

Codage audio perceptuel à bas débit par Décomposition Modale Empirique (EMD)

Kais KHALDI^{1,3}, Abdel-Ouahab BOUDRAA², Monia TURKI¹, Thierry CHONAVEL³

¹Unité Signaux et Systèmes, ENIT
BP 37, Le Belvédère 1002 Tunis, Tunisie

²IRENav, Ecole Navale, E³I²(EA3876), ENSIETA
Groupe ASM, Lanvéoc Poulmic, BP600, 29240 Brest–Armées, France

³Département Signal et Communications, TELECOM Bretagne
BP 832, 29285 Brest Cedex, France

kais.khalidi@gmail.com, m.turki@enit.rnu.tn
boudra@ecole-navale.fr
thierry.chonavel@enst-bretagne.fr

Résumé – Cet article présente une nouvelle technique de compression audio à bas débit, basée sur la décomposition EMD (Empirical Mode Decomposition), en association avec un modèle d’audition (seuil de masquage). Par un processus de tamisage, le signal audio est décomposé en une somme finie de composantes de type AM-FM, appelées IMF (*Intrinsic Mode Function*) parfaitement décrites par leurs extrema. En adoptant un seuillage approprié, contrôlé par le modèle psycho-acoustique, seuls les extrema pertinents d’une IMF sont codés. Le nombre de bits alloué au codage des extrema seuillés varie d’une IMF à une autre et respecte la contrainte d’inaudibilité de l’erreur de quantification. Les techniques de seuillage des extrema et d’allocation des bits sur lesquelles repose le procédé de compression proposé, garantissent un bas débit et une bonne qualité d’écoute du signal codé-décodé. Les résultats obtenus sur différents signaux audio, mettent en évidence l’intérêt de l’approche proposée. Comparé à une compression par ondelette et au codeur MP3, le codeur proposé présente un gain de performances significatif en termes de taux de compression et de qualité d’écoute.

Abstract – This paper presents a new audio compression method at low bit rates. It is based on the EMD (Empirical Mode Decomposition), in association with a psycho-acoustic model. By using a time decomposition called sifting process, the signal is decomposed adaptively into intrinsic oscillatory components called Intrinsic Mode Functions (IMFs), which can be fully described by their extrema. The extrema are encoded after a specific thresholding defined from the psycho-acoustic model. The number of bits allocated to the coding of the IMF’s maxima varies from an IMF to another and obeys to the coding error inaudibility constraint. Finally, the decoder recovers the original signal after IMF reconstruction by means of spline interpolations and IMFs summation. Simulation results clearly show the efficiency of the proposed approach. In particular, significant performance gain is achieved here, both in terms of compression ratio and of listening quality, compared to wavelet and a MP3 codecs.

1 Introduction

De nombreuses applications telles que la radiodiffusion, le stockage et la transmission dans les réseaux informatiques ou la télédiffusion par satellites mettent en oeuvre des volumes de données audio importants, ce qui continue à motiver les recherches pour obtenir des flux audio à très bas débit, tout en préservant la qualité d’écoute des signaux décodés [9].

Ainsi, différentes transformations, dont la transformée en ondelette, ont été proposées en association avec des techniques de seuillage qui permettent de ne garder que les informations les plus pertinentes d’un point de vue psycho-acoustique [1, 2, 4, 11]. Pour atteindre un bas débit

et garantir une bonne qualité d’écoute, l’allocation d’un nombre réduit de bits pour le codage des informations, relatives aux différentes composantes de la décomposition, est réalisée sous le contrôle d’un modèle d’audition [7],[9]. L’efficacité de la décomposition par ondelettes est cependant largement tributaire du choix de l’ondelette mère.

Dans ce travail, nous proposons une alternative à la décomposition par ondelettes. Il s’agit de la décomposition modale empirique (EMD) [8]. Contrairement à la décomposition par ondelettes, l’EMD est entièrement pilotée par les données. Par conséquent, l’EMD ne nécessite pas le choix *a priori* d’une famille de fonctions de base de décomposition des signaux.

