Une méthode modifiée de détection de pitch par passages par zéro en milieu interférant

François Gaillard, Frédéric Berthommier, Gang Feng, Jean-Luc Schwartz

Institut de la Communication Parlée
ICP/INPG, UPRESA 5009
46 avenue Félix Viallet 38031 GRENOBLE cedex 01
Tél : +33 76 57 47 15, E-mail : {gaillard, bertho, feng, schwartz}@icp.grenet.fr

RÉSUMÉ

Nous évaluons dans ce papier, au sens du traitement du signal, une méthode non-linéaire de détection de pitch basée sur la détection des passages par zéro des signaux (méthode PPZ), en diverses conditions d'interférences.
D'abord, l'identification de la fréquence fondamentale (F0) est évaluée sur des mélanges simples : mélanges de sons purs, bandes de bruit blanc Gaussien et paires de voyelles. Puis, nous modifions la méthode, en introduisant un paramètre de confiance de la mesure basé sur l'écart-type des intervalles inter-zéros en pente montante. Enfin, la robustesse de ce paramètre est testée en diverses conditions d'interférences.
Nous montrons que cette méthode PPZ permet la détection de périodicité sans connaissance a priori sur les sources mélangées, et l'identification de leur fréquence fondamentale.

ABSTRACT

This paper evaluates, in terms of speech signal processing, a non-linear method of pitch detection based on the detection of the zero-crossings of the signals ("PPZ method"), in adverse conditions of interference.
First, identification of fundamental frequency (F0) is evaluated on simple mixtures : mixtures of pure tones, white noise bands and pairs of vowels. Then, we modify the method by introducing a confidence measure based on the standard deviation of zero-crossing intervals. Finally, we test the robustness of this confidence measure in adverse conditions of interferences.
We show that this PPZ method allows to detect periodicity without any knowledge about the nature of the interfering sources, and then to identify their fundamental frequency.

1 Introduction

Le système auditif humain est capable de structurer son environnement sonore, et en particulier d'extraire une voix d'un mélange de signaux interférants. Dans le cadre de la séparation structurelle de sources, il est possible d'extraire de la structure même des signaux de parole des indices primitifs qui peuvent aider à caractériser chacune des sources mélangées. Cette stratégie se place dans le contexte de l'Analyse de Scènes Auditives [1], qui propose l'utilisation d'indices tels que la fréquence fondamentale (F0), ou encore le délai interaural (ITD) pour la séparation.

Dans le cadre de la séparation de sources F0-dépendante, nous avons implémenté, modifié et évalué une méthode non-linéaire de détection de pitch basée sur l'extraction des passages par zéro des signaux (méthode PPZ, souvent utilisée pour la détection de voisement), en diverses conditions d'interférences : mélanges de sinusoïdes, bandes de bruit blanc, mélanges de voyelles, sons purs et voyelles en conditions bruitées.

2 La méthode PPZ

2.1 Présentation

La méthode PPZ consiste en premier lieu à localiser tous les passages par zéro en pente montante du signal, contenus dans une fenêtre de 80ms. Il est alors possible, à partir de ces passages par zéro, de construire un histogramme d'intervalles, qui fournit un intervalle moyen μ, ainsi qu'un écart-type σ des intervalles inter-zéros contenus dans la fenêtre de 80 ms. L'estimateur de F0, que l'on note F0* (en Hertz), correspond à l'inverse de μ, et l'écart-type (σ, en bins) chiffre les variations des longueurs d'intervalles autour de μ.

2.2 Evaluation sur mélanges de sons purs

La première étape de l'évaluation de la méthode PPZ a consisté à mélanger deux sons purs à différents niveaux relatifs en énergie (NR, différence en dB entre les énergies de chacun des sons purs). La Figure 1 montre les évolutions de F0* et de σ vs. NR, NR variant de -15dB à 15dB. Afin de s'affranchir des effets de différences de phases entre les deux sons purs, ces variations ont été moyennées sur 100 phases aléatoires. Enfin, et pour comparaison, nous montrons l'évolution, dans les mêmes conditions expérimentales, d'un estimateur de F0 utilisant la fonction d'autocorrélation linéaire (ACF linéaire).

