DÉTECTION DU PITCH PAR LES ONDELETTES CONTINUES EN TEMPS RÉEL POUR UN SIGNAL PAROLE BASÉE SUR UN SEUIL ADAPTATIF POUR UNE DÉTERMINATION V/NV.

R AJGOU[1], S SBAA[1], S AOURAGH[2] & A. TALEB-AHMED[3]

[1] Département d'Electronique, Laboratoire LESIA, Université Med Khider 07000 Biskra.

[2]Département d'Electronique, Laboratoire LESIA, Université Kasdi Merbah 030000 Ouargla.

[3]Laboratoire LAMIH Université de Valenciennes France.

ajgou2007@yahoo.fr

RÉSUMÉ

Une nouvelle méthode de détection du pitch basée sur les transformées d'ondelette continue (TOC) a été développée pour un objectif d'implantation en temps réel qui nécessite un temps d'exécution réduit. Alors on a conçu un algorithme adaptatif comme un outil de détermination V/NV (présenté dans notre travail dans [1]) et se met au début d'analyse d'une trame, si cette trame est voisée on effectue l'analyse sinon on passe directement à la deuxième trame ce qui fournit un temps du calcul réduit (on élimine l'analyse pour un son non voisé d'une trame. Pour plusieurs méthodes, de détection du pitch la décision V/NV est faite après analyse ce qui signifie un vainement de calcul). On a évalué notre méthode sous condition bruité et comparé avec d'autre méthode. Le contour d'évolution du pitch et la décision V/NV présentent un problème complexe sous conditions bruités.

MOTS-CLES : ondelette. Pitch, estimation du pitch, TOC, Voisé/Non voisé, détection du pitch, évolution du pitch, FWT

1 INTRODUCTION

La détection du pitch est procédée en traitant des petites portions (trame) consécutives d'un signal pour avoir des valeurs du pitch. Ce processus est appelé le fenêtrage. Ce dernier généralement se fait avec recouvrement. La figure 1 exprime le fenêtrage avec un recouvrement d'un signal parole. Le signal parole est un processus aléatoire non stationnaire. Nous faisant l'hypothèse de quasi-stationnarité sur des périodes allants de 10 à 35 ms [2]. Les algorithmes de détection du pitch en temps réel peuvent être évalués selon quatre principaux caractéristiques et performances [3]. En premier le temps du calcul qui doit être minimisé [3].Le deuxième est la détermination de voisement ou non voisement d'un segment [3]. Le troisième est la précision pour estimer la période du pitch [3]. Un détecteur doit fournir une bonne résolution en temps et en fréquence avec moins d'erreurs. Le quatrième est la robustesse de détection vis-à-vis de la présence du bruit [4]. Les méthodes actuellement établies ont des inconvénients dans quelques uns de ces performances et caractéristiques. Pour cela dans ce travail on propose une méthode de détection du pitch basé sur les transformées d'ondelette continue (TOC). Cette méthode vérifie certains critères : la bonne résolution en temps fréquence, une bonne immunité au bruit vis-à-vis l'estimation de pitch, la précision de calcule de pitch, un temps d'exécution réduit et un contour de pitch bien claire à cause de la bonne décision V/NV où on a conçu un algorithme adaptatif de détermination de V/NV par application du TOC qui se met au début d'analyse d'une trame, si cette trame est voisé on effectue l'analyse sinon on passe directement au deuxième trame ce qui fournit un temps du calcul réduit (on élimine l'analyse pour un son non voisé où le pitch ne se trouve pas).

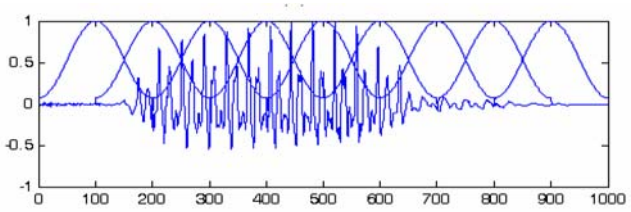


Figure 1 : Le Fenêtrages et le recouvrement procédés sur le long du signal.

