MUSICDISCOVER

1. Tâche 2. Reconnaissance des instruments de musique et indexation

1.1. Estimation de F0s multiples

1.1.1. 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. La recherche s'agit de développer un système d'estimer F0s multiples en contexte polyphonique, qui facilitera l'extraction des *features* concernant des sources quasi-harmoniques pour la reconnaissance des instruments de musique ou la séparation de sources.

1.1.2. Méthodes proposée

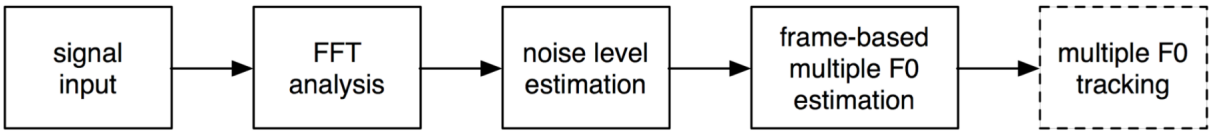

Fig.1 Aperçu du système proposé pour l'estimation de F0s

L'algorithme développé est basé sur l'analyse dans le domaine fréquentiel, c'est-à-dire, les F0s multiples vont êtres extraits à partir de la Transformée Fourier à Court Terme (TFCT). Nous représentons le modèle de signal comme une somme de signaux harmoniques plus le bruit. Un signal harmonique est représenté par une somme de sinusoïdes. Nous considérons les pics spectraux comme les composants spectraux qui seront expliqués par le modèle de signal. L'idée est à évaluer les plausibilités de toutes les combinaisons possibles parmi les hypothèses F0s et la combinaison qui *expliquent* le meilleur le spectre observé devrait être classée dans la tête.

Le système d'abord modéliser le résidu par une estimation du niveau du bruit qui caractérise l'importance de chaque pic spectral. Les pics spectraux au-dessus du niveau du bruit sont plus probables êtres sinusoïdes et ceux qui sont en dessous du niveau du bruit sont plus probables êtres bruit. Le modèle harmonique devrait accorder bien avec les pics sinusoïdaux. Nous avons proposé une fonction *score* pour évaluer une combinaison des hypothèses F0s. Le nombre de F0s se dépend des expliqués et la douceur de l'enveloppe spectrale de sources hypothétiques.

1.1.2.1. Estimation adaptive du niveau du bruit

Nous avons choisi une approche classique qui estime d'abord des pics sinusoïdaux et ensuite les soustrait en obtenant un spectre résiduel qui pourrait définir le niveau du bruit.

Nous avons utilisé une méthode récente [Röbel04] pour la classification de pics sinusoïdaux / non-sinusoïdaux. La méthode a l'avantage d'identifier non seulement les sinusoïdes stationnaires mais aussi les sinusoïdes modulés. Cependant, les signaux de musique sont souvent polyphoniques dont il existe des pics superposés desquels on n'observe pas de l'information fiable et la classification de pics spectraux reste à être corrigé. Pour raffiner le résultat de la classification de pics, on se fie à modéliser la distribution d'amplitude du bruit dans une bande étroite par la distribution Rayleigh. Le niveau du bruit est donc caractérisé par une succession des distributions Rayleigh à traverse la fréquence. Un algorithme itératif a été développé pour approcher le niveau du bruit par la vérification de la distribution et la re-classification de pics par le nouveau niveau du bruit [Yeh06b]. Le principe est à exclure des pics "*outlier*" dont les amplitudes ne correspondent pas à la distribution Rayleigh. Le niveau du bruit est défini par le cepstre lissé du spectre résiduel [Qi97]. L'avantage de cette approche est qu'elle ne dépend ni de l'observation à traverse plusieurs trames, ni de l'analyse harmonique.

1.1.2.2. Évaluation conjointe des hypothèses F0s

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 des instruments de musique [Fletcher98], et (III) la synchronisation de l'évolution de partiels d'une source.Ces principes correspondent aussi aux principes de l'analyse de scènes auditifs pour la ségrégation de sources sonores [Bregman90].

Une fonction score a été développée pour évaluer conjointement un ensemble d'hypothèses F0s [Yeh05] [Yeh04b] [Yeh04a]. Pour chaque combinaison des hypothèses F0s, nous construisons leurs Séquences Partielles Hypothétiques (HPS) par la sélection d'harmonique et le traitement de partiels superposé. 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.Afin de mesurer la plausibilité d'une combinaison d'hypothèse F0s, nous avons formulé les principes physiques en quatre critères: (1) l'accord entre spectre observé et spectre attendu, (2) la "douceur" (*smoothness*) de l'enveloppe spectrale qui résulte du spectre observé et des hypothèses F0s, (3) le centroïd spectral d'une source harmonique et (4) 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.

La fonction score a été évaluée dans le cas que le nombre de F0s est connu. Pour un nombre de F0s limité, pas plus de 5, les résultats de l'algorithme sont prometteurs. En fait, une base de données a été créée, suivant la description de l'article de A. Klapuri [Klapuri03], pour pouvoir comparer nos résultats avec les siens. 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.

