

Rapport d'avancement de l'estimation de F0s multiples

Chunghin Yeh and Axel Röbel 11/07/2005

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.. Nous avons divisé ce travail en trois étapes:

- 1) Estimer F0s en supposant que le nombre de F0s visées est connu.
- 2) Estimer le nombre de F0s dans une trame donnée.
- 3) Intégrer l'information fréquentielle-temporelle pour compléter un système de *MultiF0 tracking*.

Algorithme développé

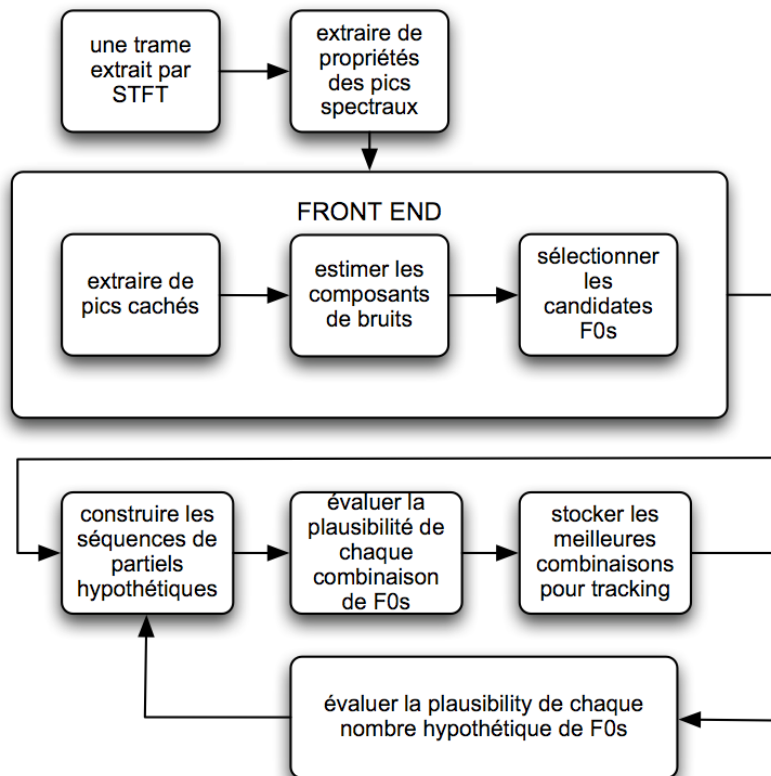


Fig.1 Aperçu du système développé pour estimer F0s dans une trame

L'algorithme de l'estimation de F0s multiples développé est guidé par trois principes physiques: (I) *spectral matching* avec bonne harmonicité, (II) des informations a priori sur les spectres des sources sonores naturelles [Fletcher98] et (III) la synchronisation de l'évolution de partiels d'une source. On résume ci-dessous les processus pour estimer des F0s dans une trame donnée pas par pas:

- 1) Extraire de propriétés de pics spectraux comme fréquences, amplitudes, phase, *bandwidth*, *duration* et *mean time*. [Cohen95]
- 2) Élimination provisoire des pics qui ne sont pas de formes symétriques et n'ont pas

de bonne sinusoïdalité pour extraire des pics cachées par ceux qui situent dans les voisinages.

3) Fondé sur le model auditif et la classification de pics sinusoïdaux/non-sinusoïdaux [Röbel04], on pourrait estimer des composants de bruits. En utilisant le *true envelope* [Röbel05], on pourrait également estimer le *noise floor* qui joue un rôle important dans l'estimation du nombre de F0s.

4) Extraire les candidates F0s en appliquant directement l'appariement harmonique

5) 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. D'ailleurs, il nous permet de définir l'énergie effective de chaque F0 dans une combinaison, qui caractérise la probabilité d'un F0 dans un ensemble hypothétique.

6) 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 appartenant à une des candidates. Les paramètres de pondération parmi les quatre critères sont optimisés par l'algorithme génétique.

La prise en compte de la fiabilité des régions du spectre observé a été étudiée. L'hypothèse est que les régions où plusieurs partiels sont superposés doivent être traitées de façon spéciale pendant l'évaluation. Comme chaque évaluation fait l'hypothèse d'un ensemble de F0s, l'ensemble correct devrait permettre de mieux exclure les régions peu fiables et donc devrait recevoir le score le plus grand.

Résultat

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

Trois 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 donnée 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 transcrits en format midi sont montrés sur la page perso de Chunghsin Yeh.

4. 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.

Travail en cours

L'estimation du nombre de F0s

Nous sommes en train de travailler sur ce problème difficile. L'idée est à observer la corrélation entre la variation de score sous les hypothèses des nombres de F0s et le *noise floor*.

F0 tracking

En l'état actuel, l'algorithme ne prend en compte que les informations d'une seule trame. Un des plus grands problèmes rencontrés est la variation naturelle des spectres instrumentaux avec parfois des résonances fortes qui engendrent des enveloppes spectrales avec des variations importantes. Un algorithme de tracking est donc nécessaire pour assurer la continuité d'une note et aussi pour corriger les *onsets*.

Références

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

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

[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.

[Röbel05] A. Röbel and X.Rodet, "*Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation*", to appear in Proc. of the 8th Int. Conf. on Digital Audio Effects (DAFx'05), Madrid, Spain, 2005.