

# Comparaison de modèles FFT et Munich pour la compression audio

Khalil Abid et Kais Ouni

Unité de Recherche Traitement du Signal, Traitement d'Images et Reconnaissance de Formes,  
École Nationale d'Ingénieurs de Tunis ,  
ENIT, BP 37, Le Belvédère 1002, Tunis, Tunisie.  
Tél : 216 71 874 700 , Fax : + 216 71 872 729  
`khalilabid06@yahoo.fr`  
`kais.ouni@enit.rnu.tn`

**Résumé** Dans cet article, nous proposons une nouvelle architecture de modèle psychoacoustique basée sur les paquets d'ondelettes au lieu de la transformée de Fourier pour le thème de la compression audio. Le principe fondamental de cette architecture est de proposer une analyse par paquets d'ondelettes sur des bandes fréquentielles se rapprochant des 24 bandes critiques de l'oreille ce qui diffère du modèle existant basé sur l'analyse par FFT.

**Mots clés** : Compression audio, Modèle psycho-acoustique, Modèle de FFT, Bandes de Munich, Ondelette de Morlet.

## 1 Introduction

Les techniques de compression audio se basent essentiellement sur codage par transformée et le codage en sous-bandes. La majorité des codeurs MP3 sont basés sur la FFT. Afin de mieux s'adapter à la perception, des algorithmes de codages par transformée ont été conçus dans des bases de paquets d'ondelettes dont la résolution fréquentielle correspond aux bandes fréquentielles critiques [10]. Cette décomposition en ondelettes donne une analyse multi-résolution du signal : une haute résolution temporelle et une basse résolution fréquentielle dans les hautes fréquences et inversement dans les basses fréquences. Elle offre ainsi une bonne modélisation du système auditif humain. Une nouvelle conception du modèle psycho-acoustique basé sur une décomposition en paquets d'ondelettes est appliquée aux codeurs audio afin d'obtenir des sous-bandes se rapprochant des bandes critiques [8]. Ce papier est présenté comme suit, après une présentation du modèle psycho-acoustique, nous présentons le modèle de Munich. L'implémentation du banc de filtres cochléaires par application de l'ondelette de Morlet est détaillée. L'évaluation des résultats de compression est présentée. Nous terminons par la conclusion et les perspectives.

## 2 Description des critères psycho-acoustiques

La psycho-acoustique rend compte de l'ensemble du phénomène perceptif, elle permet de décrire et de quantifier les diverses composants du son :

Limites perceptives : l'oreille humaine perçoit un son pur dans l'intervalle [20 Hz - 20 KHz].

Seuil d'audibilité : c'est le seuil au-delà duquel l'oreille ne perçoit pas de son. Ce seuil est aussi le zéro de l'échelle des décibels (dB), respectivement des phones. La perception des rapports d'intensité sonore (sonie) varie en proportion logarithmique avec l'échelle en phones.

Phénomène de masquage : lorsque le niveau de bruit est trop important l'oreille ne reçoit pas un son pur.

### 3 Transformée en ondelettes et modèle psycho-acoustique

L'architecture du modèle psycho-acoustique est basée sur un algorithme bien déterminé. Tout d'abord, le signal sera tronqué en 2048 points par l'utilisation de la fenêtre de Hanning. Une subdivision des bandes critiques est adoptée en utilisant une décomposition en paquets d'ondelettes (26 sous-bandes), cette approximation possède presque la même structure utilisée par Tewfik et Sinha (29 sous-bandes) [3]. Nous localisons par la suite l'ensemble des composantes tonales et non tonales. Cette étape commence par la détermination des maximums locaux, suivi d'une extraction des composantes tonales et non tonales (bruit) se trouvant dans chaque bande critique. Une suppression sélective basée sur le phénomène de masquage est adoptée, un procédé employé pour réduire le nombre de masques pris en considération pour le calcul du seuil de masquage global [1]. Le seuil de masquage individuel prend en considération le seuil de masquage pour chaque composante restante. Pour finir, le seuil de masquage global est calculé par la totalité des composantes tonales et non tonales qui sont déduites du spectre de la décomposition de paquets d'ondelettes. Le modèle psycho-acoustique analyse le contenu spectral du signal audio d'entrée afin de calculer le rapport de signal sur bruit correspondant à chaque sous-bande [1].