La Figure 1 fait apparaître une zone de transition très étroite (moins de 3dB de NR) correspondant aux faibles NR. Dans cette zone, F0* se situe entre les deux fréquences attendues, et σ est élevé. De plus, on peut noter que la transition n'est pas symétrique. En effet, dès NR=0dB apparaissent (ou disparaissent) dans le signal temporel les passages par zéro de l'une des deux sources. Ceci explique l'asymétrie de la transition, ainsi que sa très faible largeur. Ces deux propriétés distinguent très nettement la méthode PPZ de la méthode par ACF linéaire.

Ainsi, la Figure 1 montre que la méthode PPZ d'extraction de F0 est très sensible à la dominance en énergie d'une composante d'un mélange. En dehors de la zone de transition,

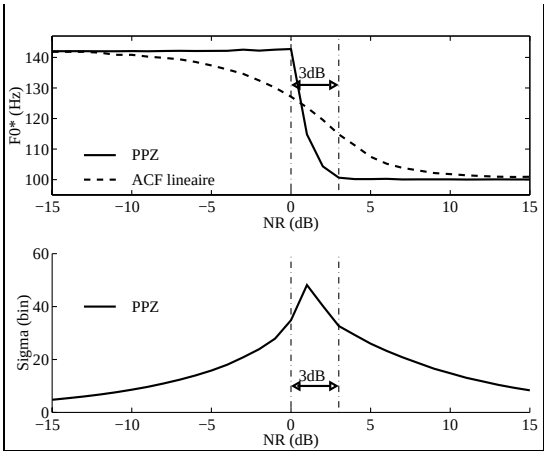

Figure 1 : Mélange de deux sons purs F01=100Hz et F02=142Hz, pour NR∈[-15..15]dB.

très étroite, la source de plus forte énergie masque les autres, sans être elle-même affectée par la présence d'autres sources. Cette propriété est le point de départ de notre évaluation; nous allons par la suite tester ce phénomène de dominance dans d'autres conditions interférantes.

3 Identification de F0 en conditions bruitées

3.1 Méthode PPZ et bandes de bruit blanc Gaussien

En traitement de la parole, la méthode PPZ sera utilisée en sortie de banc de filtres. C'est pourquoi, pour l'évaluer en conditions bruitées, nous avons construit, par filtrage rectangulaire de signaux aléatoires Gaussiens, des bandes de bruit blanc Gaussien (BBG), de fréquence centrale Fc et de largeur de bande DF, que nous avons présentées seules au détecteur de pitch par PPZ.

Le nombre moyen de passages par zéro en pente montante par unité de temps d'une telle bande [Fc-DF/2..Fc+DF/2], que l'on appellera *fréquence caractéristique de bande*, est :

$$Fb = Fc.\sqrt{1+\frac{DF^2}{12.Fc^2}}\ (Hz)$$

Pour vérifier cette expression, nous avons construit huit BBG, avec Fc∈{45, 95, 155} (Hz) et DF∈{45, 80, 180}

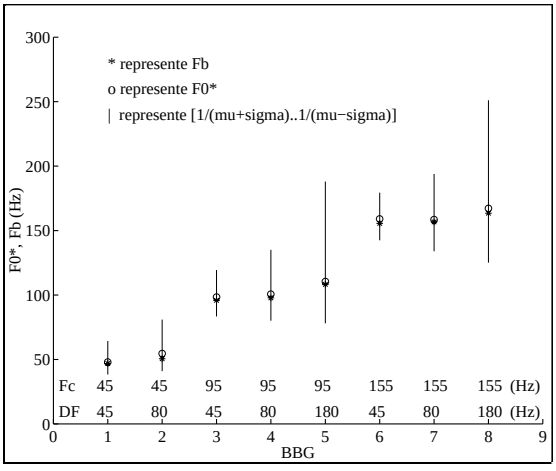

Figure 2 : Comparaison F0*/Fb pour 8 BBG.

