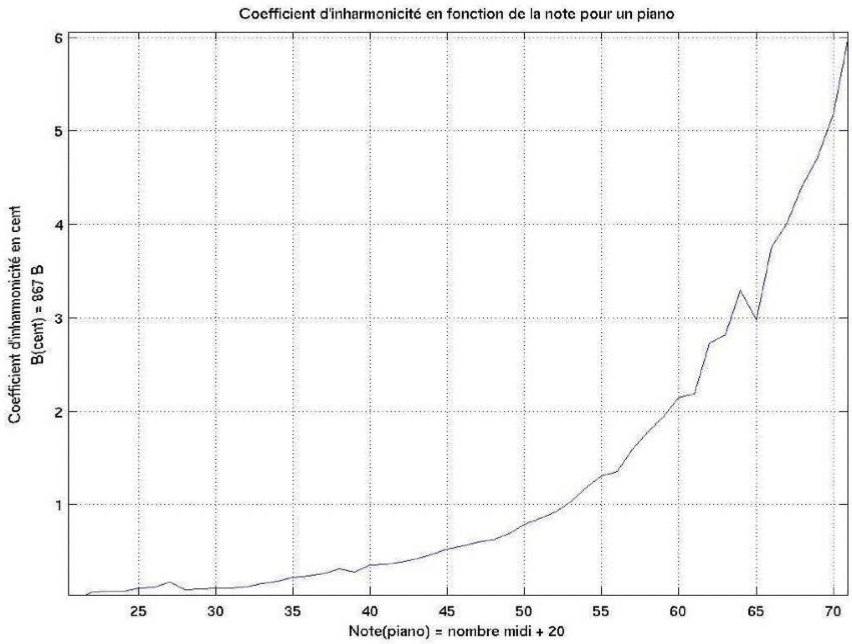

FIG. *2.11:* Coefficients d'inharmonicité déterminés de la note Gb2 (92.5 Hz) à la note A6 (1760 Hz)

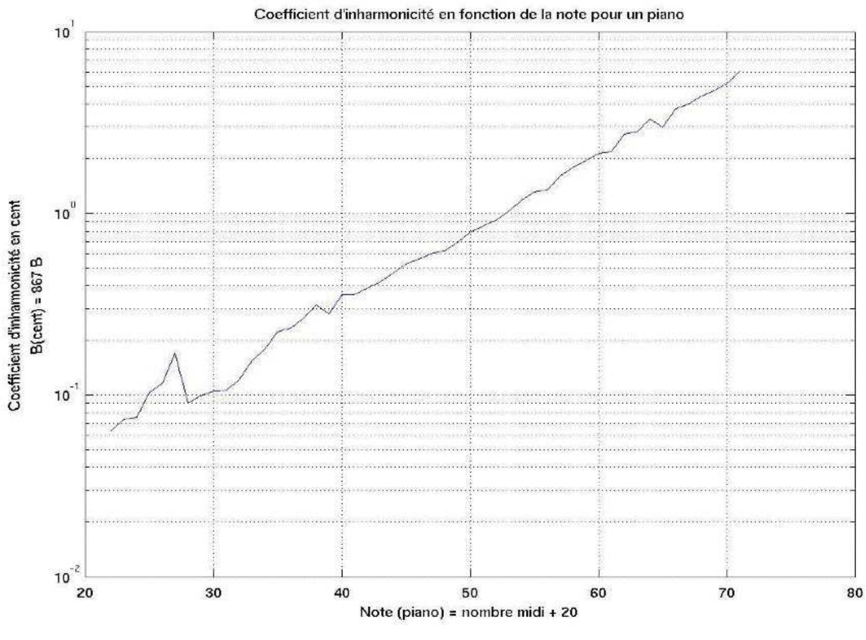

FIG. *2.12:* Coefficients d'inharmonicité déterminés de la note Gb2 (92.5 Hz) à la note A6 (1760 Hz)) en échelle logarithmique

Chapitre 3

Évaluation des algorithmes d'estimation de la fréquence fondamentale

Sommaire

3.1 Présentation des algorithmes

Nous pouvons séparer les algorithmes d'estimation de la fréquence fondamentale en deux catégories distinctes, qui ne sont pas sans rappeler la même distinction dans le domaine psycho-acoustique : les méthodes temporelles et les méthodes fréquentielles.

3.1.1 Introduction

Jusqu'à aujourd'hui, un grand nombre de méthodes ont été développées pour l'estimation de la f_0 de signaux de type parlé (application à la reconnaissance vocale et aux télécommunications) ou musicaux (application à l'analyse de signaux, à la synthèse et à la transcription).

Une simple revue permet de se rendre compte de la diversité des approches mises en jeu :

Domaine fréquentiel : Cepstrum [1] [47], Generalized spectrum [3], Filtre peigne harmonique ou non [23][44][60], High Order Spectrum [33] , Maximum Likelihood [17][18][64], Two Way Mismatch [7][19][41][56], Hidden Markov Model [14][15][18][50], Fréquence instantanée [34][49][68], Constant-Q Transform [6][63], Transformation de Hilbert [29] et méthodes psycho- acoustique pour les principales.

Domaine temporel : Autocorrelation [52], Autocorrelation normalisée, Cross-correlation [54], AMDF (pour Average Magnitude Difference Function), Fonction de différence [8][58], Fonction de ressemblance [5], Damped Oscillators [36], Zero Crossing [21], Linear Predictive Coefficients, High Order Statistics [45], et Ondelettes [42] pour les principales.

Il existe également des méthodes mixtes, essayant de tirer partie des avantages combinés des méthodes temporelles et fréquentielles, ou alors des méthodes utilisant une représentation temps-fréquence comme la distribution de Wigner-Ville [39].

Pour résumer, toutes ces méthodes tentent de donner une représentation du signal dans laquelle l'extraction de la caractéristique intéressante, ici la f_0, est facilité. Cette extraction se fait généralement par maximisation ou minimisation d'une fonction qui dépend de la méthode employée : maximisation de la probabilité, minimisation de la distance, etc...

3.1.2 Domaine temporel contre domaine fréquentiel

Nous présentons dans ce chapitre une revue succincte des avantages et des inconvénients de chacun des deux domaines utilisés lors de l'analyse.

Stationnarité/ Instationnarité :

Il a été montré que l'utilisation de la fonction d'aucorrelation ne nécessite d'avoir que deux périodes du signal pour déterminer la f_0 alors que d'un autre coté les méthodes fréquentielles nécessitent au moins 3 périodes voir 4 périodes du signal.

La réduction de la taille de fenêtre des méthodes temporelles permettent alors des applications à des signaux moins stationnaires que la méthode fréquentielle comme la voix par exemple. Par ailleurs, la précision temporelle en est d'autant augmentée.

Bruit : Les méthodes temporelles sont moins sensibles au bruit que ne le sont les méthodes fréquentielles.

Réverbération : A partir du moment où la méthode temporelle considère de facto tout le signal, c'est-à-dire tout son spectre, elle est très sensible au phénomènes de réverbération ou de f_0 multiples. Au contraire le passage dans le domaine fréquentiel permet la discrimination entre les différents partiels contenus dans le signal, et ainsi de faire de la séparation de sources.

Conclusion : il apparaît déjà de ces quelques considérations que les méthodes temporelles seront utilisées avec avantage dans le cas de signaux tels que la voix, plus instationnaire, dans un environnement plus bruité (application à la téléphonie par exemple). Tandis que les méthodes fréquentielles seront quant à elle à leur avantage dans les applications musicales, où le nombre de f_0 peut varier même dans le cas monodique, par simple effet de réverbération, ou de perduration d'anciennes notes, jusque pour des application à la détermination de f_0 multiples.

3.1.3 Le domaine temporel

La fonction d'autocorrelation

Dans le domaine temporel, la fonction d'autocorrelation a été la première [52] et est une simple solution pour l'estimation de la f_0.

La définition de la fonction d'autocorrelation est :

$$r_t(\tau) = \sum_{j=t+1}^{t+W} x_j x_{j+\tau} \tag{3.1}$$

Son utilisation est simple et intuitive : la fonction d'autocorrelation consiste en le produit du signal avec une version décalée de lui-même.

Alors dans le cas d'un signal harmonique et localement stationnaire, c'est-à-dire que le signal et sa version décalée suffisamment proche dans le temps présentent une forte corrélation, cette fonction présentera des maximums à la période fondamentale et ses multiples.

L'AMDF

L'AMDF est une version légèrement différente de la fonction d'autocorrelation : au lieu de sommer le produit du signal avec lui-même, on considère à la place la différence au carré du signal avec sa version decalée. Ainsi la recherche de la période fondamentale revient à minimiser cette différence.

$$d_t(\tau) = \sum_{j=1}^{W} (x_j - x_{j+\tau})^2 \tag{3.2}$$

Il est à noter une relation intéressante entre ces deux fonctions :

$$d_t(\tau) = r_t(0) + r_{t+\tau} - 2r_t(\tau) \tag{3.3}$$

Cette relation montre simplement que les maximums de la fonction d'autocorrelation ne correspondent pas nécessairement aux minimums de la fonction de différence.

YIN

La méthode utilisée par YIN est basée sur une nouvelle fonction temporelle prenant en compte et améliorant sensiblement les deux précédentes fonctions présentées [8].

On considère la *cumulative mean normalized difference function* définie de la façon suivante :

$$\mathrm{d}'_t(\tau) = \begin{cases} d_t(\tau)/\frac{1}{\tau}\sum_{j=1}^{\tau} d_t(j), \text{ pour } \tau \neq 0 \\ 1, \text{ pour } \tau = 0 \end{cases}$$

Cette fonction présente l'avantage de ne pas être nulle en 0, et se trouve être la fonction de différence normalisée par la moyenne de la fonction de différence prise sur un intervalle plus petit que W, et dépendant de τ .