1.1.2.3. Estimation du nombre de F0s

Un ensemble de F0s à estimer pourrait être catégorisé en deux: les F0s non-harmoniques et les F0s harmoniques. Une F0 harmonique pourrait être reliée par une autre F0 avec un multiple entier.Par contre, les F0s non-harmoniques se sont reliées par un rapport

rationnel qui n'est pas un entier. L'extraction de F0s non-harmoniques se dépend du niveau du bruit et l'extraction de F0s harmonique se dépend de l'enveloppe spectrale d'une source hypothétique. Nous se fions l'extraction de F0s non-harmoniques au niveau du bruit. Une fois le RSB du résidu devient plus petit qu'un seuil préfixé, l'algorithme pourrait arrêter d'ajuter l'hypothèse F0 non-harmonique. Comme notre approche ne dépend pas de la modélisation d'instruments de musique, nous se fions l'extraction de F0s harmoniques au critère de la fonction score qui évalue l'amélioration de la douceur de l'enveloppe spectrale pour l'ensemble des hypothèses F0s. Cette amélioration de la douceur ne doit pas dépasser l'amélioration dans le cas de lisser l'enveloppe spectrale d'une note d'un instrument de musique. Les seuils de l'amélioration sont appris pour chaque note et moyennés pour une collection des instruments de musique.

1.1.2.4. Tracking de F0s 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 [Yeh06a]. 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.

1.1.3. Base de donnée polyphonique

1.1.3.1. Mixtures artificielles d'échantillons de notes des instrument de musique

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 RWC [Goto03], l'Université McGill, l'Université Iowa et l'IRCAM avec la même énergie. Les notes de 50Hz à 2000Hz 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.

1.1.3.2. Enregistrements réels

Il existe beaucoup d'enregistrements de musique et les partitions reliées comme fichiers MIDI. Grâce aux logiciels "Alignement de Partition" [Rodet04] et "AudioSculpt" [Bogaards04], plusieurs fichiers MIDI de *RWC Classical* sont alignés automatiquement aux enregistrements réels, mais une grande parie reste à corriger à la main.

1.1.3.3. Musique synthétique

Nous avons essayé d'aligner la collection d'enregistrements polyphoniques pour gérer les références F0s, mais il nous reste toujours la question de définir la fin d'une note dans le contexte polyphonique. En conséquence, nous avons décidé de créer une base de données de musique synthétique en utilisant un "sampler" et des échantillons monophoniques des instruments de musique [Yeh07]. Une méthode a été proposée pour synthétiser la musique en gardant des notes individuelles, à la fin d'annoter des références F0s plus précises. Une quarantaine d'extraits de musique synthétisée sont prêts à servir l'évaluation de notre algorithme. Cette base de données nous permets d'étudier plus précisément des sources mixées avec l'énergie varié.

1.1.4. Évaluation internationale MIREX 2007

Dans l'année 2007, les algorithmes actuels de l'estimation de F0s a été évalués et comparés pour la première fois grâce à MIREX (Music Information Retrieval Evaluation eXchange) 2007. Nous avons participé dans la première tâche: l'estimation de F0s trame par trame. Il y a 12 participants qui ont proposé 16 systèmes. La base de données contient un enregistrement d'un quintette (le cor, la clarinette, le basson, le flûte et l'hautbois) et un morceau de la musique synthétique à partir d'échantillons monophoniques. Comme les deux morceau sont multi-track, ils sont utilisés à mixer des morceaux contenant deux, trois, quatre et cinq sources. Un système est demandé de rendre les F0s estimés pour chaque trame d'analyse. La méthode d'évaluation de la performance est basée sur celle proposé par [Poliner06]. Notre système a été classé au 2ème, après cela de Klapuri.

Publication

[Yeh07] C. Yeh , N. Bogaards and A. Röbel: Synthesized polyphonic music database with verifiable ground truth for multiple F0 estimation, 8th International Conference on Music Information Retrieval (ISMIR'07), Vienna, Austria, 2007.

[Yeh06b] C. Yeh and A. Röbel, "*Adaptive noise level estimation*", Proc. of the 9th Int. Conf. on Digital Audio Effects (DAFx'06), Montreal, 2006.

[Yeh06a] C. Yeh, A. Röbel, and X.Rodet, "*Multiple F0 tracking in solo recordings of monodic instruments*", 120th AES Convention, Paris, France, 2006.

[Yeh05] C. Yeh, A. Röbel, and X.Rodet, "*Multiple fundamental frequency estimation of polyphonic music signals*", IEEE ICASSP, pp. 225-228 (Vol. III), Philadelphia, Pennsylvvania, USA, 2005.

[Yeh04b] 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.

[Yeh04a] 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.

Références

[Bregman90] A. S. Bregman, Auditory Scene Analysis. Cambridge, Massachusetts: The MIT Press.
[Cohen95] L. Cohen, Time-frequency analysis, Prentice Hall, 1995.
[Fletcher98] N. F. Fletcher and T. D. Rossing, The Physics of Musical Instruments, Springer-Verlag, 1998.
[Qi97] Y. Qi, R. E. Hillman, "*Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals*". Journal of Acoustical Society of America, 102(1), 537–543.
[Klapuri03] A. Klapuri, "*Multiple fundamental frequency estimation by harmonicity and spectral smoothness*", IEEE Trans. Speech and Audio Processing, 11(6), 804-816, 2003.
[Goto03] M. Goto, "RWC music database: music genre database and musical instrument sound database", Proc. of the 4th International Conference on Music Information Retrieval (ISMIR 2003).
[Röbel04] A. Röbel, M. Zivanavic, 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.
[Bogaards04] N. Bogaards, A. Roebel, X. Rodet: Sound Analysis and Processing with AudioSculpt 2, International Computer Music Conference (ICMC), Miami, 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.
[Poliner06] G. Poliner, D. Ellis, "*A discriminative model for polyphonic piano transcription*". Eurasip JASP 2006.