Processing and analysis of sounds signals by Huang transform (Empirical Mode Decomposition: EMD)

Kais Khaldi

N° d'ordre : 2012telb0200

Sous le sceau de l'Université européenne de Bretagne

Télécom Bretagne

En habilitation conjointe avec l'UBO

Ecole Doctorale SICMA

Co-tutelle avec

Ecole Nationale d'Ingénieurs de Tunis

En habilitation conjointe avec l'université Tunis El Manar

Ecole Doctorale STI

TRAITEMENT ET ANALYSE DES SIGNAUX SONORES PAR TRANSFORMÉE DE HUANG (EMD)

Thèse de Doctorat

Mention : Sciences et Technologies de l'Information et de la Communication

Présentée par **Kais Khaldi**

Département : Signal et communications

Laboratoire : Lab-STICC

Directeur de thèse : Thierry Chonavel

Soutenue le 20 janvier 2012

Jury

Mme Sofia Ben Jebara, Professeur, SUP'COM (Rapporteur)

M. Laurent Daudet, Professeur, Université Paris 7 (Rapporteur)

Mme Amel Ben Azza, Professeur, SUP'COM (Présidente)

M. Ali Khenchaf, Professeur, ENSTA Bretagne (Examinateur)

Mme Monia Turki, MC (HdR), ENIT (Directeur)

M. Abdel-Ouahab Boudraa, MC (HdR), Ecole Navale de Brest (Encadrant)

...A mon cher père

...A ma tendre mère.

... A mon épouse et mes fils Ahmed et Ayhem.

...A toute ma famille.

Pour leurs soutiens et sacrifices.

A tous ceux que j'aime et qui m'aiment

Et à tout ceux qui m'ont soutenu durant ces années.

...Je dédie ce travail.

Remerciements

Ce travail a été réalisé conjointement au sein de l'Unité Signaux et Systèmes, de l'Ecole Nationale d'Ingénieurs de Tunis (ENIT), à l'Institut de Recherche de l'Ecole Navale (IRENav) de Brest et au LabSTICC de Télécom Bretagne.

En premier lieu, je tiens à exprimer mes remerciements à Madame Sofia Ben Jebara, Professeur à SUP'Com et Monsieur Laurent Daudet, Professeur à Université Paris 7, pour avoir accepté de rapporter cette thèse.

Je remercie Madame Amel Ben Azza, Professeur à SUP'Com et Monsieur Ali Khenchaf, Professeur à ENSTA Bretagne, d'avoir accepté de faire partie du jury.

Je souhaite remercier également Professeur Christophe Claramunt directeur de l'IRENav de m'avoir accueilli et fourni les moyens matériels nécessaires pour mener à bien ce travail pendant mes séjours à l'Ecole Navale. Un grand merci à l'ensemble du groupe ASM de l'IRENav.

J'adresse mes plus vifs et sincères remerciements à mes encadreurs, Madame Monia Turki Maître de Conférence à l'ENIT, Monsieur Thierry Chonavel Professeur à Télécom Bretagne et Monsieur Abdel-Ouahab Boudraa Maître de Conférences (HdR) à l'Ecole Navale, pour leur qualité d'encadrement, leur rigueure scientifique, leur soucis permanent de comprendre les problèmes traités ainsi que pour l'ambiance sympatique dans laquelle s'est passée les quatres années. La réussite de ce travail leur revient en grande partie.

Je tiens par ailleurs à remercier Bruno Torrésani Professeur à l'Université de Provence pour l'intérêt qu'il a porté à mes travaux de recherche et pour ses recommandtions et suggestions dans le domaine du codage audio.

Je remercie mon épouse, mes parents, mes frères et mes soeurs pour leur soutien et leurs encouragements à surmonter les différents problèmes rencontrés.

Un grand Merci pour tous ceux qui m'ont aidé de prêt ou de loin à réaliser ce travail.

Résumé détaillé en français de la thèse

La décomposition modale empirique (Empirical Mode Decomposition "EMD" en anglais) est une méthode caractérisée par un processus appelé Tamisage (Sifting) permettant de décomposer temporellement un signal en une somme de composantes oscillantes appelées Modes Empiriques connues sous le nom de Intrinsic Mode Functions (IMF).

Le but général de la thèse est l'exploration des possibilités de l'EMD pour traitement et l'analyse des signaux sonores avec comme application débruitage, compression et tatouage. Ainsi, mes travaux de recherche actuels s'inscrivent dans un esprit de continuité du travail effectué en mastère, qui touche particulièrement les traitements du signal. Le rapport de la thèse est écrit en anglais, il est structuré en quatre parties.

.1 Transformée de Huang : EMD

Dans ce chapitre, on propose d'étudier la technique EMD en précisant ses caractéristiques, tout en insistant sur les critères qui nous offrent une bonne décomposition du signal.

.1.1 Principe de la méthode EMD

L'EMD est une méthode algorithmique de décomposition des signaux. Elle se base sur le principe de décomposer le signal en une somme d'une composante locale haute

fréquence (oscillation rapide) et d'une composante basse fréquence (tendance). Ce principe est illustré par l'équation (1):

$$x(t) = d(t) + m(t) \tag{1}$$

où $x(t)$ constitue le signal à décomposer, $d(t)$ est l'oscillation rapide, $m(t)$ est le signal tendance et t indique le temps discret.
De même le signal tendance peut être aussi décomposé en deux termes (2).

$$m(t) = d_1(t) + m_1(t) \tag{2}$$

où $d_1(t)$ est la composante haute fréquence et $m_1(t)$ est la composante basse fréquence.
Pour calculer un mode relatif à un signal, on suit le principe suivant :

1. Identifier tous les extrema locaux de $x(t)$.
2. Interpoler les minima (resp. les maxima) de manière à construire une certaine enveloppe: EnvMin (resp. EnvMax).
3. Calculer la moyenne de deux enveloppes m(t) = (EnvMin(t) + EnvMax(t))/2.
4. Extraire le détail $d(t) = x(t) - m(t)$. Le signal d(t) n'est consideré IMF qu'après un certains nombre d'itérations nécessaires afin que d(t) obéisse à un critère d'arrêt donné.

En itérant ce principe, on obtient une décomposition du signal décrite comme suit:

$$x(t) = \sum_{j=1}^{N} IMF_j(t) + r(t) \; avec \; N \in \mathbb{N}^* \tag{3}$$

où IMF_j est l'IMF d'ordre j qui est de type plus haute fréquence que l'IMF_{j+1}.
Le signal $r(t)$ est appelé résidu, il correspond à la composante la plus basse fréquence du signal.
D'aprés (3) et en supposant que N est fini, on reconstruit linéairement le signal

original sans perte ou distorsion de l'information [34].
Toutefois, on ne parle d'une IMF que si elle vérifie les critères suivants [34]:

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 défini par l'algorithme décrit dans ce qui suit.

.1.2 Procédure algorithmique de l'EMD

Notations :
ϵ : indique le seuil prédéfinie, c'est un critère de condition de la boucle indicée par i.
j : représente l'indice de l'IMF.
i : constitue l'indice de l'itération appliquée sur le résidu pour vérifier le critère d'une IMF.
r_j : désigne le résidu aprés l'obtention de la j^{eme} IMF
$h_{j,i}$: c'est une variable intermédiaire de calcul qui prend la valeur du nouveau résidu à la première itération, puis, elle prend la différence entre le résidu et la valeur de l'enveloppe moyenne aux itérations suivantes.
$U_{j,i}$: représente l'enveloppe supérieure de $h_{j,i}$, construite par interpolation des maxima.
$L_{j,i}$: représente l'enveloppe inférieure de $h_{j,i}$, construite par interpolation des minima.
$\mu_{j,i}$: désigne l'enveloppe moyenne, obtenu à partir des deux enveloppes de $h_{j,i}$.
SD (i) : indique le critère d'arrêt à la $i^{ème}$ itération.

L'algorithme correspondant à la méthode EMD peut s'écrire sous la forme du pseudo - code suivant :
Etape1: fixer ϵ, $j \leftarrow 1$ ($j^{ème} IMF$).
Etape2 : $\mathrm{r}_{j-1}(\mathrm{t}) \leftarrow x(t)$ (résidu).
Etape3 : extraire la $j^{ème} IMF$:
(a) : $\mathrm{h}_{j,i-1}(\mathrm{t}) \leftarrow \mathrm{r}_{j-1}(\mathrm{t})$,$i \leftarrow 1$ (i;itération de la boucle sifting).
(b) : extraire les maxima et les minima locaux de $\mathrm{h}_{j,i-1}(\mathrm{t})$.

(c) : calculer les enveloppes supérieure et inférieure : $U_{j,i-1}(t)$ et $L_{j,i-1}(t)$ par interpolation (splines cubiques par exemple) des maxima et des minima de $h_{j,i-1}(t)$ respectivement.
(d) : calculer l'enveloppe moyenne : $\mu_{j,i-1}(t)$ =$(U_{j,i-1}(t) + L_{j,i-1}(t))/2$.
(e) : mettre à jour $h_{j,i}(t) \leftarrow h_{j,i-1}(t)$- $\mu_{j,i-1}(t)$, $i \leftarrow i+1$.
(f) : calculer le critère d'arrêt (par exemple) : $SD(i) = \sum_{t=0}^{T} \frac{|hj,i-1(t)-hj,i(t)|^2}{(hj,i-1(t))^2}$,
où T représente le nombre d'échantillons du signal.
(g) : décision : répeter l'étape (b),(f) tant que $SD(i)<\epsilon$.
à la sortie de l'étape(3), on met $IMF_j \leftarrow \mathbf{h_{j,i}(t)}$ ($j^{ème} IMF$).
Etape4 : mettre à jour le résidu $r_j(t) \leftarrow r_{j-1}(t)$ - $IMF_j(t)$.
Etape5 : répéter l'étape(3) avec $j \leftarrow j+1$ jusq'u à ce que le nombre d'extrema dans $r_j(t) \leq 2$.
L'algorithme décrit ci-dessus, comporte deux boucles imbriquées l'une dans l'autre, celle indicée par j permet d'extraire l'IMF, qui nous détermine le niveau de profondeur de décomposition et l'autre indicée par i conditionne la fonction IMFj(t) de manière à respecter les critères requis; avoir deux enveloppes symétriques afin que le signal extrait IMF_j soit bien une IMF.
Une bonne décomposition donnée par cet algorithme est conditionnée par le choix de certains paramètres.

.1.3 Paramètres pertinants de la décomposition

Généralement, le choix des paramètres repose sur le critère d'arrêt. Comme il existe deux boucles dans l'algorithme, il faut s'assurer que les deux doivent s'arrêter. La boucle principale indicée par j s'arrête lorsqu'il n'est plus possible de décomposer le résidu courant càd que $r_j(t)$ possède moins de deux extrema. La boucle indicée par i est liée à un critère d'arrêt qu'il convient de définir de manière précise.
La $2^{ème}$ boucle (indicée par i) va s'arrêter lorsque $h_{j,i}(t)$ vérifie les critères de définition d'une IMF (de moyenne nulle). Théoriquement, cette hypothèse n'est pas démontrée, pour cela en pratique on ajoute à ce critère un autre qui évite au proccesus de tamisage entrer dans une boucle infinie. La définition d'un critère d'arrêt du processus de tamisage est alors nécessaire:

Ainsi dans [34], les auteurs proposent un critère d'arrét SD(i) reposant sur la

deviation standard et défini par :

$$SD(i) = \sum_{t=0}^{T} \frac{|hj, i-1(t) - hj, i(t)|^2}{(hj, i-1(t))^2} \tag{4}$$

Le test d'arrêt est validé lorsque la différence entre deux tamisages consécutifs est inférieur à un seuil prédéfinie ϵ. Typiquement, la valeur ϵ permettant de stopper le tamisage est comprise entre 0.2 et 0.3 [1]. Cette valeur réalise un certain compromis. En effet si ϵ est trop grand, l'EMD ne permet pas de séparer les différents modes présents dans le signal, cependant si ϵ est trop petit, l'EMD risque d'aboutir a des composantes dont l'amplitude est quasiment constante et modulée par une seule fréquence (sur-décomposition de signal).

Un autre critère local a été proposé par P.Flandrin [28] et notamment choisi en pratique. Ce critère est défini comme suit :

$$\sigma(t) = 2|\frac{\mu_{i-1}(t)}{U_{i-1}(t) - L_{i-1}(t)}| \tag{5}$$

En adoptant le critère $\sigma(t)$, trois conditions nécessaires sont définies pour que $\mathrm{h}_{i,j}$(t) soit bien une IMF [28].

- La différence entre le nombre de zéros de h_i(t) et les nombres d'extrema de h_i(t) est infèrieure ou égale en valeur absolue à 1.
- $\sigma(t) < \theta_1$ pour t $\leq$ (1 - α)T
- $\sigma(t) < \theta_2$ pour (1 - α)T < t <T

où T : la taille de la fenêtre d'analyse, $\theta_1$et θ_2 deux réels tels que $0 \leq \theta_1 \leq \theta_2$ et $0 \leq (\alpha \equiv (Tolerence)) \leq 1$

La première condition revient à dire qu'une IMF doit être une fonction oscillante autour de zéro : entre un maximum et un minimum, il doit y avoir un passage par zéro. Les deux dernières conditions exigent que le paramètre $\sigma(t)$ soit faible. Toute fois, il peut dans une certaine mesure prendre des valeurs élevées.

Dans [28]le bon copromis du choix des valeurs des seuils $\theta_1$et θ_2 est le suivant :

$\theta_1 \approx 0.05$ et $\theta_2 \approx 10 * \theta_1$ et $\alpha \approx 0.05$. On conclut que tous les critères d'arrêt sont exigés pour que $\mathrm{h}_{j,i}$(t) vérifie bien les propriétés d'une IMF.

Dans notre travail, nous adoptons le critère choisie par [28], car il nous permet d'obtenir des modes qui correspondent bien à la définition d'une IMF.

.2 Débruitage des signaux de la parole par EMD

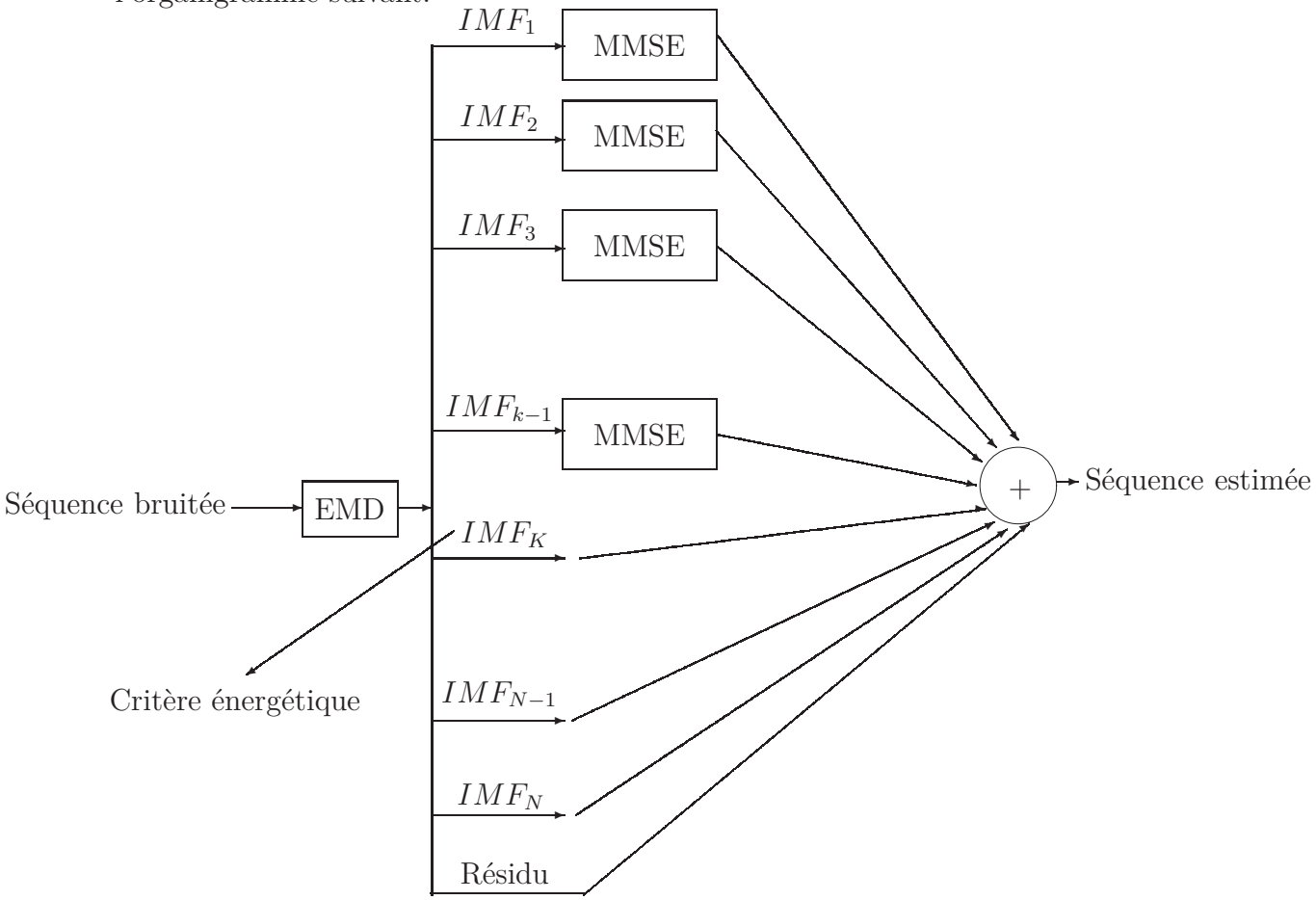
Nous présentons dans cette partie une procédure basée sur l'EMD pour le rehaussement du signal de la parole. En particulier le traitement proposé tiendra compte du caractère voisé ou non voisé de la séquence de parole considérée. Puisque le signal de parole est constitué de séquences voisées et non voisées, on a été amené à considérer séparément les deux types de séquences.
L'idée du débruitage d'un signal de parole bruité se présente selon le principe suivant:

1. Découper le signal bruité en trames.
2. Pour chaque trame on fait appelle à l'EMD pour la décomposer.
3. Après avoir décomposer la trame bruitée en modes, on calcule l'énergie de chacun des modes et suivant la variation des énergies, on déduit le type de la trame.
4. Suivant le type de la trame, on applique le procédé du débruitage, c'est à dire s'il s'agit d'une séquence voisée, on applique l'approche du filtrage puis on débruite seulement les modes qui ne sont pas pris lors du filtrage par EMD, alors que dans le cas d'une séquence non voisée, on débruite tous les modes.
5. Le signal estimé est reconstruit en utilisant les séquences débruitées.

.2.1 Débruitage de séquences voisées

La séparation entre le bruit et le signal original est possible. En fait, cette séparation se base sur l'hypothèse que les premières IMF (les modes de plus hautes fréquences) sont majoritairement dominés par le bruit et sont peu représentatives de l'information propre au signal initial. Cependant, les modes qui correspondent au signal non bruité contiennent quand même un peu du bruit. Le débruitage de ces modes va engendrer une distorsion au niveau de reconstitution du signal estimé. Ainsi, le débruitage d'une séquence voisée revient à débruiter seulement les modes qui ne sont pas filtrés par EMD. Enfin, le signal débruité est la somme des

modes filtrés par EMD et les modes débruités. L'approche proposée est résumée par l'organigramme suivant:



Organigramme de débruitage d'une séquence voisée par approche EMD-MMSE

.2.2 Débruitage de séquences non voisées

Lors de la décomposition d'un signal de type non voisé bruité par EMD, qu'il est difficile de séparer le signal original du bruit. Cependant, l'hypothèse que le bruit est uniquement réparti sur les premières IMF n'est pas vérifiée sur les séquences non voisées. Ainsi, les informations qui correspondent au signal original seraient intégrées dans tous les modes, donc l'approche du débruitage se basera sur un traitement de tous ces modes un par un. La procédé consiste à reconstruire le signal estimé avec toutes les IMF préalablement filtrés.

.3 Codage des signaux audio par EMD

Dans cette partie, nous proposons une alternative à la décomposition par ondelettes, il s'agit de la décomposition modale empirique (EMD) [34]. 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 [28]. 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 ses extrema au moyen de fonctions spline[47], permet la la reconstruction presque parfaite de l'IMF considérée. L'analyse du processus du 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, un codage pour la compression est possible. Ainsi, la compression du signal correspond à celle des extrema des IMF. 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.

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

.3.1 Décomposition par EMD

On découpe tout d'abord le signal audio en trames de taille 512 échantillons [63]. 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

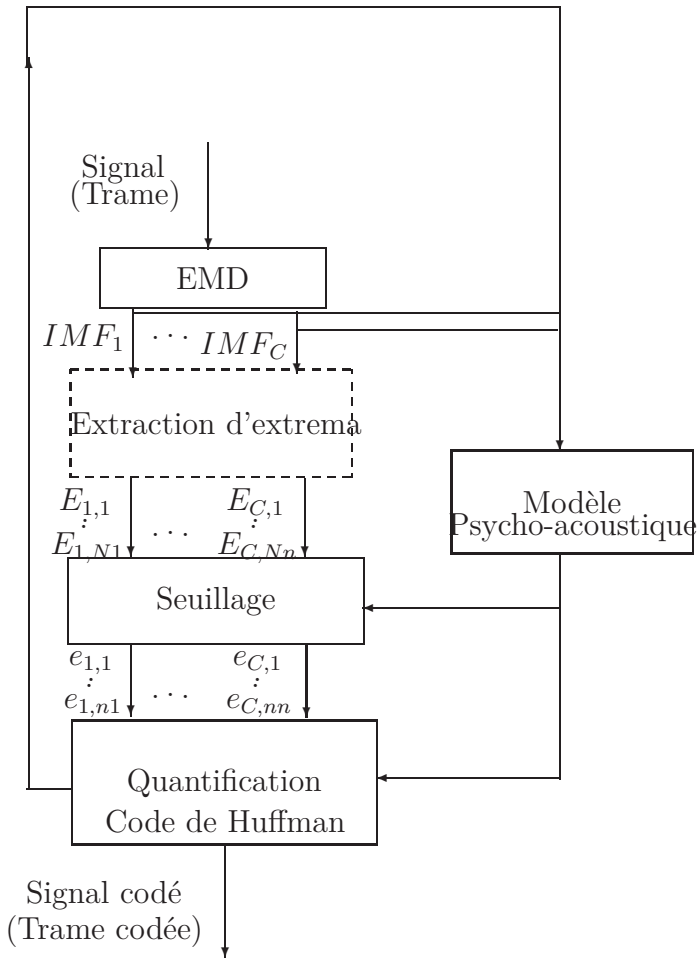


Figure 1: Organigramme de la compression par EMD.

complètement représentées par leurs extrema $(E_{i,N_i})_{i=1,C}$, avec $E_{a,b} = (X_{a,b}, Y_{a,b})$ la position du $b^{ème}$ extremum de l'IMF a.

.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. La technique de seuillage utilisée ici est de type dur [55]. On obtient ainsi un jeux réduit d'extrema $(e_{i,n_i})_{i=1,C}$.

.3.3 Quantification des extrema seuillées

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 être ensuite 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 bits), la mise à jour du nombre de bits est obtenue en addition par un l'ancienne valeur du nombre de bits. Dés 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, car 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.

.3.5 Résultats de simulation

L'approche de 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 $fe = 44.1KHz$. La Figure 2 présente les signaux originaux. Chaque signal est découpé en trames de taille 512 échantillons [63]. Ensuite en utilisant le processus de tamisage, chaque trame du signal est decompsée en ensembles d'IMFs et un résidu. Les positions des extrema sont codés sur 9 bits, alors que leurs valeurs sont codés selon le procédé de quantification décrit ci-dessus.

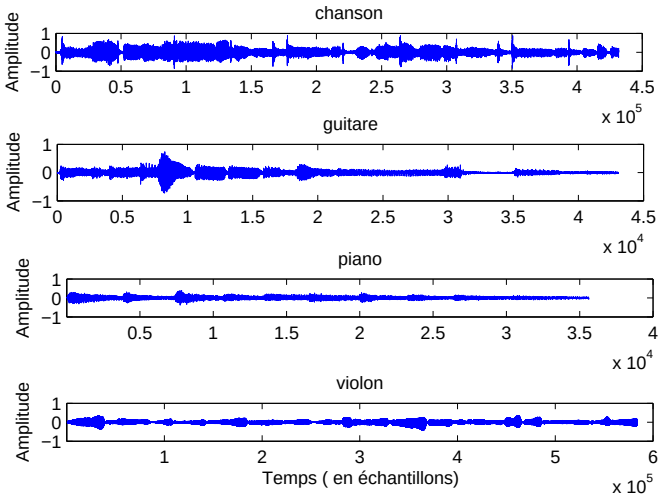


Figure 2: Signaux audio (chanson, guitare, piano et violon)

Les résultats obtenus par la méthode proposée sont comparés à ceux obtenus par la méthode à base de l'ondelette (Daubechies 8) [20] et le codeur MP3 [**?**]. En fait, nous avons choisi db8 de 5 niveaux de décomposition, parce qu'elle donne de meilleurs résultats par rapport aux autres types d'ondelettes [20]. Comme critère d'évaluation des performances de la compression des signaux audio, nous avons opté pour le Taux de Compression (TC), Rapport Signal à Bruit (RSB), Subjective Difference Grade (SDG) et instantaneous Perceptual Similarity Measure (PSMt). Ces deux derniers critères sont offrent une évaluation de la qualité d'écoute du signal. subjective
Les valeurs de TC, RSB, SDG et PSMt obtenues par ces différentes méthodes sont présentées dans le tableau VI.1.

Le tableau VI.1 montre que notre approche présente des performances meilleures

Table 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	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

que celles des autres techniques testées. En effet, 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. En particulier cette amélioration est clairement visible surtout pour les signaux guitare et violon.

.4 Tatouage des signaux audio par EMD

L'EMD consiste à décomposer un signal en une somme finie de composantes de type AM-FM, appelées IMF (*Intrinsic Mode Function*). L'analyse du processus du tamisage qui génère les IMF montre qu'on peut envisager un schéma de tatouage qui consiste à insérer la marque dans la dernière IMF. En effet, la dernière IMF peut être vue comme la composante la plus basse fréquence, par conséquence la plus résistante aux attaques. Ainsi, on propose d'insérer la marque en association avec le code de synchronisation dans les extrema de la dernière IMF.

.4.1 Algorithme de tatouage proposé

Code de synchronisation

Le code de synchronisation est introduit pour localiser la position de la marque dans le signal et par conséquence facilite l'extraction de la marque du signal tatoué.

Etant donné un code de sunchronisation U et une séquence inconnue V qui sont de même longeur. La séquence V est définie comme étant un code de synchronisation si seulement la valeur de simularité entre U et V (bit par bit) est supérieur ou égale à un seuil prédefini τ.

Procédure d'insertion

Après combination de la marque avec le code de synchronisation pour former un flux binaire m_i, la procédure d'insertion de la marque est illustré dans les étapes suivantes:

Etape 1: Segmenter le signal audio en trames.

Etape 2: Decomposer chaque trame en IMFs, en utilisant l'EMD.

Etape 3: Insérer P fois la séquence binaire m_i dans les extrema de la dernière IMF. L'insertion des bits se fait par modulation d'amplitude des extrema, ainsi chaque bit de sequence binaire doit être inséré comme suit:

$$e'_i = \begin{cases} \lfloor e_i/S \rfloor .S \ sgn \ 3S/4 & \text{si } m_i = 1 \\ \lfloor e_i/S \rfloor .S \ sgn \ S/4 & \text{si } m_i = 0 \end{cases} \tag{6}$$

e_i et e'_i désigne les extrema de la dernière IMF de signal audio respectivement le signal audio tatoué. sgn est égale à "+" si e_i est un maximum, et "-" si'il est un minimum. $\lfloor \ \rfloor$ est la fonction partie entière, et S répresente le facteur d'insertion, que doit être choisi de telle sorte que le signal tatoué respecte la contraine d'inaudibilité.

Etape 4: Reconstruire la trame (EMD^{-1}) en utilisant la dernière IMF modifiée puis on concaténe la trame tatouée pour construire le signal audio tatoué.

Procédure d'extraction

Etant donné N_1 et N_2 est le nombre de bits de code de synchronisation respectivement le nombre de bits de la marque. L'extraction de la marque est décrit comme suit:

Etape 1: Segmenter le signal en trames.

Etape 2: Decomposer chaque trame en IMFs, en utilisant l'EMD.

Etape 3: Extracter les extrema e^*_i de la dernière IMF.

Etape 4: Extracter la sequence m_i^* de e_i^*.

$$m_i^* = \begin{cases} 1 & \text{si } e_i^* - \lfloor e_i^*/S \rfloor .S \geq \ sgn\ S/2 \\ 0 & \text{si } e_i^* - \lfloor e_i^*/S \rfloor .S < \ sgn\ S/2 \end{cases} \tag{7}$$

sgn est "+" si e_i^* est un maximum,et "-" s'il est un minimum.
Etape5: Grouper toute sequence de bits y_i.
Etape 6: $I \leftarrow 1$ et $L \leftarrow N_1$ (taille fenêtre)
Etape 7: Evaluer la similarité entre le premier code de synchronisation extracté, V = y(I:L), et le code de synchronisation original U bit par bit. Si la valeur de similarité est $\geq \tau$, Donc V est considéré comme étant le code de synchronisation et sauter à l'étape 9, sinon aller à l'étape 8.
Etape 8: $I \leftarrow I + 1$ et $L \leftarrow L + 1$ et revenir à l'étape 7.
Etape 9: Evaluer la similarité entre le second code de synchronisation extracté, V'=y(I+N_1+N_2: I+2N_1+N_2) et le code de synchronisation original. si la valeur de similarité $\geq \tau$, donc V' est considéré comme étant le code de synchronisation code, et extracter la marque N2 bits (y(I+N_1: I+N_1+N_2-1)) à partir de la position I+N1 et aller à l'étape 10, sinon revenir à l'étape 8.
Etape 10: $I \leftarrow I + N_1 + N_2$, si la nouvelle valeur I est égale à la longeur de séquence y_i, aller à l'étape 11, sinon revenir à l'étape 8.
Etape 11: Extracter le P marques et fait comparison bit par bit entre ces marques, pour la correction, et finalement extracter la marque désirée.

.4.2 Principaux résultats

Pour illustrer les performances de l'algorithme de tatouage par EMD, nous avons effectué des simulations numériques sur des signaux audio de natures différentes. Les signaux sont tous échantillonnés à la fréquence fe = 44.1KHz. La marque est une image logo binaire.
Pour evaluer la performance de l'algorithme proposé, nous avons utilisé les deux critères suivants : TEB et NC (Normalised Cross-correlation).
Table VI.1 montre la robustesse de l'algoritme proposé pour le signal audio "Rock". Les Valeurs de TEB et NC reflette la bonne performance de notre algorithme pour differents type d'attaques.
Le tableau B.1 montre que notre approche présente de meilleures performances que les autres techniques testées. En effet, elle offre une amélioration en termes de débits et robustesse en fonction de codeur MP3 par rapport aux autres algorithmes

Table 2: TEB et NC de la marque extractée pour le signal audio "Rock" par l'approche proposée.

Type d'attaque	TEB %	NC
Pas d'attaque	0	1
AWGN	0	1
Débruitage	0	1
Cropping	0	1
Rechantillonage	1	0.9989
MP3(64 kb/s)	0	1
MP3 (32 kb/s)	0	1
Requantification	0	1

de tatouage.

Table 3: Performance des algorithmes de tatouage, trier par débits.

Référence	Débits (b/s)	Robustesse avec MP3 (kb/s)
Algorithme proposé	46.9-50.3	32
Bhat K	45.9	32
Lie	43	80
Cvejic	27.1	32
Yeo	10	96
Tachibana	8.5	96
Li	4.2	32
Mansour	2.3	56
Xiang	2	64
Kirovski	0.5-1	32

.5 Conclusion

Dans cette thèse on a exploré l'apport de l'EMD en traitement et en analyse des signaux audio et de parole. Cette décomposition du signal en IMF est adaptative et ne fait pas d'hypothèses (stationnarité et linéarité) sur le signal à analyser. Le comportement en banc de filtre dyadique de l'EMD ainsi que la quasi-symétrie des modes et leur représentation via leurs extrema sont les propriétés qui sont l'origine des outils qu'on a développés: débruitage, codage et tatouage. Ces contributions sont illustrées sur des données synthétiques et réelles et les résultats comparés à ceux de méthodes éprouvées telles que le filtre MMSE, l'approche ondelettes et les codecs AAC et MP3 montrent les bonnes performances des outils développés autour de l'EMD. Ces résultats montrent les capacités de l'EMD comme outils de traitement et d'analyse de façon adaptative des signaux audio et de parole.