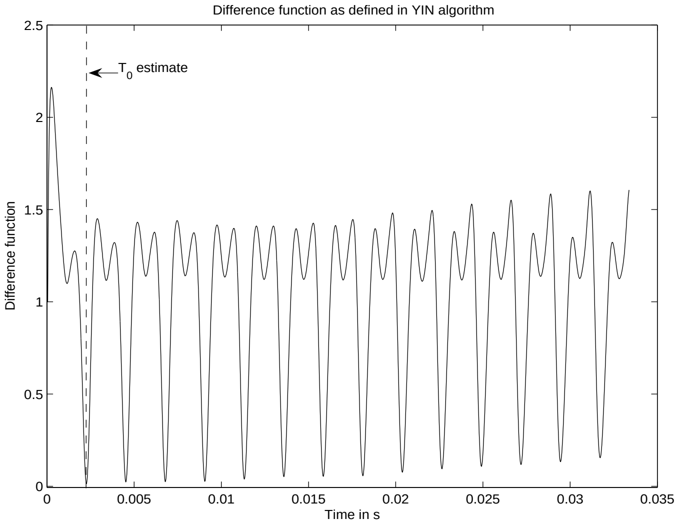

FIG. *3.1:* Fonction de différence telle qu'utilisée dans YIN

«Robust On-the-Fly Pitch Estimation»

L'algorithme *Robust On-the-Fly Pitch Estimation* a été développé à la suite de YIN, s'inspirant largement des bases de celui-ci [58]. Cet algorithme propose une approche théorique unifiée de la fonction d'autocorrelation et de la fonction de différence, pour proposer une analyse par réduction de dimension.

On considère la matrice de covariance associée au vecteur aléatoire réel $x = [x_j x_{j+\tau}]^T$

$$C_x(\tau) = E[(x-m)(x-m)^T] \tag{3.4}$$

Dans le cas idéal d'une stationnarité au sens large, et d'un signal périodique, nous pouvons montrer que cette matrice possède les vecteurs propres suivants : $[11]^T$ et $[1-1]^T$

En faisant l'hypothèse d'une distribution uniforme ainsi que de moyenne nulle, il est montré que les fonctions :

$$\begin{cases} \lambda_1(\tau) \propto \sum_{j=1}^{W-\tau}(x_j + x_{j+\tau})^2 \\ \lambda_2(\tau) \propto \sum_{j=1}^{W-\tau}(x_j - x_{j+\tau})^2 \end{cases}$$

sont des mesures effectives de comment la base $[11]^T$ et $[1-1]^T$ peuvent effectivement constituer séparément une base de l'espace à deux dimensions.

Ces considérations sur la réduction d'espaces encastrés amène alors à une meilleure définition de la fonction d'autocorrelation et de différence comme suivant :

$$\begin{cases} r'_t(\tau) = \frac{\lambda_1(\tau)-\lambda_2(\tau)}{\lambda_1(\tau)+\lambda_2(\tau)} \\ d'_t(\tau) = \frac{\lambda_2(\tau)}{\lambda_1(\tau)+\lambda_2(\tau)} \end{cases}$$

Ce sont ces deux dernières fonctions qui sont utilisées en lieu et place des fonctions d'autocorrelation et de différence initiales.

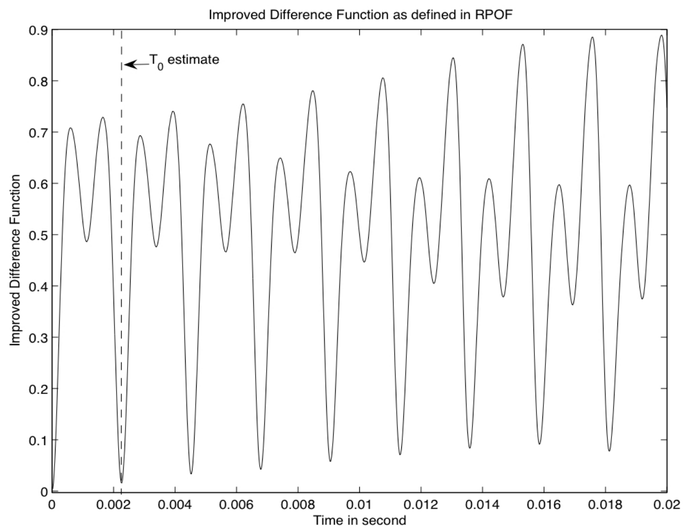

FIG. *3.2:* Fonction de différence telle qu'utilisée dans ROTFP

Le choix de ces deux algorithmes dans le domaine temporel s'est fait naturellement comme étant étant représentatives de l'état actuel en ce qui concerne le domaine temporel.

3.1.4 Le domaine fréquentiel

Il existe actuellement de nombreuses représentations du signal dans le domaine fréquentiel, de la plus courante : la transformée de Fourier, jusqu'à la fréquence instantanée qui pare aux limitations de la transformée de Fourier. L'intégralité des méthodes utilisés par les algorithmes que nous avons choisi sont basés sur la transformations de la transformée de Fourier, seuls ici se distinguent le traitement de cette transformation et les moyens utilisés pour en extraire l'information f_0.

Les filtres peignes

La méthode utilisée par cet algorithme repose sur une idée très simple [23] : maximiser l'énergie contenue dans la sortie de la transformée de Fourier passée au travers d'un filtre peigne harmonique, voir non harmonique (ou alors deux paramètres doivent être déterminés)

On considère un signal dont les pics spectraux sont prononcés et pouvant être identifiés comme partiels. On suppose que les fréquences des partiels sont définis par une relation connue comme fonction Φ

$$f_n = \Phi(n, H_1, H_2, H_3, ...) \tag{3.5}$$

où n est le numéro de partiel et $H_i$sont des paramètres inconnus déterminant l'influence de la source et les propriétés de transmission sur les fréquences des partiels.

En appliquant un filtre peigne inharmonique de paramètres ajustables dont les fréquences centrales ϕ_k du passe bande suivent la même relation que les fréquences des partiels :

$$\phi_k = \Phi(k, h_1, h_2, h_3, ...) \tag{3.6}$$

On obtient une réponse maximale du filtre lorsque les paramètres du filtre égalent les paramètres du signal.

Dans le cas qui nous intéresse, nous avons la relation suivante :

$$f_n = n f_0^1 \sqrt{1 + Bn^2} \tag{3.7}$$

où B est le coefficient d'inharmonicité, nul dans le cas harmonique.

Les paramètres inconnus à déterminer sont alors f_0^1 et B.

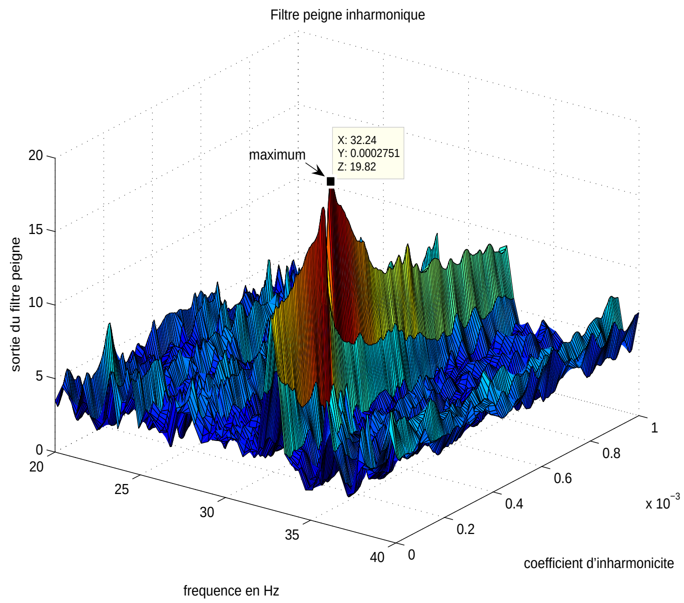

FIG. *3.3:* Sortie du filtre peigne inharmonique : le maximum donne à la fois la f_0 et le coefficient d'inharmonicité

L'algorithme f_0, tel qu'utilisé dans *Additive* **et** *Audiosculpt*

Première version : Boris Doval (1)

La première version de f_0 a été développée par Boris Doval lors d'un travail mené à l'IRCAM. Il est basé sur un modèle probabiliste visant à «expliquer» une observation définie comme le spectre du signal. Pour une observation donnée, trouver les partiels qui expliquent «le mieux» le spectre. Cette première version repose sur une optimisation d'une quantité définie par 2 critères fondamentaux

Critère 1 : Somme des amplitudes en dB expliquées

Ce premier critère correspond à un histogramme pondéré par les amplitudes.

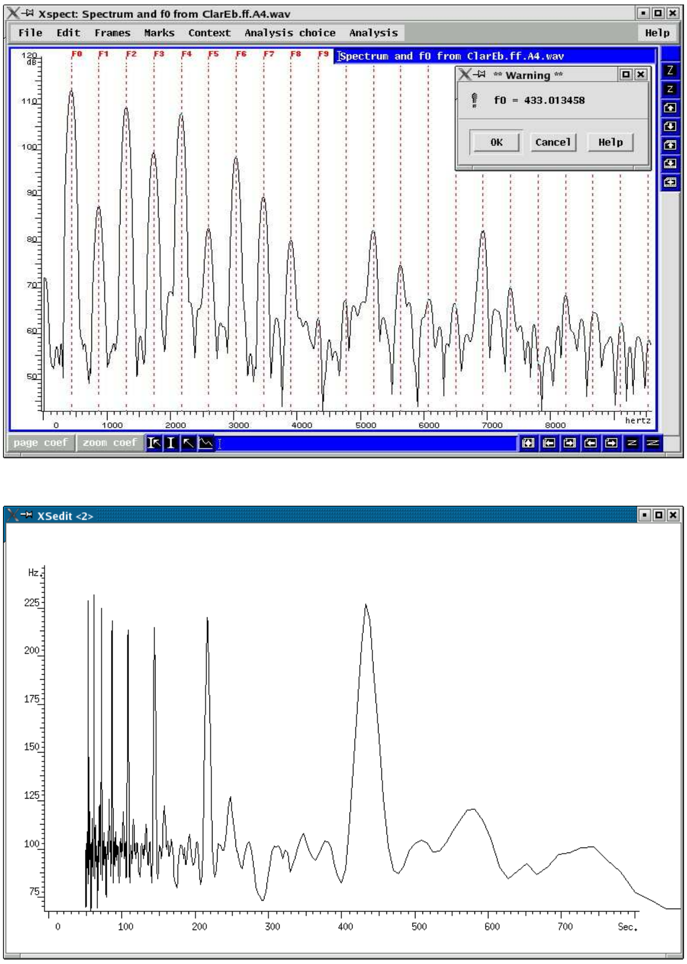

FIG. *3.4:* Histogramme tel que dans f_0 pour un son de clarinette A4 (en bas) et le spectre de ce dernier (en haut).

Plus la somme des amplitudes en dB est grande, plus la fréquence fondamentale considérée explique «mieux» le spectre considéré.

Dans ce modèle, il est évident que les sous harmoniques de cette fondamentale expliqueront bien aussi ce même spectre : il est alors nécessaire de discriminer parmi ces candidats.

On utilise alors un second critère, qui est le critère dit de «lissittude» du spectre.