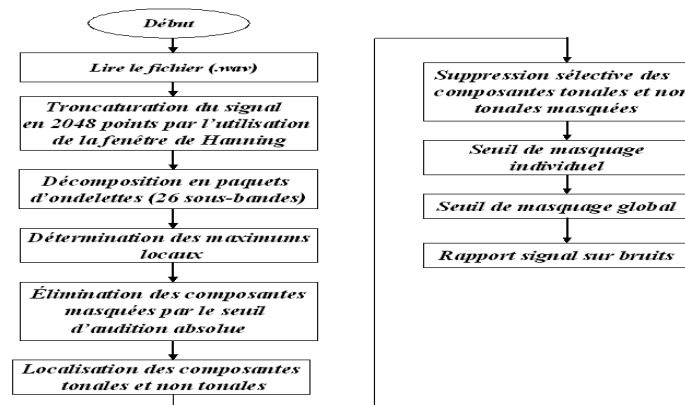


Fig. 1. Architecture du nouveau modèle psycho-acoustique

## 4 Le modèle de Munich

Le système auditif est équivalent à un banc de filtres [7] de bandes passantes égales à la largeur des bandes critiques [10]. Pour modéliser ces largeurs, nous avons adopté les bandes de Munich [10] vu qu'il y a une similitude entre les largeurs de ces bandes avec celles des bandes critiques aux fréquences correspondantes [9]. Les bandes critiques de Munich sont définies pour une fréquence centrale  $F$ (Khz) par l'expression suivante [4] :

$$B_M = 75(1.4F^2 + 1)^{0.69} + 25 \quad (1)$$

## 5 Implementation du banc de filtres cochléaires

Pour ce travail, nous avons opté pour l'ondelette de Morlet. Cette dernière est performante pour de nombreuses applications en traitement du signal, notamment le codage audio et elle a de bonnes propriétés dans la localisation temps/fréquence [2]. Elle est modulée par une enveloppe gaussienne [5]. Il s'agit d'une forme analytique simple, résolue en temps et en fréquence, régulière, localement périodique, à support non compact.

Pour  $\lambda = \frac{\sqrt{0.7}}{B\pi}$ , cette ondelette est définie par :

$$\Psi(t) = \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{t^2}{2\lambda^2}} e^{-i2\pi f_o t} \quad (2)$$

L'ondelette mère doit être caractérisée par une énergie finie :

$$E = \int_{-\infty}^{+\infty} |\Psi(t)|^2 dt < \infty \quad (3)$$

$\Gamma(f)$  est la transformée de Fourier de  $\Psi(t)$  :

$$\Gamma(f) = \sqrt{\lambda} \cdot e^{-\frac{1}{2}(2\pi \cdot f - 2\pi \cdot \lambda \cdot f_o)^2} \quad (4)$$

La condition suivante doit être vérifiée :

$$C_g = \int_0^{+\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty \quad (5)$$

Finalement, cette ondelette est définie par :

$$\psi(t) = \sqrt{\frac{\pi}{1.4}} \cdot B \cdot e^{-\frac{B^2 t^2 \pi^2}{1.4}} e^{-i2\pi f_o t} \quad (6)$$

Où  $e^{-\frac{B^2 t^2 \pi^2}{1.4}}$  est l'enveloppe gaussienne,  $f_o$  est la fréquence de modulation et  $B$  est la largeur de bande rectangulaire équivalente des filtres cochléaires [4][6].

Par application du modèle de Munich, l'ondelette de Morlet est donnée par l'expression suivante :

$$\Psi(t) = \sqrt{\frac{\pi}{1.4}} (75(1.4F^2 + 1)^{0.69} + 25) e^{-i2\pi f_o t} e^{-\frac{(75(1.4F^2 + 1)^{0.69} + 25)^2 \pi^2 t^2}{1.4}} \quad (7)$$

Le nouveau modèle analyse le signal d'entrée et l'étape fondamentale consiste à réaliser une décomposition en paquets d'ondelettes afin d'imiter étroitement les bandes critiques de l'oreille. Une décomposition en paquets d'ondelettes est développée pour une fréquence d'échantillonnage de 44.1 khz (voir figure2).