2 MÉTHODE

2.1 Introduction

La définition de la transformée d'ondelette continue est comme suit [5] :

$$TOC(t,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t)\psi\left(\frac{t-\tau}{a}\right)dt \qquad (1)$$

Les coefficients de TOC sont calculés par la convolution d'un signal f(t) avec une ondelette, d'échelle a et au temps τ. La TOC se reporte souvent par un vecteur de tous les coefficients pour une échelle donnée. Un grand nombre d'informations redondantes est chiffrées dans la TOC parce que les coefficients ne se changent pas considérablement sur faibles changements dans le petit temps ou échelle. L'information est extraite habituellement uniquement par des maximums du Coefficients TOC [5]. Le traitement de grands signaux et des signaux multidimensionnels ainsi le calcul des données redondantes peuvent conduire aux difficultés pour le processus de calculs. L'algorithme de la transformée d'ondelette rapide (FWT) [7] est inspiré par l'algorithme de pyramide, élimine la redondance à travers l'orthogonalité [8]. Cela implique que toute information à chiffrer par la première ondelette n'est pas chiffré par la seconde et vice versa. Le FWT utilise l'échelle dyadique, l'idée générale derrière l'algorithme de FWT représenté par la figure 2. La décomposition se continue jusqu'au l'ordre (n).

2.2 Le choix de la forme d'ondelette

La forme d'ondelette est le facteur qui définit le choix d'ondelette. Certain type d'ondelette sont très utiles pour extraire le pitch. Mallât [6] a prouvé que l'analyse avec une ondelette d'ordre1 donne des maximums au point de changement de signal. Comme vue à la figure 3 qui présente une ondelette de Haar (Daubechies d'ordre1), on remarque qu'elle est positif coté droite de centre et négative coté gauche. La TOC avec cette forme accentue le passage par zéro. On peut alors utiliser les maximums de TOC pour identifier le passage par zéro dans le signal originale et calculer les fréquences correspondent. Chaque maximum correspond à un passage par zéro [6]. Il est plus utile de chercher des maximums dans les coefficients de TOC que le passage par zéro dans le signal original.

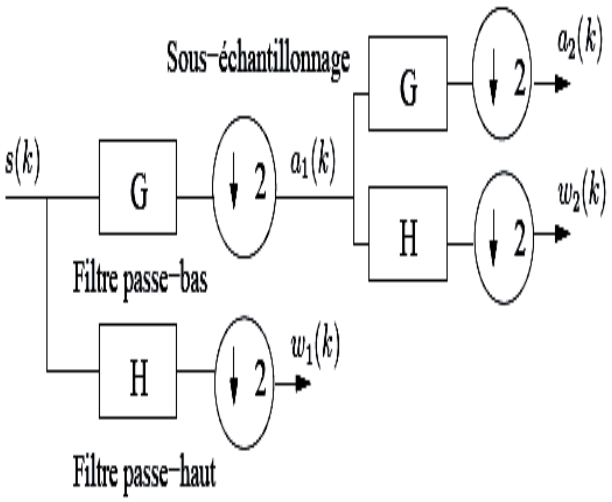


Figure 2 : Transformée dyadique d'ondelette avec 2 niveaux de décomposition.

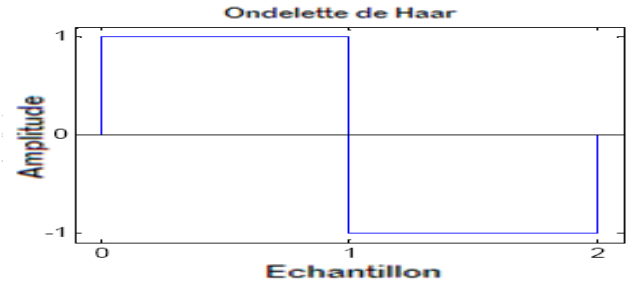


Figure 3: Ondelette de Haar.

2.3 Niveau de décomposition

En principe, l'analyse devrait être exécutée avec une ondelette sur toutes les échelles dyadiques. Chaque échelle donne différentes fréquences. Il ya alors des hauteurs qui limite le nombre d'échelle à considérer dans l'analyse. L'auteur dans [9] suggère que l'analyse peut se faire sur trois échelles adjacentes pour le cas d'une guitare. L'auteur dans [10] suggèrent que le nombre d'échelles à considérer se calcule par :

$$a = 2.[\log(n)/\log(2)] \quad (2)$$

Où : n est le nombre d'échantillons pour le signal à analyser. On note qu'après calcul, on prend le nombre entier de 'a'. Cette dernière méthode est la plus logique à considérer à cause que cette formule dépend de nombre d'échantillon du signal à analyser.