Critère 2 : «Lissitude» ou *smoothness* des partiels avec amplitude en dB - notion de convexité

La lissitude minimale est alors obtenue pour un spectre plat. Ainsi la lissitude associée à une sous harmonique de la fréquence fondamentale sera plus élevée que celle de cette dernière.

La détermination de la fréquence fondamentale est définie de la manière suivante :

$$f_0 = \underset{f_{0,k}}{max} \frac{c_1(f_{0,k})}{c_2(f_{0,k})} \tag{3.8}$$

Une fois une valeur «approximative» de la fréquence fondamentale obtenue, celle - ci est affinée par une régression linéaire menée sur le plus grand partiel déterminé par appariement des partiels.

Deuxième version : Boris Doval (2)

La deuxième version a été développée encore par Boris Doval lors de son travail de thèse. Elle repose sur un modèle probabiliste *Maximum Likelihood* associé dans un second temps à des chaînes cachées de Markov (*Hidden Markov Model*).

Troisième version : Roebel - Krauledat - Prudham (3) & Chunghsin Yeh - Axel Roebel(4)

Cette dernière version n'ayant pas abouti à des résultats réellement satisfaisant les derniers développements ont été menés à partir d'un retour aux principes de la première version. Cette première version présentait cependant l'inconvénient de ne pas pouvoir pondérér l'importance des différents critères. Le critère 2 a été redéfini alors par Chunghsin Yeh.

Un troisième critère est alors venu en complément :

Critère 3 : *Spectral centroid*

Ce critère est défini comme la somme des amplitudes des partiels pondérés par le numero des partiels, ce qui constitue une *spectral centroid* appliqués sur les partiels.

$$c_3(f_{0,k}) = \sum_{n=1}^{N} n.A(partiel(n)) \tag{3.9}$$

Ainsi ce critère défini, et pour une fréquence fondamentale donnée, une sous harmonique aura une *spectral centroid* toujours plus grande que celle associée à cette fréquence fondamentale.

Enfin, ces trois critères sont pondérés par des coefficients optimisés par apprentissage.

$$f_0 = \underset{f_{0,k}}{min}(a_1 c_1(f_{0,k}) + a_2 c_2(f_{0,k}) + a_3 c_3(f_{0,k})) \tag{3.10}$$

Les versions utilisées lors de notre évaluation sont les suivantes :
• Additive - Doval (1)
• Additive pondéré Axel Roebel (4)

3.2 Constitution d'une base de données

3.2.1 Vers une classification des sons musicaux

Nous avons défini pour la constitution d'une base de donnée d'analyse une nouvelle classification des sons, d'après la physique propre qui leur est liée. Notre classification se présente de la façon suivante : nous définissons la structure harmonique comme la distribution dans le domaine fréquentiel des partiels et de leur amplitude respectives.

- Une seule note : le spectre de la note possède une structure harmonique unique, indépendante du temps ou peu dépendante (quasi stationnaire).

- Phrase monophonique : structure harmonique unique, dépendante du temps
- Phrase duophonique : structure harmonique double et dépendante du temps
- Plusieurs timbres : structure harmonique multiples et dépendantes du temps.

Chaque structure harmonique dépend de la physique des instruments considérés.
Note : un instrument seul peut avoir également plusieurs timbres, c'est-à-dire plusieurs structures harmoniques : par exemple, le timbre des instruments à cordes frappées ou pincées dépend de la force avec laquelle on attaque la corde. - Modes multiples : structure harmonique complexe et inharmonique. C'est le cas des membranes, des instruments à percussion métalliques plans, et des cloches.
- Oscillateurs libres : régime pseudo harmonique et inharmonique. Ces oscillateurs reçoivent une impulsion initiale et sont alors en régime libre. Dans le cas des cordes, le régime est alors pseudo harmonique (c'est-à-dire avec une relation inharmonique connue entre les partiels)

Ils présentent une inharmonicité due à la raideur de la corde.

- cordes pincées :
    - famille de la guitare
    - famille de la harpe
    - famille du violon dans le cas de pizzicati
    - famille du clavecin
- cordes frappées
    - piano

Dans le cas des instruments à percussion, la relation est alors plus complexe, et il n'existe aucune relation formelle connue aujourd'hui dans la distribution des partiels.

On y retrouve les membranes, les métallophones plans, et la famille des cloches.

- Oscillateurs forcés : régime harmonique. Ces oscillateurs sont forcés par une action du joueur. La relation entre les partiels est alors harmoniques ou quasi harmonique (déplacement de certains partiels), même dans le cas des instruments à cordes ou l'harmonicité est alors forcées.

- cordes frottées
    - famille du violon
- «*pipe resonator* s»
    - famille des flûtes
    - famille de l'orgue

- famille des «horns»
    - famille des cuivres
    - famille des bois

C'est donc selon cette présente classification que nous avons construit notre base de donnée initiale.

Remarques : Les trois premières catégories sont considérées comme monotimbral, ce qui signifie que la structure harmonique correspondant à chacune des notes suivent les mêmes lois physiques - en terme de distribution fréquentiel et d'amplitude des composants harmoniques. Toutes les sortes d'instruments appartiennent aux 4 premières catégories.

<u>Note relative à la classification de la base de données</u>

La classification choisie pour notre base de données n'est évidemment pas unique, et nous souhaiterions définir une seconde classification qui elle aussi serait intéressante pour l'évaluation des différents algorithmes. Un certain nombre de problèmes rencontrés par les algorithmes lors de l'analyse sont bien connus aujourd'hui, chaque problème correspondant à une caractéristique soit de la méthode employée par l'algorithme lui-même, soit une caractéristique propre du signal.

Une étude du comportement des algorithmes dans le cadre de ces problèmes serait également très intéressante à réaliser dans un travail ultérieur :

- jeu rapide (non stationnaire : problème pour tous les fenêtrages)
- jeu à différent ranges (basse fréquence, moyenne fréquence, haute fréquence)
- large ambitus (chaque fenêtrage ayant sa zone privilégiée, nécessairement source de problème)
- réverbération :

Ce que nous rangeons dans la réverbération provient de 2 cas différents :
- la fenêtre est à cheval entre deux notes.
- une ou plusieurs notes est entretenue par réverbération

Quelle qu'en soit l'origine, le signal ne contiendra donc plus une seule fréquence fondamentale, mais au moins deux. Il serait peut être alors intéressant de ne plus définir une seule f_0 pour ce signal, mais autant qu'il y a réellement de présentes : c'est l'unique moyen de rendre compte totalement du signal considéré. Cela est alors le contraire pour le premier cas, ou une seule est présente : la seconde n'existant dans le signal que par l'erreur temporelle introduite par le fenêtrage. Ce problème est très courant, et en fait il n'existe dans la musique que très peu de cas de phrases musicales réellement monodiques, au sens où rigoureusement une seule note est présente au même instant dans le signal.
- problème pseudo harmonicité ou même inharmonicité (peu prise en compte par les algorithmes)
- bruit (extrinsèque et intrinsèque comme pour le shakuachi)
- problèmes des transitoires d'attaque

3.2.2 Présentation de la base de données

Notre base de données comporte dans un premier temps un ensemble de notes simples isolées pour un très grand nombre d'instruments prises à partir de différentes bases déjà existantes telles que la RWC Music Database, McGill University, IOWA University et des sons provenant des bases de l'IRCAM et du CNMAT.

Instruments à une seule note
- instruments à oscillations libres : 5 pianos, 2 violons, alto, 2 violoncelles, contrebasse, 2 guitares cordes en acier, guitare cordes nylon, guitare 12 cordes, banjo, mandoline, 2 basses électriques, harpe, shamisen.
- instruments polymodaux : celesta, 2 glockenspiel, gong, japanese block, 2 marimba, cloches, cloches tubulaires, tabla, temple block, timpani, 2 vibraphones, instruments à percussion du gamelan. instruments à oscillation forcées : orgue, saxophone alto, 2 bassons, clarinette basse, clarinette en mib, cornet, didgeribou, 2 trompettes, flûte traversière, flûte de pan, flûte à bec, hautbois, tuba, pelog, salendro, suling, 2 shakuachi.
- voix : soprano femme, ténor homme, chant pakistanais, chant mongole.

Chaque fichier son est d'une durée moyenne de 2 secondes.

3.2.3 Méthodologie utilisée pour l'évaluation des algorithmes

3.2.3.1 Définition du procédé d'évaluation

Dès lors que nous voulons comparer des méthodes qui peuvent être très différentes les unes des autres que ce soit par leur constitution et par les paramètres utilisés, nous nous heurtons nécessairement au problème de l'objectivité des résultats obtenus. Deux possibilités s'offrent alors à nous : soit on décide d'évaluer les algorithmes avec des paramètres «égaux» pour tous, soit on teste les versions optimisées de chacune des méthodes, c'est-à-dire chacune dans les meilleures conditions. Nous avons choisi pour notre évaluation la première de ces deux solutions. Dans le but d'une

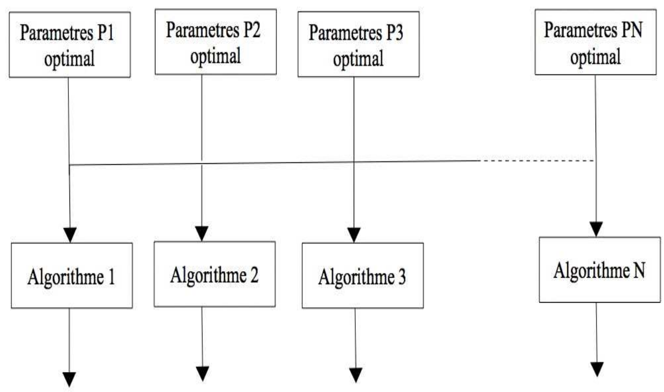

FIG. *3.5:* Evaluation «optimale»

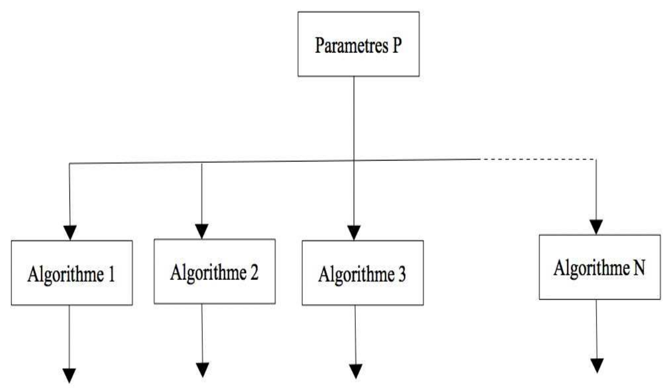

FIG. *3.6:* Evaluation «neutre»

comparaison valable et pertinente des différentes méthodes, nous avons défini les critères suivants :
- La méthode seule est testée, c'est-à-dire que nous avons enlevé tout post-processing lorsque cela était possible.

