

# Rapport d'avancement: l'estimation de F0s multiples

Chunghin Yeh et Axel Röbel 15/05/2006

## Description de tâche

L'estimation de fréquences fondamentales, dite F0s, est une recherche reliée à deux tâches de MusicDiscover: (1) reconnaissance des instruments de musique et indexation et (2) séparation de sources. Cette recherche s'agit de développer un système d'estimer F0s multiples en contexte polyphonique, qui fournira des *features* concernant des sources quasi-harmoniques pour faciliter la reconnaissance des instruments de musique, la séparation de sources, etc.

## Système proposé

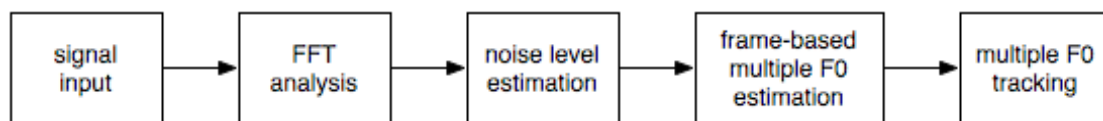


Fig.1 Aperçu du système proposé pour l'estimation de F0s

Le système proposé est montré dans la figure ci-dessus. L'algorithme développé est basé sur l'analyse dans le domaine fréquentiel. Pour estimer le nombre de sources harmoniques, un algorithme modéliser le résidu est développé qui caractérise l'importance de chaque pic spectral. Nous avons proposé une fonction *score* pour l'estimation de F0s multiples avec une hypothèse du nombre de F0s. Le travail actuel se focalise sur le développement d'un algorithme qui sert à déterminer le nombre de sources harmoniques. Une fois des F0s multiples sont estimées, un algorithme tracking sera essentiel pour corriger des erreurs et de la discontinuité des trajectoires F0s.

## Algorithmes développés

### 1. Estimation adaptative du niveau bruit

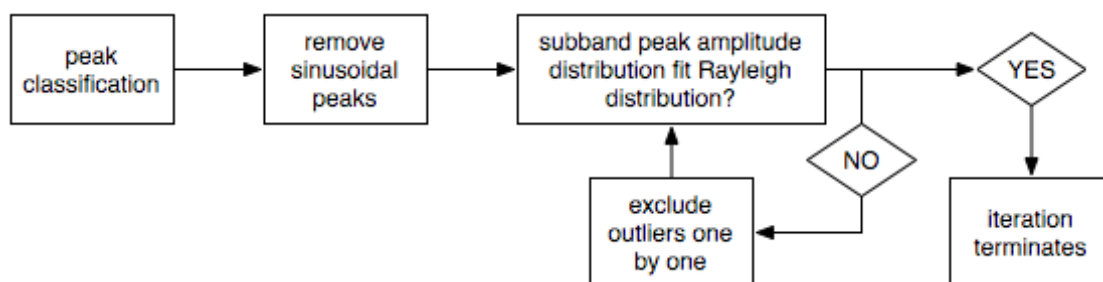


Fig.2 L'estimation du niveau bruit

Nous avons choisi une approche classique qui estime d'abord des pics sinusoïdaux et ensuite les soustrait en obtenant un spectre résiduel qui peut définir le niveau bruit. Nous avons utilisé une méthode récente [Röbel04] pour la classification de pics sinusoïdaux/non-sinusoïdaux. Comme cet algorithme vise aux signaux polyphoniques, il existe des pics superposés desquels on n'observe pas de l'information fiable. Pour raffiner le résultat de la classification de pics, on se fie à

modéliser la distribution d'amplitude du bruit dans une bande étroite. Nous avons choisi la distribution Rayleigh pour développer un algorithme itératif qui estime le niveau du bruit. Le principe est d'exclure des pics dont les amplitudes ne correspondent pas à la distribution Rayleigh. Le niveau du bruit est défini par le cepstre lissé du spectre résiduel.

## 2. Estimation de F0s multiples en supposant que le nombre de F0s visées est connu

L'algorithme de l'estimation de F0s multiples développé est guidé par trois principes physiques: (I) *spectral matching* avec bonne harmonicité, (II) l'enveloppe spectrale d'une source harmonique est lisse pour la plupart [Fletcher98], et (III) la synchronisation de l'évolution de partiels d'une source.

Les candidates F0s sont sélectionnées par *harmonic matching* qui est appliqué souvent aux signaux monophoniques. La combinaison linéaire de quatre critères forme une fonction score qui mesure la plausibilité des candidates F0s: l'accord entre spectre observé et spectre attendu, la "douceur" (*smoothness*) de l'enveloppe spectrale qui résulte du spectre observé et des hypothèses F0s, la concentration d'énergie d'une séquence harmonique et la variance du centre de gravité des pics [Cohen95] appartenant à une des candidates. Les paramètres de pondération parmi les quatre critères sont optimisés par l'algorithme génétique.

En prenant en compte des partiels superposés, on a développé une stratégie fondée sur Principe II pour attribuer ces composants aux meilleures candidates. Cette stratégie est très importante pour résoudre l'ambiguïté de pics observés.

## 3. Tracking de F0s multiples dans des enregistrements des instruments monodiques

À cause de réverbération, des enregistrements des instruments monodiques sont presque toujours polyphoniques. Sous l'hypothèse que l'un instrument monodique ne produit qu'une note à chaque l'instant, nous avons développé un algorithme pour le tracking de F0s multiples.