2.4 Détection des maximums

Le processus de détection des maximums implique que la distance entre deux maximums doit être supérieur où égale à δ. La distance minimum δ dépend de la fréquence maximale de pitch (F= 500Hz) 'F' et la fréquence d'échantillonnage Fs [3]

$$\delta = \frac{Fs\ [Hz]}{F\ [Hz]} \quad (3)$$

2.4.1 Estimation de la fréquence fondamentale

Dans la même échelle de la TOC, on estime la fréquence fondamentale comme la fréquence entre la première paire des deux maximums. La fréquence fondamentale est définit donc pour chaque échelle par [11] :

$$F_0 = \frac{Fs}{m_1 - m_2}\ [Hz] \quad (4)$$

F0 : fréquence fondamentale.

m1 : l'échantillon qui correspond au premier maximum.

m2 : l'échantillon qui correspond au deuxième maximum en respectant δ.

Fs : fréquence d'échantillonnage.

2.5 Énergie calculable sur les coefficients d'ondelette

L'énergie d'un signal est la valeur de la somme carrée de ses échantillons. Intuitivement, pour le cas des ondelettes, l'énergie est la somme de carré des coefficients pour chaque échelle. Pour une échelle « a » l'énergie se calcule par la relation suivante :

$$\overline{E}(a) = \sum_{i}^{k} |c_i|^2 \quad (5)$$

Où : a : échelle, a=1,2

$\overline{E}(a)$: L'énergie moyenne.

C : coefficient d'ondelet

K : le nombre des coefficients.

2.6 Qu'elle est la fréquence qu'on adopte comme fréquence fondamentale

On calcul la TOC pour une tranche voisée avec différentes échelles. Il résulte donc un vecteur des coefficients pour chaque échelle, ce qui signifie que pour chaque vecteur des coefficients résultent une fréquence. La question est donc sur la détermination de la fréquence fondamentale (pitch) parmi les fréquences calculées.

Par conséquent la fréquence qu'on adopte comme fréquence fondamentale ou pitch est la fréquence pour laquelle l'énergie des coefficients soit maximale et qui correspond à une échelle bien définie.

2.7 Décision voisé/non voisé

La détermination voisé /non voisé est basé sur le rapport d'énergie par le passage par zéro (EZR) pour chaque segment (tranche). Le principe d'application d'EZR dans la détermination de V/NV s'explique par le fait que l'énergie du signal voisé est importante tandis que le taux de passage par zéro est faible, ce qui donne une valeur importante d'EZR, le contraire est vrai pour un signal non voisé. Le but est alors de préciser tout d'abord avec efficacité et avec minimum de temps si la trame est voisée ou non, si le rapport d'énergie par passage par zéro (EZR) calculé pour une trame supérieur à un seuil, cette trame est alors considéré voisé, donc on peut procéder l'analyse pour la recherche de pitch sinon on considère cette trame non voisé. On montre que le seuil de détermination V/NV est estimé par l'algorithme d'une façon automatique et adaptatif. Le rapport EZR est défini par :

$$EZR[m] = \frac{\overline{E}[m]}{ZCR[m]} \quad (6)$$

Où ZCR [m] présente le taux de passage par zéro pour une trame [m] et $\overline{E}[m]$ est l'énergie moyenne pour une trame. Son principe est qu'on calcul le EZR pour une 'i'emme trame qui se compare avec le EZR de tout les 'i-n'emme trame pour avoir le minimum EZR et le maximum de EZR afin de calculer le seuil. Le seuil se calcul par [1] :

$$Seuil = \min(EZR) + 0.2(\max(EZR) - \min(EZR)) \quad (7)$$

On note que pour la première trame sa détermination de V/NV est basée sur un seuil de référence est égale à 1. Alors la détermination de V/NV est susceptible d'entaché d'erreur pour les premières trames.