- Les paramètres d'entrée sont les mêmes pour tous les algorithmes, à savoir : la taille de fenêtre, le rang de f_0 dans lequel chercher, et le pas d'avancement de la fenêtre, afin de comparer sur les mêmes temps.
- Les paramètres de sortie : estimation de la f_0 , plusieurs candidats possibles lorsque cela était possible, et une mesure de la confiance que l'on peut avoir dans le résultat donné. Malheureusement la normalisation de cette mesure de confiance n'était pas possible, et nous avons donc conservé cette mesure différente pour chacun des algorithmes.

A partir de ces critères, nous pouvons dire que les algorithmes partent tous sur un pied d'égalité lors de l'estimation de la f_0 les résultats obtenus seront donc comparables.

Problème de l'évaluation : Qu'évalue - t - on ?

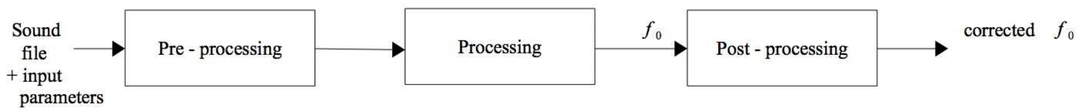

FIG. *3.7:* Structure générale d'un algorithme d'estimation de la f_0

3.2.3.2 Définition des critères d'évaluation

Nous avons défini les critères suivants pour l'évaluation et la comparaison des algorithmes :
- Chaque algorithme tourne en parallèle sur un certain fichier son
- Les résultats rendus sont utilisés pour la détermination de la référence de la façon suivante : règle de la majorité et médiane.

Règle de la majorité

Face à la taille conséquente de la base de données, il était nécessaire de définir une référence qui soit relativement rapide à déterminer, car la référence à la main demande beaucoup trop de temps par rapport à la durée de notre stage.

Définition de la référence

C'est dans cet objectif que nous avons décider d'utiliser la méthode de la règle de la majorité : chaque algorithme est amené à voter pour une certaine f_0 . La f_0 retenue comme référence est alors celle qui aura reçue le plus grand nombre de votes, moyennant une bande de tolérance, et utilisant ensuite la médiane comme référence.

Remarque : il a été montreé que cette méthode de la règle de la majorité est valide dans le cas ou les méthodes utilisées par chaque votant est indépendante des autres. Il est clair que cela n'est pas le cas dans notre présente étude, nous devons donc apporter quelque retenue sur les résultats donnés par cette méthode, même si généralement elle est convaincante.

Une fois définie cette référence, les erreurs seront calculées en fonction de leur écart à la référence.

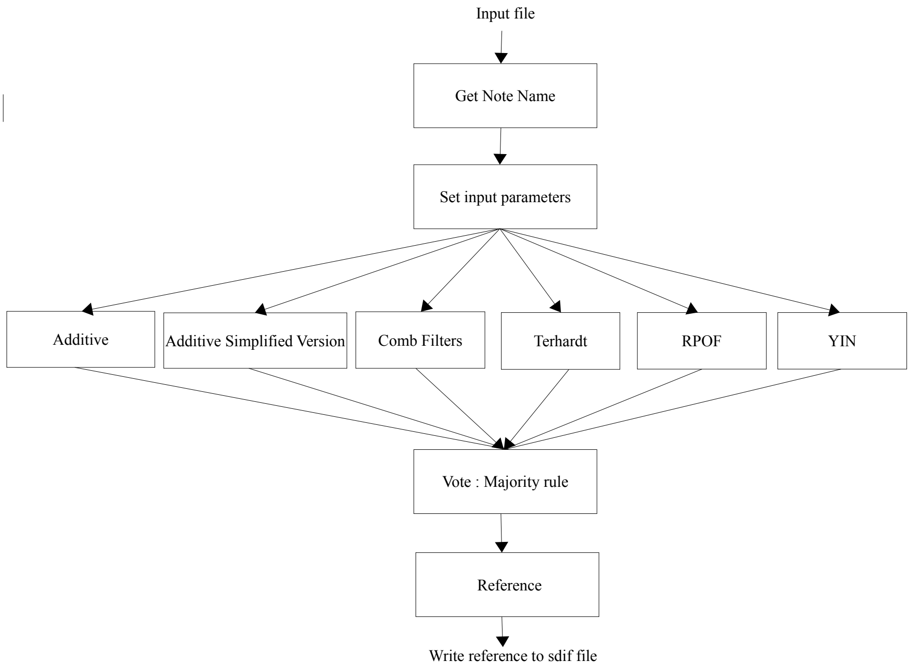

FIG. *3.8:* Structure d'estimation de la référence

3.2.3.3 Typologie des erreurs

Il est important dans l'analyse des résultats que nous obtenir de définir différentes classes de types d'erreurs : il évident que toutes les erreurs ne sont pas équivalentes.

Soit l'ensemble des erreurs :
- grosses erreurs
- erreurs fines

Nous avons respecté cette classification en utilisant les critères suivants :

- Une grosse erreur est une erreur supérieure à 1/4 de ton - critère utilsé pour la transcription notamment c'est-à-dire environ 3%.
- Une erreur fine est alors une erreur se trouvant dans une bande 1/8 de ton et ? de ton, c'est-à-dire entre 1.5% et 3%.
- Par ailleurs, dans l'ensemble des grosses erreurs, nous considérerons une erreur remarquable, qui est l'erreur dite «d'octave» (ce qui est assez impropre puisque justement les partiels ne sont pas nécessairement dans un rapport harmonique) : l'erreur d'octave est donc une erreur comprise entre 97% et 103%.

Toutes ces erreurs seront signées pour pouvoir discriminer dans un même type d'erreur les erreurs dites «trop basses» des erreurs dites «trop hautes». Qui plus est, et pour aller plus loin dans

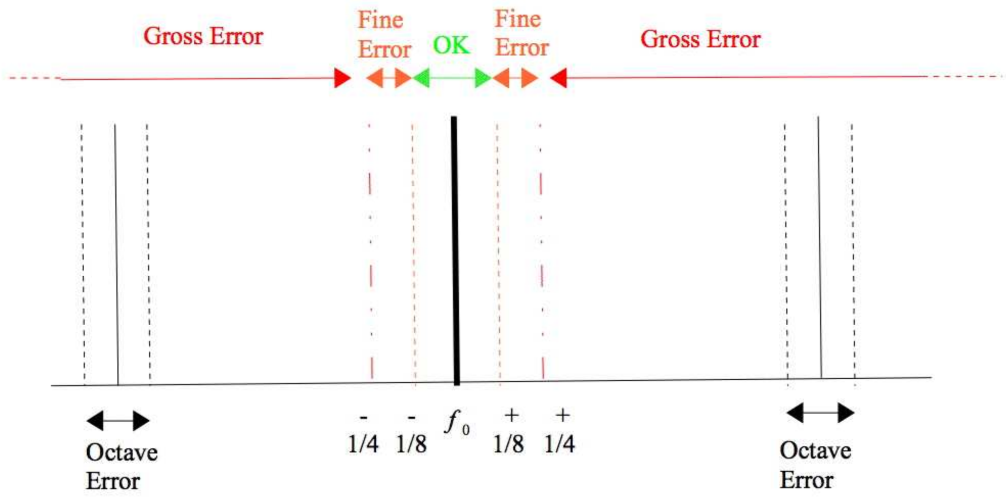

FIG. *3.9:* Classification des erreurs

l'analyse des résultats, nous avons utilisés un ensemble de descripteurs du signal audio pour tenter d'expliquer les erreurs éventuelles. Ces descripteurs sont les suivants :
- centroïde
- énergie
- *zero crossing rate*
- énergie et position des principaux pics

et sont utilisés pour détecter certaines caractéristiques du signal pouvant être cause d'erreur, à savoir :
- l'attaque : centroïde, énergie élévées

- le decay : énergie, énergie et position des principaux pics faible
- décision voisée /non voisée : énergie des principaux pics et zero crossing rate

3.2.3.4 Définition d'un nouveau type d'analyse au format SDIF

Pour cela, nous avons trouvé avantage à utiliser les fichiers d'analyse SDIF développés à l'IRCAM et au CNMAT. Ces fichiers constituent une norme pour l'écriture et la lecture d'analyses audio. Dans chacun de ces fichiers, nous utilisons pour chaque temps une frame donnée, c'est-à-dire pour un temps discret donné, un ensemble de paramètres que nous voulons utiliser par la suite.Tout d'abord la fréquence fondamentale estimée bien sur, ainsi qu'à chaque fois plusieurs candidats avec une mesure de la confiance qui leur est attachée. Ensuite nous faisons l'extraction d'une série d'autre paramètres qui vont pouvoir être utilisés lors de l'analyse des résultats.

Ce fichier contient dans l'ordre :
- la f_0 et un ensemble de candidats avec leur indice de confiance
- le facteur d'inharmonicité si nécessaire
- l'énergie du signal
- la centroïde
- le taux de *zero crossing*
- l'énergie et la fréquence des 5 principaux pics
- un indicateur de correction si il y a correction lors du post processing, et la différence relative entre la fréquence de l'erreur et de sa correction

1PRM	7	1	0.0522562
1FQ0	0x0004	5	2
446.14	0		
658.594	0.1513		
329.149	0.0992908		
293.784	0.0236407		
372.008	0.0591017		
1INC	0x0004	1	1
0.000500397			
1NRG	0x0004	1	1
19.8513			
1CTR	0x0004	1	1
1969.4			
1ZCR	0x0004	1	1
1PNR	0x0004	5	2
473.73	42.4568		
904.395	37.0464		
1335.06	30.3261		
581.396	30.0846		
344.531	28.8087		
1PER	0x0004	1	1
0			

TAB. *3.1:* Exemple d'analyse au format SDIF

3.3 Application

Nous avons pratiqué la comparaison sur un ensemble de fichiers audio correspondant à une seule note : cela étant, nous ne souhaitons évaluer que la capacité intrinsèque de la méthode sans regard pour des problèmes d'un autre ordre que nous susceptible de rencontrer sur des phrases musicales. Cela constitue donc une première étape de l'évaluation qui correspond dans notre classification proposée aux sons monophoniques seuls (note à note) de type harmoniques, pseudo harmonique, et inharmonique.