(Hz). Pour simuler une détection sur fenêtre d'analyse infinie, nous avons effectué 1000 estimations sur 1000 BBG différentes pendant 80ms pour chaque couple (Fc,DF). Les résultats présentés Figure 2 montrent que F0* correspond bien à Fb; avec un écart-type important, lié à la largeur de la bande, et peu sensible aux variations de Fc.

La méthode PPZ permet donc une bonne caractérisation d'un bruit blanc Gaussien en sortie de banc de filtres, non seulement par la détection d'une fréquence caractéristique du canal du banc de filtres (d'un point de vue identification de F0), mais aussi par le grand écart-type de la détection en présence de bruit.

3.2 Identification de F0 d'un son pur en conditions bruitées

Nous avons mélangé, à différents RSB (différence en dB entre l'énergie du son pur et l'énergie de la BBG) compris entre -15dB et 15dB, une bande de bruit 50-200Hz et un son pur de fréquence Fs, située dans la même bande de fréquences (Fs∈{50, 98, Fb, 168, 200}(Hz)). Pour se rapprocher d'une détection sur fenêtre infinie, nous avons effectué pour chaque fréquence de son pur 1000 estimations avec 1000 BBG différentes, pendant 80ms. La Figure 3 présente les variations de F0* et de σ en fonction de RSB.

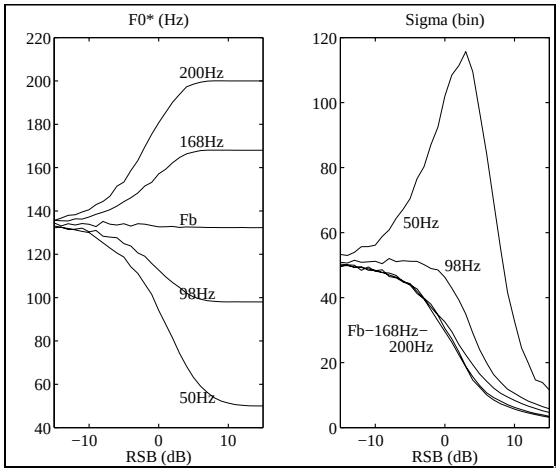

Figure 3 : Identification de F0 d'un son pur en conditions bruitées : F0* et σ vs. RSB.

On retrouve dans ces résultats la zone de transition correspondant aux faibles valeurs de RSB. Ici, l'évolution de l'écart-type en fonction de RSB n'est plus symétrique, car les composantes du mélange ne sont plus de même nature : lorsque le bruit est dominant en énergie, σ est fort; lorsque le son pur est dominant en énergie, σ est très faible. Enfin, dans la zone de transition, on remarque que σ ne prend de très grandes valeurs que si Fs est plus petite que Fb. Dans le cas contraire, σ diminue jusqu'aux valeurs attendues pour un son pur dominant, sans présenter de maximum.

4 Méthode PPZ et double-voyelles

4.1 Expérience

Nous avons préparé des mélanges de deux parmi six voyelles synthétiques du Français, [a, e, i, o, u, y], avec NR=0dB (égale énergie pour les deux voyelles de la paire). Pour chaque paire de voyelles, une fréquence fondamentale

est fixée à 100Hz, la seconde étant choisie parmi 12 autres valeurs de 106Hz à 200Hz par demi-tons. On dispose alors de 360 paires.

Chaque paire est présentée à l'entrée d'un système composé de trois étages : le premier étage est constitué par un banc de filtres gammatones de 32 canaux [2]; il est suivi d'un étage de démodulation des hautes et moyennes fréquences, par rectification puis filtrage passe-bande dans la bande 50-200Hz. Le dernier étage enfin est l'étage d'identification de F0 par la méthode PPZ, qui fournit, pour 80ms de signal, une estimation (μ,σ) par canal et par paire. Enfin, nous avons calculé sur chaque canal et pour chaque paire les niveaux relatifs $NR_{i=1..32}$ à la sortie du banc de filtres. Sur un canal donné, pour une paire de voyelles donnée, nous considérons alors que la voyelle de plus forte énergie est dominante.