2.8 Algorithme pour l'estimation du pitch

On résume la procédure d'estimation de pitch par :

- Extraction de la 'i' ème trame de 30ms.
- Détermination de V/NV, si cette trame est voisée, on passe à l'étape qui suit sinon on revient à l'étape 1.
- Exécuter la TOC pour chaque échelle chacun. Extraire les vecteurs des coefficients avec calcul des énergies sur les coefficients.
- Extraire les maximums des coefficients pour chaque échelle.
- Estimer la fréquence fondamentale pour chaque échelle en considérant les deux premiers maximums (on peut considérer puiseurs maximums c'est à dire estimation de plusieurs F0, la fréquence fondamentale est estimé par la moyenne des F0).
- Estimer la fréquence fondamentale spécifique pour cette tranche en considérant l'énergie maximale des coefficients.
- On revient à l'étape 1 pour la trame qui suit.

Algo1 résume l'algorithme général de détection du pitch par la TOC. (L'algorithme adaptatif de détermination V/NV est dans Algo2).

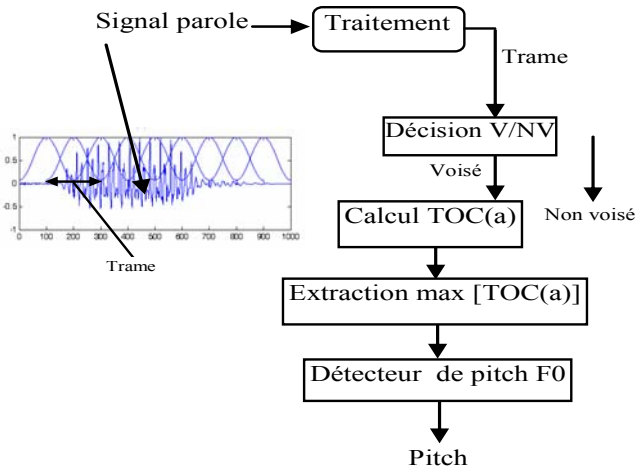
2.9 Evaluation de la méthode

2.9.1 Estimation de pitch :

On prend une tranche d'un signal vocale 'a' de 30ms avec Fs=11025Hz, où on voudrait estimer la fréquence fondamentale. L'ondelette utilisée est l'ondelette Daubechies d'ordre 1. Alors le nombre d'échantillons n est de :

n=11025*(30/1000) donc n=330.75, on prend n=330 échantillons. Alors le nombre d'échelles à considérer dans ce cas est de : a=2(log(330)/log(2)) alors : n=16 échelles. La figure 4 nous montre les coefficients d'ondelette pour l'échelle a=6, où : max1=53, max2=123 : F0=Fs/ (max2-max1). F0=11025/ (123-53), F0=157,5 Hz. Le tableau .1 présente les résultats obtenus pour les échelles a=16, l'énergie maximale dans ce cas est égale à $\overline{E}$ =62.2827 ce qui correspond à F0=157.5Hz. Cette fréquence est considérée comme pitch pour cette tranche. La figure5 présente l'évolution de pitch superposée sur la forme d'énergie (en rouge), où le maximum de la courbe

d'énergie correspond au pitch pendant 30ms pour les échelles a=1,…30. La figure 6 présente le contour de pitch d'un signal parole 'Bon' avec l'évolution de seuil adaptatif. La figure 7 présente un signal parole 'Bonjour' enregistrée par un logicielle winPitchPro [12].



Algo1 : Algorithme général de détection de pitch par la TOC.

```
%valeurs initiales
Rapporter =1;
Seuil=1;
(10) Détection V/NV pour la ''i''eme trame:
Calcul de l'énergie 'E'
Calcul de 'ZCR'      %tau de passage par zéro
Calcul de 'EZR'      %le rapport d'énergie ZCR.
Si EZR>seuil
  Un son voisé
                  Estimation de pitch
        Sinon
    Un son non voisé
Fin si
% début de Création des vecteurs de sorties
max_rapport=max(rapport_EZR);
min_rapport=min (rapport_EZR);
delta=max_rapport-min_rapport;
seuil=min_rapport+0.2*delta;
rapport_EZR = [EZR];
% fin de Création des vecteurs de sorties
Aller vers la trame suivante (aller à (10))
```

Algo2 : Algorithme principal de détermination VINV basé sur EZR [1]

Tableau 1 : Fréquences fondamentale en fonction d'échelles.