3.3.1 Description de la procédure

Toute notre programmation a été faite en MATLAB, et nous avons construit un ensemble de fonctions suffisamment souples pour pouvoir accepter facilement à l'avenir de nouveaux algorithmes, et d'autres références, par exemple des références faites à la main.

Pre-processing : Choix des Paramètres d'entrée

Lorsque la note est inscrite dans le nom du fichier, nous utilisons cette information pour déterminer les paramètres d'entrée des algorithmes, à savoir la plage de recherche de la f_0, la taille de fenêtre et le pas d'avancement de la fenêtre. Pour éviter toute erreur par rapport au système de notation (une octave de différence entre le système américain et le système français), le rang de recherche de la est entré un demi ton en- deça de la note et une octave et un demi ton au dessus de la note. Lorsque la note n'est pas accessible, nous utilisons une mesure approximative de la pour déterminer les paramètres d'entrée.

Les algorithmes tournent alors en parallèle sur ce fichier son pour en extraire la f_0.

Post- Processing : analyse des résultats obtenus

Une fois les résultats d'analyse obtenus pour notre base de données, nous procédons alors en plusieurs étapes :
- établissement de la référence par règle de la majorité et médiane
- calcul de l'écart entre chaque résultat et la référence
- tentative d'explicitation des erreurs par d'autres paramètres extraits du signal. Par exemple, si l'erreur se produit sur une attaque, sur un zone ou le signal devient très faible, si la sélection des pics a été mauvaise, etc...

3.3.2 Résultats

3.3.2.1 Pourcentage d'erreurs par famille d'instruments sur notre base de donnée

L'algorithme YIN est constitué d'une partie permettant la détection de pitch, c'est-à-dire d'une décision de voisement. Or lors de l'évaluation, nous avons constaté que dans un certain nombre de cas non négligeable, l'algorithme prend une décision non - voisée sans apparaître nécessaire pour autant puisque le reste des estimations aboutit quand même à une référence, c'est-à-dire que la majorité des algorithmes sont d'accord pour une fréquence fondamentale dans ce cas - ci. Cette particularité biasant nécessairement l'évaluation, nous avons décidé de présenter les résultats en deux parties, la première en faisant l'évaluation malgré cela sur le reste des algorithmes, mais en précisant le pourcentage de cette décision. La deuxième partie en ne menant cette fois-ci plus l'évaluation sans ces cas là.

Résultats avec la décision de voisement de YIN

Harmonique	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Errors %	0,2628	0,1566	15,2677	2,3710	1,5336	0,5837
«Octave» errors %	0,1307	0,0013	7,9723	0,2575	0,7830	0,5798

TAB. *3.2:* Synthèse des pourcentages d'erreurs et d'erreurs d'octave avec décision de voisement pour les instruments harmoniques

Harmonique	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Fine errors %	0,0518	0,0453	2,7282	0,9577	0,0505	0,0039
Fine errors too low %	0,0414	0,0336	0,0207	0,4776	0,0052	0,0000
Fine errors too high %	0,0104	0,0116	2,7075	0,4801	0,0453	0,0039
Gross errors %	0,2110	0,1113	12,54	1,4133	1,4831	0,5798
Gross errors too low%	0,0867	0,0039	11,2932	0,7895	0,0142	0,0000
Gross errors too high%	0,1242	0,1074	1,2463	0,6238	1,4689	0,5798
«Octave» errors %	0,1307	0,0013	7,9723	0,2575	0,7830	0,5798
«Octave» errors too low %	0,0155	0,0000	7,9723	0,2485	0,0013	0,0000
«Octave» errors too high %	0,1152	0,0013	0,0000	0,0091	0,7817	0,5798
«NaN» Errors %	0,0000	0,0000	0,0000	0,0000	0,0000	4,4054

TAB. *3.3:* Pourcentages de petites erreurs, de grosses erreurs et d'erreurs d'octave avec décision de voisement pour les instruments harmoniques

Piano	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Errors %	6,6001	6,2205	-	-	8,3471	3,2267
«Octave» errors %	4,1023	1,7076	-	-	4,9640	1,6070

TAB. *3.4:* Synthèse des pourcentages d'erreurs et d'erreurs d'octave avec décision de voisement pour les instruments pseudo-harmonique : piano

Piano	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Fine errors %	0,2899	1,5791	-	-	0,1182	0,3135
Fine errors too low %	0,0776	0,4918	-	-	0,0825	0,1273
Fine errors too high %	0,2122	1,0873	-	-	0,1862	0,1862
Gross errors %	6,3102	4,6414	-	-	8,2289	2,9532
Gross errors too low%	6,2902	3,9859	-	-	7,8220	2,4869
Gross errors too high%	0,0200	0,6555	-	-	0,4069	0,4663
«Octave» errors %	4,1023	1,7076	-	-	4,9649	1,6785
«Octave» errors too low %	4,0835	1,6100	-	-	4,8300	1,6070
«Octave» errors too high %	0,0188	0,0976	-	-	0,1340	0,0716
«NaN» Errors %	0,0000	0,0000	-	-	0,0000	2,3571

Tab. *3.5:* Pourcentages de petites erreurs, de grosses erreurs et d'erreurs d'octave avec décision de voisement pour les instruments pseudo-harmonique : piano

Cordes pincées	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Errors %	9,3065	2,7796	-	-	5,0366	1,7832
«Octave» errors %	7,9144	1,3930	-	-	2,09647	0,7025

Tab. *3.6:* Synthèse des pourcentages d'erreurs et d'erreurs d'octave avec décision de voisement pour les instruments pseudo-harmonique : cordes pincées

Cordes pincées	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Fine errors %	0,1185	0,4665	-	-	0,3912	0,2966
Fine errors too low %	0,0826	0,3260	-	-	0,1726	0,0735
Fine errors too high %	0,0358	0,1405	-	-	0,2185	0,2231
Gross errors %	9,1880	2,3131	-	-	4,6454	1,4866
Gross errors too low%	9,1871	2,2405	-	-	3,5389	0,4463
Gross errors too high%	0,0009	0,0725	-	-	1,1065	1,0404
«Octave» errors %	7,9144	1,3930	-	-	2,0964	0,7025
«Octave» errors too low %	7,9144	1,3921	-	-	1,7823	0,2782
«Octave» errors too high %	0,0000	0,0009	-	-	0,3140	0,4242
«NaN» Errors %	-	-	-	-	-	4

Tab. *3.7:* Pourcentages de petites erreurs, de grosses erreurs et d'erreurs d'octave avec décision de voisement pour les instruments pseudo-harmonique : cordes pincées

Percussion	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Errors %	22,7551	3,8150	-	28,6238	2,3014	0,0889
«Octave» errors %	14,4139	0,0000	-	0,1067	1,5192	0,0096

Tab. *3.8:* Synthèse des pourcentages d'erreurs et d'erreurs d'octave avec décision de voisement pour les instruments inharmoniques

Percussion	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Fine errors %	0,8355	2,8975	-	0,0315	0,0602	0,0479
Fine errors too low %	0,4239	0,7220	-	0,0191	0,0150	0,0109
Fine errors too high %	0,4116	2,1755	-	0,0123	0,0451	0,0369
Gross errors %	9,1880	2,3131	-	-	4,6454	1,4866
Gross errors too low%	21,9196	0,4034	-	27,9170	2,2384	0,0096
Gross errors too high%	0,0000	0,5141	-	0,7753	0,0027	0,0315
«Octave» errors %	14,4139	0,0000	-	0,1067	1,5192	0,0096
«Octave» errors too low %	14,4139	0,0000	-	0,0889	1,5192	0,0096
«Octave» errors too high %	0,0000	0,0000	-	0,0178	0,0000	0,0000
«NaN» Errors %	-	-	-	-	-	1,5000

TAB. *3.9:* Pourcentages de petites erreurs, de grosses erreurs et d'erreurs d'octave avec décision de voisement pour les instruments inharmoniques

Résultats sans la décision de voisement de YIN

Harmonique	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Errors %	0,1173	0,0543	9,97	1,0231	0,9823	0,4374
«Octave» errors %	0,0912	0,0000	5,45	0,0922	0,5426	0,4349

TAB. *3.10:* Synthèse des pourcentages d'erreurs et d'erreurs d'octave sans décision de voisement pour les instruments harmoniques

Harmonique	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Fine errors %	0,0019	0,0019	1,7443	0,5290	0,0039	0,0029
Fine errors too low %	0,4239	0,7220	0,0000	0,0191	0,0150	0,0109
Fine errors too high %	0,0000	0,0019	1,7433	0,3184	0,0029	0,0029
Gross errors %	0,1154	0,0524	8,2233	0,4941	0,9755	0,4349
Gross errors too low%	0,0233	0,0009	7,4070	0,2990	0,0029	0,0000
Gross errors too high%	0,0932	0,0514	0,8163	0,1951	0,9755	0,4349
«Octave» errors %	0,0912	0,0000	5,4502	0,0922	0,5426	0,4349
«Octave» errors too low %	0,0049	0,0000	5,4502	0,0922	0,0000	0,0000
«Octave» errors too high %	0,08647	0,0000	0,0000	0,0000	0,5426	0,4349

TAB. *3.11:* Pourcentages de petites erreurs, de grosses erreurs et d'erreurs d'octave sans décision de voisement pour les instruments harmoniques

Piano	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Errors %	4,7363	4,0960	-	-	5,8265	2,4499
«Octave» errors %	2,9844	1,1698	-	-	3,5957	1,2589

Tab. *3.12:* Synthèse des pourcentages d'erreurs et d'erreurs d'octave sans décision de voisement pour les instruments pseudo-harmonique : piano

Piano	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Fine errors %	0,1433	1,0074	-	-	0,0769	0,2351
Fine errors too low %	0,0559	0,3611	-	-	0,0550	0,0955
Fine errors too high %	0,0873	0,6463	-	-	0,0218	0,1396
Gross errors %	4,5930	3,0886	-	-	5,7496	2,2148
Gross errors too low%	4,5784	2,9171	-	-	5,6123	1,8651
Gross errors too high%	0,0145	0,1714	-	-	0,1373	0,3497
«Octave» errors %	2,9844	1,1698	-	-	3,5957	1,2589
«Octave» errors too low %	2,9703	1,1698	-	-	3,5065	1,2052
«Octave» errors too high %	0,0141	0,0000	-	-	0,0891	0,0537

Tab. *3.13:* Pourcentages de petites erreurs, de grosses erreurs et d'erreurs d'octave sans décision de voisement pour les instruments pseudo-harmonique : piano

Cordes pincées	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Errors %	5,1719	1,5509	-	-	3,2692	1,3374
«Octave» errors %	5,0391	0,85197	-	-	1,4607	0,5268