L’EMD consiste à décomposer un signal en une somme

finie d'IMF. L'analyse du processus du tamisage qui génère les IMF montre qu'on peut envisager un schéma de compression à bas débit basé sur le codage des IMF du signal audio à coder. En effet, chaque IMF peut être vue comme la composante du signal dans une certaine sous-bande, implicitement définie par l'EMD [10]. Du fait du caractère oscillant et de moyenne nulle des signaux à bande étroite, le codage de chaque IMF peut être réalisé en ne considérant que ses extrema. Notons, en particulier qu'une simple interpolation de ces extrema au moyen de fonctions spline[5], permet la reconstruction presque parfaite de l'IMF considérée. L'analyse du processus de tamisage qui génère les IMF montre qu'on peut envisager un schéma de compression des signaux à bas débit en utilisant l'approche EMD. En effet, comme chaque IMF est représentée uniquement par ses extrema et un modèle d'interpolation spline, des solutions efficaces peuvent être envisagées pour leur compression. Donc, le décodeur aura besoin uniquement des extrema préalablement stockés pour reconstruire les IMF et par conséquent le signal initial. L'association du modèle psycho-acoustique dans le procédé de codage des extrema des différents IMFs obtenus, garantira une bonne qualité d'écoute du signal décodé pour un débit fixé.

Le reste du papier est organisé comme suit. Au paragraphe 2, nous rappelons le principe de l'EMD. Le codeur audio perceptuel basé sur l'EMD proposé fait l'objet du paragraphe 3. Dans le paragraphe 4, nous présentons une évaluation des performances du codeur proposé.

2 Principe de l'EMD

L'EMD repose sur la décomposition adaptative du signal en une série d'IMF, au moyen du processus de tamisage. Chacune des IMF peut être considérée comme un niveau d'échelle distinct de la décomposition [8]. Cette notion d'échelle est locale et la décomposition est non-linéaire [8]. La décomposition décrit localement le signal comme une superposition de composantes harmoniques allant des hautes vers les basses fréquences et d'une tendance. Contrairement aux ondelettes, l'EMD n'utilise aucun filtre ou fonction prédéterminée. Notons que l'EMD nécessite le choix d'une fonction interpolatrice (spline cubique) afin d'estimer les enveloppes supérieures et inférieures du signal nécessaires à la reconstruction de l'IMF [8]. En fait, on ne parle d'une IMF que si elle vérifie les critères suivants [8] :

1. Une moyenne nulle.
2. La différence entre le nombre d'extrema et le nombre de passage à zéros est au plus de un (c'est à dire qu'entre un minimum et un maximum successif, l'IMF passe par zéro).

Le principe de décomposition de l'EMD est assuré par le processus de tamisage, suite auquel un signal $x(t)$ s'écrit :

$$x(t) = \sum_{j=1}^C IMF_j(t) + r_C(t), \quad (1)$$

où IMF_j est l'IMF d'ordre j , qui est de plus hautes fréquences que l' IMF_{j+1} , C est le nombre des IMF obtenus. Le signal $r_C(t)$ est appelé résidu, il correspond à la composante de plus basses fréquences du signal. D'après l'équation (1) et en supposant que C est fini, on reconstruit le signal original sans distorsion de l'information [8].

3 Compression audio par EMD

La nouvelle technique proposée ici se décompose en plusieurs modules liés les uns aux autres. Le principe de l'approche proposée est résumé par l'organigramme de la Figure 1.