Echelle	Max 1	Max 2	Energie	F0 [Hz]
a=1	73	142	0.00001	159.7826
a=2	74	143	3.5654	159.7826
a=3	74	143	8.4654	159.7826
a=4	74	144	23.3658	157.5000
a=5	74	143	37.1512	159.7826
a=6	74	144	62.2827	157.500

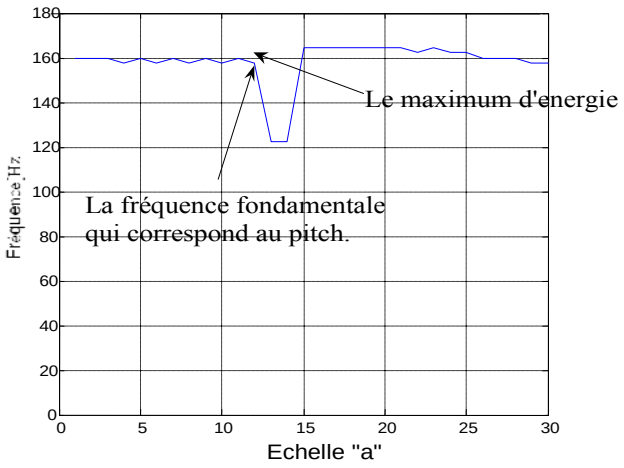


Figure 4 : Coefficients d'ondelettes en fonction des échantillons pour l'échelle a=6, dont Fo=157.5 Hz

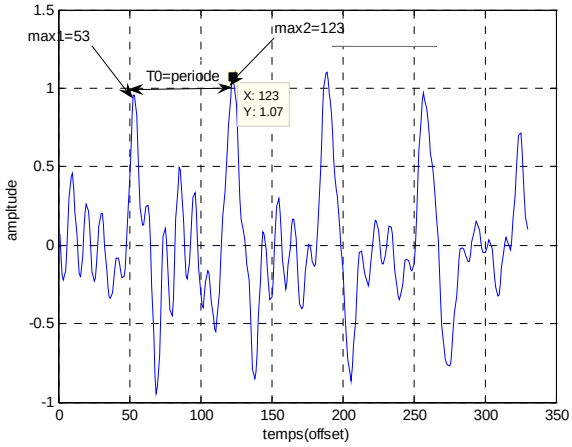


Figure 5 : l'évolution de pitch en bleu superposé sur la forme d'énergie(en rouge), a=1,.30.

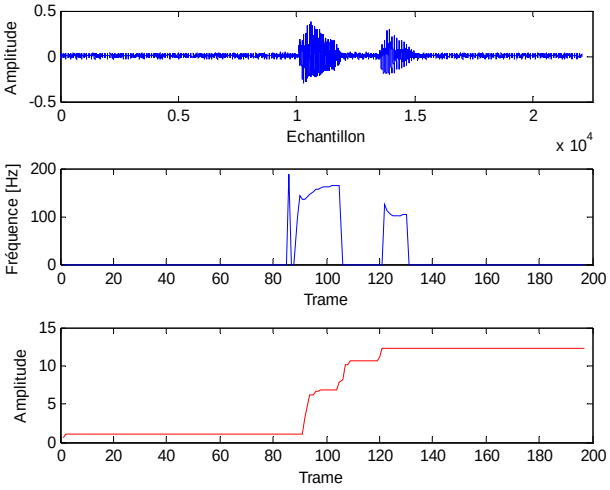


Figure 6 : le signal parole 'Bon' avec son contour de pitch De l'évolution du seuil adaptatif.

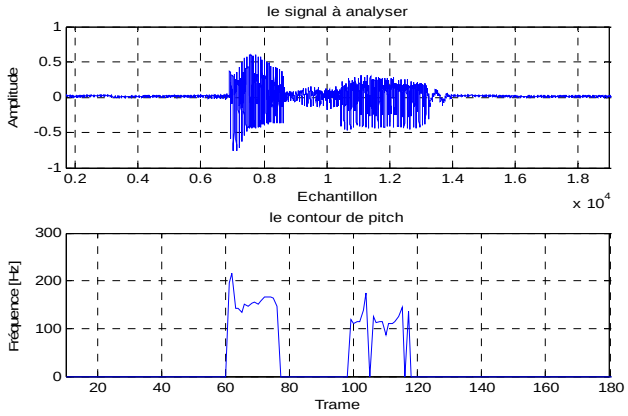


Figure 7 : « Bonjour»en haut et son contour de pitch en bas par application de TOC par Daubechies 1

2.9.2 Comparaison par les ondelettes « Daubechies », « Symlet », « Coifflet »

On estime la fréquence fondamentale et voir le contour de pitch pour une tranche de 30ms d'un son voisé par trois ondelettes « Daubechies1, Symlet1, Coifflet 1». Le tableau 2 présente les valeurs de pitch calculées. La figure 8 présente l'évolution du pitch par les trois ondelettes.