Tab. *3.14:* Synthèse des pourcentages d'erreurs et d'erreurs d'octave avec décision de voisement pour les instruments pseudo-harmonique : cordes pincées

Cordes pincées	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Fine errors %	0,0620	0,2700	-	-	0,2121	0,2224
Fine errors too low %	0,0379	0,1797	-	-	0,0937	0,0551
Fine errors too high %	0,0241	0,0902	-	-	0,1185	0,1674
Gross errors %	5,7099	1,2809	-	-	3,0571	1,1150
Gross errors too low%	5,7099	1,2506	-	-	2,3498	0,0335
Gross errors too high%	0,0000	0,0303	-	-	0,7072	0,7803
«Octave» errors %	5,0391	0,8519	-	-	1,4607	0,5268
«Octave» errors too low %	5,0391	0,8512	-	-	1,2637	0,2087
«Octave» errors too high %	0,0000	0,0006	-	-	0,1970	0,3182

Tab. *3.15:* Pourcentages de petites erreurs, de grosses erreurs et d'erreurs d'octave avec décision de voisement pour les instruments pseudo-harmonique : cordes pincées

Percussion	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Errors %	16,2325	2,5587	-	0,0144	20,6394	0,0667
«Octave» errors %	10,2627	0,0000	-	0,0564	1,1271	0,0072

Tab. *3.16:* Synthèse des pourcentages d'erreurs et d'erreurs d'octave avec décision de voisement pour les instruments inharmoniques

Percussion	f_0	f_0 **«pondéré»**	**ICF**	**Therardt**	**ROTFP**	**YIN**
Fine errors %	0,6143	2,0593	-	0,0144	0,0277	0,0359
Fine errors too low %	0,3056	0,4646	-	0,0072	0,0113	0,0082
Fine errors too high %	0,3087	1,5947	-	0,0072	0,0164	0,0277
Gross errors %	15,6182	0,4994	-	20,6250	1,6450	0,0308
Gross errors too low%	15,6182	0,2000	-	27,9170	2,2384	0,0096
Gross errors too high%	0,0000	0,5141	-	0,7753	0,0027	0,0315
«Octave» errors %	14,4139	0,0000	-	0,1067	1,5192	0,0096
«Octave» errors too low %	14,4139	0,0000	-	0,0889	1,5192	0,0096
«Octave» errors too high %	0,0000	0,0000	-	0,0178	0,0000	0,0000

TAB. *3.17:* Pourcentages de petites erreurs, de grosses erreurs et d'erreurs d'octave avec décision de voisement pour les instruments inharmoniques

3.3.2.2 Résultat global

FINAL	**Pourcentage d'erreur**	**Classement**
f_0	6.7150	4
f_0 «pondéré»	2.065	2
ROTFP	2.9377	3
YIN	**1.0676**	**1**

TAB. *3.18:* Synthèse des résultats sur l'ensemble de la base de données

Harmonique	**Pourcentage d'erreur**	**Classement**
f_0	0.1174	2
f_0 «pondéré»	**0.0543**	**1**
ROTFP	0.9823	4
YIN	0.4378	3

TAB. *3.19:* Synthèse des résultats dans le cas harmonique

Pseudo - harmonique : cordes	**Pourcentage d'erreur**	**Classement**
f_0	5.7739	4
f_0 «pondéré»	1.5509	2
ROTFP	3.2692	3
YIN	**1.3374**	**1**

Tab. *3.20:* Synthèse des résultats dans le cas pseudo-harmonique : cordes pincées

Pseudo - harmonique : piano	**Pourcentage d'erreur**	**Classement**
f_0	4.7363	3
f_0 «pondéré»	4.0960	2
ROTFP	5.8265	4
YIN	**2.4499**	**1**

Tab. *3.21:* Synthèse des résultats dans le cas pseudo-harmonique : piano

3.3.3 Discussion

3.3.3.1 Commentaire sur le filtre peigne

Le Comb Filter présente un grand nombre d'erreurs, ne serait-ce que pour le cas harmonique, et qui sont principalement des erreurs d' «octaves» trop basses. Ceci s'explique facilement par le fait que, pour une fondamentale donnée, la «sous - harmonique» explique le même spectre que celle-ci. Par ailleurs, le coût de calcul élevé de cet algorithme - et qui est dû au fait que la recherche de la fondamentale se fait par incrémentation avec un pas fréquentiel qui correspond avec la résolution fréquentielle souhaitée. Ce qui conduit à un grand nombre de calculs, et dans le cas pseudo harmonique, ce coût est multiplié par le nombre d'incrémentations nécessaire pour décrire l'espace des coefficients d'inharmonicité possibles. Donnons un exemple : On va rechercher la fondamentale sur un rang entre 200 Hz et 800 Hz, et l'on veut une résolution de 0.1 Hz sur l'estimation de la fréquence fondamentale. On a alors $\frac{800-200}{0.1} = 6000$ boucles à décrire. Si en plus, on recherche le coefficient d'inharmonicité entre 10^{-5} et 10^{-3} avec une résolution de 10^{-5}, ce qui est raisonnable pour une corde de piano, le nombre de calcul s'élève alors à $6000 \times 100 = 600000$ tours de boucle, ce qui n'est pas du tout satisfaisant. Une réponse à cet inconvenient a été présentée par Alexander Galembo, et consiste d'abors à déterminer des estimation grossières de la fréquence fondamentale, et si besoin, du coefficient d'inharmonicité, avant d'affiner l'estimation dans un second temps. Cette méthode présente l'avantage de réduire considérablement le temps de calcul, mais augmente nécessairement le nombre d'erreurs. Ce compromis n'est donc pas non plus satisfaisant.

Ceci explique l'abandon de cette méthode dans le reste de l'évaluation.

Inharmonique	**Pourcentage d'erreur**	**Classement**
f_0	16.2325	4
f_0 «pondéré»	2.5587	3
ROTFP	1.6727	2
YIN	**0.0454**	**1**

Tab. *3.22:* Synthèse des résultats dans le cas inharmonique

3.3.3.2 Commentaire sur le modèle de Terhardt

Pour des raisons similaires au Comb Filter, et malgré des résultats sensiblement meilleurs, le modèle de Terhardt a été abandonné lors de l'évaluation. Son évaluation sera finalisée ultérieurement.

3.3.3.3 Commentaires sur f_0 et f_0 pondéré

Forte diminution des erreurs

Comme le montre le tableau ci-dessous, la dernière version dite f_0 pondéré présente une diminution significative des erreurs par rapport à la version de référence, et ce dans tous les cas de figure.

Pourcentage d'erreur	f_0	f_0 **«pondéré»**
Harmonique	0.1174	0.0543
Pseudo - harmonique : piano	4.7363	4.0960
Pourcentage d'erreur : cordes	5.7739	1.5509
Inharmonique	16.2325	2.5587
TOTAL	**6.7150**	**2.0650**

TAB. *3.23:* Comparaison des résultats de f_0 et de f_0 pondéré

f_0 pondéré est biaisé vers le bas : les erreurs «d'octaves» sont à l'octave inférieure.

La plupart des erreurs occurant dans f_0 pondéré sont des erreurs trop basses, et particulièrement les erreurs d' «octaves» sont des erreurs de sous - «harmoniques», ce qui nous est montré par la distribution des erreurs ci - après. Ceci peut s'expliquer que les critères utilisés et notament les deux premiers, c'est-à-dire l'histogramme et la lissitude sont succeptible de favoriser les erreurs trop basses. En revanche, ils ne sauraient favoriser des erreur trop hautes : les sous - «harmoniques» expliquent le même spectre, les sur - «harmoniques», non.

Les erreurs sont inversées

Nombre d'erreurs en fonction du range - piano	**A0 / A1**	**A1 / A2**	**A2 / A3**	**A3 / A4**	**A4 / A5**	**A5 / A6**	**A6 / A7**	**A7 / A8**
f_0	1	8	0	51	809	1707	7022	808
f_0 **«pondéré»**	**640**	14	0	20	173	294	6114	399

TAB. *3.24:* Distribution des erreurs des algorithmes f_0 et f_0 «pondéré» en fonction du rang de hauteurs

3.3.3.4 Commentaire sur ROTFP et YIN

Les deux méthodes temporelles que nous avons utilisées, et bien sur plus particulièrement YIN, ont montré de manière générale un bon comportement lors de notre évaluation. Ceci peut *a priori* paraître étonnant dans le cas des instruments pseudo - harmoniques, et surtout inharmoniques pour lesquels ces méthodes ne devraient pas être optimale ; cependant, ce résultat s'expliquer par le fait simple que par exemple pour les instruments inharmoniques - de type percussion - les

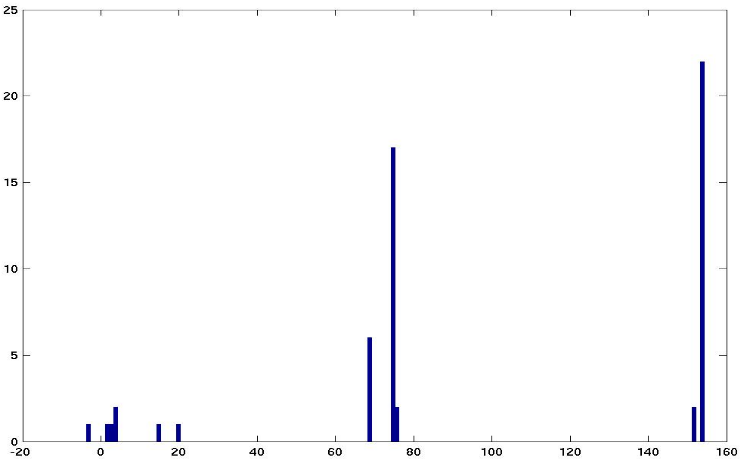

FIG. *3.10:* Distribution des erreurs de l'algorithme f_0 dans le cas harmonique

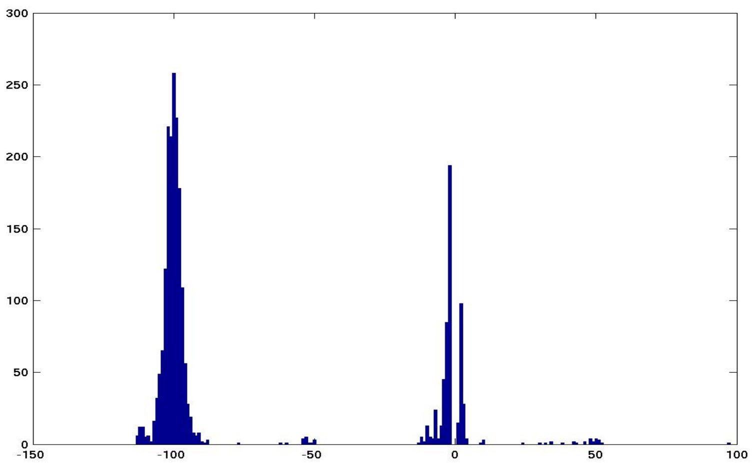

FIG. *3.11:* Distribution des erreurs de l'algorithme f_0 dans le cas pseudo-harmonique, cordes pincées