Contents

.1 Transformée de Huang : EMD iv

.1.1 Principe de la méthode EMD iv

.1.2 Procédure algorithmique de l'EMD vi

.1.3 Paramètres pertinants de la décomposition vii

.2 Débruitage des signaux de la parole par EMD ix

.2.1 Débruitage de séquences voisées ix

.2.2 Débruitage de séquences non voisées x

.3 Codage des signaux audio par EMD xi

.3.1 Décomposition par EMD xi

.3.2 Seuillage des extrema selon le modèle psycho-acoustique xii

.3.3 Quantification des extrema seuillées xiii

.3.4 Codage xiii

.3.5 Résultats de simulation xiv

.4 Tatouage des signaux audio par EMD xv

.4.1 Algorithme de tatouage proposé xv

Code de synchronisation xv

Procédure d'insertion xvi

Procédure d'extraction xvi

.4.2 Principaux résultats xvii

.5 Conclusion xviii

List of Figures **6**

List of Tables **11**

Abbreviations **14**

Introduction **15**

I Huang transform **24**

I.1 Introduction 25

I.1.1 Principle of EMD 25

I.1.2 EMD algorithm 26

I.1.3 Meaningful parameters of EMD 27

I.1.3.1 Stopping criterion 28

I.1.3.2 Interpolation 29

I.2 IMFs properties 31

I.2.1 IMFs orthogonality 31

I.2.2 PDE for IMFs characterization 33

I.3 EMD: a time-frequency description tool 33

I.3.1 Importance of the sampling frequency 33

I.3.2 Tones separation 35

I.3.3 EMD acts as a Filter bank: Gaussian white noise case 37

I.3.4 Comparison with wavelets 38

I.4 Conclusion 40

II Speech enhancement by EMD **42**

II.1 Introduction 43

II.2 EMD based white noise reduction 43

II.2.1 EMD-MMSE filter 44

II.2.2 EMD-Shrinkage 45

II.2.3 EMD-MMSE versus EMD-Shrinkage 46

II.3 EMD-ACWA filtering of white and colored noises 50
II.3.1 Interest of ACWA filter 50
II.3.2 Performance analysis of EMD-ACWA 54
II.4 Conclusion 59

III Speech denoising using EMD and local statistics 63
III.1 Introduction 65
III.2 Frames classification 65
III.2.1 Voiced frames detection 66
III.2.2 Transient frames detection 68
III.3 Proposed speech denoising method 70
III.3.1 Voiced sequence denoising 72
III.3.2 Unvoiced sequence denoising 73
III.3.3 Transient sequence denoising 73
III.4 Performance analysis 73
III.4.1 Voiced frames 73
III.4.2 Speech signal 78
III.5 Conclusion 82

IV Signal coding schemes in EMD framework 85
IV.1 Introduction 87
IV.2 Why IMFs coding? 87
IV.2.1 IMF extrema 87
IV.2.2 Quasi-symmetry of IMF 88
IV.2.3 IMF modelling 89
IV.3 EMD based encoder architecture 90
IV.3.1 IMF extrema coding basics: $IMF_{extrema}$ 90
IV.3.1.1 Segmentation and decomposition 90
IV.3.1.2 Extrema thresholding 90
IV.3.1.3 Extrema quantification 91

IV.3.1.4 Coding 92
IV.3.1.5 Decoding process 92
IV.3.2 IMF envelope coding basics : $IMF_{envelope}$ 92
IV.3.2.1 Encoding scheme 92
IV.3.2.2 Decoding process 93
IV.4 HHT based encoder architecture 94
IV.4.1 IA and IP coding basics: $IA - IP$ 94
IV.4.1.1 IA encoding 94
IV.4.1.2 IP encoding 94
IV.4.1.3 Decoding scheme 94
IV.4.2 IA and IF coding basics: $IA - IF$ 95
IV.4.2.1 IF encoding 95
IV.4.2.2 Decoding approach 96
IV.5 conclusion 96

V Encoding schemes: Application to audio signals 99

V.1 Introduction 100
V.2 Encoders architecture 100
V.2.1 Transient detection 100
V.2.2 Thresholding step for $IMF_{extrema}$ coder 103
V.2.3 Quantization step for $IMF_{envelope}$ and $IMF_{extrema}$ 104
V.3 EMD based audio coders performance 105
V.4 Conclusion 108

VI Audio watermarking based on the EMD 110

VI.1 Introduction 112
VI.2 Proposed watermarking algorithm 112
VI.2.1 Synchronization code 114
VI.2.2 Watermark embedding 115
VI.2.3 Watermark extraction 116

VI.3 Performance analysis 117
VI.4 Results 118
VI.5 Conclusion 124

Conclusion and perspectives **126**

Bibliography **132**

Appendix **141**

A Chapter III **141**

B Chapter V **142**

List of Figures

1 Organigramme de la compression par EMD. xii

2 Signaux audio (chanson, guitare, piano et violon) xiv

I.1 Interpolation of the signal ($x(t) = \cos(t) + \sqrt{(}t)$) by different methods and the corresponding error. 30

I.2 Decomposition the signal $x(t) = \sin(8t) + \sin(3t) + 2t$ by EMD. . . . 31

I.3 Decomposition of the tone signal (Eq. I.13) by EMD. 34

I.4 Estimation and behavior of $E(\nu)$ associated with the first IMF for a tone. 35

I.5 Decomposition of the signal (Eq.I.15) by EMD. 36

I.6 Estimation and behavior of the error $E(\nu_1, \nu_2)$ Eq. I.16 for signal $x(t) = x_{\nu_1}(t) + x_{\nu_2}(t) = cos(2\pi\nu_1 t) + cos(2\pi\nu_2 t)$. 37

I.7 IMFs spectra for a white noise. 38

I.8 Signal $x(t) = \sin(3t) + \sin(0.3t) + \sin(0.03t)$ and its theoretical components. 39

I.9 Comparison of decomposition by the EMD to wavelet. 39

I.10 Error estimates with EMD and wavelet. 40

II.1 Original signals "speech1", "speech2", "speech3" and "speech4". 46

II.2 Noisy version of signals "speech1", "speech2","speech3" and "speech4" (input SNR = 5 dB). 47

II.3 Denoising results of signals "speech1", "speech2","speech3" and "speech4" by the EMD-MMSE and the MMSE filter. 48

II.4 Final SNR values obtained from different initial noise levels of signals "speech1", "speech2", "speech3" and "speech4". The results are averages over 100 instances of the noisy signals. They are reported for EMD-MMSE and the MMSE filter. 49

II.5 PESQ values obtained from different initial noise levels of signals "speech1", "speech2", "speech3" and "speech4". The results are averages over 100 instances of the noisy signals. They are for EMD-MMSE and the MMSE filter. 50

II.6 Noisy versions of signals "speech1", "speech2", "speech3" and "speech4" (input SNR =-5 dB). 51

II.7 Denoising results of signals "speech1", "speech2", "speech3" and "speech4" by the EMD-Shrinkage and the wavelet approach (Daubechies 4). 52

II.8 Final SNR values obtained from different initial noise levels of signals "speech1", "speech2", "speech3" and "speech4". The results are averages over 100 instances of the noisy signals. They are reported for EMD-Shrinkage and for three different Wavelets (Haar, Symmlet 4, Daubechies 4). 53

II.9 Variations of the PESQ values versus from the input SNR for signals "speech1", "speech2", "speech3" and "speech4". The results are averages over 100 instances of the noisy signals. They are reported for EMD-Shrinkage and for three different wavelets (Haar, Symmlet 4, Daubechies 4). 54

II.10 Clean and filtered signals by the ACWA and the MMSE filters (input SNR=2 dB). 55

II.11 Noise power spectral density 56

II.12 Original signals ("speech1" and "speech2") and their noisy versions (f16 noise with SNR =-2 dB). 56

II.13 Variation of the output SNR relating to the noisy signal "speech1" versus the size L of the ACWA filter window (f16 noise with SNR=-2 dB and SNR=0 dB). 57

II.14 Denoised version of the signals "speech1" and "speech2" obtained by the EMD-ACWA, the wavelet (db4) and ACWA filter (f16 noise with input SNR =-2 dB) 58

II.15 Variation of the output SNR versus the input SNR relating to the denoising of the signals "speech1" and "speech2" corrupted by a white Gaussian noise. The results are averages over 100 instances of the noisy signals. They are reported for EMD-ACWA, ACWA filter and wavelet(db4) 58

II.16 Variation of the output SNR versus the input SNR relating to the denoising of the signals "speech1" and "speech2" corrupted by the f16 noise. The results are reported for EMD-ACWA, ACWA filter and wavelet (db4) 59

II.17 Variation of the output SNR versus the input SNR relating to the denoising of the signals "speech1" and "speech2" corrupted by the factory noise. The results are reported for EMD-ACWA, ACWA filter and wavelet(db4) 59

II.18 PESQ values obtained from different initial noise levels of signals "speech1" and "speech2". The results are an average of 100 instances signal. It's reported for EMD-ACWA, ACWA filter and wavelet(db4) 60

II.19 PESQ values obtained from different initial noise levels of signals "speech1" and "speech2" corrupted by the f16 noise. The results are reported for EMD-ACWA, ACWA filter and wavelet(db4) 60

II.20 PESQ values obtained from different initial noise levels of signals "speech1" and "speech2" corrupted by the factory noise. The results are reported for EMD-ACWA, ACWA filter and wavelet(db4) 61

III.1 Frames classification scheme. 66

III.2 Voiced sequence, noisy voiced sequence and the energy variations of their noisy IMFs. 69

III.3 Not voiced sequence, noisy not voiced sequence and variations of the energies of its noisy IMFs. 70

III.4 Unvoiced sequence, noisy unvoiced sequence and variations of the energies of its IMFs. 71

III.5 The stationarity index of a noisy transient frame. 72

III.6 Energy variations of the IMFs of the sub-frames. 72

III.7 Original signals /o/, /a/, /e/ and /i/. 74

III.8 Noisy versions of signals /o/, /a/, /e/ and /i/ (input SNR=2 dB). . . 74

III.9 Decomposition of noisy signal /o/ into IMFs (input SNR= 2dB) . . . 75

III.10 Variations of CMSE (energy) values versus the number of IMFs for the four noisy signals. 75

III.11 Enhanced signals obtained by the proposed method, Wavelet (db4), ACWA filter and EMD-ACWA (input SNR=2 dB). 77

III.12 Variations of output SNR versus input SNR for signals /o/, /a/, /e/ and /i/. The results are average over 100 noise realizations. The reported results correspond to the proposed method, Wavelet(db4), ACWA filter and the EMD-ACWA. 78

III.13 Variations of PESQ values versus input SNR for the signals /o/, /a/, /e/ and /i/. The results are average over 100 noise realizations. The reported results correspond to the proposed method, wavelet(db4), ACWA filter and the EMD-ACWA. 79

III.14 Original signals "speech1", "speech2", "speech3" and "speech4". 80

III.15 Noisy version of signals "speech1", "speech2", "speech3" and "speech4" (input SNR = 2 dB). 80

III.16 Denoising of noisy signals "speech1", "speech2", "speech3" and "speech4" (input SNR=2 dB) by the proposed method, Wavelet (db4), ACWA filter and EMD-ACWA. 81

III.17 Final SNR values obtained from different initial noise levels of signals "speech1", "speech2", "speech3" and "speech4". The results averages over 100 Monte Carlo simulations of the additive noise. It is reported for the proposed method, wavelet(db4), ACWA filter and the EMD-ACWA. 82

III.18 PESQ values obtained from different initial noise levels of signals "speech1", "speech2","speech3" and "speech4". It is reported for proposed method, Wavelet(db4), ACWA filter and the EMD-ACWA. . . 83

IV.1 Original IMF and its estimated version by spline interpolation. 88

IV.2 IMF mean envelope offset 89
IV.3 IA, IP and IF of an IMF. 90
IV.4 Encoding scheme. 91
IV.5 Encoding scheme. 93
IV.6 Decomposition of an audio frame by EMD. 95
IV.7 Partial autocorrelation coefficient for IF of IMF generated by audio frame (figure IV.6). 96

V.1 $IMF_{extrema}$ encoder architecture in context of audio signals. 101
V.2 LEC variation for an audio frame. 102
V.3 Example of segmentation for an audio frame. 102
V.4 Quantization scheme. 104
V.5 Original audio signals (gspi, harp, quar, song, trpt and violin). 105

VI.1 Decomposition of an audio frame into IMFs. 113
VI.2 Data structure $\{m_i\}$. 113
VI.3 Watermark embedding. 113
VI.4 Decomposition of the watermarked audio frame by EMD. 114
VI.5 Watermark extraction. 115
VI.6 Illustration of the last IMF of an audio frame before and after watermarking. 115
VI.7 Binary watermark. 119
VI.8 A portion of the pop audio signal and its watermarked version. 119
VI.9 P_{FPE} versus synchronization code length. 123
VI.10P_{FNE} versus the length of embedding bits. 124

List of Tables

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

2 TEB et NC de la marque extractée pour le signal audio "Rock" par l'approche proposée. xviii

3 Performance des algorithmes de tatouage, trier par débits. xviii

I.1 Matrix of orthogonality of the signal (Eq. I.7) 32

II.1 Variations of the output SNR and of the PESQ over the input SNR for the MMSE and ACWA filters. 55

III.1 C and j_s values of each signal 76

III.2 Denoising results, based on the output SNR, of four noisy voiced different signals (input SNR=2 dB) 76

IV.1 Offset values of IMFs extracted from an audio frame. 88

IV.2 Order of AR model for IF of IMFs (figure IV.7). 96

V.1 Compression results of audio signals (gspi, harp, quar, song, trpt and violin) by $IMF_{extrema}$, $IA-IP$, AAC, MP3 and the wavelet. 106

V.2 Compression results of audio signals (gspi, harp, quar, song, trpt and violin) by $IMF_{envelope}$, $IA-IF$, AAC, MP3 and wavelet methods. . 107

VI.1 SNR and ODG between original and watermarked audio. 120

VI.2 BER and NC of extracted watermark for pop audio signal by proposed approach. 121

VI.3 BER and NC of extracted watermark for different audio signals (Classical, Jazz, Rock) by proposed approach. 122

VI.4 BER and NC of extracted watermark for different audio signals (Classical, Jazz, Rock) by proposed approach. 123

B.1 Impairment grade. 142

Abbreviations

AAC	Advanced Audio Coding
ACWA	Adaptive Center Weighted Average
AM	Amplitude Modulation
AR	Auto Regressive
BER	Bit Error Rate
BR	Bit Rate
CMSE	Consecutive Mean Square Error
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
EMD	Empirical Mode Decomposition
FFT	Fast Fourier Transform
FM	Frequency Modulation
FNE	False Negative Error
FPE	False Positive Error
FT	Fourier Transform
HHT	Hilbert-Huang Transform
IA	Instantaneous Amplitude
IF	Instantaneous Frequency
IFPI	International Federation of the Photographic Industry
IP	Instantaneous Phase
IMF	Intrinsic Mode Function
LEC	Local Entropic Criterion
MAE	Mean Absolute Error
MSE	Mean Square Error
NC	Normalized Cross-correlation
NMR	Noise to Mask Ratio
ODG	Objective Difference Grade
PESQ	Perceptual Evaluation of Speech Quality
QIM	Quantization Index Modulation
PDE	Partial Differential Equation
SC	Synchronized Code

Introduction

Signals can be derived from different sources, but most of them, arising from physical phenomena, are non-stationary. Locally, these signals can be regarded as stationary and thus decomposed as a superposition of sine waves, the frequency of which evolves over time. Among non-stationary signals, we can distinguish speech and audio signals. Conventional tools such as Fourier Transform (FT) and Discrete Cosine Transform (DCT) are unsuitable to analyze non stationary signals. In fact, when the signal spectrum is time varying, such as for music, speech and biomedical signals, time-frequency analysis approach is more relevant. The results of a time-frequency analysis depend on the choice of the time-frequency decomposition tool used, such as Short-Time Fourier Transform (STFT), Wigner distribution or Wavelet Transform (WT).

In several scenarios, it is preferable to take advantage of multi-resolution characteristics of WT. A limit of the wavelet approach is that first, the basis function must be specified and, second a specific basis function may not be able to catch all the non stationarity of the analyzed signal. To overcome this drawback time-frequency atomic signal decomposition can be used [31],[56]. As for wavelet packets, if the dictionary is very large and rich enough with a large collection of atomic waveforms which are located on a much finer grid in time-frequency space than wavelet and cosine packet tables, then it should be possible to faithfully represent a wide range of real signals. Furthermore, the ideal is to find an adaptive decomposition of the signal, so that it does not require a priori information about the signal time varying characteristics.

Recently, a new data-driven technique, referred to as Empirical Mode Decomposition (EMD) has been introduced by Huang et *al.* [34] for analyzing data resulting from non-stationary and nonlinear processes. EMD has received much attention in terms of applications [3],[7]-[9] interpretation [33]-[35], and improvement [16],[84]. Major advantage of EMD is that the basis function is derived from

the signal itself. Hence, the analysis is adaptive, in contrast to the traditional methods where the basis functions are fixed. The EMD is based on the sequential extraction of energy associated with various intrinsic time scales of the signal, called Intrinsic Mode Functions (IMFs), starting from finer temporal scales (high frequency IMFs) to coarser ones (low frequency IMFs). The superposition of the extracted IMFs matches the signal very well and therefore ensures completeness [34].

Characteristics of EMD and its effectiveness as a decomposing tool, have been addressed by different research. Indeed, an improvement in terms of signal decomposition has been shown in [68],[69]. The combination of EMD with Hilbert transform demonstrated the interest of EMD as a tool to investigate time-frequency domain representations [6],[11]. In [21] it has been shown that, provided some hypothesis, the extraction of IMF is reduced to the resolution of partial differential equation (Heat equation). Further, EMD has demonstrated its usefulness and effectiveness in many applications such as biomedical signals filtering and sonar target tracking [6],[11].

Main motivation of this thesis is to investigate the potential of EMD as an analyzing method for both speech and audio signals. More particularly, we address the problems of denoising, coding and watermarking. Also the goal of this work is to explore the limit of self-adaptive nature of the EMD process as signal analyzing tool in speech and audio processing.

Outline of the thesis

The dissertation is organized chapter by chapter as follows

chapter I is devoted to a presentation of the Huang transform, known as EMD. In particular, interest is focused on the relevant parameters which have influences on extracted IMFs, such as interpolation and sampling [34],[69]. The capability of EMD for separation of components is also studied and illustrated.

In the first part of the thesis, we are interested in techniques of noise reduction (filtering and denoising). Particularly in the case of additive white Gaussian noise,

different approaches have been proposed [72],[75]. When the noise distribution can be estimated accurately, then filtering yields acceptable results. However, these methods are not so effective when the noise level is difficult to estimate. Linear methods based on Wiener filtering [67] are sometimes preferred because linear filters are easy to implement and design. However, linear filtering methods are not so effective when signals contain sharp edges and impulses of short duration. Furthermore, real signals are often non-stationary. In order to overcome these shortcomings, nonlinear methods have been proposed and especially those based on wavelets thresholding [22],[23]. The idea of wavelet thresholding relies on the assumption that signal magnitudes dominate the magnitudes of the noise in a wavelet representation, so that wavelet coefficients can be set to zero if their magnitudes are less than a pre-determined threshold [22]. Using the same strategy as in wavelets thresholding approach, we propose in this thesis new techniques of speech denoising based on EMD. Our contribution related to these techniques are organized into two chapters.

In **Chapter II**, different denoising strategies based on EMD that address both additive white and colored noise are presented. In fact, it has been shown in [7]-[9], that EMD can be used for signal denoising. The proposed denoising method reconstructs the signal from all the IMFs previously filtered or thresholded as in wavelet analysis [7]-[9]. In this chapter, firstly two new denoising strategies for white noise context are presented. The first strategy combines EMD and Minimum Mean Squared Error (MMSE) filter [75], and the second one associates EMD with hard shrinkage [7]-[9]. The two methods, effective for a large class of signals, are applied to speech signals corrupted with different white noise levels.

The third denoising technique, called EMD-ACWA, consists in filtering IMFs by Adaptive Center Weighted Average (ACWA) filter [52], which exploits some local statistics of the signal. This technique is efficient both in the context of white noise and colored one. The use of ACWA filter is motivated by two important reasons. First, it operates in the time domain as the EMD. So, there is no need to use of FT as in the case of the MMSE filter [75]. Second, the ACWA filter operates regardless of the nature of the signal and noise. In particular, the assumptions of signal stationarity and white noise are not required.

Chapter III deals with a new noise reduction technique dedicated to speech signal. This technique, which combines EMD with ACWA filter, takes into account

the characteristics of speech signal. The proposed approach takes into account the class of the processed speech frame (voiced/unvoiced and transient). Indeed, in the IMF filtering step the number of denoised IMFs depends on whether the noisy frame is voiced or unvoiced. An energy criterion detects voiced frames while the stationarity index [51] is used to distinguish between unvoiced and transient frames .

The second part of the thesis is devoted to audio coding. The coding process is a central topic in the fields of audio and image processing [39],[82] and particularly in audio domain where different strategies have been proposed [40],[64]. When applications are not limited by low bit rate constraints, coding usually leads to acceptable results. However, in many applications such as digital audio broadcasting or multimedia, low bit rate and high fidelity are required. In order to reduce the bit rate, sub-band coding [10],[78] and transform coding approaches [20],[74] have been used to design efficient coding algorithms. These methods use basically a subband decomposition of the signal followed by perceptual encoding of significant coefficients at each subband which appeals to the following principle: *do not code what the ear can't listen*. Applying this principle enables good results at low bit rate. Unfortunately, using a decomposition strategy based on the representation on a fixed basis prevents the decomposition from being parsimonious for any kind of audio signal. Indeed, even if a decomposition tool is well suited for a large class of audio signals, in the sense that it yields compact descriptions with only a few significant terms, there are audio signals for which the basis under consideration performs poorly [20]. The EMD can be seen as a type of subband decomposition whose subbands are able to automatically separate the different components of a signal. Each IMF replaces the signal details, at a certain scale or frequency band. Thanks to IMF properties, the EMD seems to be a very interesting decomposition tool to use for a low bit rate audio coding. The presentation of our contribution to audio coding is organized in two chapters.

In **chapter IV**, a new signal coding based on EMD is introduced. The first step consists in encoding the IMFs extrema, since the IMFs are fully described by their local extrema [34]. To further reduce the bit rate, only one of the IMFs envelops is encoded. This is motivated by the quasi-symmetrical property of the IMF. In a second step, a waveform coding approach based on EMD in association with Hilbert transform is presented. Based on the Hilbert and Huang transforms,

we can calculate the Instantaneous Amplitude (IA), Instantaneous Phase (IP) and the Instantaneous Frequency (IF) for IMFs. The idea is then to encode the IA and IF by linear prediction, while the IP is encoded by a scalar quantization.

Chapter V is devoted to apply the proposed coding approaches, described in the previous chapter to audio signals. We show that it is interesting to introduce a psychoacoustic model, in extrema thresholding and bit allocation; and detector for transient sequence in these approaches. The performance of the proposed methods are analyzed and compared to the MPEG1 layer3, known as MP3, to AAC codecs and to the wavelet based compression.

Watermarking is as a solution to control unapproved copying and redistribution of multimedia data, where many bit streams can be transmitted by taking the audio signal as a transmission medium. Various constraints must be considered in the watermarking process such as inaudibility of the watermarked signal, higher transmission bit rate and robustness against distortions. The detection of the inserted message is the subject of several research [2],[79] where several watermarking techniques have been proposed [13],[41]. The watermarking approach of Malvar [50] is among of the recent algorithms in the context of audio signals. This approach has shown good robustness to a wide variety of attacks but it imposes a very limited transmission bit rate. So, to increase the bit rate, many watermarking algorithms based on the wavelet has been presented [41],[86]. A limit of the wavelet approach is that the basis functions are fixed, and thus may not be effective for all real signals. The IMFs are fully described by their local extrema [34], thus, they can be constructed from only their extrema [47]. The superposition of extracted IMFs matches the signal very well and therefore ensures completeness [34]. Based on these interesting proprieties, we considered a watermarking scheme based on EMD. The proposed watermarking approach is the subject of chapter VI.

Chapter VI introduces a new audio watermarking approach, based on EMD, dedicated to control unapproved copying. The watermark and the synchronization codes are embedded into the extrema of the last IMF, a low frequency mode stable under different attacks and preserving an audio perceptual quality of the host signal. Relying on exhaustive simulations, we show the robustness of the hidden watermark data to additive noise, low-pass filtering, MP3 compression, re-quantization and denoising. The reported results are compared to watermarking

schemes reported recently.

Finally, the **conclusion** will review all the work done and presents several suggestions and extensions to improve and optimize the contributions of this work.

Main contributions of the thesis

In the following we list the main contributions of the dissertation

- Introduction of a new noise reduction scheme operating in adaptive way. Different strategies of filtering and denoising of audio signals are developed. Improvement in terms of output SNR (Signal to Noise Ratio) and PESQ (Perceptual Evaluation Speech Quality) are obtained compared to MMSE filter and wavelet approach.

- Introduction of a new signal coding framework based on the extrema of IMFs. Different coding strategies are presented. No assumptions concerning the linearity or the stationary are made about the signal to be coded. The new scheme can be extended to encode any signal and from any source. Improvement in terms of BR (Bit Rate), ODG (Objective Difference Grade) and NMR (Noise to Mask Ratio) are obtained compared to MP3 and AAC codecs, and wavelet based compression.

- Introduction of a new adaptive watermarking scheme based on the EMD. Watermark is embedded in very low frequency mode (last IMF), thus achieving good performance against various attacks. Data bits of the synchronized watermark are embedded in the extrema of the last IMF of the audio signal based on quantization index modulation. Extensive simulations over different audio signals indicate that the proposed watermarking scheme has greater robustness against common attacks than nine recently proposed algorithms. The new scheme has higher payload and better performance against MP3 compression compared to these earlier audio watermarking methods.

List of publications

Denoising

- **K. Khaldi**, A.O. Boudraa, A. Bouchikhi, and M. Turki, "Speech Enhancement via EMD", EURASIP Journal Advances in Signal Processing, vol. 2008, Article ID 873204, 8 pages, 2008.

- **K. Khaldi**, M. Turki and A.O. Boudraa, "Voiced speech enhancement based on adaptive filtering of selected Intrinsic Mode Functions", Advances in Adaptive Data Analysis (AADA), vol. 2, n°. 1, pp. 65-80, 2010.

- **K. Khaldi**, A.O. Boudraa, A. Bouchikhi, M. Turki and E. Diop, "Speech signal noise reduction by EMD", IEEE International Symposium on Communications, Control and Signal Processing (ISCCSP), march 2008, Malta.

- **K. Khaldi**, A. Adib et M. Turki, "Amélioration des techniques de séparation de sources par débruitage via EMD", Colloque Africain sur la Recherche en Informatique et en Mathématiques Appliquées (CARI), octobre 2008, Rabat, Maroc.

- **K. Khaldi**, M. Turki and A.O. Boudraa "A new EMD denoising approach dedicated to voiced speech signals", IEEE Signals, Circuits and Systems (SCS), november 2008, Hammamet, Tunisia.

- **K. Khaldi**, M. Turki and A.O. Boudraa "Speech enhancement by adaptive weighted average filtering in the EMD framework", IEEE SCS, november 2008, Hammamet, Tunisia.

- **K. Khaldi**, M. Turki and A.O. Boudraa, "Speech denoising using modal decomposition and local statistics", Digital Signal Processing (submitted).

Coding

- **K. Khaldi**, A.O. Boudraa, M. Turki, Th. Chonavel and I. Samaali, "Audio encoding based on the Empirical Mode Decomposition", IEEE European Signal Processing Conference (EUSIPCO), august 2009, Glasgow, Scotland.

- **K. Khaldi**, A.O. Boudraa, M. Turki et Th. Chonavel, "Codage audio perceptuel à bas débit par Décomposition en Modes Empiriques (EMD)", Colloque Groupe de Recherche et d'Etudes du Traitement du Signal et des Images (GRETSI), septembre 2009, Dijon, France.

- **K. Khaldi**, A.O. Boudraa, B. Torrésani, Th. Chonavel and M. Turki, "Audio encoding using Huang and Hilbert transforms", IEEE ISCCSP,march 2010, Limassol, Cyprus.

- **K. Khaldi**, A.O. Boudraa, Th. Chonavel, and M. Turki "Empirical Mode Compression (EMC) of audio signals", Signal Processing Journal (First Revision).

- **K. Khaldi**, A.O. Boudraa, B. Torrésani, M. Turki and Th. Chonavel, "HHT-Based Audio Coding", International Journal of Wavelets, Multiresolution and Information Processing (submitted).

Watermarking

- **K. Khaldi** and A.O. Boudraa, "Audio watermarking via EMD", IEEE Transactions on Audio, Speech and Language Processing (submitted).

CHAPTER I

Huang transform

Contents

I.1 Introduction 25

I.1.1 Principle of EMD 25

I.1.2 EMD algorithm 26

I.1.3 Meaningful parameters of EMD 27

I.2 IMFs properties 31

I.2.1 IMFs orthogonality 31

I.2.2 PDE for IMFs characterization 33

I.3 EMD: a time-frequency description tool 33

I.3.1 Importance of the sampling frequency 33

I.3.2 Tones separation 35

I.3.3 EMD acts as a Filter bank: Gaussian white noise case . . 37

I.3.4 Comparison with wavelets 38

I.4 Conclusion 40

This chapter presents the Huang transform, known as Empirical Mode Decomposition (EMD), introduced by Huang et al. [34]. The EMD is a data driven method, defined by an algorithm, that enables the adaptive decomposition of a signal into finite sum of components, called Intrinsic Mode Functions (IMFs). The principle of EMD is presented and is illustrated on synthetic signals. Some parameters, such as interpolation and sampling frequency, which influence the results of the decomposition are point out. Finally, some aspects of the EMD considered as a time-frequency description tool are presented and discussed.

I.1 Introduction

In this chapter, we introduce the Huang transform known as Empirical Mode Decomposition (EMD). The EMD is introduced by Huang *et al*, to overcome the limitations of Fourier based methods when applied to non-stationary signals. The EMD decomposes adaptively a signal into a sum of oscillating components. Unlike the Fourier Transform (FT) or Wavelet Transform (FT), the EMD is a data driven decomposition technique. It has been introduced for analyzing data deriving from non-stationary and nonlinear processes. The major advantage of the EMD is that the basis functions are derived from the signal itself. Hence, the analysis is adaptive in contrast to the traditional methods where the basis functions are fixed. The EMD is based on the sequential extraction of energy associated with various intrinsic time scales of the signal or oscillating components, called Intrinsic Mode Functions (IMFs), starting from finer temporal scales (high frequency IMFs) to coarser ones (low frequency IMFs). The total sum of the IMFs matches the signal very well and therefore ensures completeness [34].

I.1.1 Principle of EMD

The EMD is an algorithmic signal decomposition method. It is based on the principle of decomposing a signal into the sum of a high frequency component (fast oscillation) and a low frequency component (trend). This principle is illustrated by equation (I.1),

$$x(t) = d(t) + m(t), \tag{I.1}$$

where t denotes the discrete time, $x(t)$ is the signal to decompose, $d(t)$ is the fast oscillation and $m(t)$ is the signal trend. Similarly, the signal trend can also be decomposed into two terms,

$$m(t) = d_1(t) + m_1(t), \tag{I.2}$$

where $d_1(t)$ is the high frequency component of $m(t)$, and $m_1(t)$ is its low frequency component.
To extract the mode of a signal $x(t)$, the following principle is considered:

- identify all extrema of $x(t)$.
- interpolate between minima (resp. maxima), ending up with some envelope

$e_{min}(t)$ (resp $e_{max}(t)$).

- compute the average $m(t) = (e_{min}(t) + e_{max}(t))/2$.
- extract the detail $d(t) = x(t) - m(t)$.

The signal $d(t)$ is considered as IMF after a number of iterations needed to satisfy a given stop criterion. By iterating this principle to the obtained trends, we get a signal decomposition described as follows:

$$x(t) = \sum_{j=1}^{C} IMF_j(t) + r_C(t), C \in N^* \tag{I.3}$$

where IMF_j is the j^{th} order IMF. IMF_j contains higher frequency oscillations than the IMF_{j+1}. The signal $r_c(t)$ is called the residual, it is the lower frequency component of signal $x(t)$. According to Eq. I.3 and assuming that C is finite, we can construct linearly the original signal without loss of any information [34].
By definition, a component is considered as a true IMF if it satisfies the following criteria [34]:

1. the number of its extrema and the number of its zero crossings may differ by no more than one.
2. the average value of the envelope defined by the local maxima and the envelope defined by the local minima, is zero.

I.1.2 EMD algorithm

The principle of IMFs extraction is ensured by the *sifting* process, which is implemented by the following generic algorithm.

Notations:
ϵ: predetermined threshold, that is used to specify the loop exit condition.
j: IMF index.
i: index of current iteration in the loop for extracting an IMF.
T: length of the decomposed signal: x= $x(t)_{t=1...T}$.
r_j: residual after obtaining the j^{th} IMF

$h_{j,i}$: intermediate variable, equal to the value of the new residual at the first iteration. It is equal to the difference between the residual and the value of the average envelope in the following iterations.
$U_{j,i}$: upper envelope of $h_{j,i}$ constructed by maxima interpolation.
$L_{j,i}$: lower envelope of $h_{j,i}$ built by minima interpolation.
$\mu_{j,i}$: average envelope, obtained from both envelopes of $h_{j,i}$.
$SD(i)$: stopping criterion at i^{th} iteration.