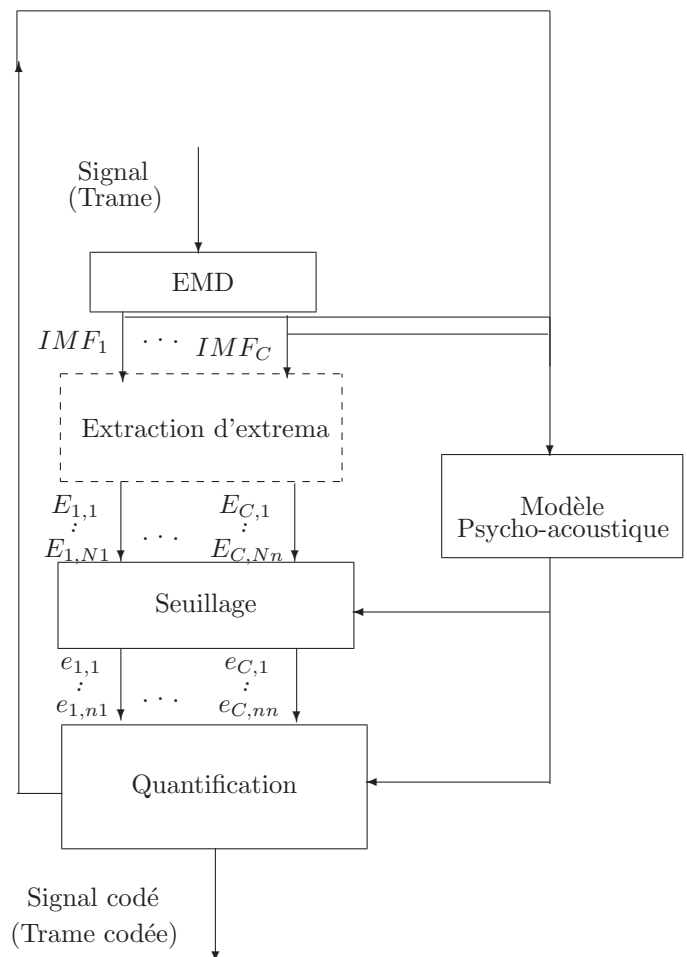


FIG. 1 – Organigramme de la compression par EMD.

3.1 Décomposition par EMD

On découpe tout d'abord le signal audio en trames. Dans la pratique on a considéré lors des tests des trames de longueurs 512 échantillons [9]. En utilisant le processus de tamisage, chaque trame du signal est ensuite décomposée temporellement en une somme de composantes modales $(IMF_i)_{i=1,C}$, qui sont complètement représentées par leurs extrema $(E_{i,N_i})_{i=1,C}$, avec $E_{a,b} = (X_{a,b}, Y_{a,b})$ les coordonnées (position dans le temps et amplitude respectivement) du $b^{ème}$ extremum de l'IMF a [5].

3.2 Seuillage des extrema selon le modèle psycho-acoustique

Notre objectif dans cette partie est de réduire au maximum le nombre d'extrema d'une IMF, tout en assurant que l'erreur entre l'IMF estimée à partir des extrema restants et la vraie IMF reste au-dessous de son seuil de masquage. Ce dernier est calculé en se basant sur le modèle psycho-acoustique utilisé dans le codeur MPEG1 [3],[7]. Il s'agit de déterminer pour chaque IMF le seuil de masquage et la puissance spectrale. Les extrema situés en dessous du seuil de masquage ne sont pas codés. La technique de seuillage utilisée ici est de type dur [13]. On obtient ainsi un jeu réduit d'extrema $(e_{i,n_i})_{i=1,C}$.

3.3 Quantification des extrema seuillés

Puisque le nombre des extrema seuillés décroît d'une IMF à la suivante (les IMF successives sélectionnent des composantes du signal de fréquences de plus en plus basses), le nombre de bits alloués varie d'une IMF à l'autre afin d'optimiser l'allocation de débit, comme c'est le cas dans les codeurs en sous-bandes de type MPEG. Ainsi, le nombre réduit de bits utilisés pour coder les extrema de chaque IMF doit garantir l'inaudibilité de l'erreur de quantification de l'IMF.