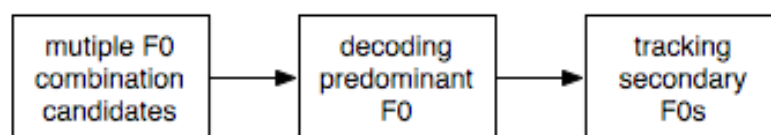


Fig.3 Tracking de F0s multiples

Toutes les hypothèses du nombre de F0s sont prises en compte jusqu'à un nombre maximal. Ce nombre maximal de F0s est déterminé par l'amélioration de score quand on évalue la plausibilité d'une hypothèse à l'une autre. Si l'amélioration de score ne dépasse pas un seuil probabiliste, l'évaluation de F0s multiples termine pour la trame actuelle. Après avoir obtenu toutes les combinaisons possibles, la trajectoire de F0s principales est construite selon deux propriétés: (1) la probabilité individuelle d'une F0, et (2) la continuité de la trajectoire. Enfin, la partie réverbérée est estimée par la prolongation de trajectoires de F0s principales.

## Résultats

### 1. Une base de donnée pour évaluer l'estimation de F0s

Quatre ensembles d'échantillons polyphoniques sont préparés pour évaluer notre système. TWO, THREE, FOUR et FIVE correspondant aux mixtures de deux notes, trois notes, quatre notes et cinq notes sont générées en mixant des échantillons monophoniques de l'Université McGill, l'Université Iowa et l'IRCAM. Les notes de 65Hz à 1980Hz sont sélectionnées sémi-aléatoirement pour le mixage en assurant que la probabilité de choisir un des 12 tons chromatiques soit égale. Cette base de données contient environ 30 instruments de musique. Les F0s des échantillons monophoniques sont d'abord estimées à donner des références F0s pour des échantillons polyphoniques. Nous avons publié cette base de données sur le site MusicDiscover.

2. L'évaluation de l'algorithme proposé dans le cas que le nombre de F0s est connu

window size	TWO	THREE	FOUR	FIVE
186ms	0.93%	3.13%	4.46%	7.60%
93ms	1.82%	5.10%	8.90%	13.84%

Pour un nombre de F0s limité, pas plus de 5, les résultats de l'algorithme sont prometteurs. En fait, cette base de données a été créée, suivant la description de la thèse de A. Klapuri, pour pouvoir comparer nos résultats avec les siens.[Klapuri03] Comme la sélection des sons utilisés pour l'évaluation est aléatoire, la comparaison avec les chiffres de A. Klapuri doit être interprétée avec précaution. Mais nous avons constaté que notre algorithme se compare plutôt favorablement avec le sien.

3. Quelques enregistrements polyphoniques simples ont été testés. Les exemples transcrit en format midi sont montrés sur la page perso de Chunghsin Yeh.

4. Grâce aux logiciels "Alignement de Partition" [Rodet04] et "AudioSculpt", plusieurs fichiers MIDI de *RWC Classical* sont alignés aux enregistrements réels. Ils sont accessibles du site MusicDiscover.

5. Collaboration avec A. Livshin pour la reconnaissance des instruments de musique.

6. Publication

[1] C. Yeh, A. Röbel, and X.Rodet, "*Multiple fundamental frequency estimation of polyphonic music signals*", IEEE ICASSP, pp. 225-228 (Vol. III), Philadelphia, Pennsylvania, USA, 2005.

[2] C. Yeh and A. Röbel, "*A new score function for joint evaluation of multiple F0 hypothesis*", Proc. of the 7th Int. Conf. on Digital Audio Effects (DAFx'04), pp. 234-239, Naples, 2004.

[3] C. Yeh and A. Röbel, "*Physical principles driven joint evaluation of multiple F0 hypotheses*", Proc. ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing SAPA'04, pp. , Jeju, 2004.

[4] C. Yeh and A. Röbel, "*Adaptive noise level estimation*", Workshop on Computer Music and Audio Technology (WOCMAT'06), Taipei, 2006.

[5] C. Yeh, A. Röbel, and X. Rodet, "*Multiple F0 tracking in solo recordings of monodic instruments*", 120th AES Convention, Paris, France, 2006.

## **Références**

[Fletcher98] N. F. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer-Verlag, 1998.

[Cohen95] L. Cohen, *Time-frequency analysis*, Prentice Hall, 1995.

[Klapuri03] A. Klapuri, "*Multiple fundamental frequency estimation by harmonicity and spectral smoothness*", *IEEE Trans. Speech and Audio Processing*, 11(6), 804-816, 2003.

[Röbel04] A. Röbel, M. Zivanovic, and X. Rodet, "*Signal decomposition by means of classification of spectral peaks*", *Proc. Int. Computer Music Conference (ICMC'04)*, pp. 446-449, Miami, 2004.

[Rodet04] X. Rodet, J. Escribe, S. Durigon, "*Improving score to audio alignment: Percussion alignment and Precise Onset Estimation*", *ICMC*, 2004.

[Rodet06] H. Kaprykowsky and X. Rodet, "*Globally Optimal Short-Time Dynamic Time Warping Application to Score to Audio Alignment*", *IEEE ICASSP*, Toulouse, France, 2006.

[Bogaards04] N. Bogaards, A. Roebel, X. Rodet: *Sound Analysis and Processing with AudioSculpt 2*, *International Computer Music Conference (ICMC)*, Miami, 2004.