The sifting can be summarized as follows :
Step 1: Fix the threshold ϵ and set $j \leftarrow 1$ (j^{th}IMF)
Step 2: $\mathrm{r}_{j-1}(\mathrm{t}) \leftarrow x(t)$ (residual)
Step 3: Extract the j^{th}IMF :
(a) : $\mathrm{h}_{j,i-1}(\mathrm{t}) \leftarrow \mathrm{r}_{j-1}(\mathrm{t})$,$i \leftarrow 1$ (i number of sifts)
(b) : Extract local maxima/minima of $\mathrm{h}_{j,i-1}(\mathrm{t})$
(c) : Compute upper and lower envelopes $\mathrm{U}_{j,i-1}(\mathrm{t})$ and $\mathrm{L}_{j,i-1}(\mathrm{t})$ by spline, interpolation of local maxima and minima of $\mathrm{h}_{j,i-1}(\mathrm{t})$ respectively
(d) : Compute the mean of the envelopes : $\mu_{j,i-1}(t)$ =($\mathrm{U}_{j,i-1}(\mathrm{t})$ + $\mathrm{L}_{j,i-1}(\mathrm{t})$)/2
(e) : Update : $\mathrm{h}_{j,i}(t) := h_{j,i-1}(t) - \mu_{j,i-1}(t)$, $i := i + 1$
(f) : Calculate the stopping criterion : $\mathrm{SD(i)} = \sum_{t=1}^{T} \frac{|hj,i-1(t)-hj,i(t)|^2}{(hj,i-1(t))^2}$
(g) : Repeat Steps (b)-(f) until SD(i)$< \epsilon$ and then put $\mathrm{IMF}_j(\mathrm{t}) \leftarrow \mathrm{h}_{j,i}(\mathrm{t})$ (j^{th}IMF)
Step 4: Update residual : $r_j(t) := r_{j-1}(t) - \mathrm{IMF}_j(t)$.
Step 5: Repeat Step 3 with $j := j + 1$ until the number of extrema in $r_j(t)$ is ≤ 2.

The sifting is repeated several times (i) in order to guarantee that the computed IMF $h_{j,i}$ fulfills the required conditions (1) and (2).
The sifting has two effects: (a) it eliminates riding waves and (b) it smoothes uneven amplitudes.

I.1.3 Meaningful parameters of EMD

Decomposition result of EMD depends on the choice of two important parameters: the stopping criterion and the interpolation technique used.

I.1.3.1 Stopping criterion

Since there are two loops in the EMD algorithm, we must ensure that both must stop. The main loop indexed by j stops when it is impossible to decompose the current residual, i.e., that $r_j(t)$ has less than two extrema. The second loop indexed by i is linked to a stopping criterion that should be defined precisely.
In fact, the second loop will stop when $h_{j,i}(t)$ satisfies the criteria defining an IMF. Theoretically, this assumption is not proven. So in practice a stopping criterion of the sifting process is imposed, in order to prevent the sifting processes from coming into an infinite loop.
Thus in [34], the authors propose a stopping criterion based on the standard deviation $SD(i)$ defined by,

$$SD(i) = \sum_{t=0}^{T} \frac{|h_{j,i-1}(t) - h_{j,i}(t)|^2}{(h_{j,i-1}(t))^2} \tag{I.4}$$

The stopping test is validated when $SD(i)$ is below a predefined threshold ϵ. Typically, the value ϵ to stop the sifting is between 0.2 and 0.3 [34],[35]. In fact, if ϵ is too high, the EMD does not separate the different modes present in the signal, however if ϵ is too small, the EMD extracts components whose amplitudes are almost constant and modulated by a single frequency (over decomposition of the signal).
Another stopping criterion was proposed by Rilling et *al* [69], and is particularly chosen in practice. This criterion is defined from the function:

$$\sigma(t) = 2|\frac{\mu_{i-1}(t)}{U_{i-1}(t) - L_{i-1}(t)}|, \tag{I.5}$$

By adopting the criterion (Eq. I.5), three conditions must be satisfied so that $h_{i,j}(t)$ is an IMF [69].

- The number of extrema of $h_i(t)$ and the number of the zeros crossings of $h_i(t)$ may differ by no more than one.
- $\sigma(t) < \theta_1$ for t $\leq$ (1 - α)T.
- $\sigma(t) < \theta_2$ for (1 - α)T < t <T.

θ_1 and θ_2 are real numbers such that $0 \leq \theta_1 \leq \theta_2$ and α is chosen in [0,1].

The first condition means that an IMF must be an oscillating function around zero. The last two conditions require that the function $\sigma(t)$ has small values. Therefore, it can take high values to some extent. In [69], a good compromise of empirical choice of the threshold values θ_1 and θ_2 is given as follows:
$\theta_1 = 0.05$ and $\theta_2 \approx 10 * \theta_1$ and $\alpha = 0.05$.
In our work, we adopt the criteria chosen by [69]. Indeed, it gives modes that correspond well to the definition of an IMF.
Nevertheless, to improve the decomposition result, we need to find an appropriate interpolation method, which allows to estimate both envelopes (maxima, minima) with the lowest error.

I.1.3.2 Interpolation

Interpolation is an important step in the estimation and extraction of IMFs. Indeed, the envelopes are estimated by interpolation from the extrema, so the interpolation determines the shape of the IMF. There are various interpolation methods, but not all are effective for a good EMD representation. According to [34], the interpolation methods known as "nearest" and "linear" are not recommended for the estimation of IMFs, because both methods result in an excessive number of modes. However, the spline interpolation method provides better results than those obtained by other approaches. To illustrate the efficiency of spline interpolation compared to other methods, we consider the signal described by:

$$x(t) = \cos(t) + \sqrt{(t)} \tag{I.6}$$

where $t \geq 0$ represents the discrete time.
Figure I.1 shows the curves presenting the interpolated version of the signal by four interpolation methods (cubic, spline, nearest and linear). The corresponding errors are also presented. The error corresponds to the difference between the original signal $x(t)$ and its interpolated version. This figure shows a much higher quality obtained with the spline method. Indeed, with a spline interpolation the error is ten times lower than other methods (nearest, linear and cubic).

The spline interpolation method coupled with a good choice of stopping criterion ensures a good result of the signal decomposition. This is illustrated by the following example.

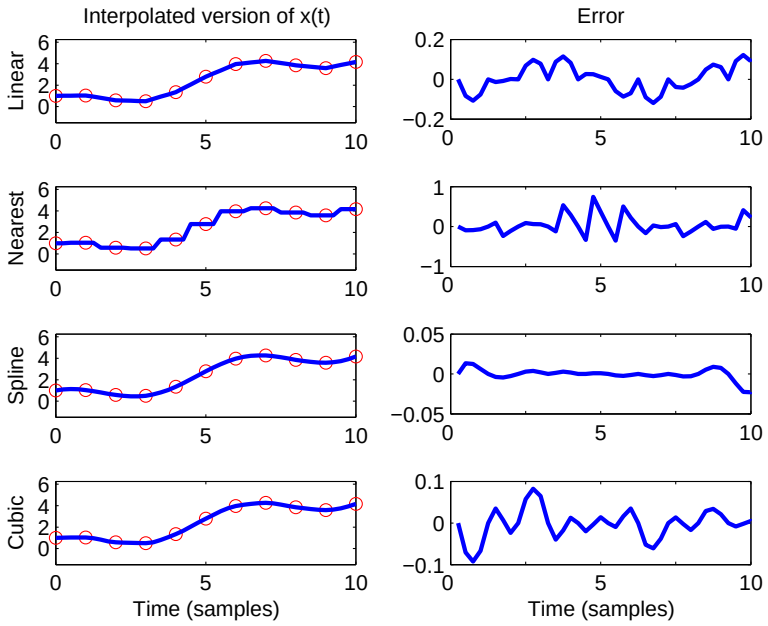


Figure I.1: Interpolation of the signal ($x(t) = \cos(t) + \sqrt{(t)}$) by different methods and the corresponding error.

Example:

Consider the signal described by,

$$x(t) = \sin(8t) + \sin(3t) + 2t \tag{I.7}$$

The threshold values are identical to those chosen in [69], i.e $\theta_1 = 0.05$, $\theta_2 \approx 10 * \theta_1$, $\alpha = 0.05$. The spline interpolation method is used.

The decomposition of this signal by EMD is shown in figure I.2.

The stopping criterion chosen in [69] and the spline interpolation method offer in this case a very accurate decomposition, since we extracted two sinusoidal signals corresponding to $\sin(8t)$ and $\sin(3t)$, and trend signal corresponding to $(3t)$.

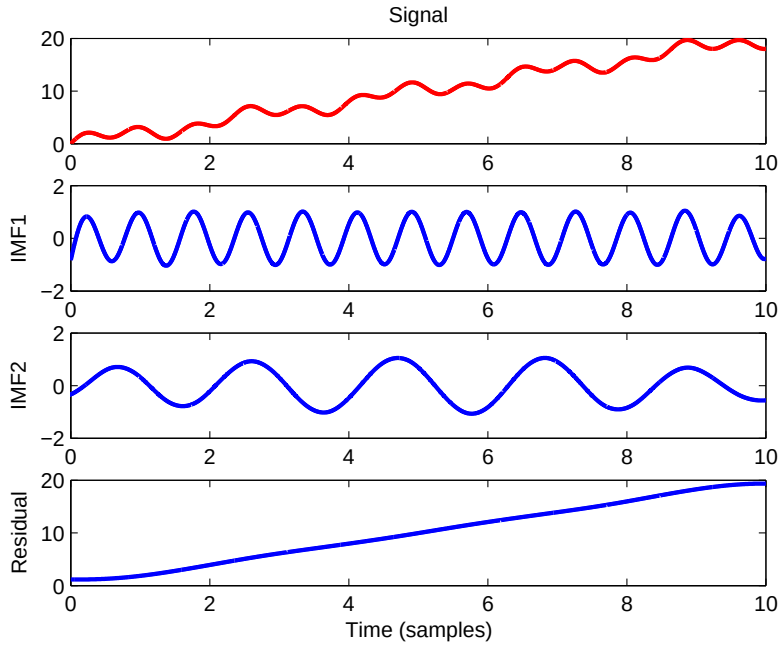


Figure I.2: Decomposition the signal $x(t) = \sin(8t) + \sin(3t) + 2t$ by EMD.

I.2 IMFs properties

I.2.1 IMFs orthogonality

The EMD decomposes a signal into a finite sum of components. According to [34], the IMFs of a signal are orthogonal. For a signal $x(t)$, one can write that

$$\langle IMF_i | IMF_j \rangle = 0 \ \forall \ i \neq j \tag{I.8}$$

where $\langle \ | \ \rangle$ denotes the scalar product in L^2, i and $j \in \{1, ..., C\}$ and C is the number of IMFs obtained .

Theoretically this orthogonality cannot be proved. In practice, the equality (Eq. I.8) is not strictly verified because the average envelope is derived from two envelopes which are estimated by interpolation [34]. Note that in general the residual is not orthogonal to IMFs. It is a non-oscillating function (it is a trend increasing or decreasing or null). As a measure of orthogonality between different IMFs, we propose the use of the orthogonality index.

The orthogonality index OI_{IMF_i, IMF_j} is defined as the normalized version of product $\langle \ IMF_i \ | IMF_j \rangle$ [11],

$$OI_{IMF_i,IMF_j} = \frac{\langle IMF_i \mid IMF_j \rangle}{\parallel IMF_i \parallel . \parallel IMFj \parallel} \forall (i,j) \tag{I.9}$$

where $\parallel IMF \parallel$ corresponds to the standard Euclidean norm. If $(i = j)$ $OI_{IMF_i,IMFj}$ =1.

$$OI_{IMF_i,IMF_j} = \frac{\sum\limits_t IMF_i(t).IMF_j(t)}{\sqrt{\sum\limits_t IMF_i^2(t)}.\sqrt{\sum\limits_t IMF_j^2(t)}} \forall (i,j), \tag{I.10}$$

OI can also be interpretable as a correlation coefficient between IMF_i and IMF_j as the cosine of the angle between these two signals. According to Eq. I.9, it is also possible to define a matrix of orthogonality, that embodies all calculated indices [11]. We denote this matrix by OI_{EMD}; its entry (i,j) is defined as OI_{IMF_i,IMF_j}

In practice the matrix OI_{EMD} is symmetric and its main diagonal is unitary. Ideally, in the case of exact estimation of IMFs, the matrix OI_{EMD} is equal to the identity matrix. An overall orthogonality index oi can be defined from the matrix IO_{EMD} as follows [34]:

$$oi = \sum_{1 \leq i < j \leq C} (OI_{IMF_i,IMFj})^2, \tag{I.11}$$

If we consider the signal described by Eq. I.7, then the table I.1 shows the matrix of orthogonality computed considering the components IMF_1 and IMF_2 obtained by EMD. In this example, the overall index of orthogonality oi = 3.14 10^{-5} Despite the good decomposition seen in the figure I.2, both components of

OI_{EMD}	IMF_1	IMF_2
IMF_1	1.0000	0.0056
IMF_2	0.0056	1.0000

Table I.1: Matrix of orthogonality of the signal (Eq. I.7)

the matrix OI_{EMD} (Table I.1) are not strictly orthogonal. Indeed, the overall orthogonality index (3.14 10^{-5}) is very low but not 0. Obtained orthogonality errors are due to the IMFs estimation error. This is attributed to the envelope calculation by the interpolation method. In fact, given that the estimated IMFs are obtained from one another by subtraction, then there is an error propagation [34].

I.2.2 PDE for IMFs characterization

Since the EMD is defined by a sifting process, recently many studies have been focused on the comprehension of the EMD [18],[60],[73],[81]. These different studies have tried to find a mathematical framework for the IMF description. In [18], the sifting process is modeled by a fourth order Partial Differential Equation (PDE), such approach was validated by numerical simulations.
In [21], a mathematical characterization of IMFs is obtained. The IMFs are the solutions of the PDE as follows:

$$\begin{cases} \frac{\partial h}{\partial s} + \frac{1}{\delta^2} h + \frac{1}{2}\frac{\partial^2 h}{\partial t^2} = & 0 \\ h(t,0) = & x(t) \end{cases} \tag{I.12}$$

where s is a PDE variable, t is a time variable, δ is the adjusted parameter and x denote the signal. This mathematical model depends on the sifting process, in contrast to the works [73],[81], where the mathematical relationship of IMFs is independent of the sifting process. This model (Eq. I.12) holds only for harmonic signals and require a good choice of the adjusting parameter δ.

I.3 EMD: a time-frequency description tool

I.3.1 Importance of the sampling frequency

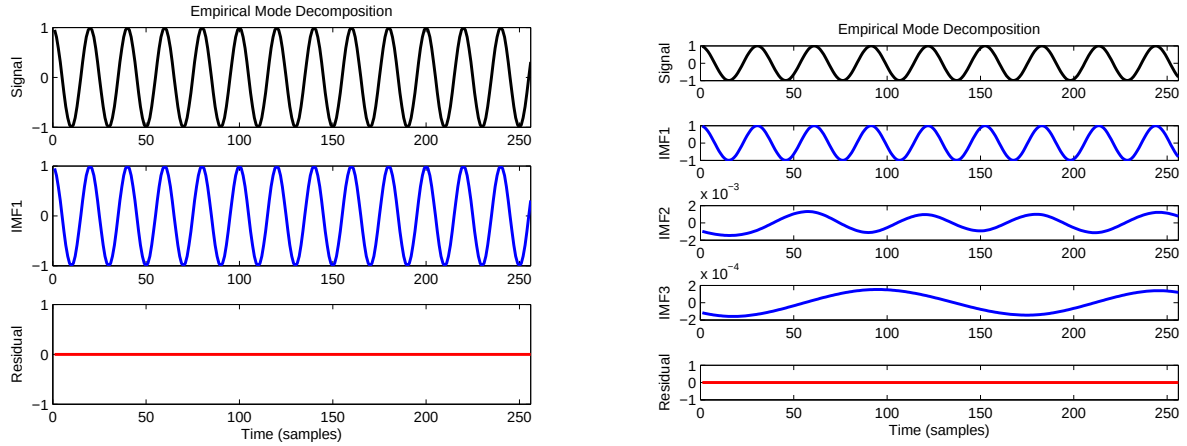
Sampling frequency has a big influence on the results of the decomposition. It can influence the number of IMFs obtained. We propose to display the effect of the sampling frequency for a pure frequency signal (or tone). This study is based on the work of Rilling et al. [69],[68] and that of Stevenson et al. [77].

Let us consider the following signal:

$$x(t) = \cos(2\pi\nu t) \tag{I.13}$$

where t is the discrete time $\in \{1.....N\}$, $\nu = \frac{f}{fe}$ is the normalized frequency and fe is the sampling frequency.
The study consists in varying the normalized frequency of the signal, and the results of the EMD are compared to the theoretical sinusoidal component of frequency ν, i.e the original signal.

Figure I.3 illustrates perfectly the problem of sampling frequency. Thus, although the number of samples is the same in both cases (N = 256 samples), the decomposition of $x(t)$ is different depending on the chosen normalized frequency. Only the signal with normalized frequency $\nu = 0,050$ is decomposed correctly by EMD (the residual is null). However, for signal of normalized frequency $\nu = 0,032$, the decomposition by EMD gives three IMFs and a residual not null. To illustrate



(a) ν=0.05 . (b) ν=0.032.

Figure I.3: Decomposition of the tone signal (Eq. I.13) by EMD.

the phenomenon, the relative error $E(\nu)$ associated to the first IMF is defined as follows [69]:

$$E(\nu) = \sqrt{\frac{\sum_{t=1}^{N}[x_\nu(t) - IMF_1(t)]^2}{\sum_{t=1}^{N} x_\nu^2(t)}} \tag{I.14}$$

where ν is the normalized frequency, and $IMF_1(t)$ is the first IMF of signal $x(t)$. Figure I.4 shows the variation of $E(\nu)$ as a function of ν in a log-log (base 2) plane. We see that the overall error is raised by a quadratic function of ν : $E(\nu) \leq \lambda\nu^2$. More precisely the error $E(\nu)$ is modeled as follows [11],[68],[69] (Fig I.4):

- the error is raised by : $E(\nu) \leq \frac{1-\cos(\pi\nu)}{\sqrt{2}} \leq \frac{\pi^2\nu^2}{2\sqrt{2}}$,
- the errors are increased for frequencies $\nu = \frac{1}{2k+1}$ and $\nu = \frac{2}{2k+1}$, where $k \in N^*$,
- the error is zero for frequencies $\nu = \frac{1}{2k}$, where $k \in N^* \Rightarrow f_e = 2kf$, where f and f_e denotes the frequency and the sampling frequency respectively.

The EMD of the tone depends strongly on the normalized frequency ν, consequently the sampling frequency for a fixed frequency f [11],[68],[69].

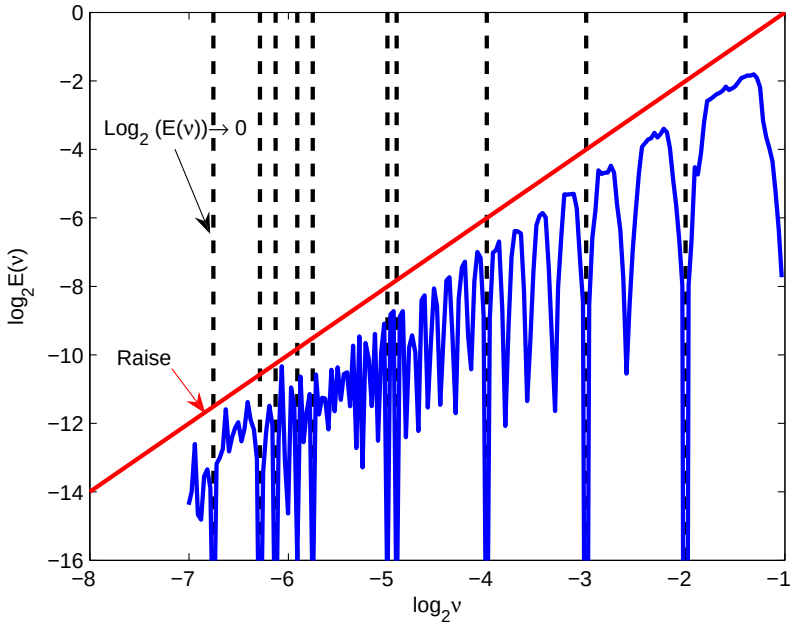


Figure I.4: Estimation and behavior of $E(\nu)$ associated with the first IMF for a tone.

I.3.2 Tones separation

In this section we study the ability of EMD to separate two sinusoidal components, according to the ratio of their frequency. This study is inspired by the work of Rilling *et al.* [68],[69].

We consider a signal composed of two tones. It is defined as follows:

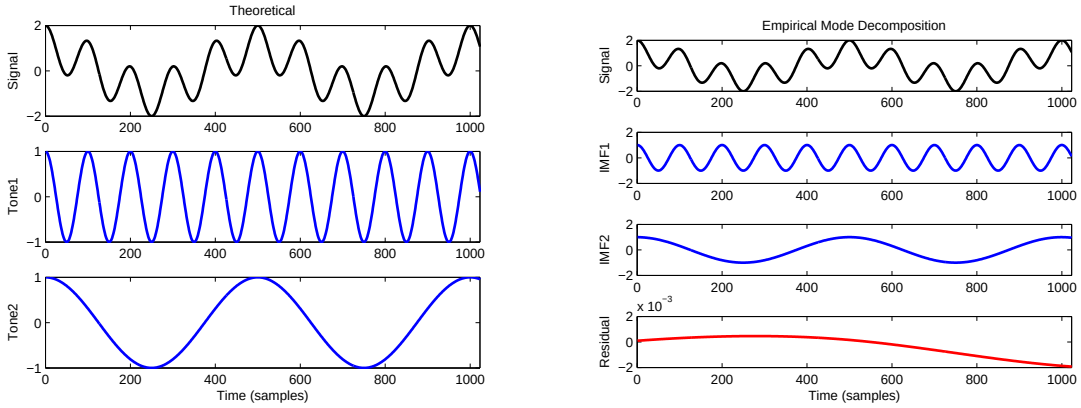
$$x(t) = x_{\nu_1}(t) + x_{\nu_2}(t) = cos(2\pi\nu_1 t) + cos(2\pi\nu_2 t) \tag{I.15}$$

where $t \in \{1, ..., N\}$, and (ν_1, ν_2) the pair of distinct normalized frequencies, such that $\nu_1 > \nu_2$. For simplicity, we assume that the amplitudes both signals $x_{\nu_1}(t)$ and $x_{\nu_2}(t)$ are equal.

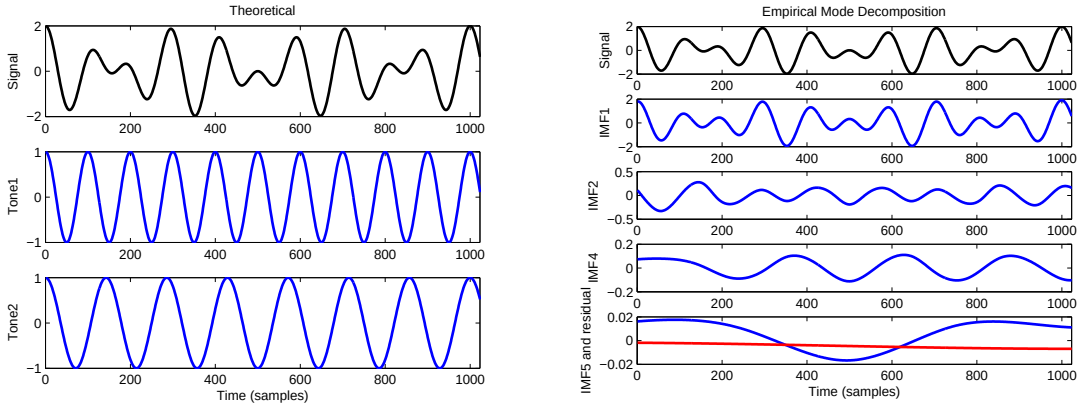
We expect the EMD to produce two IMFs at least: one associated to the highest frequency and the other to the lowest one.

Figure I.5 shows the decomposition of signal $x(t)$ by EMD. We remark that the decomposition depends on the ratio of the two frequencies ν_1 and ν_2. For two frequencies sufficiently distinct from each other, the decomposition of a signal $x(t)$ gives two IMFs (Fig. I.5(a)). Otherwise, the EMD approach considers the signal $x(t)$ as a single component amplitude modulated (Fig. I.5(b)) [68],[69].

According to [69], the estimation error of the first and second IMF is given by:



(a) (ν_1, ν_2)=(0.01,0.002).



(b) (ν_1, ν_2)=(0.01,0.007).

Figure I.5: Decomposition of the signal (Eq.I.15) by EMD.

$$E(\nu_1, \nu_2) = \sqrt{\frac{\sum_{t=1}^{N} x_{\nu_1}^2(t) \sum_{t=1}^{N} (x_{\nu_1}(t) - IMF_1(t))^2 + \sum_{t=1}^{N} x_{\nu_2}^2(t) \sum_{t=1}^{N} (x_{\nu_2}(t) - IMF_2(t))^2}{(\sum_{t=1}^{N} (x_{\nu_1}^2(t) + x_{\nu_2}^2(t))) \sum_{t=1}^{N} x^2(t)}} \tag{I.16}$$

where (ν_1, ν_2) both frequencies are varied in the interval]0,0.5[and satisfying $\nu_1 > \nu_2$.

Figure I.6 shows the variation of $E(\nu_1, \nu_2)$ versus frequencies ν_1 and ν_2. Fig.I.6 shows that the separation is very good over a certain range of values (ν_1,ν_2) where $|\nu_1$-$\nu_2|$ is large enough. We note that in case of very low error, the decomposition of $x(t)$ by EMD gives two IMFs, each IMF corresponds to a tone. However, when the error $E(\nu_1, \nu_2)$ is high, the signal $x(t)$ is analyzed as modulated in frequency and amplitude.

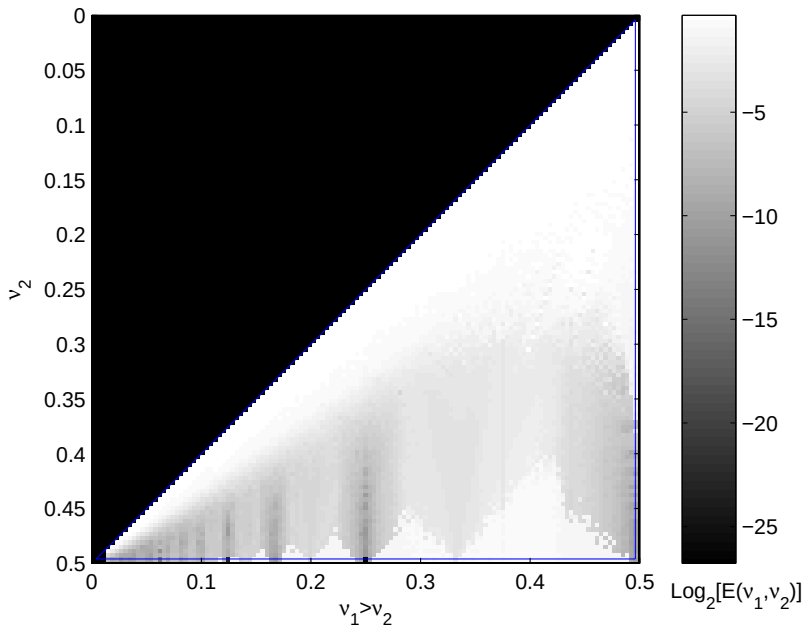


Figure I.6: Estimation and behavior of the error $E(\nu_1, \nu_2)$ Eq. I.16 for signal $x(t) = x_{\nu_1}(t) + x_{\nu_2}(t) = cos(2\pi\nu_1 t) + cos(2\pi\nu_2 t)$.

I.3.3 EMD acts as a Filter bank: Gaussian white noise case

The previous section showed that the EMD behaves as a self-adaptive filter bank. Thus, EMD decomposes locally a signal into a sum of IMFs from the highest frequencies to the lowest. In some well-controlled cases (Gaussian white noise for example), this decomposition is organized in a structure of a filter bank [11],[29],[87].
This filter bank structure is illustrated in figure I.7. The considered signal is a Gaussian white noise with zero mean and a variance equal to 1. The reported results correspond to averages over three thousand realizations [11]. We have plotted in the log-log plane the standardized spectral of the seven IMFs obtained for all realizations.

We note that the EMD behaves as well as a diadic filter bank for modes higher than 2. The first mode corresponds to the high-pass filter of the filter bank.
Based on exhaustive simulations, similar behavior of diadic filter bank is also proven for fractional Gaussian noise [26],[28],[68]. It is also proven that for fractional Gaussian noise, the spectral power is distributed over all IMFs under a law of exponential type [29],[28],[68]. All of these studies characterized the behavior of the EMD towards different noise types.

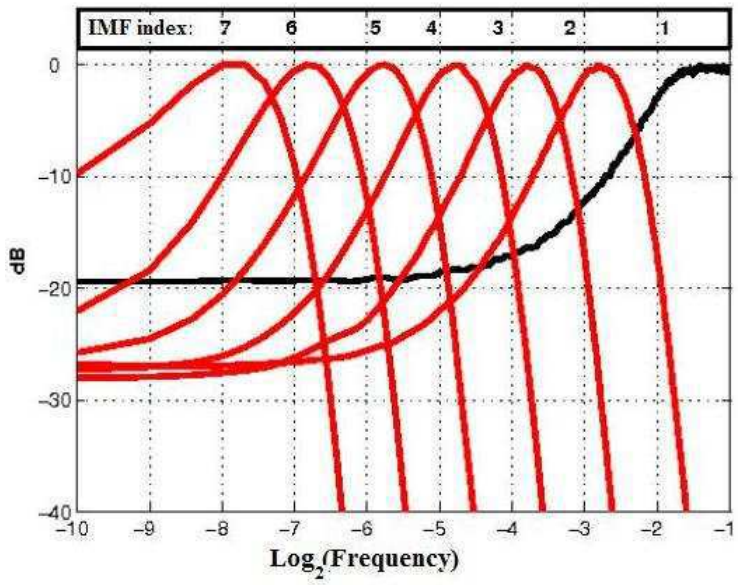


Figure I.7: IMFs spectra for a white noise.

I.3.4 Comparison with wavelets

The EMD is similar to a multi-resolution analysis, so it explores the signal from the highest to the lowest frequencies, i.e from the smallest details to the largest. However, the EMD is different from the wavelet decomposition in the way it describes the signal. Indeed, in a multi-resolution analysis by wavelet, the decomposition goes from low frequencies to higher frequencies. The EMD is an auto-adaptive method, contrary to the wavelet, where a mother function is needed to decompose a given data.

To further illustrate this difference, we compare the results of decomposition by the EMD to the orthogonal Daubechies wavelet (db3) over four levels (Fig. I.9). The analyzed signal is given by:

$$x(t) = \sin(3t) + \sin(0.3t) + \sin(0.03t) \tag{I.17}$$

This signal is constituted of three different tones (Fig. I.8). The EMD manages to separate correctly the three components that constitute the signal. A low residual signal is still obtained.