Tableau 2 : Le Pitch calculés par trois ondelettes : .Daubechies, Symlet, Cofflet.

Ondelette d'ordre 1	**F0 [Hz]**
Daubetchies	157.500
Coifflet	159.7826
Symlet	159.7826

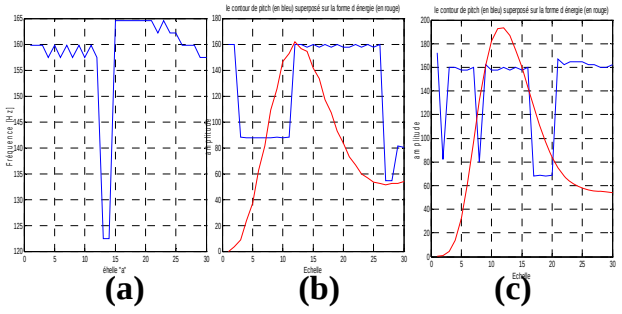


Figure 8 : L'évolution de pitch pour une tranche d'un phonème 'a'' par : (a) Daubetchies 1, (b) symlet 1, (c) Coiflet 1 avec forme d'énergie en rouge.

2.9.3 Influence de bruit sur l'estimation de pitch :

On a évalué notre méthode en comparaison avec d'autres méthodes connues de détermination du pitch sous conditions bruitées. Le tableau 3 résume la comparaison des différentes méthodes de détermination du pitch par différents SNR. Le bruit considéré est un bruit blanc gaussien.

Tableau 3 : Comparaison de plusieurs méthodes de détection du pitch avec notre méthode TOC en fonction de (SNR) vis-à-vis l'estimation du pitch.

SNR	**STFT [Hz]**	**CEPS [HZ]**	**HPS [Hz]**	**Autoc [Hz]**	**TOC [Hz]**
Sans bruit	139.32	132.10	131.93	131.25	133.11
38	139.2	132.10	131.93	131.25	133.11
27	139.32	132.10	131.93	131.25	133.11
18	139.32	132.10	136.76	131.25	133.11
14	132.32	132.10	159.86	131.25	133.11
6	120.01	355.65	135.60	131.25	133.32
2	113.14	157.50	138.18	136.61	134.60
-1	83.59	136.11	138.18	139.45	137.50

2.9.4 La précision d'estimation

Concernant la précision d'estimation. Les valeurs du pitch obtenus par Praat [13] (logicielles utilisé pour le traitement de la parole utilisé par plusieurs auteurs) sont prise comme référence pour évaluer la méthode. Pour cela on compare plusieurs méthodes vis-à-vis l'estimation du pitch. Les résultats obtenus sont enregistrés dans le tableau 4. Le tableau 5 exprime la grosse erreur d'estimation du pitch pour différentes méthodes en respectant la relation suivante :

$$\text{erreur} = \frac{\text{pitch Praat référence - pitch estimé}}{\text{pitch Praat référence}} \% \qquad (8)$$

Tableau 4 : Le pitch estime par différentes méthodes en prenant les valeurs du pitch obtenus par Praat [13] comme référence.