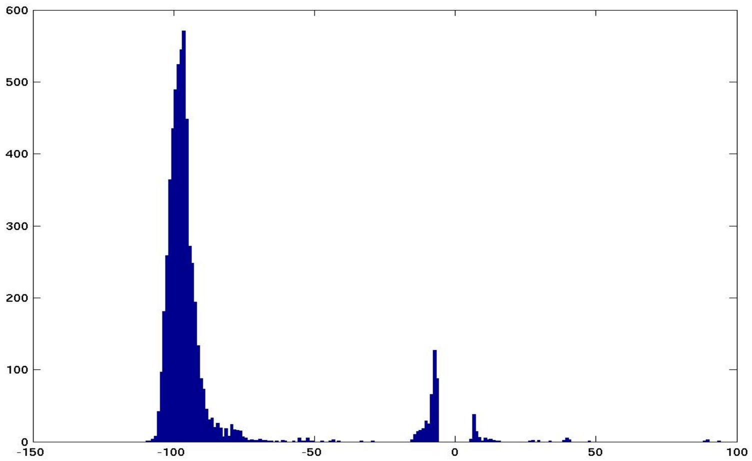

FIG. *3.12:* Distribution des erreurs de l'algorithme f_0 dans le cas pseudo-harmonique, piano

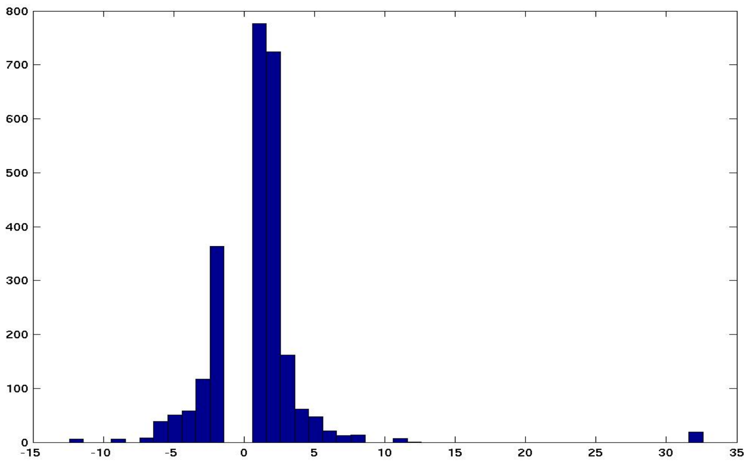

FIG. *3.13:* Distribution des erreurs de l'algorithme f_0 dans le cas inharmonique fréquences fondamentales sont relativement élevées, et par conséquent que le nombre de partiels est relativement faible, ce qui constitue un terrain favorable pour les méthodes employées. Il doit cependant être remarqué que le gros problème des méthodes temporel devrait apparaître lors de l'estimation de la fréquence fondamentale pour des phrases musicales, car les méthodes temporelles ne sont pas constiuée pour répondre aux phénomènes de réverberations lors du passage d'une note à l'autre, ou autrement dit pour faire de la séparation des constituants d'un spectre donné. YIN est systématiquement meilleur que ROTFP.

Pourcentage d'erreur	**ROTFP**	**YIN**
Harmonique	0.9823	0.4378
Pseudo harmonique	4.2778	0.8876
Inharmonique	1.6727	0.0454

Tab. *3.25:* Comparaison des résultats de YIN et de ROTFP

Conclusions

Nous venons de présenter un nouvel algorithme basé sur le modèle pseudo-harmonique (et développé en un temps record !) qui a montré de très bon résultats. Nous avons également développé tout un arrière plan flexible permettant l'évaluation des algorithmes d'estimation de la sur une large base de données musicale, et enfin avons procédé à une évaluation des algorithmes : notre algorithme a encore montré de très bons résultats.

A l'avenir, il serait avantageux de poursuivre le développement de cet algorithme en :
- améliorant l'extraction des pics par le critère de sinusoïdalité proposé par Axel Roebel : nous gagnerons à la fois en précision et en temps.
- développant une méthode de détermination couplée de la et du coefficient d'inharmonicité.
- permettant une estimation de f_0 multiples rendue possible par le modèle de Terhardt.

Par ailleurs, notre évaluation en cours de réalisation intéresse de plus en plus de monde, et nous pourrons ajouter très prochainement les algorithmes de Xavier Serra (Two Way Mismatch, UPF, Espagne) et de Chunghsin Yeh (IRCAM).

Perspectives et travaux futurs

Développements possibles de notre algorithme

Pre - processing
- Résolution :
    - Implémenter l'interpolation par régression en log amplitude, qui est le tout dernier outil d'interpolation [48]
    - Implémenter le modèle d'estimation de la fréquence et de l'amplitude des pics spectraux par convergence à partir d'une fenêtre sans sidelobe proposé par Thomas Hélie [16]
- Extraction des pics :

Deux possibilités :
- Utiliser le critère de sinusoidalité des pics spectraux pour améiorer l'extraction des pics [67]
- Ajouter au modèle de Terhardt les critères psycho- acoustique, notamment les facteurs de masques qui permettent de discriminer les différents pics spectraux [61]

Modèle de Terhardt
- Permettre l'estimation de f_0 multiples dans le cas simple d'un instrument inharmonique à plusieurs f_0.

Least Square Curve Fitting
- La résolution de l'estimation du coefficient d'inharmonicité dépend de deux choses : la résolution sur l'estimation des partiels, et le choix des partiels pour entrée du least square.
- Le premier est une question de résolution fréquentielle, ce qui nous renvoie aux solutions présentées précedemment.
- On peut améliorer la sélection des partiels en incluant à l'algorithme un modèle d'enveloppe spectrale, celle du piano par exemple.

Développements possibles de l'évaluation
- On peut à partir des références déterminées lors de notre étude s'engager dans une étude de l'influence des paramètres sur l'estimation de la f_0. Dans un premier temps, nous allons relancer une évaluation en prenant cette fois- ci comme bornes fréquentielles de recherche un espace de 3 octaves centré autour de notre référence, ce qui va nous permettre une évaluation plus réaliste des algorithmes, en incluant notamment la possibilité d'erreurs d'»octave» qui constituent la plus grande source de grosses erreurs lors de l'estimation.
- Poursuivre notre démarche en constituant une base de données de phrases monodiques, et donc étudier par cela des problèmes d'un autre ordre rencontrés lors de l'évaluation de la , à savoir essentiellement les problèmes liés à la réverbération, à l'entrechevêtrement de fenêtre d'analyse, ainsi qu'au problème de résolution temporelle.

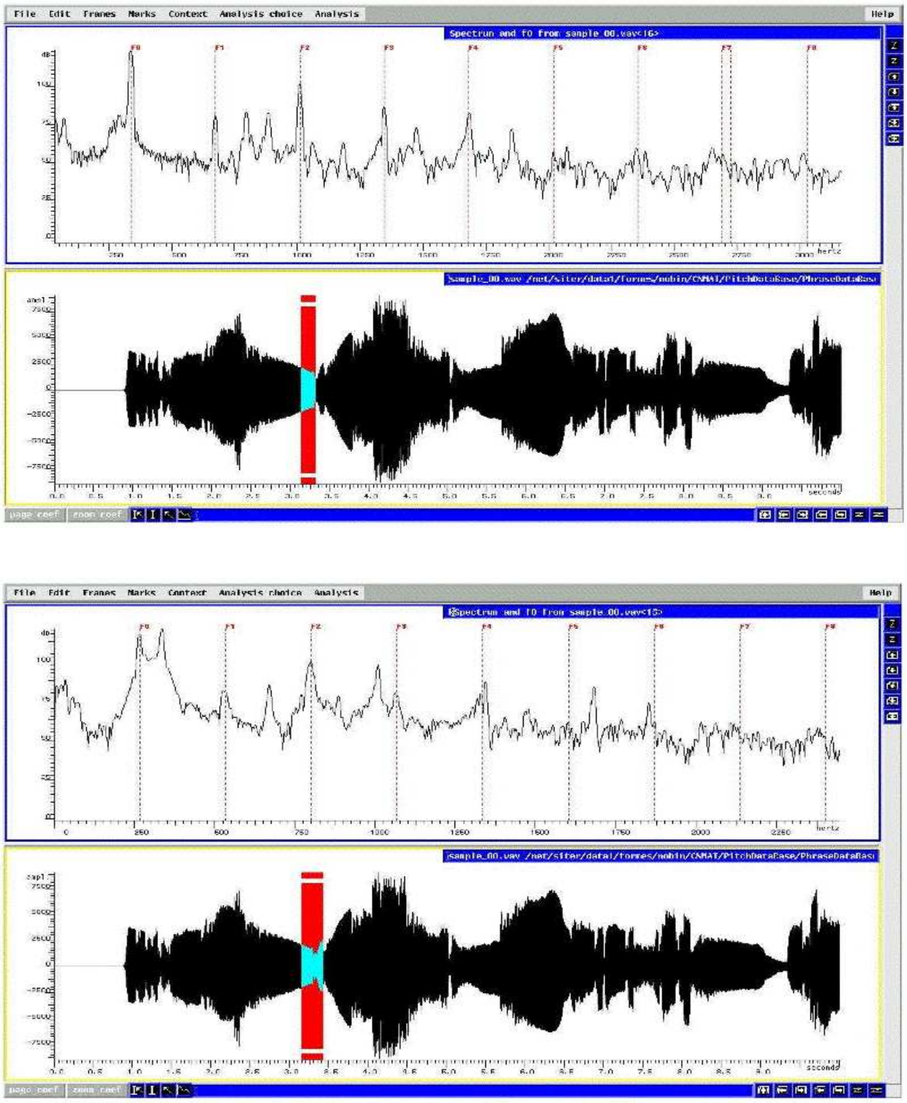

FIG. *3.14:* Exemple de problème de réverbération ou de fenêtre à cheval sur deux notes : le spectre montre clairement deux spectres superposés