The wavelet decomposition is shown in figure I.9(b), where An and Dn represent

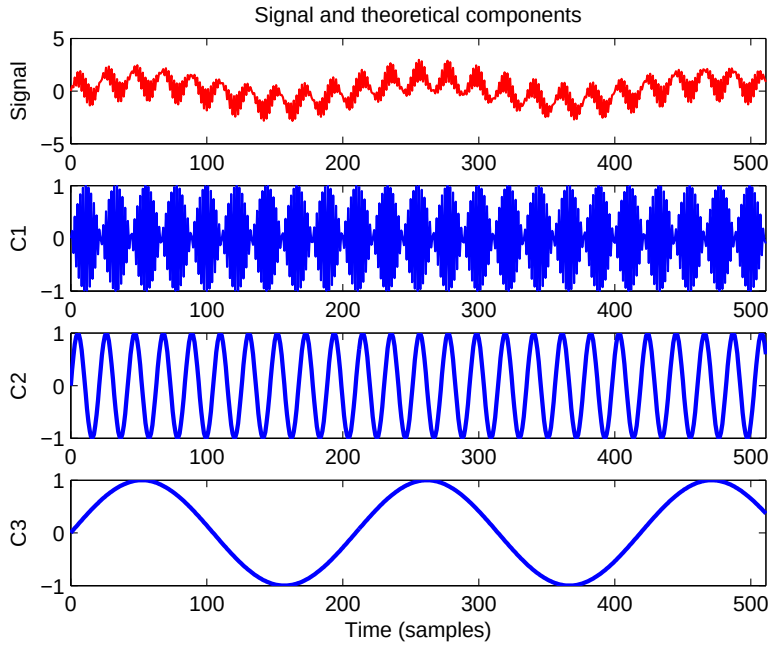
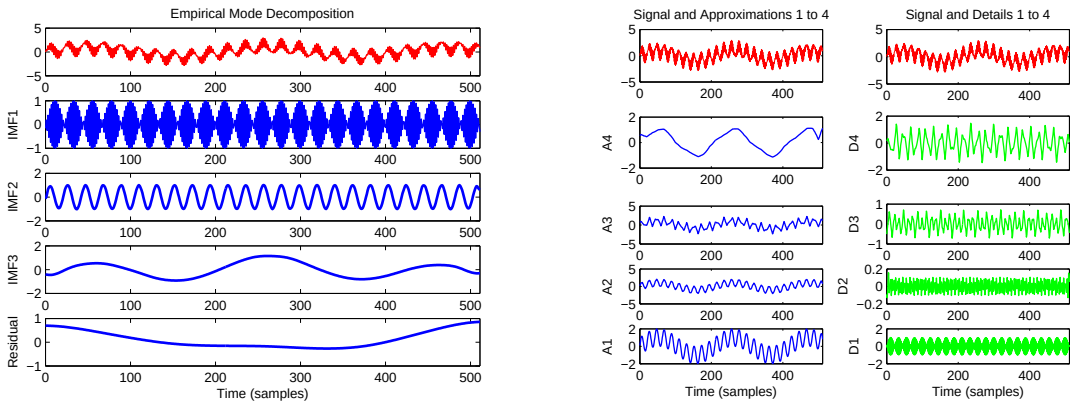


Figure I.8: Signal $x(t) = \sin(3t) + \sin(0.3t) + \sin(0.03t)$ and its theoretical components.

respectively the approximations and details of signal. This figure shows that the first component $\sin(3t)$ is presented in D1, the second component $\sin(0.3t)$ in D4, and finally A4 roughly corresponds to the third component $\sin(0.03t)$.

As shown by Fig. I.10, we note that in this case the estimated errors for different

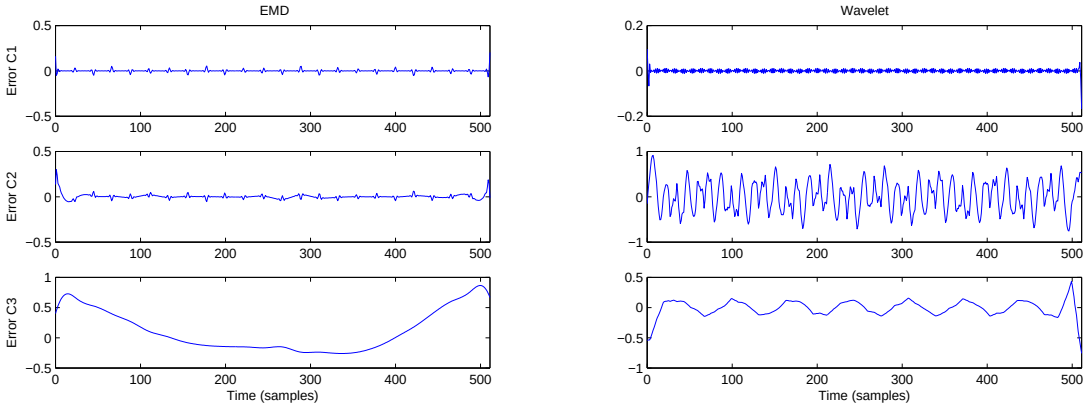


(a) EMD . (b) Wavelet.

Figure I.9: Comparison of decomposition by the EMD to wavelet.

components are generally more important in the case of wavelet approach than for the EMD approach. All the differences and errors in the EMD decomposition are mainly due to the sampling frequency of the signal and the spectral differences

between the components of the signal. For the wavelet, the lack of accuracy of the choice of the base function respect to $x(t)$ explain its less effective behavior.



(a) EMD . (b) Wavelet.

Figure I.10: Error estimates with EMD and wavelet.

I.4 Conclusion

In this chapter, we studied some aspects of the EMD [34]. This temporal and nonlinear decomposition is defined as the output of an algorithm. We have shown that the signal can be decomposed into a finite number of components of oscillating nature and named IMF. IMFs are centered modes and type AM-FM. We have checked, based on simulations, that the extraction of IMFs is nonlinear, but that their linear recombination is accurate. Each IMF is obtained by a process called *sifting*, which is iterative, sequential and local. We have shown that the decomposition results supplied by the EMD is conditioned by the interpolation technique used and the sampling frequency of the signal. Finally, we noted that this decomposition is organized in a dyadic filter bank structure, in particular for Gaussian white noise.

The analysis of the behavior of the EMD suggests that it could be a useful tool for many problems met with audio signal processing. Indeed, many audio signals are known to be well described as a sum of harmonics and white noise. This is the case in particular for speech signals. Thus, in the rest of the thesis, we are going to investigates how successful could be the use of EMD in audio and signal processing.

CHAPTER II

Speech enhancement by EMD

Contents

II.1 Introduction 43

II.2 EMD based white noise reduction 43

II.2.1 EMD-MMSE filter 44

II.2.2 EMD-Shrinkage 45

II.2.3 EMD-MMSE versus EMD-Shrinkage 46

II.3 EMD-ACWA filtering of white and colored noises 50

II.3.1 Interest of ACWA filter 50

II.3.2 Performance analysis of EMD-ACWA 54

II.4 Conclusion 59

In this chapter, a new speech denoising strategy is presented. Based on EMD, the method is fully data-driven approach. For additive white Gaussian noise, two strategies to denoise each extracted IMF are proposed: filtering using the Minimum Mean Squared Error (MMSE) filter [75], or thresholding using a shrinkage function. The performance of the two methods is analyzed and compared with those based on MMSE filter, and wavelet shrinking approach. To avoid frequency analysis when using MMSE filter [75], the IMFs are filtered by Adaptive Center Weighted Average (ACWA) filter [52], which operates in time domain. Finally, we show the interest of the conjunction EMD and ACWA for both white and colored noises reduction.

II.1 Introduction

In this chapter, denoising methods based on the EMD are proposed. We first propose two new noise reduction schemes dedicated to additive white noise. Actually these schemes are complementary and depend on the noise level and its estimation. The first strategy combines the EMD and the MMSE filter [75], and the second one associates the EMD with hard shrinkage [7],[9]. The MMSE filter assumes the whiteness of the noise and the stationarity of the denoised signal. In second time, a noise reduction approach combining the EMD with ACWA filter is introduced. Indeed, the ACWA filter, as the EMD, operates in the time domain, and it does not require neither the stationarity of the signal nor the whiteness of the noise. As a result, this method is effective for both white noise and colored one. Furthermore, in contrast to the classical filters, such as MMSE filter [75], all the parameters are computed in time domain and, hence, transformation to frequency domain is not necessary.

II.2 EMD based white noise reduction

Let a clean speech signal $x(t)$ be corrupted by an additive white noise $b(t)$ as follows:

$$y(t) = x(t) + b(t) \tag{II.1}$$

Noisy signal is decomposed into a sum of IMFs by the EMD, such that:

$$y(t) = \sum_{j=1}^{C} \mathrm{IMF}_j(t) + r_C(t) \tag{II.2}$$

We make assumption that each mode $\mathrm{IMF}_j(t)$ is a noisy version of the signal $f_j(t)$:

$$\mathrm{IMF}_j(t) = f_j(t) + b_j(t) \tag{II.3}$$

Based on the noisy observation $\mathrm{IMF}_j(t)$, an estimation $\tilde{f}_j(t)$ of $f_j(t)$ is given by,

$$\tilde{f}_j(t) = \Gamma[\mathrm{IMF}_j(t)], \tag{II.4}$$

where $\Gamma[\mathrm{IMF}_j(t)]$ is a filtering function applied to $\mathrm{IMF}_j(t)$ [7],[9]. Function Γ corresponds to MMSE filter [75], or to a thresholding function [7],[9]. Finally, the

denoised signal, $\tilde{x}(t)$, is given by:

$$\tilde{x}(t) = \sum_{j=1}^{C} \tilde{f}_j(t) + r_C(t) \tag{II.5}$$

Note that the use of MMSE filter requires the knowledge of the noise level.

II.2.1 EMD-MMSE filter

As mentioned previously, the EMD-MMSE strategy combines the EMD and the MMSE filter [25]. To guarantee the signal stationarity imposed by the MMSE filter a frame processing is required. Thus, each IMF is filtered in frequency domain by the MMSE filter as follows:

$$\tilde{\mathrm{F}}_j(f_d, m) = \mathrm{H}(f_d, m)\mathrm{IMF}_j(f_d, m), \tag{II.6}$$

where $\mathrm{IMF}_j(f_d, m)$ and $\tilde{\mathrm{F}}_j(f_d, m)$ are the spectral noisy IMF and the spectral denoised IMF respectively, observed at the discrete frequency f_d on the frame m. The frequency response of the MMSE filter $H(f_d, m)$ is given by [25]:

$$H(f_d, m) = \frac{SNR_{prio}(f_d, m)}{1 + SNR_{prio}(f_d, m)}, \tag{II.7}$$

where the a priori Signal to Noise Ratio (SNR), SNR_{prio}, is estimated according to the method of Ephraim and Malah [25], as following:

$$SNR_{prio}(f_d, m) = \alpha \frac{\tilde{F}^2(f_d, m-1)}{B^2(f_d, m-1)} + (1-\alpha)\max(SNR_{inst}(f_d, m), 0) \tag{II.8}$$

where α is a weighting factor (equal to 0.98, it is a compromise), SNR_{inst} is the instantaneous SNR, defined as the local estimation of SNR_{prio}, and the $B^2(f_d, m-1)$ is the noise power spectra value at the discrete frequency f_d in the frame (m-1).

$$SNR_{inst} = \frac{|\mathrm{IMF}(f_d, m)|^2}{|B(f_d)|^2} - 1 \tag{II.9}$$

Generally, noise estimation in speech is performed using the Boll's method [5]. Indeed, first, the silence periods of the signal are detected at the beginning of the signal. Then, the estimation of the noise power spectra is obtained by averaging the power spectra of the noisy signal over M frames which are considered as being

moments of silence. This method gives an estimation of the noise power spectra [5].

$$|\hat{B}(f_d)|^2 = \frac{1}{M}\sum_{k_m=0}^{M-1}|B(f_d,k_m)|^2 \tag{II.10}$$

where the k_m are the frame indices that correspond to silence periods. Extensive simulations have shown that when the speech signal presents the silence period, the first IMF also presents the silence period. Since, the first IMF is noise dominant then it can be used to estimate accurately the noise level. According to [28], behavior of EMD is like of that of wavelets [76], and thus the noise levels of the modes following the first IMF (k=1) are estimated as follows:

$$\tilde{\sigma_k} = \frac{\tilde{\sigma_1}}{\sqrt{2}^{k-1}},\ k \geq 2, \tag{II.11}$$

where $\tilde{\sigma_1}$ is the noise level of the first IMF.

II.2.2 EMD-Shrinkage

A smooth version of the input signal can be obtained by thresholding the IMFs before signal reconstruction [7],[9]. In this case, the threshold parameter of each IMF_k is estimated by the following expression [7],[9],[23]-[24]:

$$\tau_k = \sqrt{2\log(T)}\sigma_k \tag{II.12}$$

where T is the signal length and σ_k is the estimated noise level (scale level) at the IMF_k. Noise level of the first IMF is given by [7],[9],[66]

$$\tilde{\sigma_1} = 1.4826 \times \text{Median}\left\{\left|\text{IMF}_1(t) - \text{Median}\left\{\text{IMF}_1(t)\right\}\right|\right\} \tag{II.13}$$

According to [28], the noise level $\tilde{\sigma_k}$ of the k^{th} IMF can be deduced from $\tilde{\sigma_1}$ by Eq.(II.11).

There are different non-linear shrinkage functions [55]. In the present work, we use the hard shrinkage which has given interesting denoising results for speech enhancement compared to the soft shrinkage:

$$\tilde{f}_j(t) = \begin{cases} \text{IMF}_j(t), & \text{if } |\text{IMF}_j(t)| > \tau_j \\ 0, & \text{if } |\text{IMF}_j(t)| \leq \tau_j \end{cases} \tag{II.14}$$

II.2.3 EMD-MMSE versus EMD-Shrinkage

The two proposed noise reduction methods are tested on speech signals corrupted by additive white Gaussian noise with different input SNRs. Four clean speech signals, download from Brown Corpus Database, pronounced by a male and female speaker, and are sampled by a sampling frequency equal to 16 KHz. The results are compared to MMSE filter and to wavelet approach (Haar, Symmlet 4, Daubechies 4). The output SNR [1] and Perceptual Evaluation of Speech Quality (PESQ)[2] [65],[70] are used as objective measure to evaluate the denoising methods. More precisely, the PESQ criterion measures the perceptual quality of speech signal. The EMD-MMSE method is compared to the classical MMSE denoising method [75].

The EMD-MMSE denoising scheme is applied to four clean speech signals "spech1", "speech2", "speech3" and "speech4" (Figs. II.1(a)-(b)-(c)-(d)) corrupted by additive white Gaussian noise with input SNR values ranging from 4 dB to 10 dB. Noisy versions of the original signals corresponding to input SNR=5 dB are shown in figure II.2. Figure II.3 shows the denoising results obtained by the EMD-MMSE and

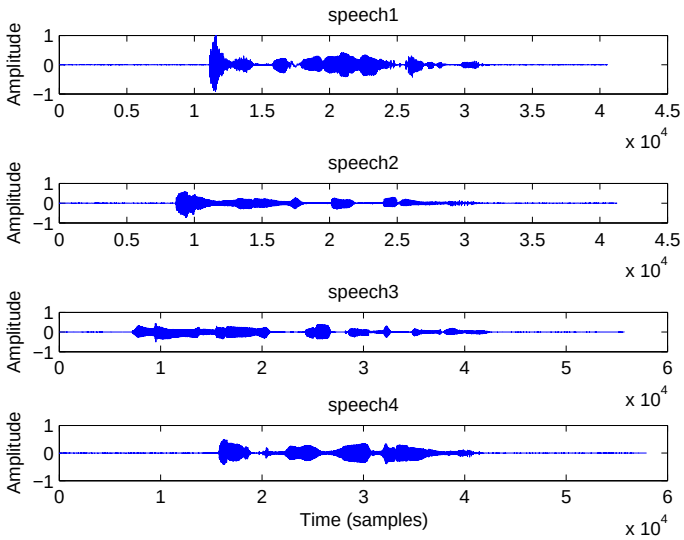


Figure II.1: Original signals "speech1", "speech2", "speech3" and "speech4".

the MMSE filter. From this figure, one can conclude that the EMD-MMSE performs better (noise reduction) than MMSE filter compared to the original signals (Figs. II.1). This fact is confirmed by the results shown in figure II.4, where more SNR gain

[1] see Appendix A
[2] see Appendix A

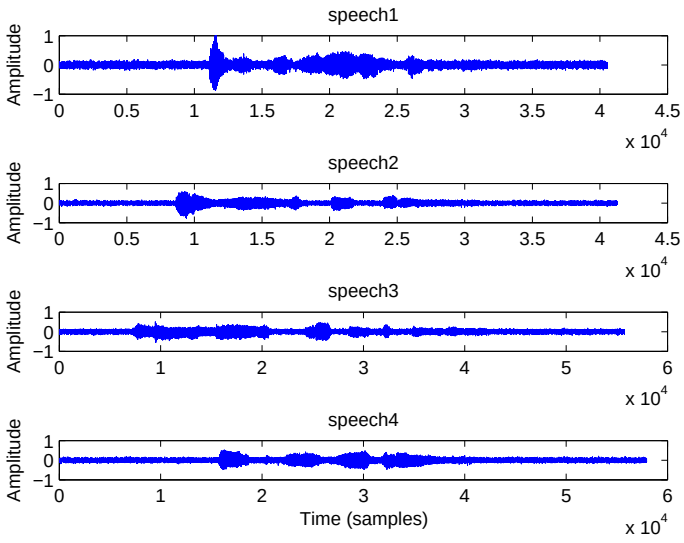
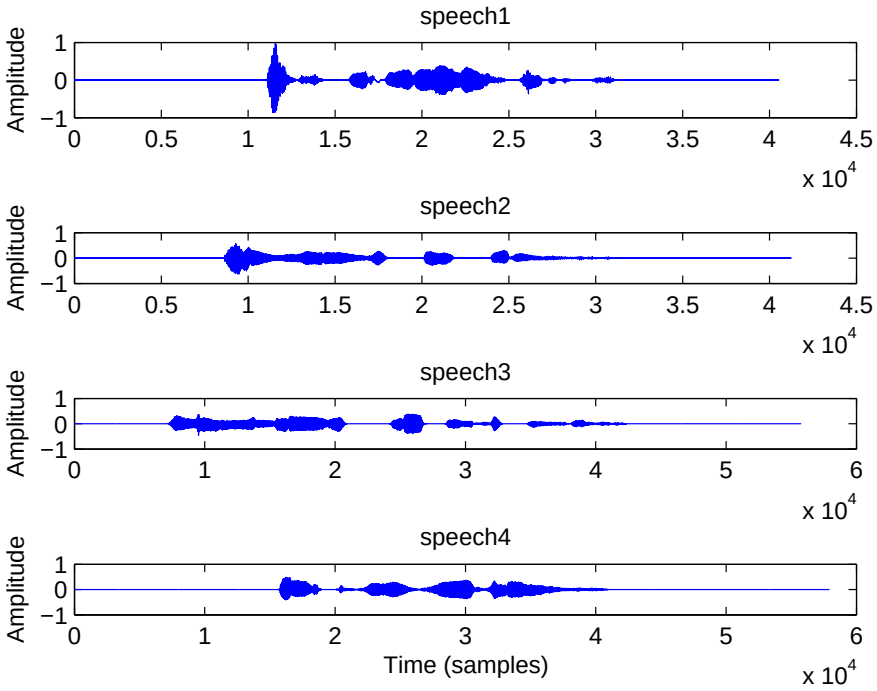


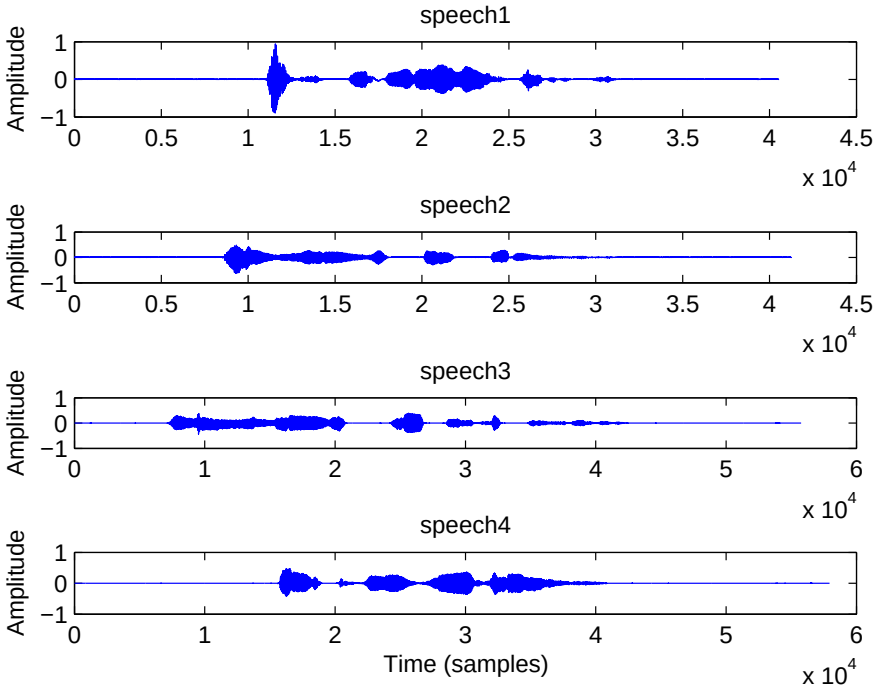
Figure II.2: Noisy version of signals "speech1", "speech2","speech3" and "speech4" (input SNR = 5 dB).

is obtained by the EMD-MMSE compared to the MMSE. For each input SNR value, 100 independent noise simulations are generated and the average of output SNR and the PESQ values are calculated. One may note that the EMD-MMSE provides an improvement about 1 dB compared to standard MMSE filter for noisy versions of all signals "spech1", "speech2", "speech3" and "speech4". The obtained results also show that it is more efficient to apply the MMSE to the different components of the signal than to the signal itself. These results are also demonstrated by Fig. II.5, where the PESQ values of the proposed approach are better than those of MMSE filter. These results also show that it is more efficient to apply MMSE filter to all IMFs, since the IMFs are more stationary than the original signal.

The EMD-Shrinkage is applied to the same clean speech signals "speech1", "speech2", "speech3" and "speech4" (Figs.II.1(a)-(b)-(c)-(d)) corrupted by additive white Gaussian noise with input SNR values ranging from -10 dB to 3 dB. Noisy versions of the original signals corresponding to input SNR=-5 dB are shown in figure II.6. Denoising results of the EMD-Shrinkage (hard thresholding) and the wavelet method (Daubechies 4) are shown in figure II.7.

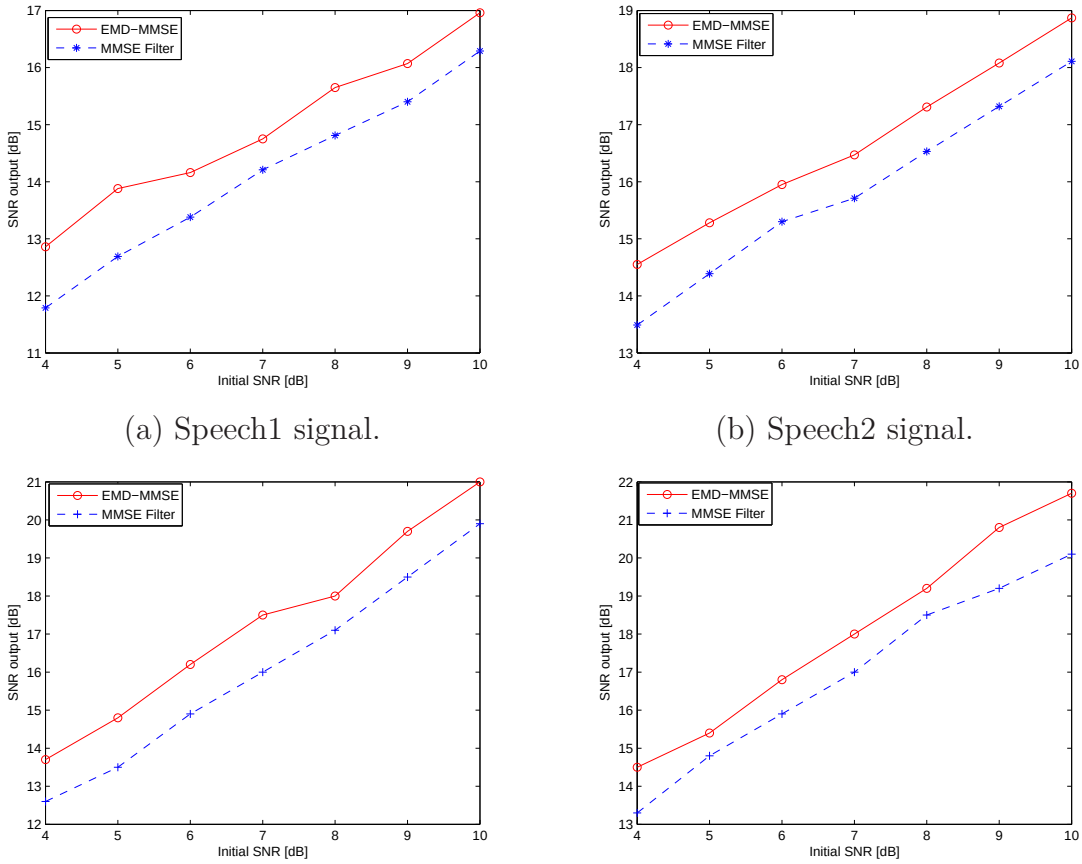


EMD-MMSE



MMSE filter

Figure II.3: Denoising results of signals "speech1", "speech2","speech3" and "speech4" by the EMD-MMSE and the MMSE filter.



(a) Speech1 signal.

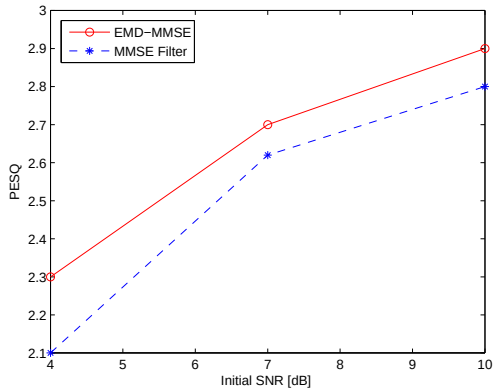
(b) Speech2 signal.

(c) Speech3 signal.

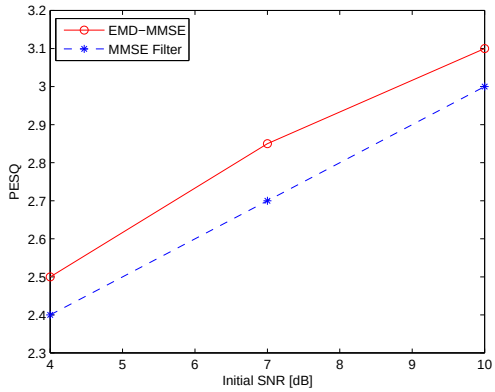
(d) Speech4 signal.

Figure II.4: Final SNR values obtained from different initial noise levels of signals "speech1", "speech2", "speech3" and "speech4". The results are averages over 100 instances of the noisy signals. They are reported for EMD-MMSE and the MMSE filter.

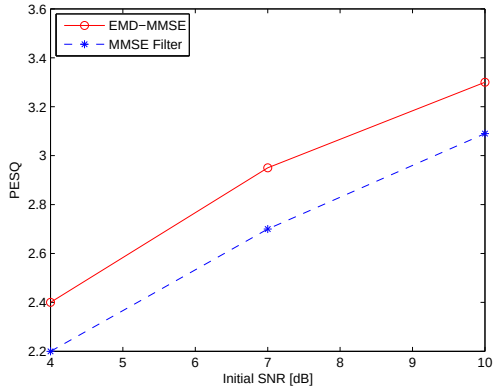
A careful examination of the signals shown in figures II.1 and II.7, shows that the EMD-Shrinkage performs better than the wavelet method in terms of noise reduction. Furthermore, signals structures or features are globally better preserved with the EMD-Shrinkage than with the wavelet method. Figure II.8 shows the improvement in SNR values obtained with different noise levels of the signals "speech1", "speech2", "speech3" and "speech4" for the EMD-Shrinkage and three wavelet methods (Haar, Symmlet 4, Daubechies 4). This figure shows that the improvement in output SNR provided by the EMD-Shrinkage varies from -0.7 dB to 11.5 dB compared to the three wavelet methods. The gain in SNR achieved by the EMD-shrinkage is much higher than with wavelets. When listening to the enhanced speeches, the EMD-shrinkage is found to produce lower residual noise and, noticeably, less speech distortion for all the signals compared to the wavelet method (Fig.



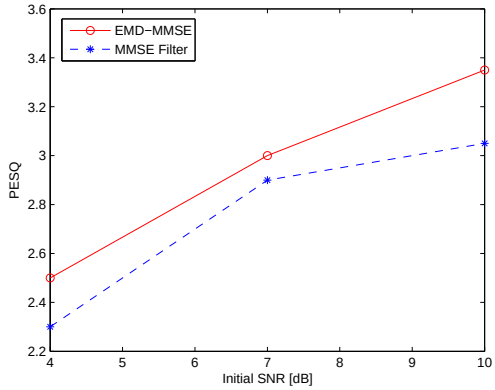
(a) Speech1 signal.



(b) Speech2 signal.



(c) Speech3 signal.



(d) Speech4 signal.

Figure II.5: PESQ values obtained from different initial noise levels of signals "speech1", "speech2", "speech3" and "speech4". The results are averages over 100 instances of the noisy signals. They are for EMD-MMSE and the MMSE filter.

II.9).

II.3 EMD-ACWA filtering of white and colored noises

II.3.1 Interest of ACWA filter

Classically, the ACWA filter has been used in image enhancement applications [52],[59],[71]. The ACWA filter operates in the time domain, and it does not require the stationarity of the signals and the whiteness of the noise. The best of our knowledge it is the first time (in this thesis) that ACWA filter is used in signal processing. The effectiveness of the ACWA filter can be improved when it is associated

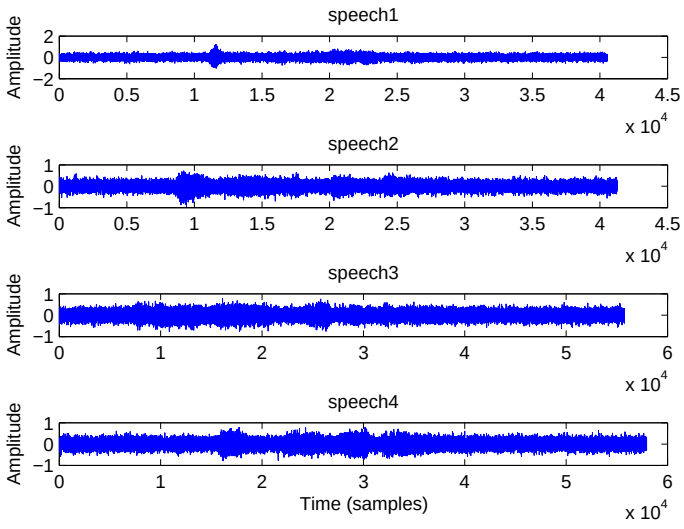


Figure II.6: Noisy versions of signals "speech1", "speech2", "speech3" and "speech4" (input SNR =-5 dB).

with the EMD. Indeed, the IMFs are less noisy and more stationary than the noisy speech signal. In contrast to the classical filters, such as MMSE filter [75], all the parameters are computed in time domain and, hence, transformation to frequency domain is not necessary. Besides, the noise variance is computed at each instant time, and this filter can adapt to more general noisy contexts: white as well as colored noise, high as well as low noise level.
The ACWA filtered signal $\tilde{x}(t)$ is described as follows [52]:

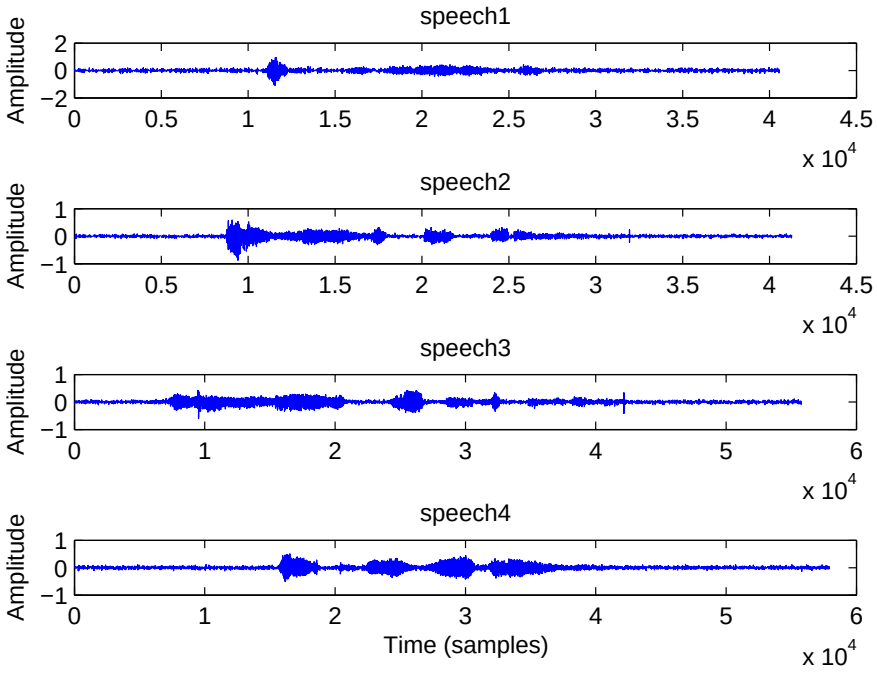
$$\tilde{x}(t) = \begin{cases} F_{\text{mean}} + K(\text{y}(t) - F_{\text{mean}}) & \text{if } F_{\text{var}} \geq \sigma^2 \\ F_{\text{mean}} & \text{otherwise} \end{cases} \tag{II.15}$$

where,

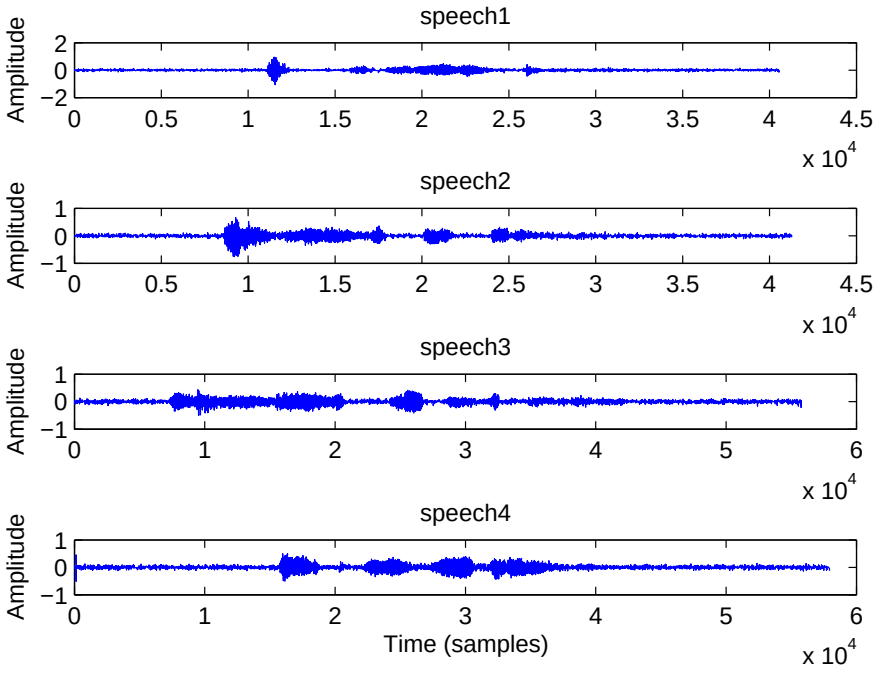
$$K = 1 - \frac{\sigma^2}{F_{\text{var}}}, \tag{II.16}$$

F_{mean} and F_{var} denote respectively the average and the variance of the noisy signal $y(t)$ computed over a sliding window of size L, and σ^2 denotes the variance of the noise. The noise variance, σ^2, is calculated as previously (Eq. II.13).