Praat [Hz]	**STFT [Hz]**	**CEPS [HZ**	**HPS [Hz]**	**Autoc [Hz]**	**TOC [Hz]**
60	62.81	78.75	60.78	56.17	60.48
80	83.93	79.12	80.40	75.25	77.36
110	113.25	108.08	109.00	103.25	110.25
150	142.56	145.06	146.03	146.35	156.50
200	182.18	190.08	191.03	193.51	200.45
250	126.50	239.67	235.31	250.38	245.00
300	283.12	266.31	267.03	290.15	299.97
350	311.44	310.32	313.13	350.11	350.50

Tableau 5 : la grosse erreur d'estimation du pitch pour différentes méthodes

Méthodes	Grosse erreur [%]
SIFT	11.01
CEPS	10.82
HPS	1053
Atocor	06.13
TOC	03.30

2.10 Discussion

Cette méthode nous montre une bonne résolution en temps-fréquence, pendant une durée de 30ms. On voit une transcription de pitch en fonction d'échelle. Parmi les fréquences calculées pour chaque échelle, la fréquence fondamentale est choisit de telle manière que cette fréquence correspond à une maximum d'énergie. L'énergie se calcul pour chaque échelle par le carré des coefficients. Cette notion d'énergie est inspiré de carrée des échantillons. On remarque que le pitch estimé par les ondelettes de mères Daubechies1, Coifflet1, Symlet1 donnent des résultats semblable. Les résultats d'estimation de pitch en présence de bruit est bonnes pour notre méthode TOC. Le contour de pitch pour un signal en fonction des trames est bien clair. On remarque aussi la bonne précision d'estimation du pitch en comparaison avec d'autres méthodes.

3 CONCLUSION

La première motivation d'utiliser les ondelettes est la bonne résolution temps fréquence. L'idée de base de cet algorithme est que, pour une ondelette convenablement choisit, la transformée d'ondelette expose les maximums locaux aux points de variation du signal ce qui nous permet d'exploiter au maximum pour la détection et l'estimation du pitch. L'estimation de pitch est basée sur l'énergie calculée sur les coefficients d'ondelette. Notre algorithme débute par une décision voisé /non voisé basée sur un seuil adaptatif pour déterminer la zone voisée où se trouve la fréquence fondamentale, c'est à dire éliminer l'analyse pour le cas non voisé, alors on résulte un temps d'exécution réduit. Les résultats d'estimation de pitch par les trois types d'ondelettes donnent presque les mêmes résultats. On a une bonne résistance à la présence de bruit justifiée par les résultats de simulation. Comme perspective, on présente le problème de détermination du pitch dans le cas des fréquences élevées. On propose pour cela l'utilisation de plusieurs maximums afin d'identifier la fréquence recherchée par le calcul de leurs moyenne.

BIBLIOGRAPHIE

[1] AJGOU Riad, and SBAA Salim, « Algorithme adaptatif de détermination de V/NV basé sur EZR ». First International Conférence on Image ans Signal Processing and their Applications, Mostaganem University, ISPA 2009.

[2] R. Boite et all, Traitement de la parole, PPUR, 2000.

[3] Eric Larson , Ross Maddox « Real-Time Time-Domain Pitch Tracking Using Wavelets » Departments of Mathematics, Physics and Philosophy, Kalamazoo College, Center for Performing Arts Technology University of Michigan School of Music.USA.2004.

[4] Bojan Kotnik1, Harald Höge, Zdravko Kacic1 "Evaluation of Pitch Detection Algorithms in Adverse Conditions", University of Maribor, Slovenia, Siemens AG, Corporate Technology, Germany 2006.

[5] A Spaargaren, MJ English « Detecting Ventricular Late Potentials using the Continuous Wavelet Transform » University of Sussex, Brighton, UK.1999.

[6] John McCullough « UsingWavelets for Monophonic Pitch » Computer Science Department, Harvey Mudd College, USA. HMC-CS-2005-01 September, 2005.

[7] Michel Misiti, Yves Misiti, Georges Oppenheim, Jean-Michel Poggi. « Wavelet Toolbox », for use with Matlab, September 2000 by The MathWorks.

[8] Emmanuel Didiot « Segmentation parole/musique pour la transcription automatique de parole continue » université Henri Poincaré Nancy 1France .NOV 2007.

[9] W.Shabana and J.Fitch « a wavelet-based pitch detector for musical signals ». Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK.2000.

[10] Valerie Perrier - Programme de Transformée en Ondelettes. Institut nationale polytechnique de Grenoble. LMD novembre 1993.

[11] John McCullough Harvey Mudd College Using Wavelets for Monophonic Pitch Detection. HMC-CS-2005-0. 1September, 2005

[12] www.winPitch.com (2009).

[13] Tutoriel Praat Jean-Philippe Goldman Université de Genève Décembre 2006. Http//www.praat.org. 2009.