Pour cela, on commence par affecter un même nombre réduit de bits pour chaque IMF. Ce nombre de bits peut ensuite être augmenté jusqu'à assurer l'inaudibilité de l'erreur de codage de l'IMF. Il s'agit d'un procédé itératif de quantification de l'IMF suivi de sa reconstruction, en augmentant progressivement le nombre de bits alloués jusqu'à satisfaire la contrainte de masquage. En fait, ce procédé consiste à quantifier l'IMF, la reconstruire puis comparer la Densité Spectrale de Puissance (DSP) son erreur par rapport à son seuil de masquage. Si la DSP de l'erreur est au dessus du seuil de masquage, on recommence la quantification en augmentant le nombre de bits alloués et ainsi de suite jusqu'à ce que la DSP de l'erreur soit au dessous de la courbe de masquage.

Au début, on fixe le nombre de bits pour tout extrema des IMFs (1 bit), la mise à jour du nombre de bits est obtenue en incrémentant de 1 la valeur précédente du nombre de bits. Dès que la nouvelle IMF reconstruite respecte le seuil

de masquage, la boucle de quantification pour cette IMF s'arrête.

Cette méthode de quantification présente un avantage : le nombre de bits utilisés pour respecter la contrainte psycho-acoustique est ici minimisé individuellement pour chaque IMF.

3.4 Codage

La réduction de l'information redondante résiduelle est alors assurée par un codage d'Huffman. Son principe est basé sur une étude statistique définie par la PDF (Probability Density Function). Le code le plus fréquent est attribué à un nouveau code contenant le nombre minimal des bits possible et ainsi de suite.

4 Résultats de simulation

La compression par EMD est appliquée à des signaux audio de natures différentes (chanson, guitare, piano et violon). Ils sont tous échantillonnés à la même fréquence $f_e = 44.1KHz$. La figure 2 présente les signaux originaux. Chaque signal est découpé en trames de taille 512 échantillons [9]. Ensuite en utilisant le processus de tamisage, chaque trame du signal est décomposée en un ensemble d'IMFs et un résidu. Les positions des extrema sont codées sur 9 bits, alors que leurs valeurs sont codées selon le procédé de quantification décrit ci-dessus.

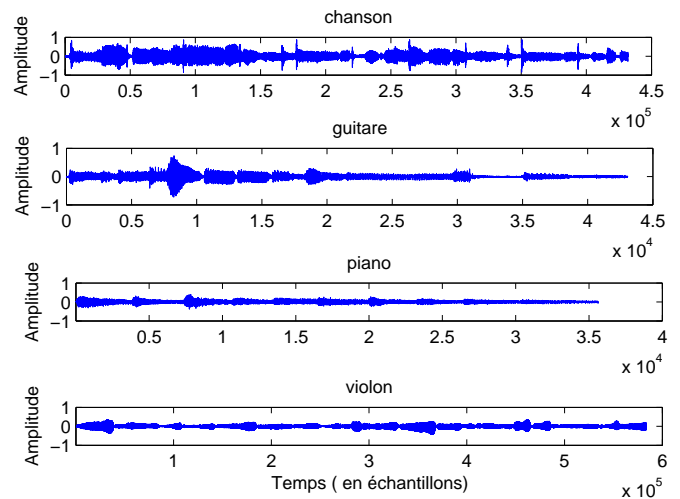


FIG. 2 – Signaux audio (chanson, guitare, piano et violon)

Les résultats obtenus par la méthode proposée sont comparés à ceux obtenus par la méthode de compression par ondelettes (Daubechies 8) [11] et le codeur MP3 [6]. En fait, nous avons choisi l'ondelette avec db8 de 5 niveaux de

décomposition, parce qu'elle donne de meilleurs résultats par rapport aux autres types d'ondelettes [11]. Comme critères d'évaluation des performances de la compression des signaux audio, nous avons opté pour le Taux de Compression (TC), le Rapport Signal à Bruit (RSB), le Subjective Difference Grade (SDG) et le Perceptual Similarity Measure instantané (PSMt) [12]. Ces deux derniers critères fournissent une évaluation subjective de la qualité d'écoute du signal.

Les valeurs de TC, RSB, SDG et PSMt obtenues par ces différentes méthodes sont présentées dans le tableau 1.

TAB. 1 – Résultats de la compression par EMD, MP3 et par ondelette.