In order to show the effectiveness of this filter in the speech context, a comparative analysis between ACWA filter and MMSE filter [75] is preformed in a context of additive white noise with $SNR_{in} = 2$dB. Figure II.10 shows the superposition

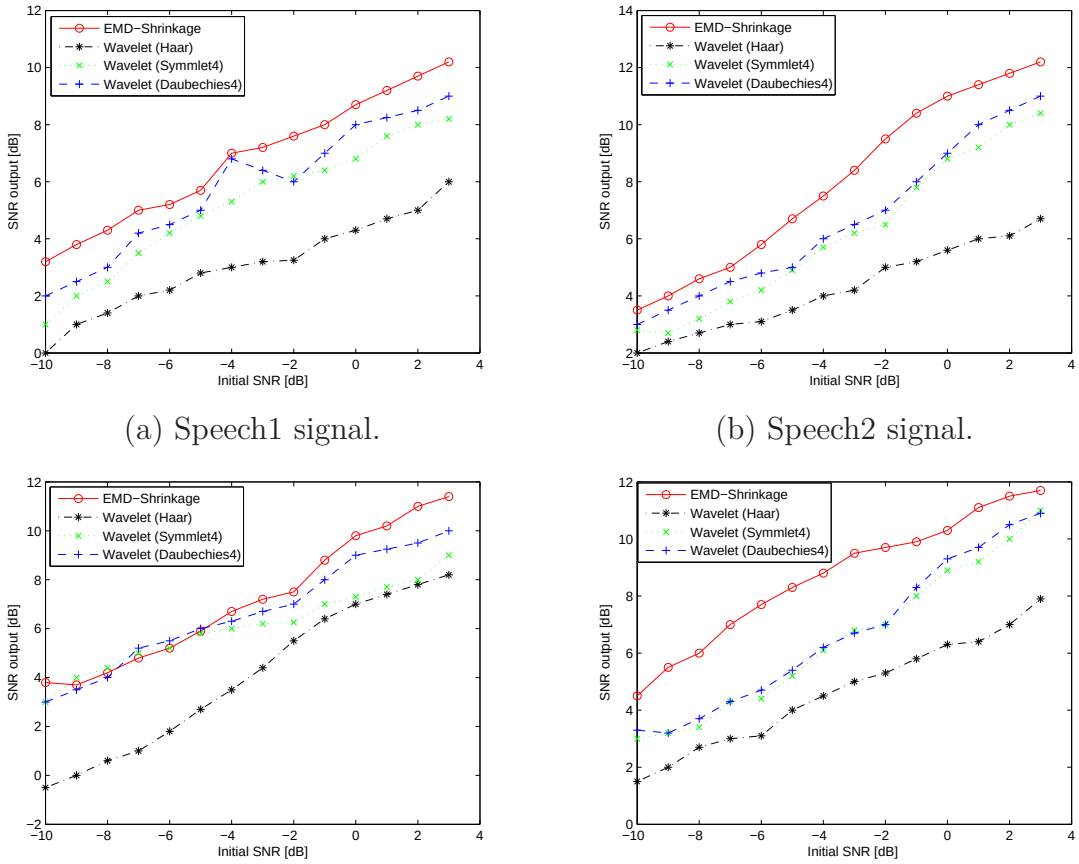


EMD-Shrinkage



Wavelet-Shrinkage (Daubechies 4)

Figure II.7: Denoising results of signals "speech1", "speech2", "speech3" and "speech4" by the EMD-Shrinkage and the wavelet approach (Daubechies 4).



(a) Speech1 signal.

(b) Speech2 signal.

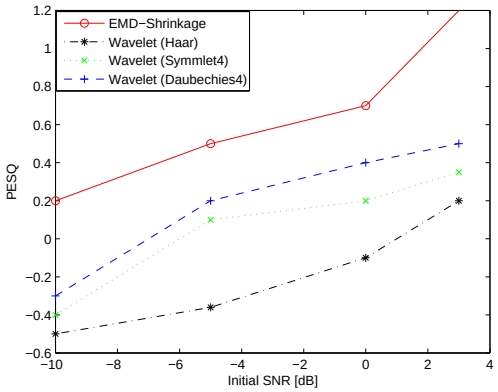
(c) Speech3 signal.

(d) Speech4 signal.

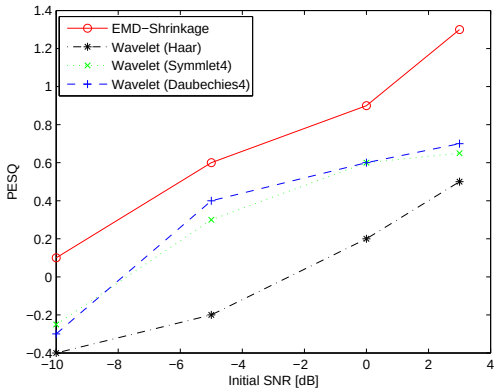
Figure II.8: Final SNR values obtained from different initial noise levels of signals "speech1", "speech2", "speech3" and "speech4". The results are averages over 100 instances of the noisy signals. They are reported for EMD-Shrinkage and for three different Wavelets (Haar, Symmlet 4, Daubechies 4).

of the clean signal and the filtered signals obtained by the ACWA and the MMSE filters.

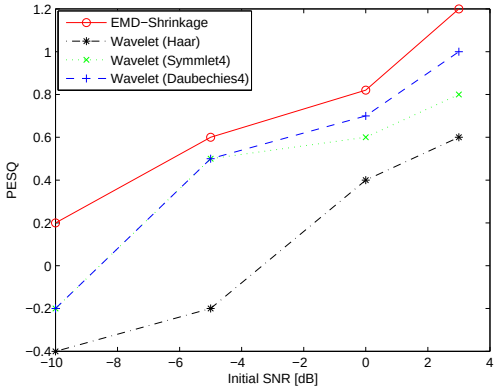
The comparative analysis of the three signals (Fig. II.10) does not clearly show the superiority of the ACWA filter over the MMSE one. Therefore, we use the output SNR and the PESQ criteria to quantify the speech enhancement quality obtained by both filters. Table II.1 reports the obtained results for different levels of the additive noise fixed through the input SNR. These results show that for very low input SNR values, the ACWA filter gives higher output SNR than the MMSE filter. In addition, for most considered input SNR values, the PESQ values given by the ACWA filter are higher than those of the MMSE filter. The PESQ results confirm that the ACWA filter guarantees better listening quality of the enhanced speech than the MMSE filter.



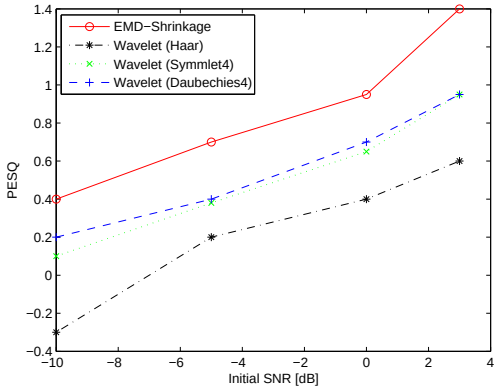
(a) Speech1 signal.



(b) Speech2 signal.



(c) Speech3 signal.



(d) Speech4 signal.

Figure II.9: Variations of the PESQ values versus from the input SNR for signals "speech1", "speech2", "speech3" and "speech4". The results are averages over 100 instances of the noisy signals. They are reported for EMD-Shrinkage and for three different wavelets (Haar, Symmlet 4, Daubechies 4).

II.3.2 Performance analysis of EMD-ACWA

The EMD-ACWA denoising technique consists on filtering all IMFs by ACWA filter. This approach is still applicable regardless of the value added noise and noise type. Note that he function Γ (Eq. II.4) can be interpretable as a kind of ACWA filter. Finally, the estimated signal, $\tilde{x}(t)$, is given by :

$$\tilde{x}(t) = \sum_{j=1}^{C} \tilde{f}_j(t) + r_C(t) \tag{II.17}$$

The denoising of the IMF by the ACWA filter is given by Eq. II.15. The noise level σ_j is calculated using equation II.13.

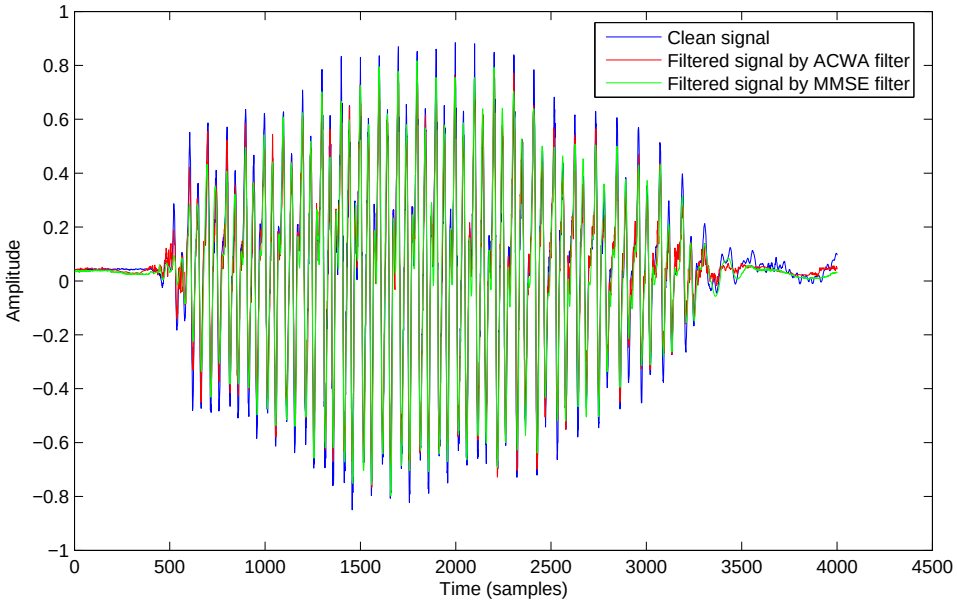


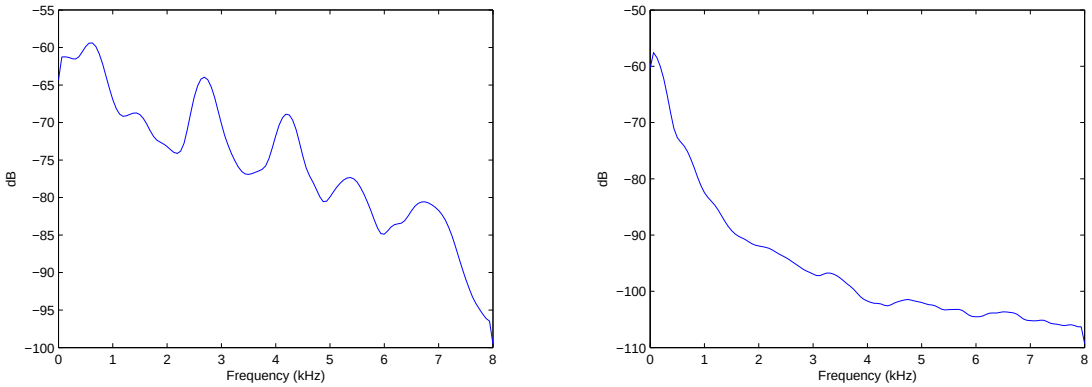
Figure II.10: Clean and filtered signals by the ACWA and the MMSE filters (input SNR=2 dB).

Table II.1: Variations of the output SNR and of the PESQ over the input SNR for the MMSE and ACWA filters.

	MMSE filter		ACWA filter	
SNR input [dB]	SNR output [dB]	PESQ	SNR output [dB]	PESQ
-10	0.87	0.70	2.7	1.05
-8	1.53	0.91	4.04	1.2
-6	3.52	1.07	5.94	1.38
-4	5.00	1.29	7.98	1.58
-2	7.37	1.51	10.18	2.15
0	9.82	2.05	11.19	2.21
2	12.63	2.35	12.08	2.27
4	13.76	2.4	13.95	2.4
6	15.88	2.51	15.67	2.49
8	16.53	2.64	16.58	2.7
10	17.23	2.8	17.18	2.73

This proposed noise reduction method is tested on speech signals corrupted by different noises, taken from Noisex-92 database, whose levels are fixed through the input SNR. The noise (f16, factory) spectrum is depicted in figure II.11.

The results obtained by the proposed method are compared to the wavelet approach



(a) f16 noise spectrum (b) Factory noise spectrum

Figure II.11: Noise power spectral density

(Daubechies 4) and ACWA filter. Here, we use the ACWA filter as a reference comparison method because it gives better results than the MMSE filter [43]. As objective criteria to evaluate the performance of the denoising method, we use the output SNR and PESQ as before.

At a first step, we take as example two speech signals "speech1" and "speech2". These signals are corrupted by a colored noise "f16" (cockpit noise) with input SNR value fixed to -2 dB. The original signals and their corresponding noisy versions are depicted in figure II.12.

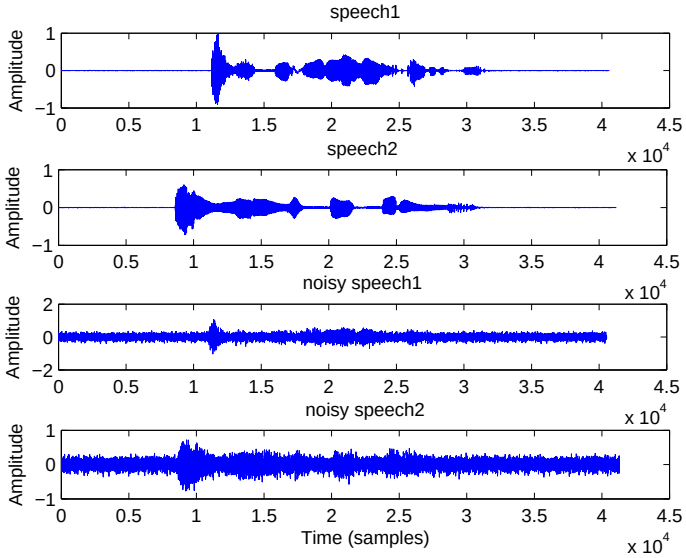


Figure II.12: Original signals ("speech1" and "speech2") and their noisy versions (f16 noise with SNR =-2 dB).

The size L of the ACWA filter window is fixed to 511. This choice is justified by the

results shown in figure II.13 where the variations of the output SNR versus L are displayed for two values of input SNR : -2 dB and 0 dB. Figure II.13, shows that for $L = 511$ the output SNR remains almost constant.

The denoised versions of signals "speech1" and "speech2" obtained by the EMD-

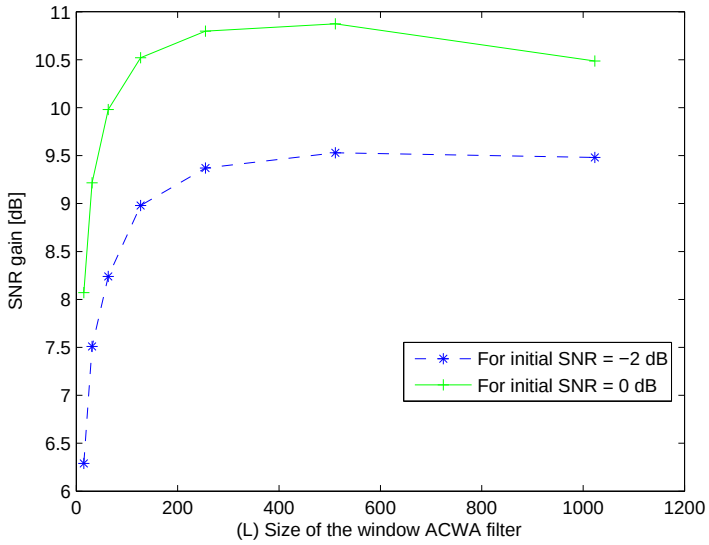
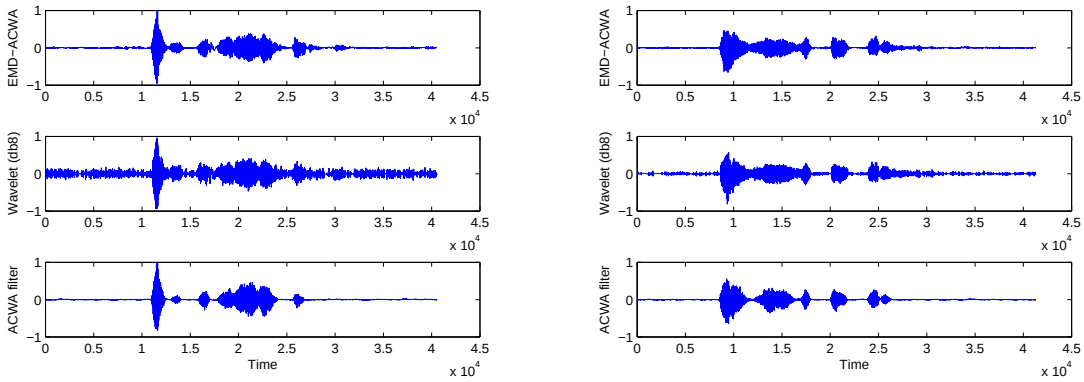


Figure II.13: Variation of the output SNR relating to the noisy signal "speech1" versus the size L of the ACWA filter window (f16 noise with SNR=-2 dB and SNR=0 dB).

ACWA, the wavelet thresholds (db4), and the ACWA filter, are shown respectively in figures II.14(a) and II.14(b). The input SNR is fixed to -2 dB. In fact, we consider db4 with a hard threshold for comparison, because it gives good results compared to other wavelets.

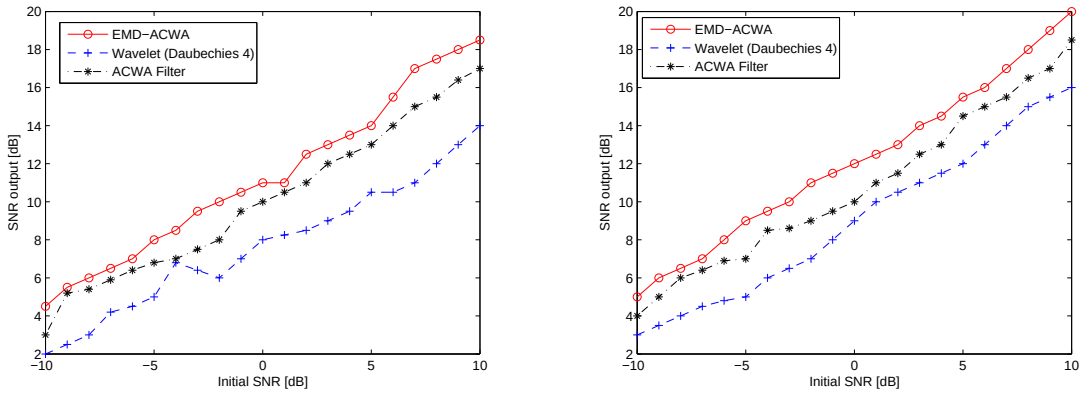
A careful comparative examination of the signals of figures II.14, shows that the EMD-ACWA performs better than the wavelet (db4) and ACWA-filter in terms of noise reduction. For deeper performance analysis, figures II.15, II.16 and II.17 show the variations of the output SNR versus the input SNR relating to the denoising of signals "speech1" and "speech2" when corrupted respectively by a white Gaussian noise, the colored f16 noise and the colored factory noise, taken from Noisex-92 database. The reported results demonstrate the effectiveness of the proposed method. Indeed, the improvement in SNR provided by the EMD-ACWA is much higher than those given by the wavelet method and the ACWA filter. Besides, a significant SNR improvement, varying from 4.2 dB to 17.4 dB, is achieved by the EMD-ACWA method. In fact, even for very low SNR values, we can still observe the effectiveness of the proposed method in removing the noise components as the



(a) Speech1 signal (b) speech2 signal

Figure II.14: Denoised version of the signals "speech1" and "speech2" obtained by the EMD-ACWA, the wavelet (db4) and ACWA filter (f16 noise with input SNR =-2 dB)

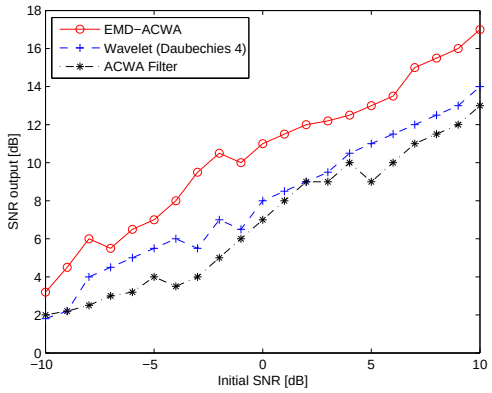
gain in SNR can go up to 14 dB.



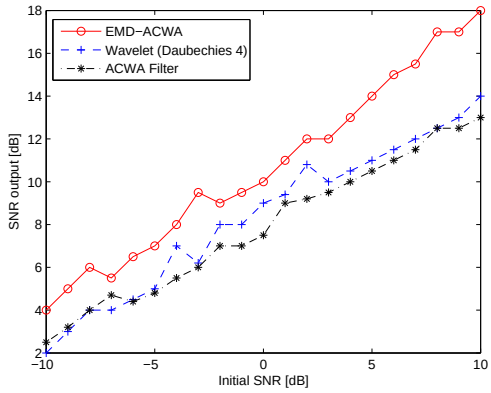
(a) Speech1 signal (b) Speech2 signal

Figure II.15: Variation of the output SNR versus the input SNR relating to the denoising of the signals "speech1" and "speech2" corrupted by a white Gaussian noise. The results are averages over 100 instances of the noisy signals. They are reported for EMD-ACWA, ACWA filter and wavelet(db4)

Figures II.18, II.19 and II.20 show the PESQ values: the obtained results show that the PESQ values achieved by EMD-ACWA are higher than those obtained by wavelet and ACWA filters. Consequently when listening to the enhanced speeches, the EMD-ACWA produces lower residual noise, noticeably less speech distortion compared to the wavelet (db4) method and ACWA filter.

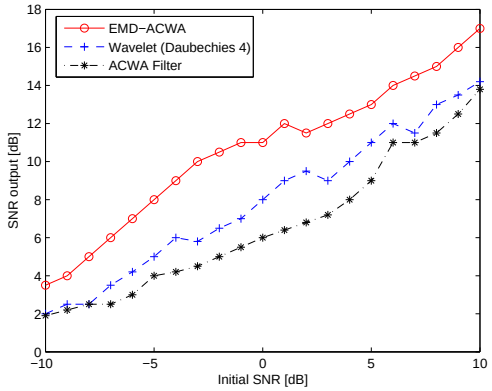


(a) Speech1 signal

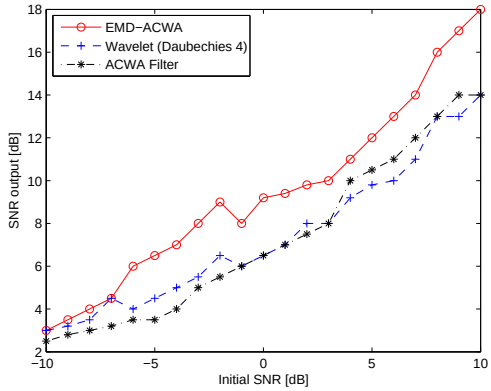


(b) Speech2 signal

Figure II.16: Variation of the output SNR versus the input SNR relating to the denoising of the signals "speech1" and "speech2" corrupted by the f16 noise. The results are reported for EMD-ACWA, ACWA filter and wavelet (db4)



(a) Speech1 signal



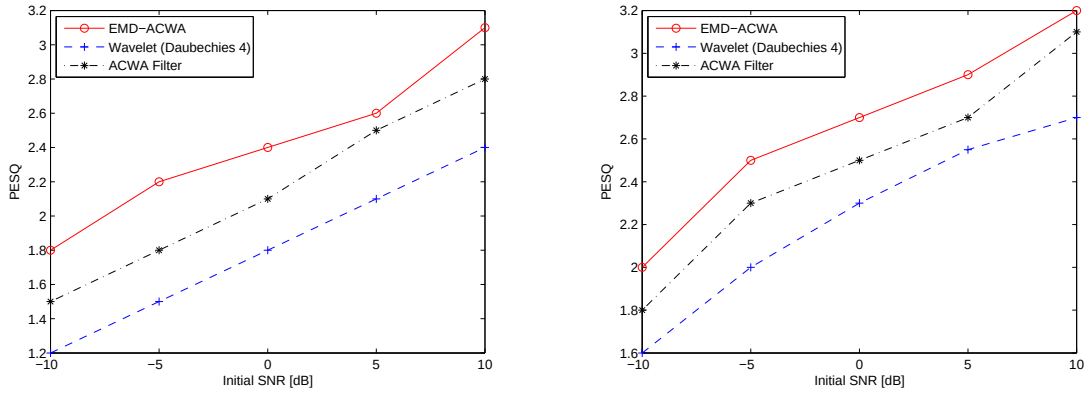
(b) Speech2 signal

Figure II.17: Variation of the output SNR versus the input SNR relating to the denoising of the signals "speech1" and "speech2" corrupted by the factory noise. The results are reported for EMD-ACWA, ACWA filter and wavelet(db4)

II.4 Conclusion

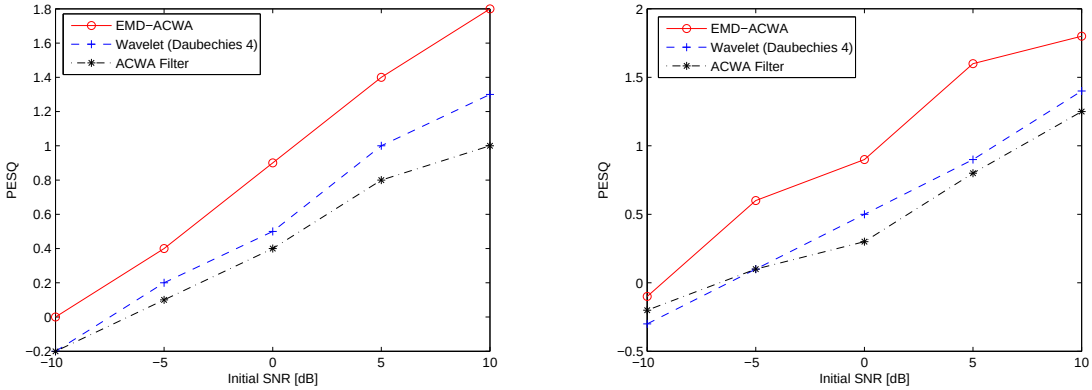
The proposed denoising schemes introduced in this chapter are based on the EMD. They are simple and fully data-driven methods. In particular, they do not require any pre- or post-processing and any use of parameters setting (except L value using ACWA).

For the two first approaches, the study is limited to signals corrupted by additive white noise. Obtained results for clean speech signals corrupted with additive Gaussian noise with different SNR values ranging from -10 dB to 10 dB show that the proposed EMD-MMSE and EMD-Shrinkage methods, perform better than the MMSE



(a) Speech1 signal (b) Speech2 signal

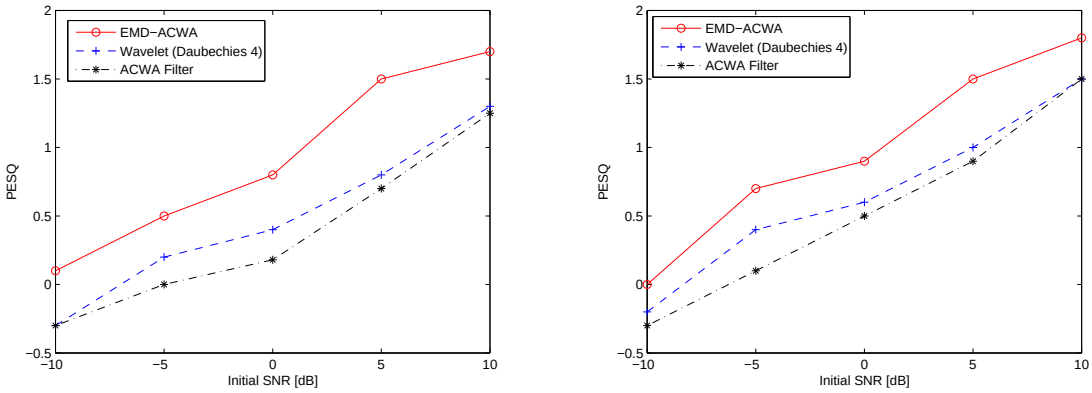
Figure II.18: PESQ values obtained from different initial noise levels of signals "speech1" and "speech2". The results are an average of 100 instances signal. It's reported for EMD-ACWA, ACWA filter and wavelet(db4)



(a) Speech1 signal (b) Speech2 signal

Figure II.19: PESQ values obtained from different initial noise levels of signals "speech1" and "speech2" corrupted by the f16 noise. The results are reported for EMD-ACWA, ACWA filter and wavelet(db4)

filter and the wavelet approaches. These results show that the EMD-denoising methods are effective for noise removal and confirm our findings presented in [7]-[9]. In particular, the obtained results also show that it is more efficient to apply the thresholding or the filtering to the different components (IMFs) of the signal than to the signal itself. Quite normal, since IMFs are more stationary than the noisy signal, and consequently the association of filter or threshold with the EMD improves the denoising results. Furthermore, the introduction of the EMD is very simple, since it is an adaptive decomposition, data driven, and does not need to define a kernel function. Thus, the results are not limited to the choice of basic functions, as in the case of wavelets.



(a) Speech1 signal (b) Speech2 signal

Figure II.20: PESQ values obtained from different initial noise levels of signals "speech1" and "speech2" corrupted by the factory noise. The results are reported for EMD-ACWA, ACWA filter and wavelet(db4)

In the case of colored and white noise, the EMD-ACWA gives better results compared to the other approaches (ACWA filter, wavelet). The effectiveness of the ACWA filter is improved when it is associated with the EMD. In addition, the ACWA filter does not require the stationarity of the signal or the whiteness of the noise. For these reasons, the ACWA filter will be used in the future techniques dedicated to denoising speech signal, subject of the next chapter.

CHAPTER III

Speech denoising using EMD and local statistics

Contents

III.1 Introduction 65

III.2 Frames classification 65

III.2.1 Voiced frames detection 66

III.2.2 Transient frames detection 68

III.3 Proposed speech denoising method 70

III.3.1 Voiced sequence denoising 72

III.3.2 Unvoiced sequence denoising 73

III.3.3 Transient sequence denoising 73

III.4 Performance analysis 73

III.4.1 Voiced frames 73

III.4.2 Speech signal 78

III.5 Conclusion 82

In this chapter, we show how to improve the performance of the EMD-ACWA method. The new scheme takes into account the class of speech frames (voiced/unvoiced and transient). The noisy signal is divided into frames and each one is decomposed adaptively into IMFs. The number of IMFs filtered by ACWA filter depends on the frame class and is selected according to an appropriate criterion. An energy criterion detects voiced frames while a stationarity index,

based on the local statistics, is used to distinguish between unvoiced and transient frames sequences. The new denoising approach performs better than ACWA filter, wavelet (db4) approach and the conventional EMD-ACWA in terms of output SNR and PESQ.

III.1 Introduction

In chapter II, three strategies for noise reduction were proposed: MMSE filtering, thresholding and ACWA filtering of the extracted IMFs from the noisy frame. In particular, the ACWA filter [52], using local statistics of the speech signal, has shown very interesting performances in speech denoising. These last methods are based on filtering of all IMFs extracted from noisy frame regardless of their speech class (voiced/unvoiced/transient). However, when the signal features are concentrated on medium and low frequencies such as voiced speech, the filtering of all IMFs introduces some distortions in the denoised signal [1],[8],[85]. As a matter of fact, when voiced speech signal is contaminated by an additive white noise, the first IMFs are much more noisy than the last ones. Consequently, in the case of voiced speech, it is more appropriate to only filter the first IMFs, and to keep unchanged the last ones which are signal dominated.