Une première inspection des résultats montre un mauvais comportement du système dans les basses fréquences, probablement parce que les harmoniques sont résolus dans ces canaux de faible largeur de bande. Les résultats présentés par la suite ne concerneront donc que les canaux de fréquences supérieures à 700Hz.

4.2 Identification de F0

Pour une paire de voyelles sur un canal, F0* peut correspondre, avec une résolution de ±2,5Hz, soit au F0 de la voyelle dominante sur ce canal ("Dominant"), soit au F0 de la voyelle non-dominante ("Non-Dominant"), soit à aucune de ces deux valeurs ("Erreurs").

Le tableau 1 présente, pour tous les canaux et toutes les paires, les pourcentages de cas correspondant à ces trois situations :

Dominant	Non-Dominant	Erreurs
63.8%	2.2%	34%

Tableau 1 : Taux de détection pour l'identification de F0 en double-voyelles.

Dans le paradigme de double-voyelles, le taux d'erreur est donc très élevé. Ainsi, la méthode PPZ, utilisée directement sur des mélanges de sons complexes tels que des voyelles, ne permet plus l'identification de F0 avec aussi peu d'erreur que dans le cas des mélanges de sons purs.

Cependant, la section 2.2 nous a montré que les erreurs s'accompagnaient de valeurs de σ élevées, information que nous n'avons pas encore utilisée pour l'identification de F0. Dans la suite, nous allons donc modifier la méthode PPZ, pour d'abord détecter une périodicité dominante sur la base d'une mesure de confiance utilisant σ, puis ensuite pour l'identification de F0.

4.3 Introduction d'une mesure de confiance pour la détection de périodicité

La Figure 4 présente les résultats de l'estimation de F0 dans le paradigme de double-voyelles, présentés dans le plan $(F0^*,\sigma)$. Dans ce plan, *un* point correspond à *une* estimation pour *une* paire de voyelles de F0 fixés dans *un* canal.

S'il semble évident qu'une détection du F0 de la voyelle dominante s'accompagne d'un faible écart-type, et qu'une erreur de détection produit de fortes valeurs de σ, on remarque ici une nette séparation entre les deux groupes de valeurs. De plus, bien que nous ne l'ayons pas encore formalisé, la frontière semble lier σ et F0* par une fonction décroissante. A l'aide de critères statistiques basés sur la construction de courbes ROC, nous avons construit une séparatrice linéaire, du type $\sigma=A\mu+B$, $(A < 0)$, que nous présentons en Figure 4 superposée aux résultats des 360 estimations de F0 sur les 23 canaux utiles, soit 8280 points.

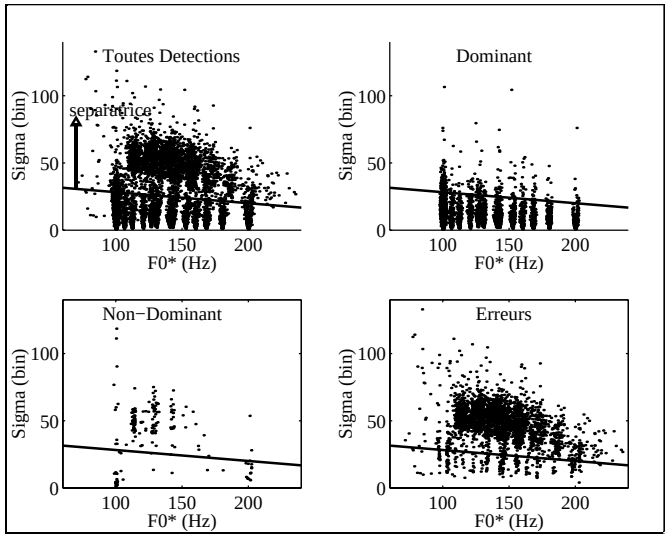

Figure 4 : Détection sur double-voyelles, représentée dans le plan $(F0^*,\sigma)$.