	Signal	chanson	guitare	Piano	violon
EMD	TC	11.62 :1	12.8 :1	11.96 :1	12.41 :1
	RSB[dB]	22.28	19.15	21.43	20.03
	SDG	-0.63	-0.70	-0.9	-0.83
	PSMt	0.96	0.94	0.88	0.91
Ondelette	TC	10.11 :1	9.42 :1	9.25 :1	9.83 :1
	RSB[dB]	23.43	20.17	21.59	19.65
	SDG	-1.94	-1.51	-2.01	-1.76
	PSMt	0.81	0.83	0.72	0.79
MP3	TC	6.92 :1	7.37 :1	8.21 :1	7.84 :1
	RSB[dB]	23.69	21.84	17.63	19.72
	SDG	-0.67	-0.79	-0.72	-1.05
	PSMt	0.96	0.92	0.94	0.86

Le tableau 1 montre que notre approche présente des performances meilleures que celles des autres techniques testées. En particulier, l'analyse des valeurs du TC et du (SDG) montre qu'elle offre une amélioration en termes de taux de compression et de qualité audio du signal décodé respectivement. Cette amélioration est clairement visible surtout pour les signaux guitare et violon.

5 Conclusion

Dans ce papier, nous avons présenté une nouvelle approche de compression audio utilisant la méthode EMD en association avec une allocation de bits contrôlée par un modèle psycho-acoustique. Les résultats d'évaluation des performances de l'approche de codage proposée, ont démontré sa capacité à offrir une bonne qualité audio du signal décodé à des débits relativement réduits. Notre approche présente aussi de meilleurs résultats en termes de taux de compression et de qualité d'écoute comparé à l'approche basée sur l'ondelette et le codeur MP3.

Références

- [1] D. Sinha and A. Tewfik, "Low bit rate transparent audio compression using adapted Wavelets," *IEEE Trans. ASSP*, vol. 41, no. 12, 1993.
- [2] G. Stoll and Y. F. Kim, "High quality audio bit rate reduction system family for different applications" in *Proc. Supercomm 1990*, Atlanta, GA.
- [3] H. Taddei, D. Massaloux and A. Le Guyader, "A scalable three bit-rates 8-14.1-24 kbit/s audio coder", *Annales des télécommunications*, Springer, Heidelberg, ALLEMAGNE, 2000, Vol. 55, no 9-10, pp. 483-492, ISSN 0003-4347.
- [4] J. D. Johnston, "Transform coding of audio signals using perceptual criteria", *IEEE J. Select Areas Commun*, vol. 6, pp. 314-323, 1988.
- [5] K.Khalidi, A.O. Boudraa, M.Turki, T. Chonavel and I. Samaali, "Audio encoding based on the Empirical Mode Decomposition," *EUSIPCO*, 2009, Glasgow, Scotland.
- [6] M. Cremmel "Principe du codage MPEG Layer3 / MP3", pp. 1-11, 2006.
- [7] N. Moreau, "Techniques de compression des signaux", *Technique et Scientifique des Télécommunications*, Masson, 1997, ISBN : 2-225-84720-7.
- [8] N. E. Huang et al., "The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. Royal Society*, vol. 454, no. 1971, pp. 903-995, 1998.
- [9] P. Noll, "MPEG digital audio coding", *IEEE signal Proc.*, Vol. 14, no. 5 pp 59-81, Sept.1997.
- [10] P. Flandrin, G. Rilling and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Sig. Proc. Lett.*, vol. 11, no. 2, pp. 112-114, 2004.
- [11] P. R. Deshmukh, "Multiwavelet decomposition for audio coding", *IE(I) Journal-ET* vol. 87, pp. 38-41, 2006.
- [12] R. Huber and B. Kollmeier, "PEMO-Q-A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception," *IEEE Trans. Audio, Speech and Language Proc.*, vol. 14, no 6, 2006.
- [13] S. Mallat, "Une exploration des signaux en ondelettes", *Ellipses, Ecole Polytechnique, Palaiseau*, 2000.