In this chapter, we further improve the speech denoising using the EMD and the ACWA filter. This is achieved by taking into account the type of the processed frame: voiced, unvoiced and transient. As for the voiced frame special consideration related to the signal characteristics must be taken into account when denoising unvoiced frame or transient one that is considered here as concatenation of two sub-frames: voiced or unvoiced.

This chapter is organized as follows. In the second section, we present the techniques adopted to determine the type of frame. A criterion based on the IMFs energy is used to detect voiced frames, while a stationarity index criterion is used to distinguish between an unvoiced and a transient frame. Section III.3 details the speech denoising technique. The idea is based on filtering selected IMFs by the ACWA filter. The number of selected IMFs depends on the frame class. Section III.4 investigates the performance of this speech denoising approach, based on exhaustive simulation results. In a first step we shall only consider voiced frames, and frames of different types in a second step.

III.2 Frames classification

The speech signal is a combination of voiced and unvoiced frames. So, to apply the denoising approach for each frame, we must firstly determine the frame type.

The principle of frames classification is depicted in figure III.1. Noisy frame is first

decomposed by EMD and the energies of the associated IMFs are calculated. An energy criterion is applied to detect whether the input frame is voiced speech or not. While a stationary index is used to classify the frame into unvoiced or transient frame formed by two adjacent sub-frames. Finally, in the case of transient frame, the energy criterion is applied to classify the two sub-frame into voiced or unvoiced speech.

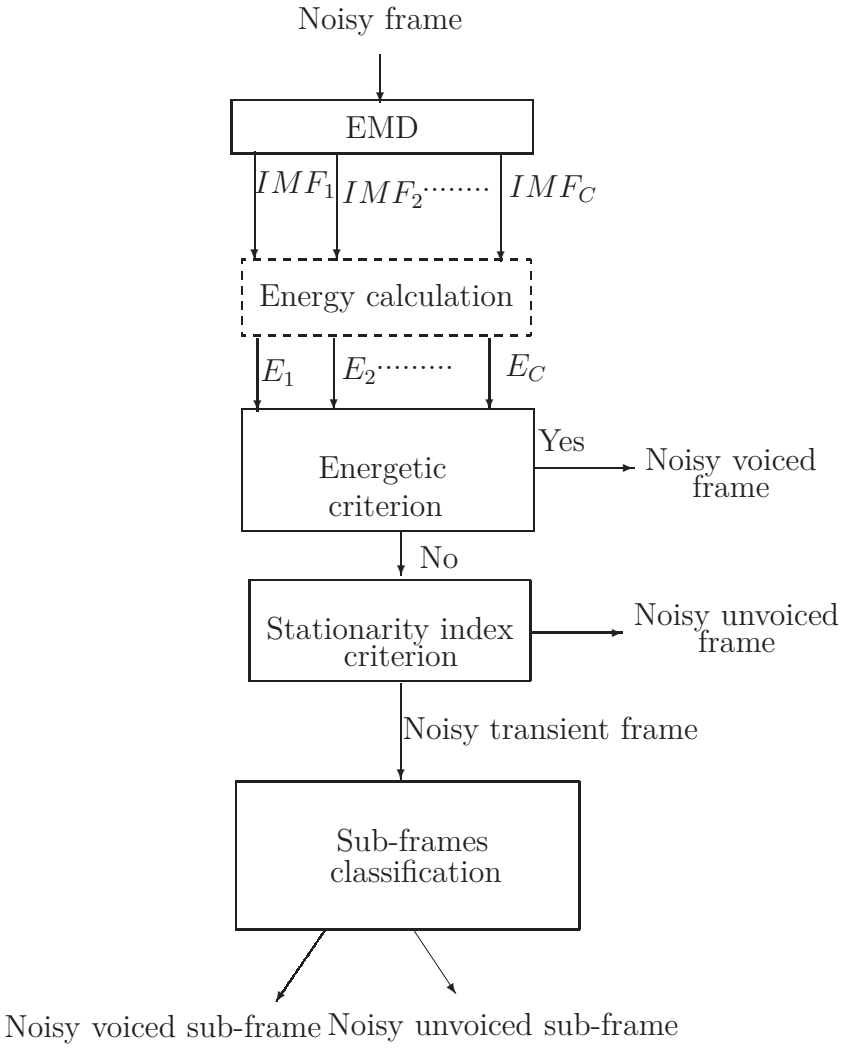


Figure III.1: Frames classification scheme.

III.2.1 Voiced frames detection

The energy criterion relies on the basic idea that most important features structures of the signal are concentrated at medium and low frequencies (last IMFs) [1][8],[11],[17],[44],[85], in particular for voiced frames. Therefore the first IMFs of

the noisy voiced frames are essentially noise dominated, while the last ones are signal dominated. According to this idea and in the case of an additive white noise, there will be a mode, indexed by j_s, from which energy distribution of the important structures of the signal overcomes that of the noise [1],[8]. Thus, a criterion based on energy density can be used to detect voiced frames [27],[87].
From the observed noisy signal $y(t)$, the objective is to find an approximation $\tilde{x}(t)$ to the original signal $x(t)$ that minimizes the Mean Square Error (MSE):

$$\mathrm{MSE}(x,\tilde{x}) \triangleq \frac{1}{N}\sum_{i=1}^{N}(x(i)-\tilde{x}(i))^2 \tag{III.1}$$

where $x=[x(1),x(2),\ldots,x(N)]^T$ and $\tilde{x}=[\tilde{x}(1),\tilde{x}(2),\ldots,\tilde{x}(N)]^T$. N is the length of the signal. Other distortion measures such as the Mean Absolute Error (MAE) can be used. Then, the signal $y(t)$ is first decomposed into $\mathrm{IMF}_j(t), j=1,\ldots C$, and a residual $r_C(t)$,

$$y(t)=\sum_{j=1}^{C}\mathrm{IMF}_j(t)+\mathtt{r}_C(t), \tag{III.2}$$

Finally $\tilde{x}(t)$ is reconstructed using $(C-k+1)$ selected IMFs starting from k to C.

$$\tilde{x}_k(t)=\sum_{j=k}^{C}\mathrm{IMF}_j(t)+\mathtt{r}_C(t),\ k=2,\ldots,C \tag{III.3}$$

The aim of the EMD filtering, which is carried out in the time domain, is to find the index $k=j_s$ that minimizes the $\mathrm{MSE}(x,\tilde{x})$. Note that Eq. (III.3) corresponds to a low-pass filtering [33]. In practice the MSE or the MAE can not be calculated because the original signal $x(t)$ is unknown. In this work, we use a distortion measure called Consecutive MSE (CMSE) that does not require the knowledge of $x(t)$ [8]. This quantity measures the squared Euclidean distance between two consecutive reconstructions of the signal. The CMSE is defined as follows [8]:

$$\mathrm{CMSE}(\tilde{x}_k,\tilde{x}_{k+1}) \triangleq \frac{1}{N}\sum_{i=1}^{N}(\tilde{x}_k(i)-\tilde{x}_{k+1}(i))^2,\ k=1,\ldots,C-1 \tag{III.4}$$

$$\triangleq \frac{1}{N}\sum_{i=1}^{N}(\mathrm{IMF}_k(t_i))^2 \tag{III.5}$$

where $\tilde{x}_k$ and $\tilde{x}_{k+1}$ are signals reconstructed starting from the IMFs indexed by k and $(k+1)$ respectively.
Thus, according to Eq. (III.5) the CMSE is reduced to the energy of the k^{th} IMF. It

is also the classical empirical variance estimate of the k^{th} IMF. Note that, if $k = 1$, $\tilde{x}_k(t) = y(t)$. Finally, the index j_s is given by

$$j_s = \underset{1 \leq k \leq C-1}{Arg\,max} \left[\text{CMSE}(\tilde{x}_k, \tilde{x}_{k+1})\right] \tag{III.6}$$

The CMSE criterion allows to identify the IMF order where there is the first significant change in energy. This empirical fact is derived from extensive experiments and simulations [8]. Figure III.2(c) shows the energies of the IMFs of a noisy voiced sequence (Figure III.2(b)). The most important features of this speech sequence begin at the fourth IMF ($j_s = 4$). This energy criterion is appropriate only to detect voiced frames [42]. In fact, as shown by figure III.3(c) the energies of IMFs decrease versus the IMF index for a noisy transient frame (Fig. III.3(b)). The same result is obtained in the case of an unvoiced frame (Fig. III.4(c)).

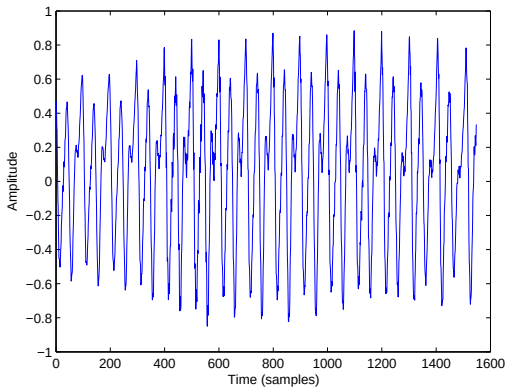
III.2.2 Transient frames detection

A transient frame can be linked to a concatenation of two sub-frames of different nature: voiced and unvoiced. The statistical properties of voiced and unvoiced speech are very different. The invariance of statistical properties over the time of a speech or audio signal can be measured using a stationarity index. Indeed, based on time-frequency analysis, this index detects fast transients of signals [51]. It was shown that both Kullback and Bhattacharyya distances are sensitive to abrupt changes of signals in the time-frequency plane [51]. In this work, Bhattacharyya distance is used as index of stationarity.
Two sub-images $\text{I}_1(n;\tau,f)$ and $\text{I}_2(n;\tau,f)$ are extracted, at each time n, from a Time Frequency Representation (TFR) of the signal [51]:

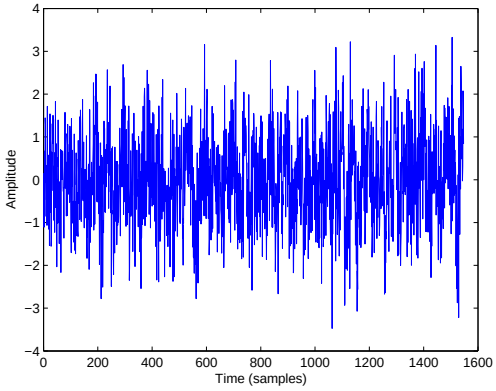
$$\text{I}_1(n;\tau,f) = \text{TFR}(n - L + \tau, f) \tag{III.7}$$
$$\text{I}_2(n;\tau,f) = \text{TFR}(n + \tau, f) \tag{III.8}$$

where L is the width of sub-images, f is the frequency and $\tau \in [0, L]$. The stationarity index is obtained by computing the Bhattacharyya distance between the two sub-images:

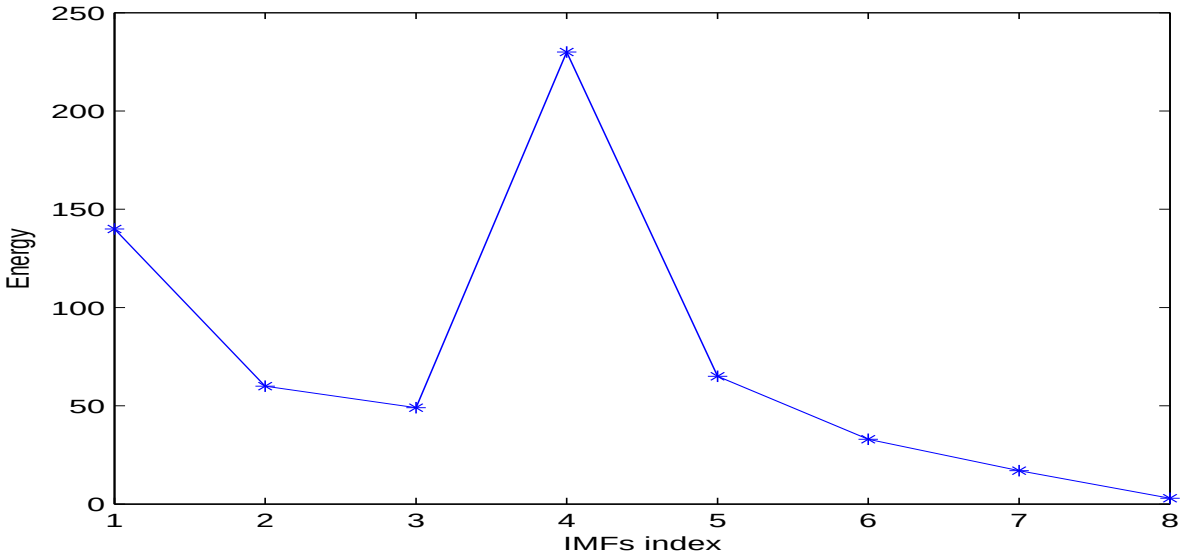
$$\text{SI}(n) = -\log\left(\int_{\tau=0}^{L}\int_{-\infty}^{+\infty}\sqrt{\text{NI}_1(n;\tau,f)\text{NI}_2(n;\tau,f)}df\,d\tau\right) \tag{III.9}$$



(a) Voiced sequence.



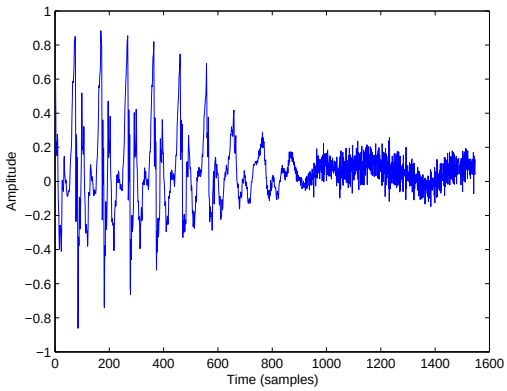
(b) Noisy voiced sequence.



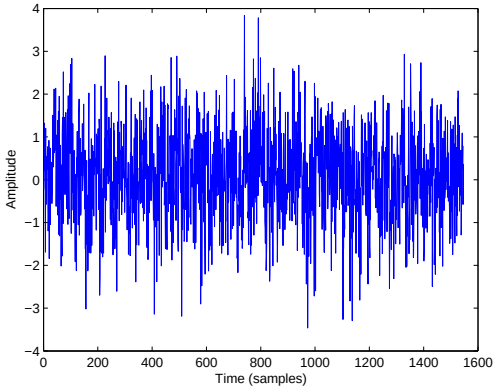
(c) Energy of noisy IMFs.

Figure III.2: Voiced sequence, noisy voiced sequence and the energy variations of their noisy IMFs.

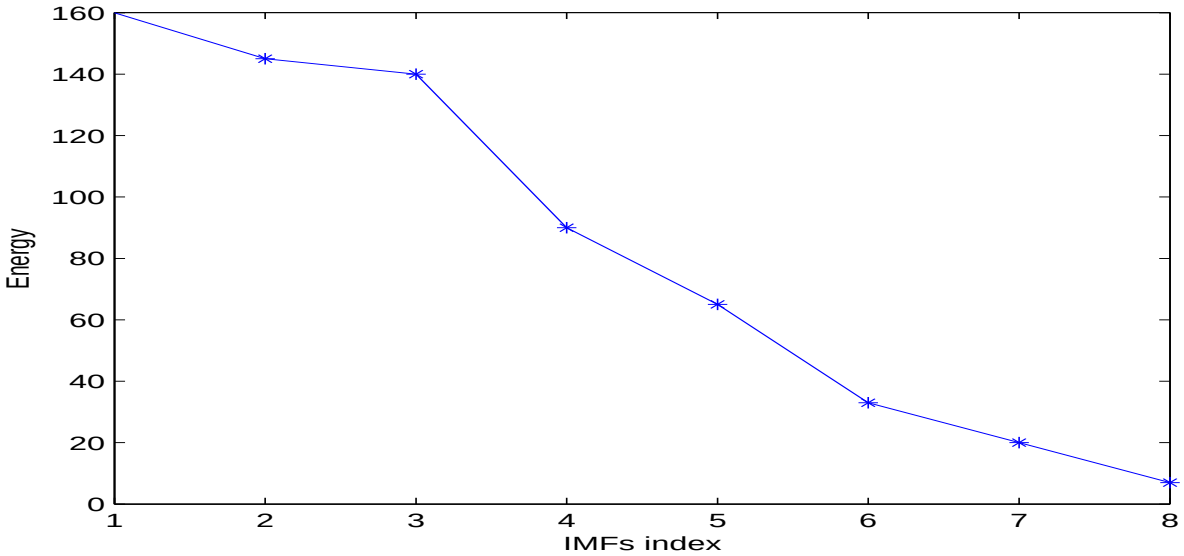
$NI_{k(k=1,2)}$ is the normalized version of the TFR of two sub-images I_k ($k = 1, 2$). In this work, the TFR used is the spectrogram which is of simple use (one parameter). A peak in the SI(n) variations indicates abrupt changing in the signal spectrum. Thus, it demonstrates the presence of transition zone. Indeed, the unvoiced frame is much more stationary than the transient one, the distinction between them can be performed using a SI index (Eq. III.9). Figure III.5 shows the behavior of the stationarity index in presence of a transient sequence (Fig. III.3(b)). Based on the stationarity index, it is possible to locate the time position of the transient that separates the transient speech frame into two sub-frames of different nature: voiced and unvoiced. A large peak is noticed at the transient instant corresponding to the



(a) Not voiced sequence.



(b) Noisy not voiced sequence.



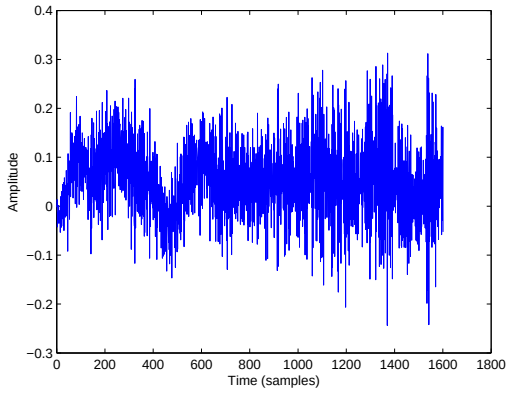
(c) Energy of noisy IMFs.

Figure III.3: Not voiced sequence, noisy not voiced sequence and variations of the energies of its noisy IMFs.

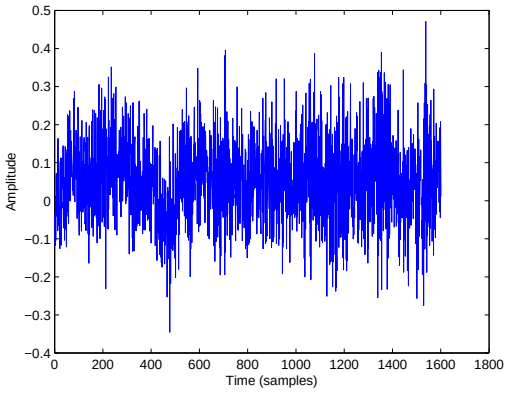
beginning of the second sub-frame. The location of the transient instant supplied by the stationarity index can be used to split the frame into sub-frames. Then, sub-frames can be classified using the energy criterion. As shown in figure III.6, the two sub-frames of different nature: the first one is voiced, while the second one is unvoiced.

III.3 Proposed speech denoising method

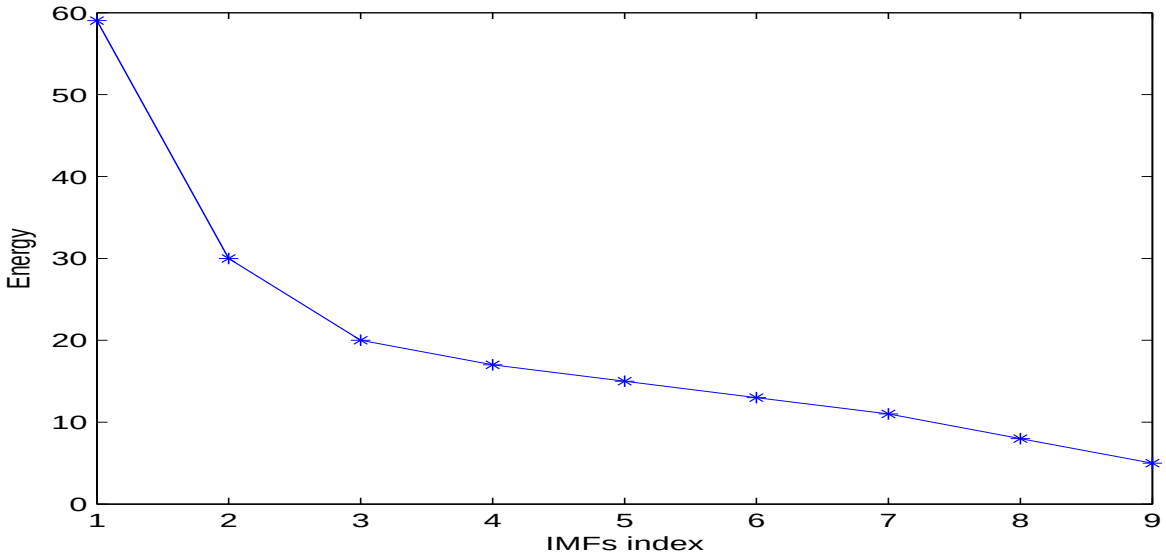
Basics of the proposed speech denoising technique are summarized as follows:



(a) Unvoiced sequence.



(b) Noisy unvoiced sequence.



(c) Energy of IMFs.

Figure III.4: Unvoiced sequence, noisy unvoiced sequence and variations of the energies of its IMFs.

1. Noisy speech signal is segmented into frames.

2. Each frame is decomposed into IMFs.

3. Detection of the frame class. For transient frames, a detection of the two sub-frames type is performed.

4. IMFs are filtered by ACWA filter depending on whether the frame or sub-frame is voiced or unvoiced.

5. The enhanced speech signal is reconstructed from denoised frames.

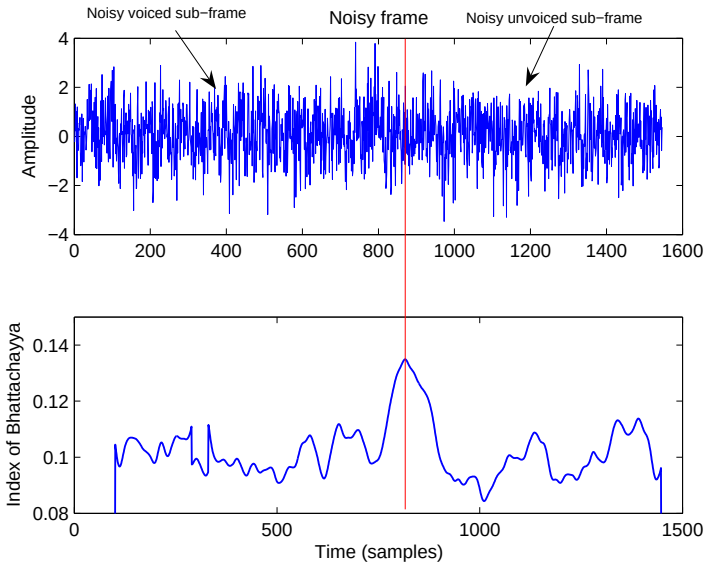
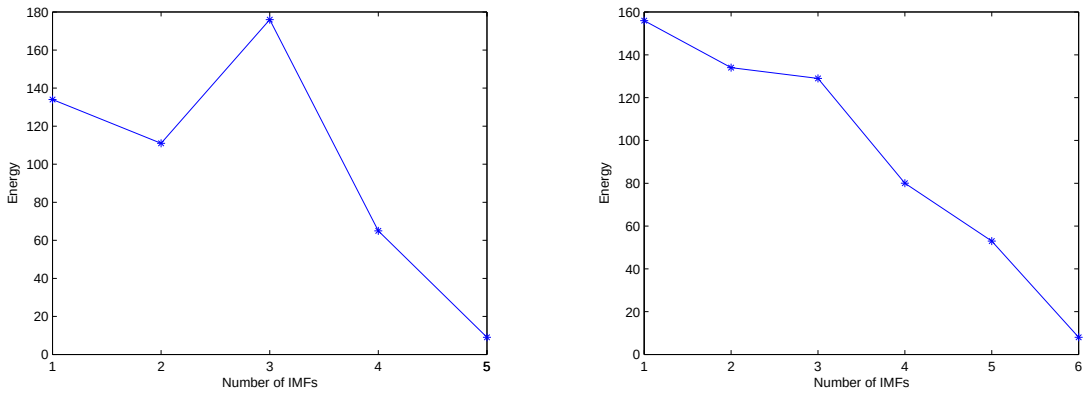


Figure III.5: The stationarity index of a noisy transient frame.



(a) First sub-frame. (b) Second sub-frame.

Figure III.6: Energy variations of the IMFs of the sub-frames.

III.3.1 Voiced sequence denoising

The denoising method dedicated to voiced frames consists in filtering a set of IMFs selected using the energy criterion (Eq. III.5) [8]. It is described in the four following steps :

Step A: Decompose $y(t)$ into j IMFs, $j \in \{1, ..., C\}$, and the residual $r_C(t)$.

Step B: Calculate the energy of each IMF and find the index j_s using equation III.6.

Step C: Denoise the shorter scale $(j_s - 1)$ IMFs with the according ACWA filter (Eq. II.15).

Step D: The denoised signal, $\tilde{x}(t)$, is reconstructed as follows:

$$\tilde{x}(t) = \sum_{j=1}^{j_s-1} \tilde{f}_j(t) + \sum_{j=j_s}^{C} \mathrm{IMF}_j(t) + r_C(t) \tag{III.10}$$

III.3.2 Unvoiced sequence denoising

For a noisy unvoiced speech frame, all the extracted IMFs are noisy. Consequently each $\mathrm{IMF}_j(t)$ must be filtered by ACWA filter. The estimated signal frame, $\tilde{x}(t)$, is given by:

$$\tilde{x}(t) = \sum_{j=1}^{C} \tilde{f}_j(t) + r_C(t). \tag{III.11}$$

III.3.3 Transient sequence denoising

A transient frame corresponds to two adjacent sub-frames of unvoiced (voiced) and voiced (unvoiced) speech. The stationarity index helps to locate the transient instant between these two sub-frames. The denoising strategy is chosen according to the sub-frame class : voiced or unvoiced.

III.4 Performance analysis

III.4.1 Voiced frames

The proposed noise reduction method is tested on voiced speech signals corrupted by varying additive white Gaussian noise levels, fixed through the input SNR. Four clean voiced speech signals vowels /o/, /a/, /e/ and /i/ (Fig. III.7) pronounced by a male speaker are analyzed.

These signals are corrupted by an additive white Gaussian noise with SNR values ranging from -10 dB to 10 dB. The results of the proposed scheme are compared to those of three methods: ACWA filtering of all IMFs (EMD-ACWA), denoising based on wavelet decomposition and ACWA filtering of the noisy voiced signal, i.e., ACWA filtering all IMFs. The performance evaluation is based on the output SNR and the PESQ measures. For each input SNR value, 100 independent noise realizations are generated and averaged values of the output SNR and the PESQ are computed. Noisy versions of the original signals corresponding to input SNR = 2 dB are shown

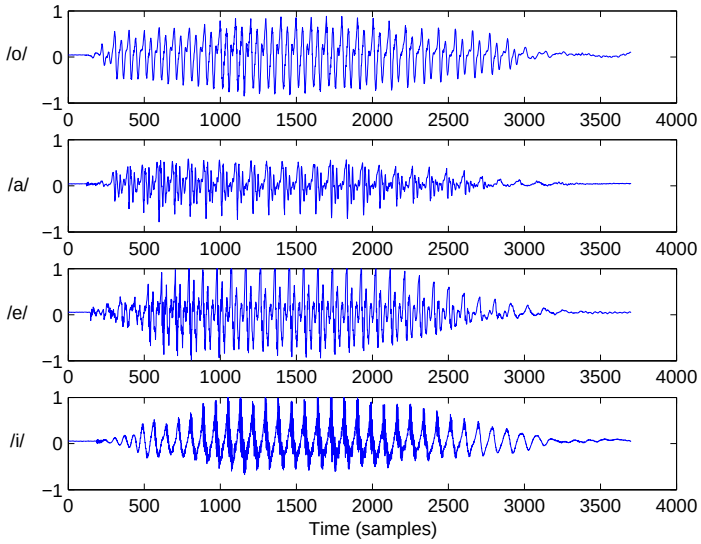


Figure III.7: Original signals /o/, /a/, /e/ and /i/.

in figure III.8. For illustration, figure III.9 shows that the EMD decomposes the

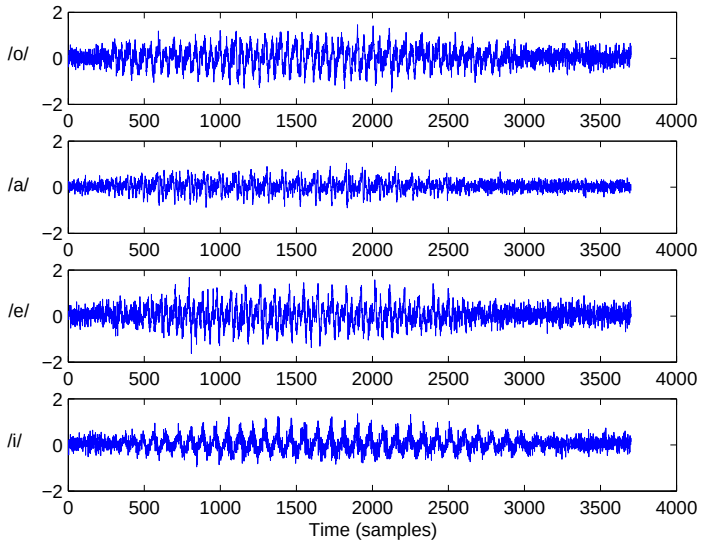


Figure III.8: Noisy versions of signals /o/, /a/, /e/ and /i/ (input SNR=2 dB).

noisy signal /o/ into 10 IMFs and a residual. According to this decomposition, we can see that from the fourth IMF, the signal energy dominates over the noise. This observation is well verified based on CMSE criterion.

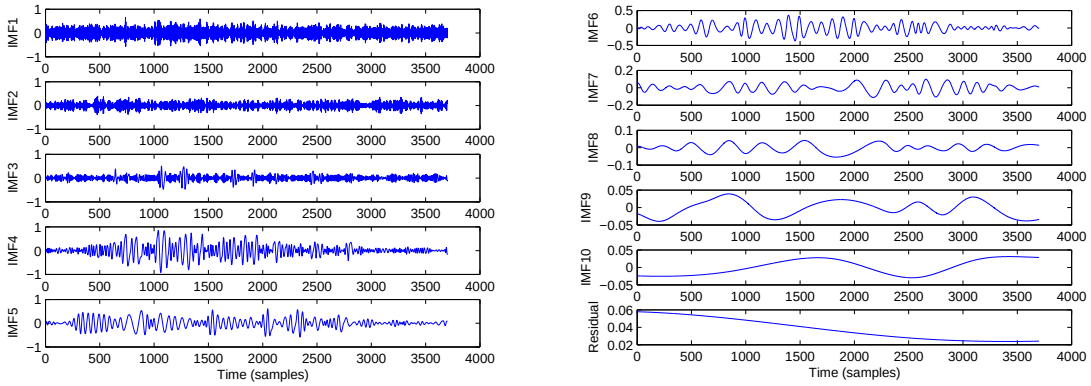


Figure III.9: Decomposition of noisy signal /o/ into IMFs (input SNR= 2dB)

Figure III.10 shows that for the sequence /o/, the maximum of CMSE corresponds to the fourth IMF. Figure III.10 shows the plots of the CMSE values versus the extracted IMFs index for the four signals. Each curve is characterized by only one maximum that defines the index j_s. Table III.1 summarizes for each signal, the

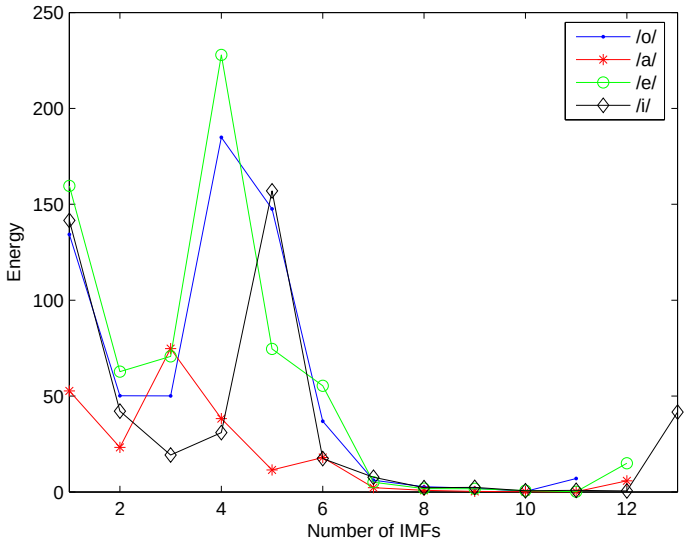


Figure III.10: Variations of CMSE (energy) values versus the number of IMFs for the four noisy signals.

number of extracted IMFs, and the index j_s corresponding to the largest CMSE or IMF energy. The second stage of the proposed method consists in filtering the $(j_s - 1)$ first IMFs using the ACWA filter. The size, L, of the sliding window of ACWA filter is set to 511. Such setting is justified by the results shown in figure II.13.