Si l'on étudie les distributions des $NR_{i=1..32}$ de part et d'autre de la séparatrice, on constate que les points situés sous la séparatrice correspondent à des grandes valeurs (positives et négatives) des NR_i, alors que les points situés au dessus ont des NR_i proches de 0dB. Ainsi, cette mesure de confiance basée sur l'utilisation de σ semble permettre de décider de la présence ou non d'une périodicité dominante, sans connaissance a priori sur la valeur attendue de F0.

La détection de périodicité peut ensuite être suivie de l'identification de F0 : en effet, en se replaçant dans la stratégie de la section 4.2, nous présentons dans le Tableau 2 le taux de détection "Dominante" (avec une résolution de ±2,5Hz), et le taux correspondant aux "Erreurs" (incluant les détections "Non-dominantes", qui restent marginales), mais en séparant les deux groupes de points.

Avec une résolution de ±2,5Hz, 94.4% des détections sont correctes sous la frontière. De plus, les erreurs situées sous la frontière sont principalement dues au choix de la résolution de ±2,5Hz. Si ce critère est remplacé par un critère relatif à la fréquence (i.e. 5% de la fréquence attendue), le taux de détection dominante sous la séparatrice passe à 99% de bonnes détections.

	Dominant	Erreurs
Au dessus	9%	**91%**
En dessous	**94.4%**	5.6%

Tableau 2 : Taux de détection en dessous et au dessus de la séparatrice, dans le paradigme de double-voyelles.

Ce critère basé sur σ permet donc, après détection de périodicité, d'identifier le F0 de la voyelle dominante. Nous allons dans la suite tester la robustesse de ce critère, construit dans le paradigme de double-voyelles, dans d'autres cas d'interférences.

5 Robustesse de la mesure de confiance

5.1 Mélanges de sons purs et de BBG

Nous avons mélangé, à différents RSB∈[-15..15]dB, une BBG 50-200Hz avec cinq sons purs de fréquences comprises entre 50 Hz et 200Hz, incluant la fréquence caractéristique de la BBG. Nous disposons pour chaque son pur de 100 estimations pour 100 bandes différentes. Les résultats de ces estimations sont représentés dans le plan (F0*,σ) en Figure 5. La séparatrice construite dans le paradigme de double-voyelles est superposée à ces résultats.

L'observation des distributions de RSB sous et sur la frontière montre que les points situés sous la séparatrice correspondent principalement aux cas où le son pur est

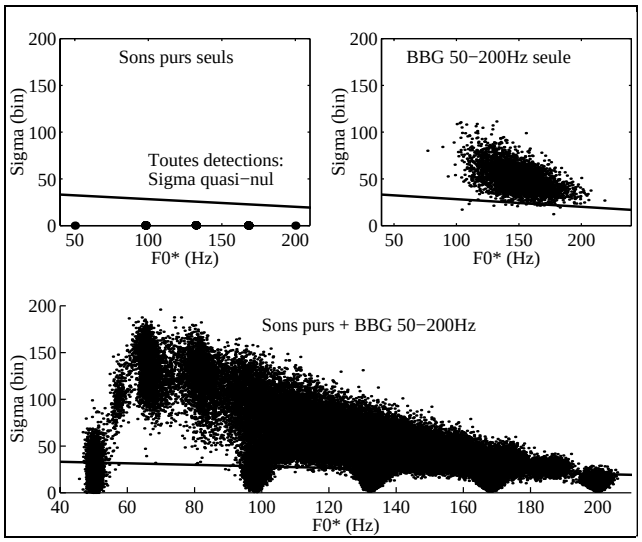

Figure 5 : Détection sur sons purs+ BBG, représentée dans le plan (F0*,σ).

dominant. La position d'un point sous la séparatrice dans le plan (F0*,σ) permet donc de décider de la présence d'une périodicité dominante.