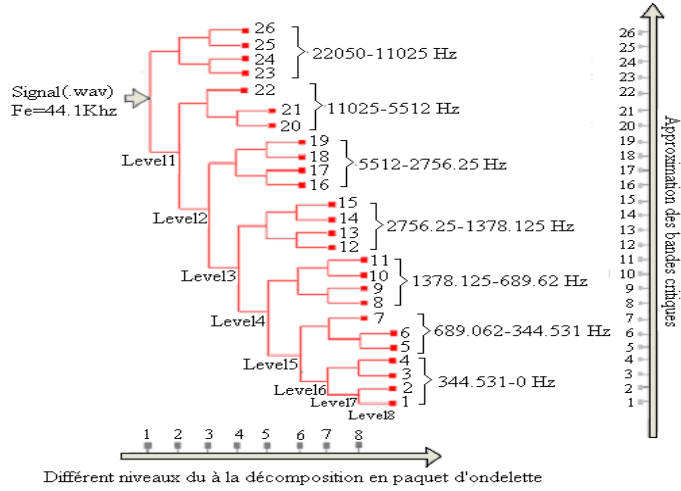


Fig. 2. La décomposition en paquets d'ondelettes proposée

La répartition fréquentielle de la figure2 est illustrée dans la table1 :

Filtre	Limite inférieure	Limite supérieure	Filtre	Limite inférieure	Limite supérieure
1	0	86.132	14	2067.187	2411.718
2	86.132	172.266	15	2411.718	2756.25
3	172.266	258.398	16	2756.25	3445.312
4	258.398	344.531	17	3445.312	4134.375
5	344.531	430.664	18	4134.375	4823.437
6	430.664	516.796	19	4823.437	5512.5
7	516.796	689.062	20	5512.5	6890.625
8	689.062	861.328	21	6890.625	8268.75
9	861.328	1033.593	22	8268.75	11025
10	1033.593	1205.859	23	11025	13781.25
11	1205.859	1378.125	24	13781.25	16537.5
12	1378.125	1722.656	25	16537.5	19293.75
13	1722.656	2067.187	26	19293.75	22050

Tab. 1. : La nouvelle répartition fréquentielle des bandes critiques

## 6 Evaluation du modèle proposé

Pour évaluer la compression par l'ondelette de Morlet et suivant le modèle de Munich, nous avons utilisé pour différents débits (64kbits/s... 192kbits/s) plusieurs type de sons. L'évaluation est basée sur le taux de compression défini comme le rapport entre la taille originale du fichier (wav) et le fichier compressé (MP3). Le tableau 2 contient les types de sons utilisés pour l'essai, leurs durées, la moyenne des taux de compression ( $\overline{Comp}$ ), leurs tailles et le taux de compression calculé pour chaque débit binaire .

<i>Son</i>	Taille(ko)	Durée(s)	64(Kbits/s)	96(Kbits/s)	112(Kbits/s)	128(Kbits/s)	160(Kbits/s)	192(Kbits/s)
<i>Soul</i>	1650	19	10.57	9.28	7.95	7.14	5.57	4.36
<i>Slow</i>	2080	24	10.56	9.27	7.93	7.13	5.54	4.36
<i>Rock</i>	2520	30	10.55	9.26	7.91	7.12	5.53	4.35
<i>Chant Arabe</i>	6380	75	10.54	9.25	7.90	7.11	5.52	4.35
<i>Dialogue</i>	1730	20	10.52	9.24	7.86	7.11	5.51	4.34
$\overline{Comp}$	—	—	10.55	9.26	7.91	7.12	5.53	4.35

**Tab. 2.** Valeurs des taux de compression en fonction du débit (Modèle de Munich)

Les plus faibles taux de compression sont donnés pour le débit de 192Kbit/s. Cependant, plus le débit binaire devient faible, plus le taux de compression devient élevé et moins sa qualité devient intelligible. Basé sur les résultats obtenus dans le tableau 2, la banc de filtres de Morlet basé sur le modèle de Munich tient compte du phénomène de masquage ainsi que des bandes critiques.

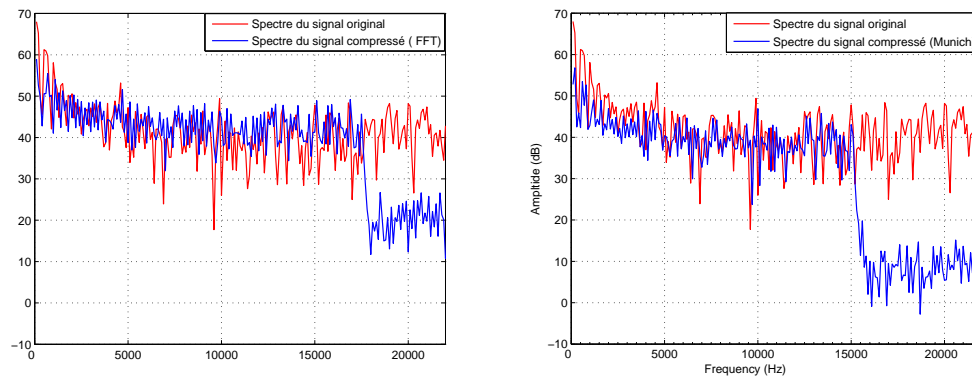
La deuxième partie de notre évaluation est focalisée sur la comparaison du taux de compression obtenu à partir du codeur basé sur l'analyse par ondelette et du codeur basé sur une analyse par FFT pour un débit binaire de 128 Kbits/s donnant un meilleur compromis entre le taux de compression et la qualité sonore. D'après le tableau 3, on constate que la compression par ondelettes est mieux que celle de la FFT puisque sa moyenne des taux de compression a évolué de 13% par rapport à la FFT. ( $\overline{Comp}$  est la moyenne des taux de compression)