Denoising results obtained by the proposed method, the ACWA filtering of the

Table III.1: C and j_s values of each signal

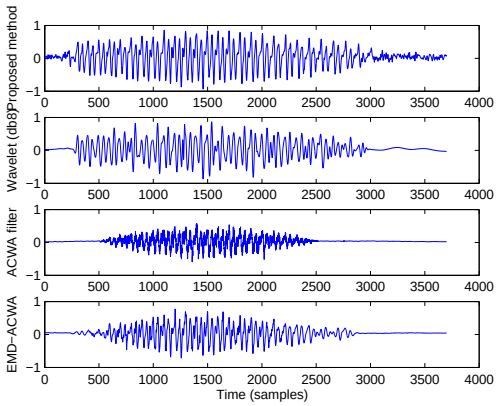
Signals	/o/	/a/	/e/	/i/
C	10	11	11	12
j_s	4	3	4	5

noisy signal, the ACWA filtering of all IMFs of the noisy signal (EMD-ACWA), and a denoising based on the wavelet (db4) thresholding [45], are shown in figure III.11 for an input SNR = 2 dB. In chapter II, we choose the db4 wavelet with a hard thresholding, because it gives good results compared to the other wavelets. A careful comparative examination between the signals shown in figures III.7 and III.11, shows that the proposed method performs better than the other three methods in terms of noise reduction. This conclusion is confirmed by the output SNR values listed in table III.2. For all voiced speech signals, the SNR gain achieved by the proposed method is the highest one.

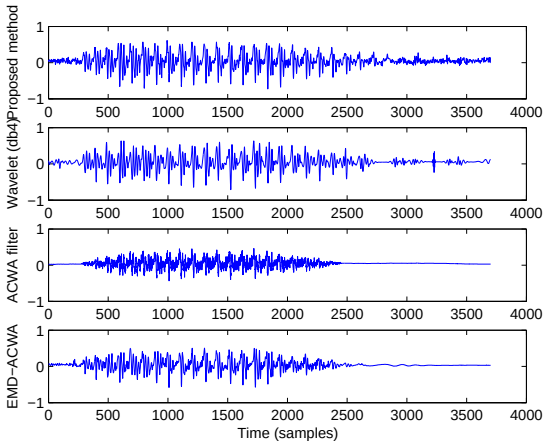
Table III.2: Denoising results, based on the output SNR, of four noisy voiced different signals (input SNR=2 dB)

Noisy signals (SNR=2dB)	/o/	/a/	/e/	/i/
Proposed method	14.82	11.87	10.55	9.44
EMD-ACWA	11.94	7.87	7.41	5.23
Wavelet (db4)	11.38	7.85	7.40	5.24
ACWA filter	9.80	8.04	7.91	7.31

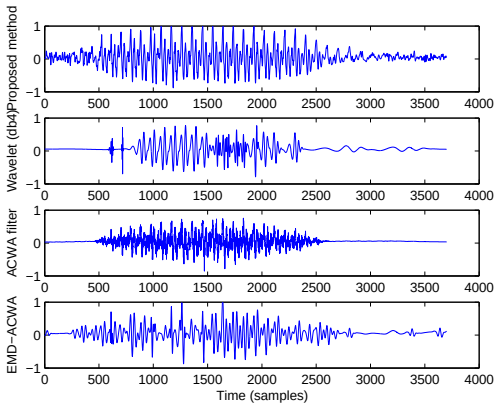
These findings are also confirmed by the results shown in figure III.12, where it is shown that for the four signals the proposed method performs remarkably better than the EMD-ACWA and the other methods. The SNR improvement achieved by the proposed method varies from 3.4 dB to 17.9 dB. For very low input SNR values, we still observe the effectiveness of the proposed method in removing the noise components.



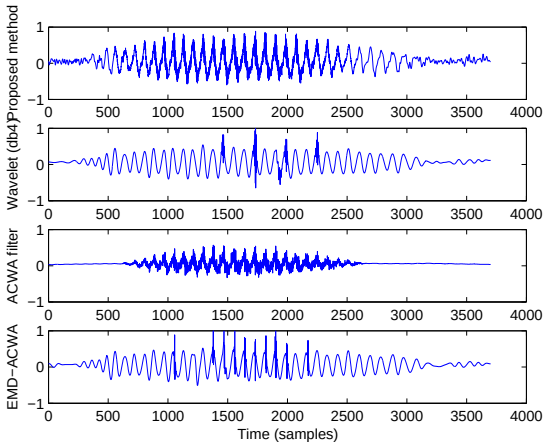
(a) Denoising of noisy signal /o/.



(b) Denoising of noisy signal /a/.



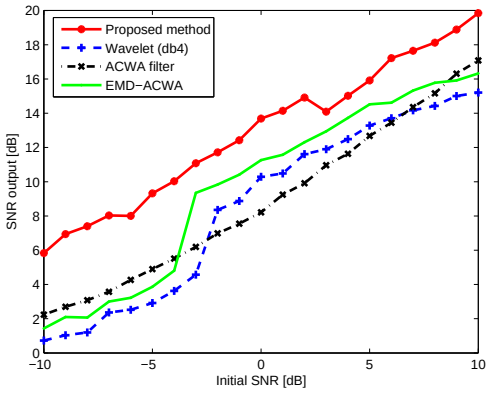
(c) Denoising of noisy signal /e/.



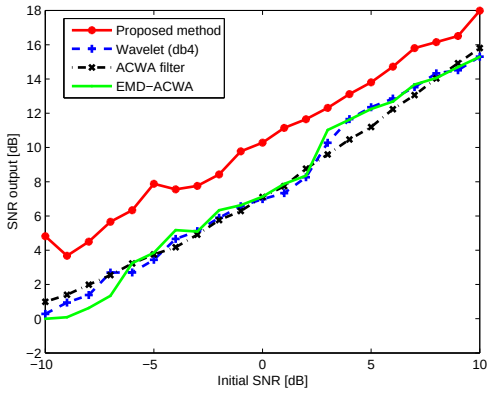
(d) Denoising of noisy signal /i/.

Figure III.11: Enhanced signals obtained by the proposed method, Wavelet (db4), ACWA filter and EMD-ACWA (input SNR=2 dB).

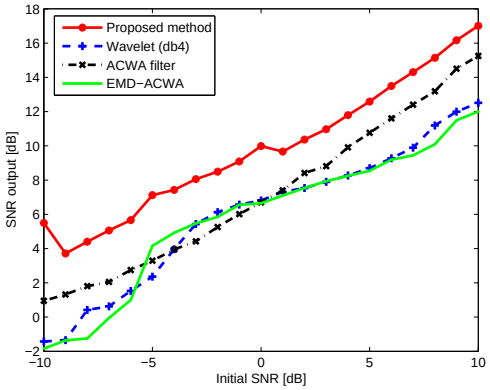
When listening to the enhanced speech signals, the proposed method produces lower residual noise and noticeably less speech distortion for all the signals. This result is confirmed by the PESQ results shown in figure III.13. These results demonstrate that our approach gives a significant enhancement in listening quality as the improvement of the PESQ values is high. Indeed, the obtained results also show that it is more efficient to apply the ACWA filter to selected IMFs of the noisy signal than to all the IMFS. These results are very logical, since the information of original signal is concentrated into last IMFs, consequently the filtering of all IMFs introduces some distortions in the denoised signal.



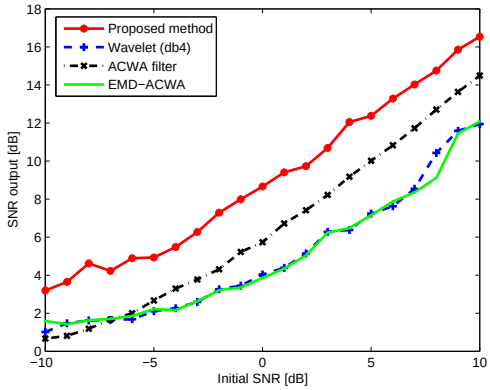
(a) Gain in SNR for noisy version of /o/.



(b) Gain in SNR for noisy version of /a/.



(c) Gain in SNR for noisy version of /e/.

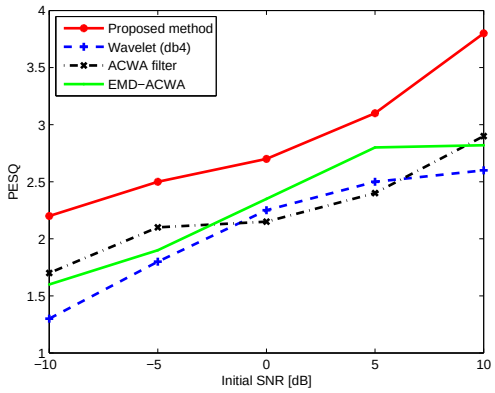


(d) Gain in SNR for noisy version of /i/.

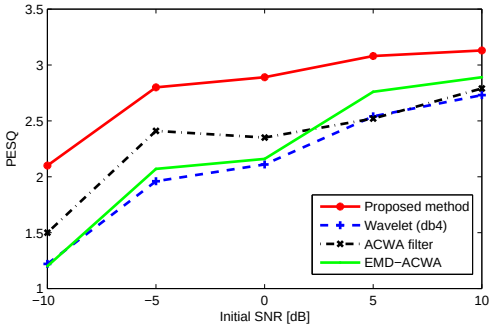
Figure III.12: Variations of output SNR versus input SNR for signals /o/, /a/, /e/ and /i/. The results are average over 100 noise realizations. The reported results correspond to the proposed method, Wavelet(db4), ACWA filter and the EMD-ACWA.

III.4.2 Speech signal

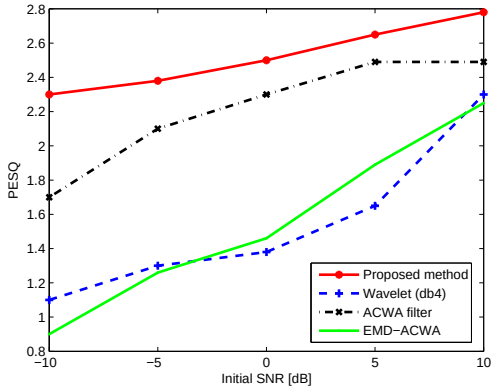
The proposed noise reduction methods are tested on speech signals corrupted by additive white Gaussian noise with different variances fixed through the input SNR. The performances of the proposed technique are compared to those of the following methods: ACWA filtering of all IMFs (EMD-ACWA), wavelet (db4) thresholding method [45], and ACWA filtering of the noisy signal. As objective criteria to evaluate the performance of the denoising method, we use the output SNR and PESQ. For our simulations, we consider four clean speech signals "speech1", "speech2", "speech3" and "speech4" (Figure III.14) corrupted by additive white Gaussian noise with input SNR values ranging from -10 dB to 10 dB. Noisy versions of the original signals corresponding to input SNR =2 dB are shown in figure III.15.



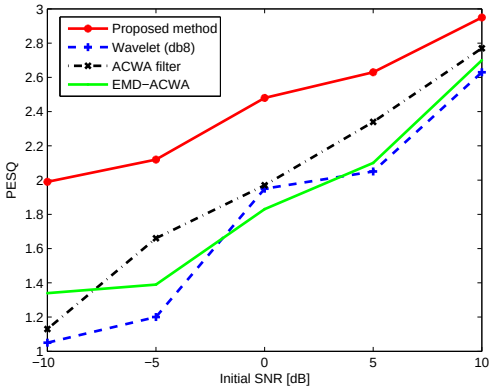
(a) PESQ for noisy version of /o/.



(b) PESQ for noisy version of /a/.



(c) PESQ for noisy version of /e/.



(d) PESQ for noisy version of /i/.

Figure III.13: Variations of PESQ values versus input SNR for the signals /o/, /a/, /e/ and /i/. The results are average over 100 noise realizations. The reported results correspond to the proposed method, wavelet(db4), ACWA filter and the EMD-ACWA.

Figure III.16 shows the denoising signals result obtained by applying the proposed method, the wavelet method, the ACWA filtering and the EMD-ACWA technique.

From figure III.16 and compared to the original signals (Fig. III.14), one can conclude that the proposed method performs better in terms of noise reduction than the other techniques. This fact is confirmed by the results shown in figure III.17 where the gain in output SNR achieved by the proposed method compared to other methods is presented. Indeed, we note that the proposed method provides an improvement of about 2 dB compared to the other methods for different considered signals "speech1", "speech2", "speech3" and "speech4". For deeper performance investigation, figure III.17 shows the variations of the output SNR versus the input SNR corresponding to the denoising of the speech signals: "speech1", "speech2", "speech3" and "speech4". For each input SNR value, averaged values are calculated over 100

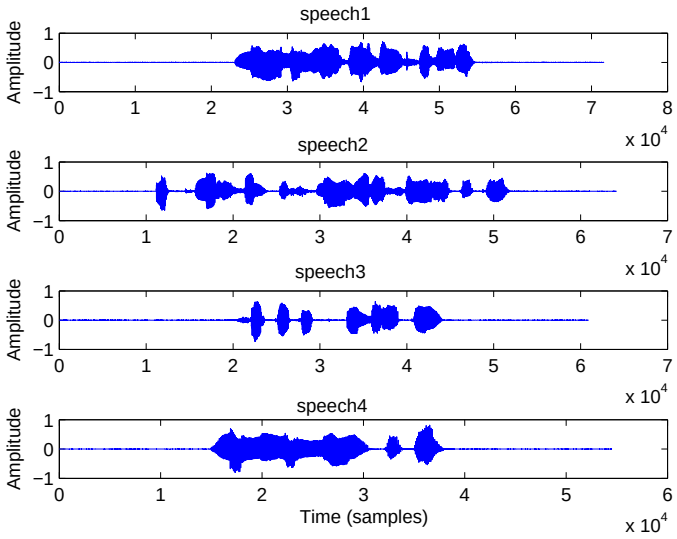


Figure III.14: Original signals "speech1", "speech2", "speech3" and "speech4".

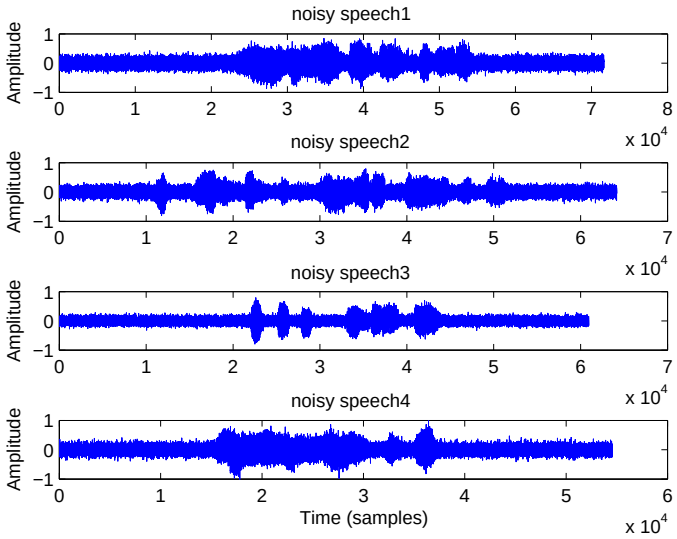
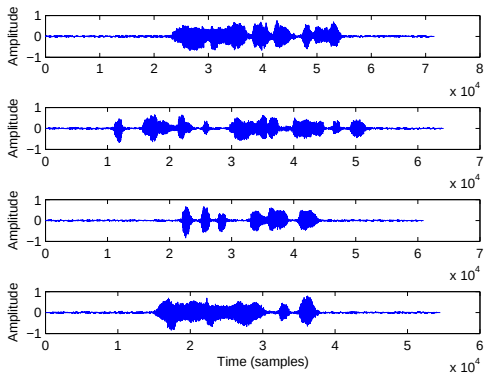
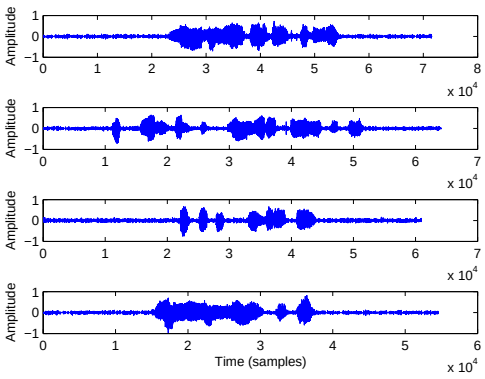


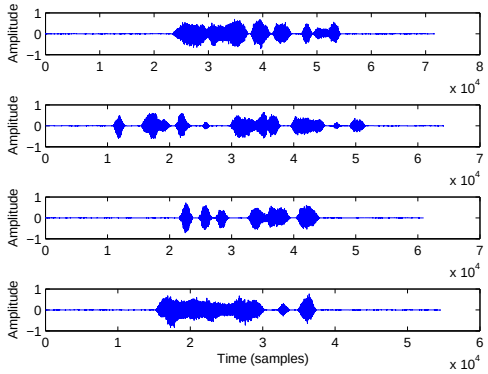
Figure III.15: Noisy version of signals "speech1", "speech2", "speech3" and "speech4" (input SNR = 2 dB).



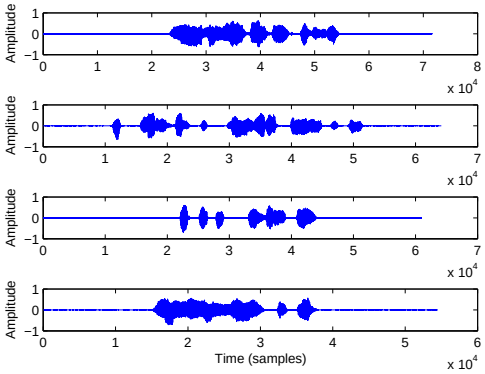
(a) Proposed method.



(b) Wavelet (db4).



(c) ACWA filter.



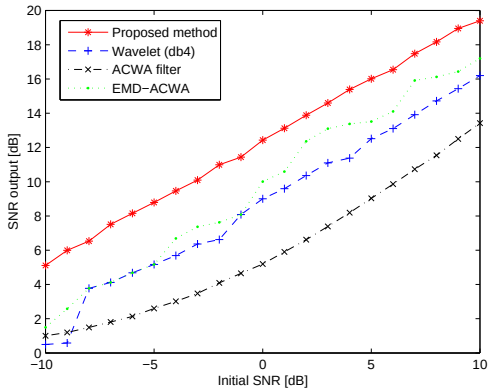
(d) EMD-ACWA.

Figure III.16: Denoising of noisy signals "speech1", "speech2", "speech3" and "speech4" (input SNR=2 dB) by the proposed method, Wavelet (db4), ACWA filter and EMD-ACWA.

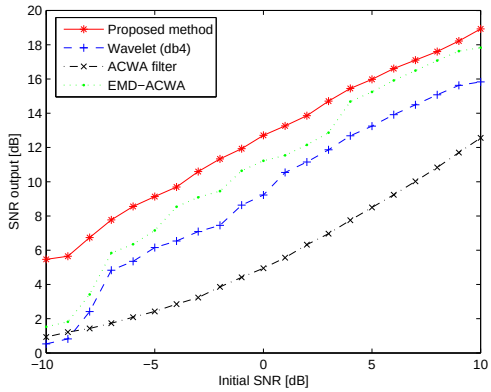
independent noise simulations.

These results demonstrate the effectiveness of the proposed method. Indeed, the output SNR values obtained by the proposed speech denoising technique are much higher than those obtained by the wavelet method, EMD-ACWA and the ACWA filtering. In particular, even for very low input SNR values, we can still observe the effectiveness of the proposed method in removing the noise components as the gain in SNR can go up to 15 dB. The PESQ measures reported by the figure III.18 also show that the proposed method offers much better speech quality than the other methods.

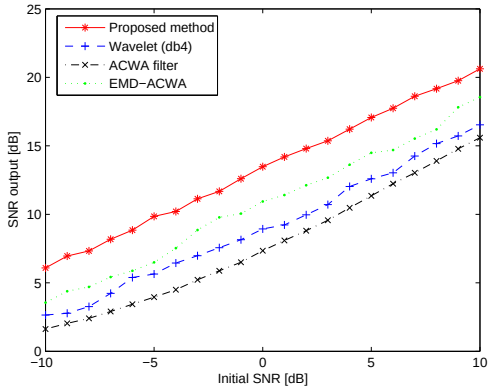
The analysis of the reported results shows the interest to take into account the frame class in the IMF filtering strategy. Indeed, the proposed method outperforms the EMD-ACWA technique where all the IMFs are filtered.



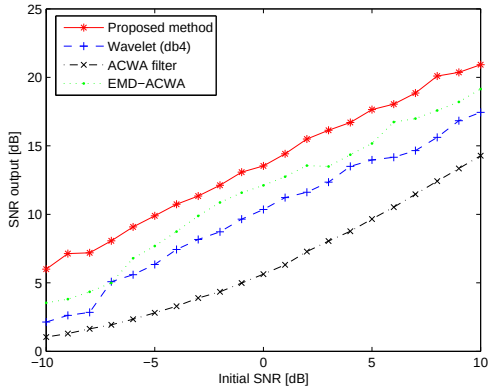
(a) Speech1 signal.



(b) Speech2 signal.



(c) Speech3 signal.

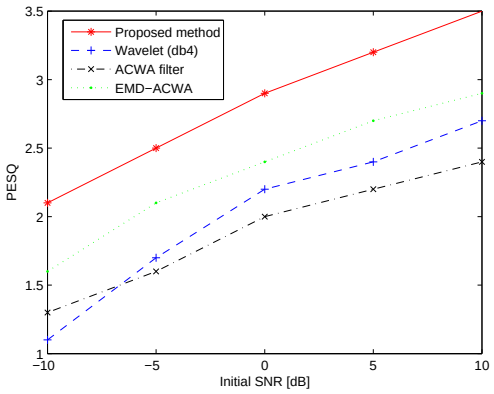


(d) Speech4 signal.

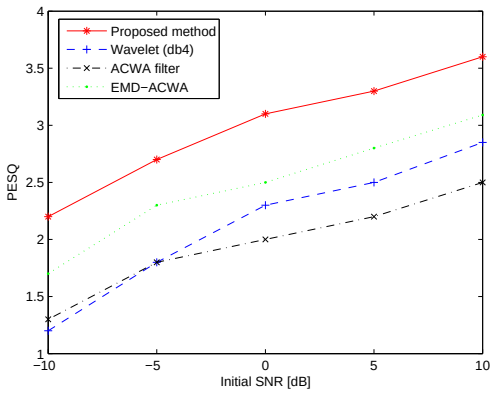
Figure III.17: Final SNR values obtained from different initial noise levels of signals "speech1", "speech2", "speech3" and "speech4". The results averages over 100 Monte Carlo simulations of the additive noise. It is reported for the proposed method, wavelet(db4), ACWA filter and the EMD-ACWA.

III.5 Conclusion

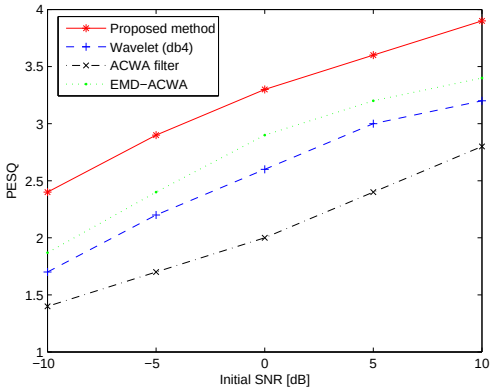
In this chapter, a new speech enhancement method that takes into account the frame class (voiced or unvoiced) is proposed. In fact, according to the frame class, a set of the IMFs of the noisy frame are filtered by the ACWA filter. Obtained results in the case of additive white Gaussian noise with varying SNR values show that the proposed method performs better than the ACWA filtering of all IMFs (EMD-ACWA), wavelet denoising approach (db4) and ACWA filtering of the noisy signal. Taking into account the frame class (voiced/unvoiced) in the filtering process, gives very interesting performance in terms of output SNR and PESQ.



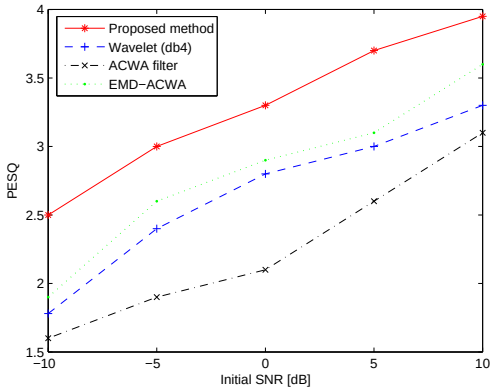
(a) Speech1 signal.



(b) Speech2 signal.



(c) Speech3 signal.



(d) Speech4 signal.

Figure III.18: PESQ values obtained from different initial noise levels of signals "speech1", "speech2","speech3" and "speech4". It is reported for proposed method, Wavelet(db4), ACWA filter and the EMD-ACWA.

CHAPTER IV

Signal coding schemes in EMD framework

Contents

IV.1 Introduction 87

IV.2 Why IMFs coding? 87

IV.2.1 IMF extrema 87

IV.2.2 Quasi-symmetry of IMF 88

IV.2.3 IMF modelling 89

IV.3 EMD based encoder architecture 90

IV.3.1 IMF extrema coding basics: $IMF_{extrema}$ 90

IV.3.2 IMF envelope coding basics : $IMF_{envelope}$ 92

IV.4 HHT based encoder architecture 94

IV.4.1 IA and IP coding basics: $IA - IP$ 94

IV.4.2 IA and IF coding basics: $IA - IF$ 95

IV.5 conclusion 96

This chapter introduces different signal coding approaches based on EMD. In the first approach, we were interested in coding the IMFs extrema since the IMFs are fully described by their local extrema [34]. In order to achieve lower BR, the coding of one out of the two IMFs envelopes was considered. This idea relies on the quasi-symmetrical property of the IMF. In a second approach, we exploit the auto-correlation of the Instantaneous Amplitude (IA) and

Instantaneous Frequency (IF) of the IMFs of the signals to be encoded. Thus a parametric coding based on linear prediction has been adopted to encode IA and IF components.

IV.1 Introduction

In this chapter we present the encoders architecture. The first encoder consists in encoding the IMFs extrema, since the IMFs are fully described by their local extrema [34]. To further reduce the Bit Rate (BR), one out of two IMFs envelops is coded. This is motivated by the quasi-symmetrical property of the IMF. In the second architecture, a parametric coding approach based on the EMD in association with Hilbert transform is presented. Based on the Hilbert and Huang Transforms (HHT), Instantaneous Amplitude (IA), Instantaneous Phase (IP) and Instantaneous Frequency (IF) of each extracted IMF are calculated. Given the relatively high autocorrelation of the IA and IF values, a linear predictive coding technique of IA and IF is used

This chapter is organized as follows. In the first section, we present the motivation of IMFs coding. Section IV.3.2.2 details the first encoder architecture that consists to encode the IMFs extrema or one of its envelopes. Finally Section IV.4 focuses on the second encoder architecture witch revolves around decoding of IA and IF of each IMF of the signal.

IV.2 Why IMFs coding?

Two main properties of the IMFs are exploited for coding purpose.

IV.2.1 IMF extrema

As earlier recalled, the IMFs are zero mean and have oscillating shape properties. With a view to compression, these are interesting features. Indeed, most relevant information of the IMF can be represented by its extrema [34]. Roughly, this amounts to sampling the IMF almost regularly at twice its original frequency. Figure IV.1 shows the plots of an IMF and its approximate obtained by spline interpolation of the extrema. A comparative examination of the true IMF and its estimate shows the effectiveness of the spline interpolation for the reconstruction of the IMF from its extrema. Indeed, we notice that the error corresponding to the difference between the true and the reconstructed IMF is negligible. So, the idea of encoding the IMFs extrema seems interesting and will be more advantageous in terms of reduction of coding rate compared to waveform coding.

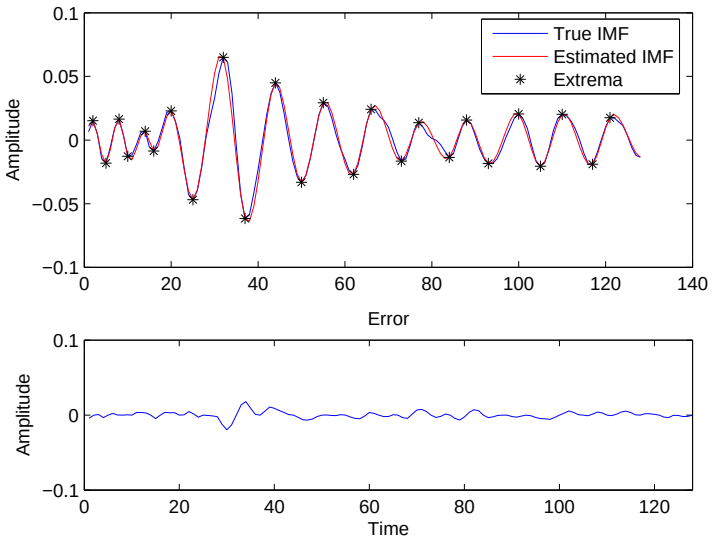


Figure IV.1: Original IMF and its estimated version by spline interpolation.

IV.2.2 Quasi-symmetry of IMF

The aim of the sifting process is to remove the dissymmetry between the upper (maxima) and lower (minima) envelopes in order to transform the original signal into an amplitude modulated signal. So exploiting the symmetry of the upper and lower envelopes, it is possible to encode only a single envelope and as a result reduce the coding rate while ensuring good quality of the encoded signal. However, extracted IMFs are, in general, not truly symmetric with respect to the time axis ($\alpha = 0$) but they are symmetric about a parallel line $y = \alpha$. This problem is illustrated by figure IV.2 where the envelopes are symmetrical with respect to line $y = 0.05$. An example of offset values obtained for five IMFs extracted from an audio frame signal is presented in table IV.1. As expected, IMFs are not all symmetric with respect to $y = 0$.

Table IV.1: Offset values of IMFs extracted from an audio frame.

IMF	1	2	3	4	5
α	0.05	0.02	0	0	0.006

Thus, provided the offset α is encoded, at the decoder the upper (lower) envelope is reconstructed and then the lower (upper) envelope is deduced by symmetry about the line $y = \alpha$.

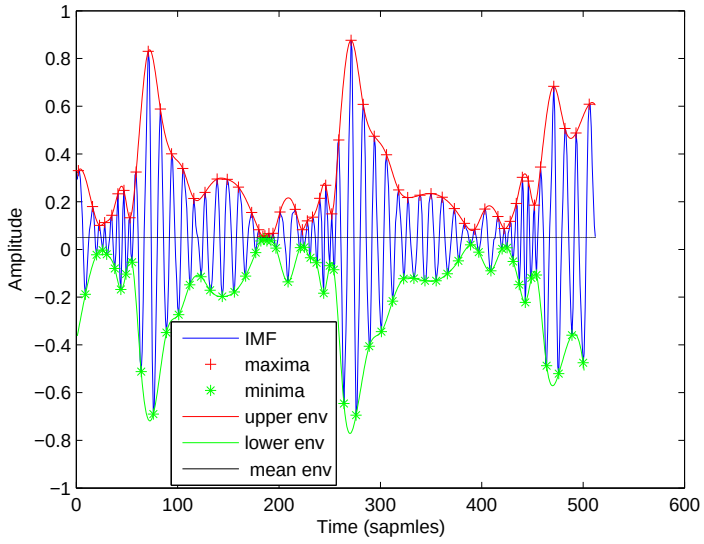


Figure IV.2: IMF mean envelope offset

IV.2.3 IMF modelling

We have shown that any signal can be decomposed, using EMD, into a finite number of IMFs. These oscillating components are centered modes and AM-FM type. Using Hilbert transform, $\mathcal{H}[.]$, the analytic signal $z(t)$ corresponding to $IMF(t)$ is given by :

$$z(t) = IMF(t) + i\mathcal{H}[IMF(t)] \tag{IV.1}$$

where the signal $IMF(t)$ is the real part of Eq. (IV.1), and the imaginary part is the Hilbert transform of $IMF(t)$,

$$\mathcal{H}[IMF(t)] = \frac{1}{\pi}\mathrm{PV}\int_{-\infty}^{+\infty}\frac{IMF(\tau)}{t-\tau}d\tau \tag{IV.2}$$

where PV is the Cauchy principal value of the integral. In the complex plane, the analytic signal z(t) can be written as follows,

$$z(t) = a(t)e^{i\theta(t)}, \tag{IV.3}$$

where $a(t)= \sqrt{[IMF(t)]^2 + \mathcal{H}[IMF(t)]^2}$ is the IA and $\theta(t) = \tan^{-1}\left(\frac{\mathcal{H}[IMF(t)]}{IMF(t)}\right)$ corresponds to the IP. Recall that $f(t)= \frac{1}{2\pi}\frac{d\theta(t)}{dt}$ is the IF.