Fréquences des notes sur l'échelle tempérée

Note	**Fréq. (Hz)**	**Note**	**Fréq (Hz)**	**Note**	**Fréq. (Hz)**	**Note**	**Fréq. (Hz)**
C0	16.35	C2	65.41	C4	261.63	C6	1046.50
Db0	17.32	Db2	69.30	Db4	277.18	Db6	1108.73
D0	18.35	D2	73.42	D4	293.66	D6	1174.66
Eb0	19.45	Eb2	77.78	Eb4	311.13	Eb6	1244.51
E0	20.60	E2	82.41	E4	329.63	E6	1318.51
F0	21.83	F2	87.31	F4	349.23	F6	1396.91
Gb0	23,5	Gb2	92.50	Gb4	369.99	Gb6	1479.98
G0	24.50	G2	98.00	G4	392.00	G6	1567.98
Ab0	25.96	Ab2	103.83	Ab4	415.30	Ab6	1661.22
A0	27,5	A2	110.00	**A4**	**440.00**	A6	1760.00
Bb0	29.14	Bb2	116.54	Bb4	466.16	Bb6	1864.66
B0	30.87	B2	123.47	B4	493.88	B6	1975.53
C1	32.70	C3	130.81	C5	523.25	C7	2093.00
Db1	34.65	Db3	138.81	Db5	554.37	Db7	2217.46
D1	36.71	D3	146.83	D5	587.33	D7	2349.32
Eb1	38.89	Eb3	155.56	Eb5	622.25	Eb7	2489.02
F1	43.65	F3	174.61	F5	698.46	F7	2793.83
Gb1	46.25	Gb3	185.00	Gb5	739.99	Gb7	2959.96
G1	49.00	G3	196.00	G5	783.99	G7	3135.96
Ab1	51.91	Ab3	207.65	Ab5	830.61	Ab7	3322.44
A1	55.00	A3	220.00	A5	880	A7	3520.00
Bb1	58.27	Bb3	233.08	Bb5	932.33	Bb7	3729.31
B1	61.74	B3	246.94	B5	987.77	B7	3951.07

Bibliographie

[1] Sassan Ahmadi, «Cepstrum-based pitch detection using a new statistical V/UV classification algorithm», IEEE, 1999.

[2] Jean Paul Berrut, «Barycentric Lagrange Interpolation», dans Society for Imdustrial and applied Mathematics, 2004.

[3] Tim Black, «Pitch determination of music signals using generalized spectrum», IEEE, 2000.

[4] J. Frederic Bonnans, «Optimisation numérique : Aspects théoriques pratiques », Springer Verlag, 1997.

[5] N.M. Botros, «Automatic estimation of speech pitch period», International Journal of Modelling and Simulation, 1998.

[6] J.C. Brown «Calculation of a constant-Q spectral transform», JASA, 1991.

[7] Pedro Cano, «Fundamental Frequency Estimation in SMS Analysis», dans DAFx, 1998.

[8] Alain de Cheveigné, «YIN, a fundamental frequency estimator for speech and Music», dans Journal of Acoustic Society of America, 2002.

[9] Alain de Cheveigné, Hideki Kawahara, «Compararive evaluation of f_0's estimation algorithms», dans Eurospeech, 2001.

[10] Alain de Ceheveigné, «Pitch perception models - a historical review».

[11] Patricio de la Cuadra, Aaron Master, Craig Sapp, «Efficient pitch detection techniques for interactive music».

[12] Huanping Daï, «On the relative influence of individual harmonics on pitch judgment», ans JASA, 2000.

[13] John E Dennis, David M Gay, Roy E Welsch, «An adaptative nonlinear least square algorithm», dans Transaction on Matehmatical Software, 1981.

[14] Phillipe Depalle, G Garcia, Xavier Rodet, «Tracking of partials for additive sound synthesis using HiddenMarkov Models», dans IEEE, 1993.

[15] Phillipe Depalle, Xavier Rodet, «Tracking of partilals for additiveÊsound synthesis using Hidden Markov Models», dans IEEE, 1993.

[16] Philippe Depalle, Thomas Helie, «Extraction of spectral peak parameters using a short-time Fourier transform modeling and no sidelobe window».

[17] Boris Doval, Xavier Rodet, «Estimation of fundamental frequency of musical sound signals», dans IEEE, 2004.

[18] Boris Doval, Xavier Rodet, «Fundamental frequency estimation and tracking using maximum likelihood harmonic matching and HMM's», dans IEEE, 1993.

[19] Nicolas Durand, Emilia Gomez, «Periodicity analysis using a Harmonic Matching Method and Bandwise processing».

[20] Neville H Fletcher, Thomas D Rossing, «The Physics of Musical Instrument», Sringer Verlag, 1988.

[21] David Friedman, «Multichannel zero-crossing-interval pitch estimation», IEEE, 1979.

[22] F. Fritsch, «Monotone piecewise cubic interpolation», dans Society for Industrial and Applied Mathematics, 1980.

[23] Alexander Galembo, Anders Askenfelt, «Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano», IEEE, 1999.

[24] David Gerhard, «Pitch extraction and fundamental frequency : history ans current techniques», Technical report.

[25] Simon Godsill, «Bayesian harmonic models for musical pitch estimation and analysis», IEEE, 2002.

[26] Masataka Goto, «A real-time music-scene description system : predominant f_0 estimation for detecting melody and bass lines in real-world audio signals», in Speech Communication, 2004.

[27] Yutaka Goto, «Highly accurate interpolation of apolized transformée de Fourier magnitude mode spectrum», dans Applied Spectroscopy, 1998.

[28] Rémi Gribonval, Phillipe Depalle, Xavier Rodet, E Bacry, Stephane Mallat, «Sound signals decomposition using a high resolution matching pursuit».

[29] Norden E. Huang, «The empirical mode decomposition and the Hilbert spectrum for non linear and non stationnary time series analysis», in The Royal society of London, 1998.

[30] Hanna Jarvelainen, «The effect of inharmonicity on pitch in string instrument sounds».

[31] Hanna Jarvelaïnen, «Perception and adjustment of pitch in inharmonic string instrument tones».

[32] Andrei Jefremov, Bastiaan Kleijn, «Spline-based continuous-time pitch estimation», dans IEEE, 2000.

[33] Xudon Jiang, «Fundamental frequency estimation by high order spectrum», IEEE, 2000.

[34] Hideki Kawahara, Alain de Cheveigné, «Fixed point analysis of frequency to instataneous frequency mapping for accurate estimation of and periodicity», in Eurospeech, 1999.

[35] Martin Kean, «Understanding the complex nature of piano tone», 2004.

[36] Azar Khurshid, «A temporal analysis based pitch estimation system for noisy speech with comparative study performance», in IEEE, 2004.

[37] Anssi Klapuri, «Wide- band pitch estimation for natural sound sources with inharmonicities».

[38] Anssi Klapuri, «Multiple fundamental frequency estimation based on harmonicity and spectral smoothness», in IEEE, 2003.

[39] Sam Kwong, «A pitch detection algorithm based on time-frequency analysis», in ICC, 1992.

[40] Jian-Yu Lin, William M Hartmann, «The pitch of mistuned harmonics : Evidence for a template model», dans Journal of Acoustic Society of America, 1998.

[41] Robert C Maher, James W Beauchamp, «Fundamental frequency estimation of musical signals using a two-way mismatch procedure», dans Journal of Acoustic Society of America, 1994.

[42] Stephane Mallat, «A theory of multiresolution signal decomposition : the wavelet representation», IEEE, 1989.

[43] Sylvain Marchand, «Modélisation informatique du son musical - anlayse, transformation, synthèse», Thèse, Université de Bordeaux.

[44] Philippe Martin, «Peigne et brosse pour f_0 : Mesure de la fréquence fondamentale par alignement de spectres séquentiels», in 23eme journées d'étude sur la parole, 2000.

[45] Asuncion Moreno, «Pitch determination of noisy speech using high order statistics», in IEEE, 1992.

[46] Kazuki Nishi, «Optimum filter for quasi-periodic signals», dans Electronic and Communication in Japan, 2003.

[47] A.M Noll, «Cepstrum pitch determination», JASA, 1967.

[48] Geoffroy Peeters, Thèse de l'université Pierre et Marie Curie, Paris 6.

[49] Lungi Qiu, «Fundamental frequency determination based on instantaneous frquency estimation», in Signal processing, 1995.

[50] Lawrence Rabiner, «A tutorial on hidden Markov Models and selected application in speech recognition», dans IEEE, 1989.

[51] Lawrence Rabiner, «A comparative study of several pitch detection algorithms», dans IEEE, 1976.

[52] Lawrence Rabiner, «On the use of autocorrelation analysis for pitch detection», dans IEEE, 1979.

[53] Lawrence Rabiner, «Applications of a non alinear smoothnig algorithm to speech processing» dans IEEE, 1975.

[54] Salima Abdul Samad, «Pitch detection of speech signals using the cross-correlation technique», in IEEE, 2000.

[55] David Schwartz, «Pitch is determined naturally occuring periodic sounds».

[56] Xavier Serra, «Musical sound modelling with sinusoid plus noise», 1997.

[57] Julius O Smith, «Mathematics of the discrete Fourrier transform», Wok, 2003.

[58] Saurabh Sood, Ashok Krishnamurthy, «A Robust On-The-Fligth Pitch Estimation Algorithm», dans MM, 2004.

[59] Joseph Tabrikian, «Maximum a-posteriori probability pitch tracking in noisy environments using harmonic model», dans IEEE, 2004.

[60] Yoshiaki Tadokoro, «Pitch detection of musical sounds using adaptative comb filters controlled by time delay», dans IEEE, 2002.

[61] Ernst Terhardt, «Calculating virtual pitch», dans Hearing Research, 1979.

[62] Tero Tolonnen, Matti Karjalainen, «Modeling of tension modulation nonlinearity in plucked Cordes pincées», dans IEEE, 2004.

[63] V. Verfaille, «LIFT : Likelihood-time-analysis for partial tracking and automatic transcription of music», Dafx, 2001.

[64] James D Wise, «Maximum Likelihood Pitch Estimation», dans IEEE, 1976.

[65] Chunghsin Yeh, Axel Roebel, « A new score function for joint evaluation of multiple », dans DAFx, 2004.

[66] A Zilinkas, «A review of statistical Models for global optimization», dans Journal of Global Optimization, 1993.

[67] Miroslav Zivanovic, Axel Roebel, Xavier Rodet, «A new approach to spectral peak classification», dans DAFx, 2004.

[68] Parham Zolfaghari, Hideki Kawahara, «A sinusoidal model based on frequency to instantaneous frequency mapping».