<i>Son</i>	FFT	Modèle de Munich
<i>Soul</i>	6.380	7.136
<i>Slow</i>	6.373	7.121
<i>Rock</i>	6.329	7.119
<i>Chant Arabe</i>	6.281	7.112
<i>Dialogue</i>	6.227	7.109
$\overline{Comp}$	6.32	7.12

**Tab. 3.** Comparaison entre la compression par FFT et celle par ondelette (Débit :128 Kbits/s)

Le nouveau modèle psycho-acoustique a été mis en oeuvre en se référant au standard MPEG1 couche 3 [1]. La figure 3 montre le spectre du signal original et compressé pour les

deux modèles (FFT et Munich). D'après les deux simulations, on constate une dégradation des hautes fréquences dans le cas des signaux compressés ce qui prouve la sensibilité de l'oreille envers les basses fréquences. De plus cette dégradation est plus importante pour le modèle de Munich. Du point de vue compression, on remarque que le modèle de Munich donne un meilleur résultat que la FFT étant donné que les hautes fréquences sont épuisées à partir de 15000 Hz pour le modèle de Munich et 17500 Hz pour le modèle de la FFT.



**Fig. 3.** Spectre du signal original et du signal compressé : (à gauche le modèle FFT) et (à droite le modèle de Munich)

## 7 Conclusion

Ce papier étudie la compression audio par l'ondelette de Morlet. Un banc de filtre de 26 bandes selon le modèle de Munich a été conçu et testé sur différents type de sons. La comparaison a été mise en se basant sur le résultat des taux de compression obtenus à partir du modèle de la FFT et de Munich. Les résultats obtenus montrent que l'ondelette de Morlet et suivant le modèle de Munich donne, dans ces essais préliminaires, une bonne performance. Une bonne sélectivité a été constatée qui pourra par la suite conduire à des perspectives intéressantes concernant le codage audio en utilisant ce type de modèle psycho-acoustique.

## Références

1. ISO/IEC 11172-3 (F), Information technology –Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3 : Audio,1999
2. Hernandez.G, Mendoza.M, Reusch.B Salinas.L 'Shiftability and filter bank design using Morlet wavelet',Computer Science Society, pp.141- 148,Nov 2004
3. D.Sinha and A.Tewfik, 'Low Bit Rate Transparent Audio Compression using adapted wavelets', IEEE Trans.Sig.Proc, pp.3463-3479,Déc 1993

4. E. Zwicker E.Terhardt, 'Analytical expressions for critical band rate and critical bandwidth as a function of frequency', J.Acoust. Soc.Am, Vol 68, No 5, Nov, 1980
5. I. Daubechies, 'Ten lectures on waveletes', SIAM, Philadelphia, PA, 1992
6. J.O Smith III J.S Abel, 'Bark and ERB Bilinear transforms', IEEE Trans. On Speech and Audio Processing, 1999.
7. H. Fletcher, 'Auditory patterns' Rev. Mod. Phys, 1940, 12,pp. 47-65.
8. M.Rosa Zurera, F.Lopez Ferreras, M.P Jarabo Amores, S.Maldonado Bascon N.Ruiz Reyes, 'A new algorithm for translating a psychoacoustic' information to the wavelet domain', Signal Processing, pp. 519-531,(2001)
9. Chao Wang, Yit-Chow Tong, 'An improved critical-band transform processor for speech applications', Circuits and Systems, pp. 461-464, May 2004
10. M.V. Wickerhauser. Acoustic signal compression with wavelet packets. Dans C.K. Chui, rédacteur, Wavelets : A Tutorial in Theory and Applications. Academic Press, New York, 1992 ;