D'un point de vue identification de F0, nous présentons les taux de détection sous et sur la frontière dans le Tableau 3, pour lequel, avec une résolution de ±2,5Hz, 90.0% des détections sont correctes. Si, comme en 4.3, nous remplaçons le critère de résolution de ±2,5Hz pour l'identification par un critère relatif à la fréquence de 5%, le taux d'identification de F0 sous la frontière passe à 96%. Comme dans le paradigme double-voyelles, la séparatrice permet la détection de périodicité, puis l'identification de F0 avec peu d'erreur.

	Son pur	Erreurs
Au dessus	7.2%	**92.8%**
En dessous	**90.0%**	10.0%

Tableau 3 : Taux de détection en dessous et au dessus de la séparatrice, avec des mélanges sons purs + BBG.

5.2 Voyelles en conditions bruitées

Dans une seconde expérience, nous avons mélangé (à niveaux rms égaux, sur les canaux utiles en sortie de banc de filtres) un bruit blanc avec une parmi six voyelles du Français [a, e, i, o, u, y], de F0 compris entre 100Hz et 200Hz, puis présenté ces mélanges au système de détection décrit en 4.1. Pour chaque couple (voyelle, F0), nous effectuons 10 estimations avec dix bruits blancs différents.

Les résultats présentés dans le plan (F0*,σ) sont similaires à ceux de la Figure 5. On retrouve dans ce plan deux groupes de points, l'un pour lequel σ est faible, et F0* correspond au F0 de la voyelle dominante dans le canal considéré, l'autre correspondant au bruit de bande (bruit filtré par banc de filtres puis démodulé), où σ est fort. Selon la même stratégie que les sections précédentes, les taux d'identification sont donnés dans le Tableau 4.

	Voyelle	Erreur
Au dessus	9.9%	**90.1%**
En dessous	**90.0%**	10.0%

Tableau 4 : Taux de détection en dessous et au dessus de la séparatrice, pour des mélanges voyelle + bruit blanc.

L'observation des distributions des RSB permet de montrer que les points situés sous la frontière correspondent principalement aux cas de dominance de la voyelle sur le bruit. Enfin, avec un critère relatif à la fréquence de 5%, 98% des détections sont correctes.

6 Conclusion

La méthode PPZ modifiée permet, en différentes conditions interférantes, de détecter une périodicité dominante avec peu d'erreur, et sans connaissance a priori sur les fréquences attendues, *puis* l'identification de F0, avec un taux d'identification de 96% à 99% pour les trois cas étudiés.

Parmi les nombreuses méthodes temporelles d'extraction de pitch [3] [4], celle-ci est intéressante dans la mesure où, outre son faible coût d'implémentation et sa simplicité, elle est fondée sur un échantillonnage du signal suivi d'une statistique simple, locale dans le domaine fréquentiel, utilisant la redondance temporelle des signaux harmoniques, et ne nécessitant pas de représentation intermédiaire exhaustive, comme par exemple les autocorrélogrammes [3] ou les AM-MAPS [5].

De plus, nous avons constaté l'efficacité de cette méthode pour l'identification des F0 après détection de périodicité. La suite de notre étude portera sur son efficacité pour la séparation des voyelles. Il serait possible par exemple de regrouper les canaux selon les F0 détectés, pour alimenter un processus de reconnaissance partielle [6] afin d'utiliser dans le plan temps-fréquence la redondance spectrale.

Remerciements : Le calcul de Fb nous a été apporté par Pierre Chenevier (ENSERG).

7 Références

[1] Bregman A.S. (1990), *ASA*, MIT Press, London.

[2] Patterson R.D. et al. (1992), in *« The auditory processing of speech »*, Schouten, M. (Ed.), Mouton de Gruyter, 67-83.

[3] Meddis R., Hewitt M. (1992), *JASA* ***91****, 233-245.*

[4] De Cheveigne A. (1993), *JASA*, ***93****, 3271-3290.*

[5] Berthommier F., Meyer G. (1995), *Proc. Eurospeech Madrid, 135-138.*

[6] Cooke M., Morris A., Green P. (1996), *Proc. of Workshop on the Auditory basis of speech perception, Keele, 297-300.*