Figure IV.3 shows the time variations of IA, IP, and IF of an IMF. EMD roughly implements filter bank decomposition [28] and its IMFs are oscillatory type. Thus,

IMFs are strongly correlated. Consequently, IA and IF values are very strongly correlated, while the IP values are slowly varying. The basic idea consists to encode the IA and IF by linear prediction, or IP by scalar quantization [58],[19].

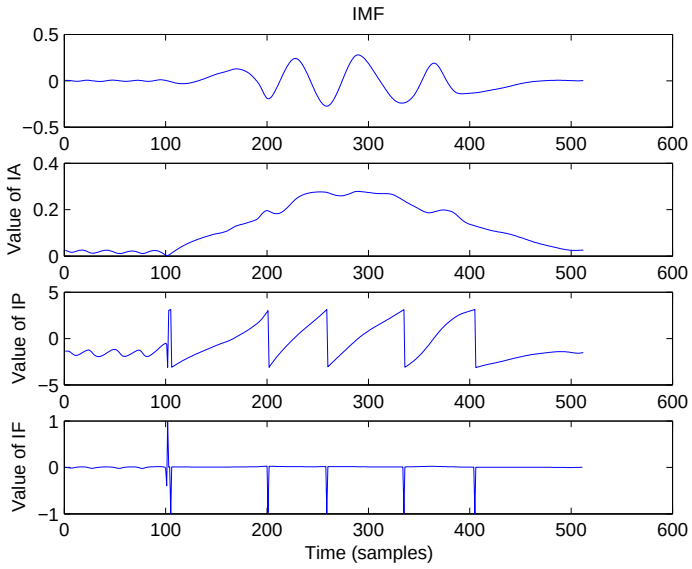


Figure IV.3: IA, IP and IF of an IMF.

IV.3 EMD based encoder architecture

IV.3.1 IMF extrema coding basics: $IMF_{extrema}$

The proposed coding scheme is shown in figure IV.4.

IV.3.1.1 Segmentation and decomposition

The first step consists in a segmentation of the signal into frames and each one is decomposed into IMFs and a residual. These IMFs are completely represented by their extrema $(E_{i,N_i})_{i=1,C}$. Each extrema is characterized by a time position and an amplitude.

IV.3.1.2 Extrema thresholding

The number of extrema for each IMF is reduced by using an appropriate threshold fixed according to the signal type. For example in the case of audio signal, the

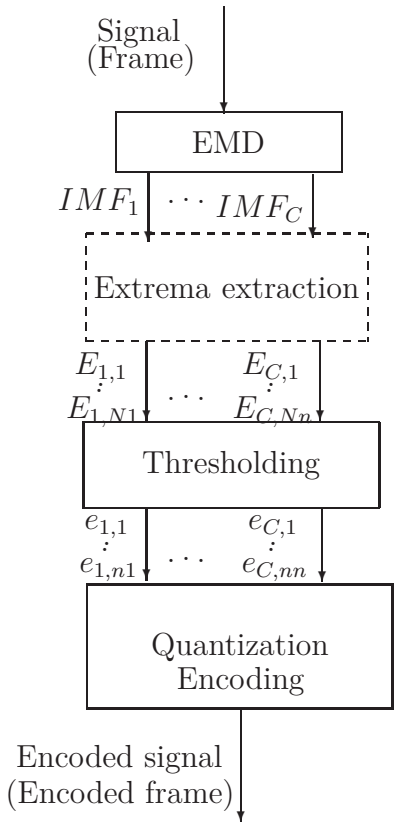


Figure IV.4: Encoding scheme.

threshold can be fixed using the psychoacoustic model which corresponds the behavior of the human ear. In coding audio application [4], the threshold is chosen depending to the fixed compression ratio. The objective of this step is to diminish the number of extrema to be coded, in order to reduce the BR.

IV.3.1.3 Extrema quantification

The extrema amplitudes of each IMF are scaled by their maximum of value. We quantize the positions of the extrema, the scaling factor and the scaled extrema amplitudes. The extrema's positions are quantified, in a fixed way, by a scalar quantization. We note that the number of extrema, $(e_{i,n_i})_{i=1,C}$, selected for coding decreases from one IMF to the next. Consequently, in order to optimize the BR, the number of bits allocated to the quantization extrema's amplitudes must vary from an IMF to another while ensuring a minimum quantization error of the IMF.

IV.3.1.4 Coding

Better performance can be achieved by using lossless compression such as Huffman or Lempel-Ziv encoding techniques. These techniques account for probability of occurrence of encoded data to reduce the number of bits allocated to. Although Lempel-Ziv is not optimum, the decoder does not need to know the encoding dictionary [83].

IV.3.1.5 Decoding process

Firstly, we begin by decoding the extrema positions. Then we decode the extrema amplitude. Finally, the IMFs are recovered thanks to a spline interpolation among the extrema [46],[47], and the sum of IMFs yields the original signal [34].

IV.3.2 IMF envelope coding basics : $IMF_{envelope}$

Based on quasi-symmetry property of IMF, and in order to reduce further the BR, we propose to encode one out of the two envelopes of each IMF. In this approach, we focus on the coding of the upper (maxima) envelope.

IV.3.2.1 Encoding scheme

The block diagram of the proposed encoding scheme $IMF_{envelope}$ is presented in figure IV.5. The signal is segmented into frames. The windowed signal frame is decomposed into IMFs and a residual. These modes are encoded under the two following constraints.

- BR: the number of bits used to encode maxima must be as small as possible.
- Encoding noise: the difference between the true IMF and the reconstructed one must be negligible.

Each maxima is presented by both position index and amplitude. The quantization of the maxima amplitudes is performed as in presented in section IV.3.1.3. Finally the maxima positions, the scaling factors and the offset values are also encoded. The different steps of the proposed approach are summarized as follows:

- Divide the original signal into frames.

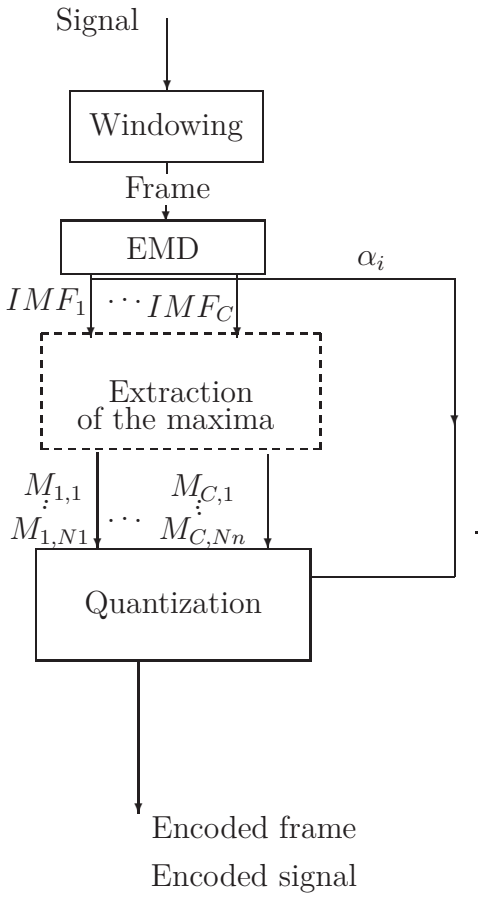


Figure IV.5: Encoding scheme.

- Using EMD, extract the j$^{\text{th}}$ IMF, $j \in \{1, ..., C\}$, and the associated residual $r_C(t)$.
- For the j$^{\text{th}}$ IMF, determine all the maxima and the offset α_j.
- Quantize and encode maxima positions, maxima amplitudes value, scaling factors and offsets ($\alpha_j æ = 1, ..., C$).

IV.3.2.2 Decoding process

For each IMF, the decoder first recovers the upper (lower) envelope using the corresponding scaling factor and the encoded maxima (positions and amplitudes). Then the lower (upper) envelope of the IMF is determined from the upper (lower) envelope by symmetry using the corresponding decoded offset value. Finally, the IMFs are recovered thanks to a spline interpolation between the extrema [46],[47], and the sum of IMFs yields the original signal [34].

IV.4 HHT based encoder architecture

The signal is segmented into frames. Then, using EMD each frame is decomposed into sum of IMFs. For each IMF, IA $a(t)$, IF $f(t)$ and IP $\theta(t)$ using Hilbert transform are calculated.

IV.4.1 IA and IP coding basics: $IA - IP$

IV.4.1.1 IA encoding

As mentioned before, IA $a(t)$ values are strongly correlated. So, Auto Regressive (AR) model can be used to exploit efficiently this temporally correlated information.

$$a(t) = \sum_{k=1}^{p} c(k)a(t-k) + \epsilon(t) \tag{IV.4}$$

where $[c(1), c(2), ..., c(p)]$ are the coefficients of the AR model and $\epsilon(t)$ is assumed to be a white noise process. The determination of the coefficients is based on the minimization of the prediction Mean Square Error (MSE). The order, p, of AR model for IA coding is fixed depending to the signal type. For example, in [48], the order is fixed at 9 for a speech signal. In the proposed approach, at encoder we encode the coefficients and the noise variance.

IV.4.1.2 IP encoding

The analysis of IP variations shows that only IP's extrema can be encoded by classical scalar quantization. As shown in figure IV.3, the phase variations is almost linear as it undergoes variation of 2π. This suggests encoding only zero crossings of the phase, together with its initial and final values.

IV.4.1.3 Decoding scheme

IP $\theta(t)$ is decoded from zero crossing by linear interpolation. IA $a(t)$ is recovered by linear prediction. The estimated IMF is calculated as follows:

$$\hat{\mathrm{IMF}}(t) = |\hat{a}(t)| \cos(\hat{\theta}(t)) \tag{IV.5}$$

The signal frame is constructed from estimated IMFs summation and the decoded signal is obtained by frames concatenation.

IV.4.2 IA and IF coding basics: $IA - IF$

The principle of the proposed approach consists in encoding IA and IF by linear prediction. The encoding of IF instead of IP allows a decrease of the BR without increasing of decoding error.

IV.4.2.1 IF encoding

The IF encoding is done in the same way as the coding of IA. But, the only change is at the choice of the order of the model. Since each IMF contains lower frequency oscillations than each previously extracted ones, the order of the AR model for IF is varies from one IMF to another. Therefore, for each IMF we determine the order of the AR model of the IF. The determination of the order of the AR model is based on the estimation of the partial autocorrelation coefficients that fits the variations of the corresponding IF. As illustration example, let consider an audio frame. This last is decomposed into IMFs and residual by EMD (Fig. IV.6). The

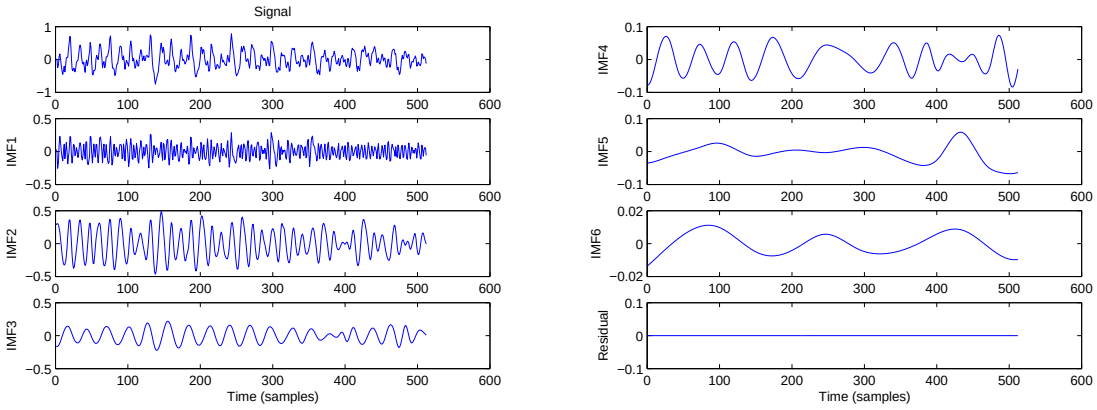


Figure IV.6: Decomposition of an audio frame by EMD.

partial autocorrelation coefficients corresponding to the IF of IMF (Fig. IV.6) are shown in figure IV.7.

Table IV.2 resumes the orders of the AR models for the IFs. The order of AR model is determined according to the plotted of Partial autocorrelation coefficient for IF, that is constant. The transmitted information corresponds to the coefficients and the variances of the excitation of the AR models of the IA and IF of the IMFs. The

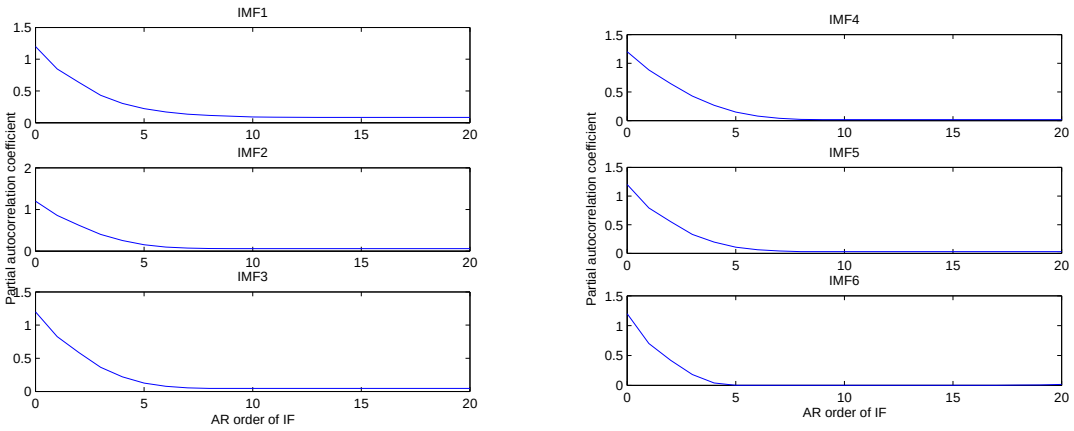


Figure IV.7: Partial autocorrelation coefficient for IF of IMF generated by audio frame (figure IV.6).

Table IV.2: Order of AR model for IF of IMFs (figure IV.7).

IMF	1	2	3	4	5	6
Order of AR model	11	7	8	7	8	4

value from which the partial autocorrelation curve is constant, is identified as the order for IF modeling (Fig. IV.7) Finally, we encode the coefficients, variances, and order of the model.

IV.4.2.2 Decoding approach

IA $a(t)$ and IF $f(t)$ are decoded using linear prediction. The estimated IMF for the IF coding approach (HHT_{IF}) is calculated as follows:

$$\hat{\mathtt{IMF}}(t) = |\hat{a}(t)| \cos(\int 2\pi \hat{f}(t)dt) \tag{IV.6}$$

The signal frame is constructed from estimated IMFs summation, and the decoded signal is obtained by frames concatenation.

IV.5 conclusion

In this chapter coding approaches, based on the EMD, are presented. The properties of IMF allow the introduction of different coding schemes which are applicable for

wide range of signals. In particular, for both techniques $IMF_{extrema}$ and $IMF_{envelope}$, the bit allocation is done in accordance with a constraint that depends on the nature of the signal. In the case of audio signals, this constraint is the audibility. In the next chapter we illustrate the proposed schemes on audio signals.

CHAPTER V

Encoding schemes: Application to audio signals

Contents

V.1 Introduction 100

V.2 Encoders architecture 100

V.2.1 Transient detection 100

V.2.2 Thresholding step for $IMF_{extrema}$ coder 103

V.2.3 Quantization step for $IMF_{envelope}$ and $IMF_{extrema}$ 104

V.3 EMD based audio coders performance 105

V.4 Conclusion 108

In this chapter, we illustrate the developed coding schemes based on EMD on audio signals. We show the interest to introduce psychoacoustic model and transient sequences detector to guarantee good performances of the proposed codings. The performance of the proposed methods are analyzed and compared to the MPEG1 layer3 known as MP3 and AAC codecs, and to the wavelet based compression.

V.1 Introduction

In this chapter we show the effectiveness of introduced EMD-based coding strategies on audio signals. For $IMF_{extrema}$ and $IMF_{envelope}$ coders, a masking threshold related to the psychoacoustic model is used. To guarantee good performances, specificity of the transient sequences is taken into account in the coding process. Also, to ensure a good audio quality and a reduced BR, the encoders based on the IMF waveform coding, are slightly modified according to the specificity of the audio signal. Thus the structure of the $IMF_{extrema}$ coder is depicted by Fig.∨.1.
This chapter is organized as follows. In the second section, we present the encoders architecture for an audio signal, consequently we detail the segmentation step, thresholding procedure and the quantization step for $IMF_{extrema}$ and also $IMF_{envelope}$ approaches. Finally, Section V.3 presents the performance of the audio encoding approaches, based on exhaustive simulation results.

V.2 Encoders architecture

V.2.1 Transient detection

We have shown that the first step of the proposed coders ($IMF_{extrema}$,...) is to divide the signal into frames. Indeed, insofar as all the approaches compute parameters depending on the signal statistics, these frames must be stationary. To guarantee the statistics invariance of each frame, a non parametric detector [30] is used to test this stationarity. Thus, when a transient is detected, the frame is divided into two sub-frames. The detection of transient sequence is based on the Local Entropic Criterion (LEC) which is a non parametric detector. The LEC of signal $x(t)$ is given by [30]:

$$\mathrm{LEC}_x(t) = \frac{\mathrm{E}_{xc}(t) - [\mathrm{E}_{xl}(t) + \mathrm{E}_{xr}(t)]}{|\mathrm{E}_{xc}(t)|} \tag{V.1}$$

where $\mathrm{E}_{xc}(t)$, $\mathrm{E}_{xl}(t)$ and$\mathrm{E}_{xr}(t)$ denote the Shannon entropies of the principal window and of the left and right sub-windows respectively.
$\mathrm{E}_{xc}(t) = \mathrm{E}_{x[t-\frac{N}{2},t+\frac{N}{2}-1]}$,
$\mathrm{E}_{xl}(t) = \mathrm{E}_{x[t-\frac{N}{2},t-1]}$,
$\mathrm{E}_{xr}(t) = \mathrm{E}_{x[t,t+\frac{N}{2}-1]}$.

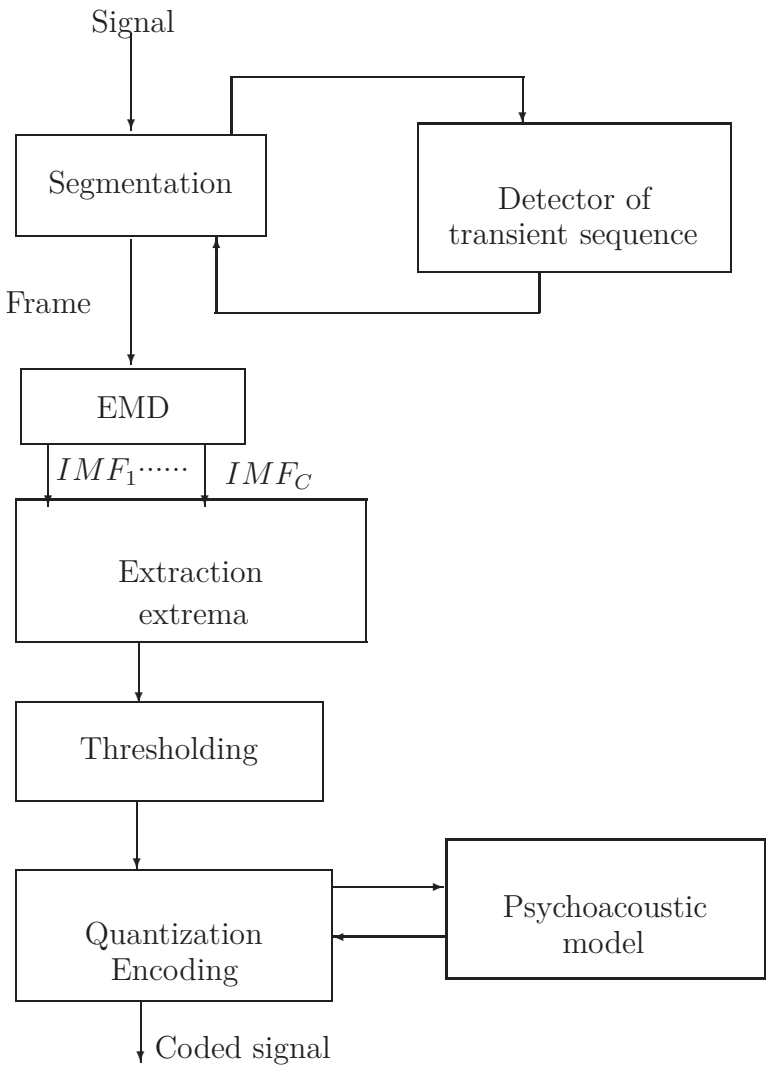


Figure V.1: $IMF_{extrema}$ encoder architecture in context of audio signals.

The Shannon entropy of a signal $x(t)$ in the interval $[0, N-1]$, $\mathrm{E}_{x[0,N-1]}$, is defined by :

$$\mathrm{E}_{x[0,N-1]} = -\sum_{k=0}^{N-1} |X(k)|^2 \log |X(k)|^2 \tag{V.2}$$

with $X(k)$ the discrete FT of $x(t)$. Thus, the LEC takes its values in the range of -1 to 1. A transient in the signal that occurs at time t is characterized by a LEC value which is greater to 0. An example of LEC variations for an audio frame is shown in figure V.2, with N set to 64 [30]. Figure V.3 shows an example of segmentation of an audio frame of 1500 samples (zoom for audio frame signal figure V.2).

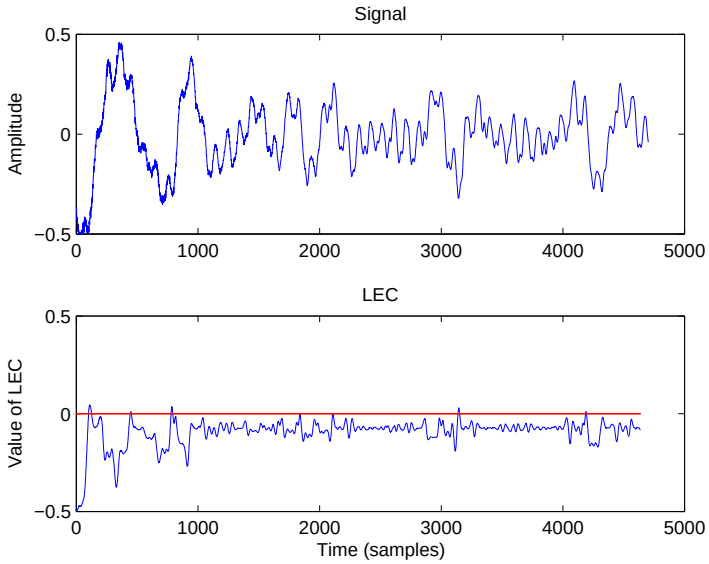


Figure V.2: LEC variation for an audio frame.

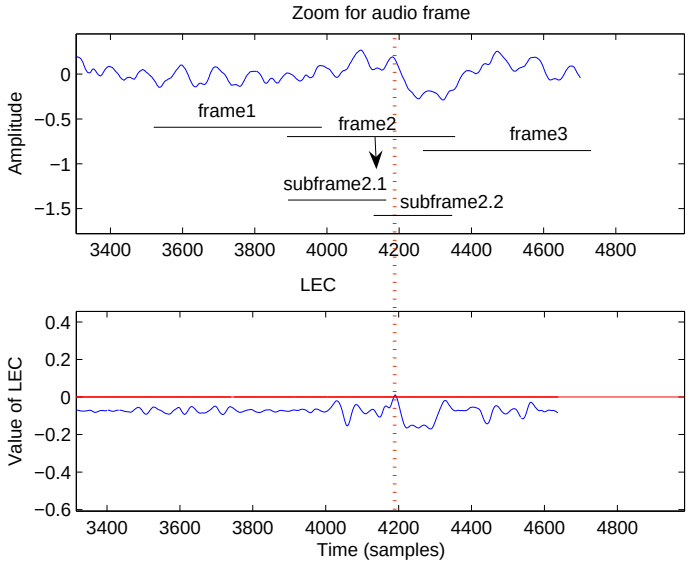


Figure V.3: Example of segmentation for an audio frame.

V.2.2 Thresholding step for $IMF_{extrema}$ coder

To decrease the BR, we have shown that in the second step of $IMF_{extrema}$ coding the extrema must be thresholded. Further, the error between the estimated IMF, from the selected extrema, and the true IMF must respect some constraint. For audio coding, this constraint relies on the masking threshold. In fact the number of extrema of each IMF is reduced while ensuring that the PSD of IMF's estimating error remains below the masking curve of the IMF. This reduction of the number of extrema controlled by the masking curve provides significant compression gain while maintains a good listening quality. Clearly, the thresholding procedure is an iterative process: aiming at estimating an IMF from a reduced number of extrema, while ensuring the inaudibility of the reconstruction error. Since EMD behaves as a wavelet decomposition [28], the threshold parameter is very soon given by a standard wavelet coefficients thresholding procedure [7]. The following e xpression gives for an IMF the initial value of the threshold ($\tau_{j,0}$) [9],[4],[62]:

$$\tau_{j,0} = \begin{cases} 0.05 \ \max \ |\mathrm{IMF}_j(t)|, & \text{if } \tilde{\sigma_j} = 0 \\ \tilde{\sigma_j}, & \text{else} \end{cases} \tag{V.3}$$

where $\tilde{\sigma_j}$ is given by [7]:

$$\tilde{\sigma_j} = \mathrm{Median}\left\{|\mathrm{IMF}_j(t) - \mathrm{Median}\left\{\mathrm{IMF}_j(t)\right\}|\right\}. \tag{V.4}$$

Although there are different non linear thresholding functions [55], in the present work, hard thresholding is used:

$$e_j = \begin{cases} \mathrm{E}_j, & \text{if } |\mathrm{E}_j| > \tau_j \\ 0, & \text{if } |\mathrm{E}_j| \leq \tau_j, \end{cases} \tag{V.5}$$

where e_j et E_j correspond respectively to the thresholded and the initial extrema values. To confirm the efficiency of the initial value of the threshold (Eq. V.3), the estimated IMF is reconstructed from non zero thresholded extrema by using spline interpolation. If the error's DSP is under the masking curve of the IMF[63], we iterate the thresholding procedure by reducing the threshold value, as follows [4],[61].

$$\tau_{j,i} = \frac{\tau_{j,i-1}}{2}, \tag{V.6}$$

where $\tau_{j,i}$ is the threshold parameter of the IMF_j at the iteration number i ($i \geq 1$).

V.2.3 Quantization step for $IMF_{envelope}$ and $IMF_{extrema}$

In order to reduce the BR, a perceptual coding controlled by the psychoacoustic model [63] is used to encode the scaled maxima (or extrema) amplitudes. Initially, the number of the allocated bits is fixed according to the coding BR. However, the number of bits allocated to each IMF is adjusted in order to ensure that the PSD of the quantization error of the IMF is below its masking curve [46]. We start by allocating the same number of bits to all IMFs maxima (or extrema) amplitude. Since each IMF contains lower frequency oscillations than each previously extracted ones, we start firstly the quantification of the last IMF. If the number of bits does not exceed the starting number of bits allocated, we will keep the number of remaining bits in the previous IMF, i.e., the new starting number of allocated bits for previous IMF becomes the old number of bits allocated added to the remaining bits of next IMFs. Since direct optimization is unfeasible, bit allocation is done in iterative way. A loop is intended to quantize the scaled maxima (or extrema) amplitude, to reconstruct IMF, and then to compare the reconstruction error PSD to the masking threshold: if it remains under the masking curve, the quantization is restarted with an increased number of bits, and so on until the masking constraint is satisfied. The quantization loop is shown in figure V.4. This loop is stopped for IMF respecting the inaudibility constraint. Initially, we allocate one bit for each maximum (or extrema). At each iteration of the quantization loop the number of bits is increased by one.

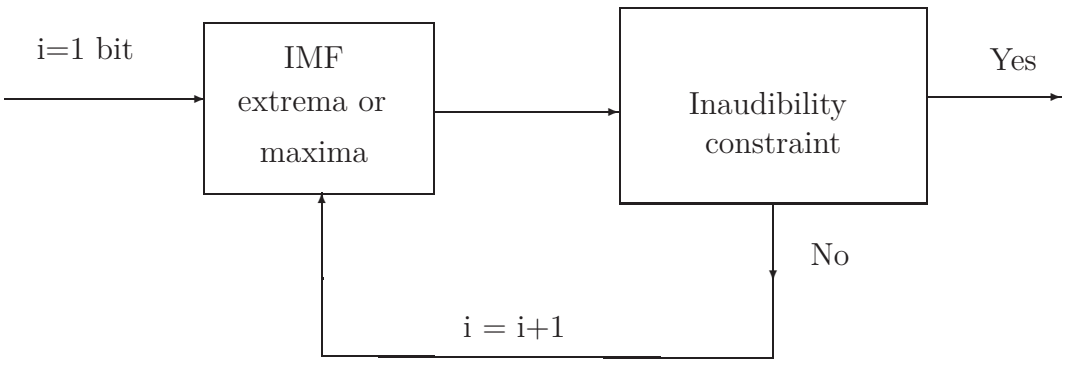


Figure V.4: Quantization scheme.

V.3 EMD based audio coders performance

The coding approaches, described in the previous chapter, are tested on different audio signals sampled at 44100 HZ. In particular, gspi, harp, quar and trpt recordings are taken from the SQAM database. The results are compared to the MP3 (ISO/IEC 11172- 3 MPEG Layer 3) and the AAC (ISO/IEC 13818-7 Advanced Audio Coding) codecs, and to the wavelet compression approach. We used Daubechies wavelet of order 8 which, in general, gives good results in comparison to other wavelets [20]. The obtained performances are analyzed using the BR, the Noise to Mask Ratio (NMR), and the Objective Difference Grade (ODG)[1] [38], which is a perceptual criterion, using the algorithm of Huber [36]. The original tested audio signals are depicted in figure V.5. Firstly, the audio signal is segmented into frames, of size

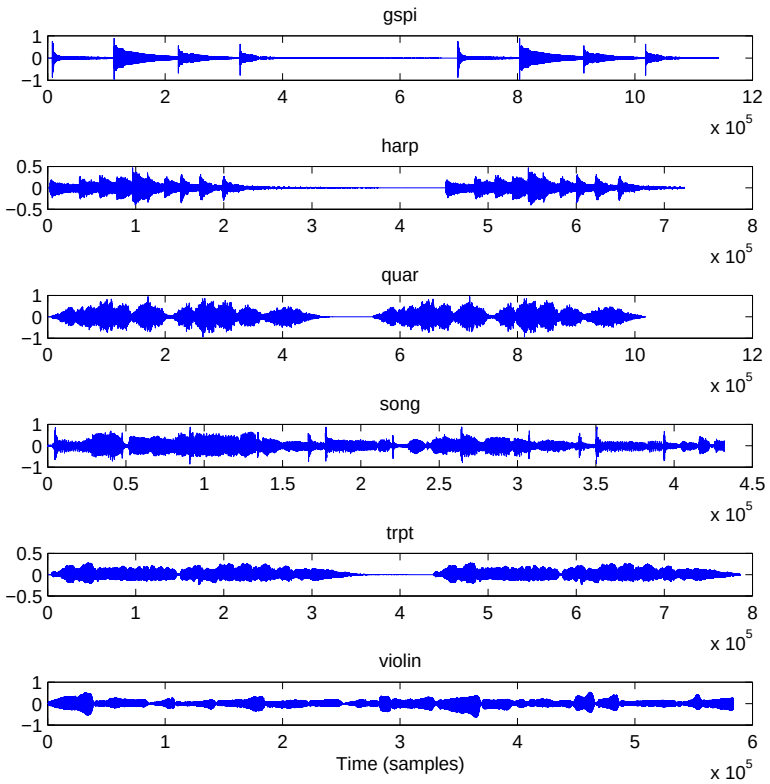


Figure V.5: Original audio signals (gspi, harp, quar, song, trpt and violin).

512 samples, with an overlap is equal to 64 samples. Using the LEC, the transient frame is divided into sub-frames. In our approaches, we have focused essentially

[1]see Appendix B

on the quality of encoding/decoding signal rather than ratio compression. Indeed all loop in the proposed algorithms are stopped when the quality is satisfied. We essentially focus on the quality of encoding/decoding signal rather than on the BR. For $IMF_{extrema}$ and $IA-IP$ coders, it is not possible for comparison purpose to fix the BR to 64 kb/s, so both proposed approaches are compared to AAC and MP3 codecs with a BR set to 96kb/s. In $IMF_{extrema}$ coder, the time index of extrema is encoded by 5 bits, the scaled factor is encoded by 8 bits, while the amplitude index is encoded variably, such that the final BR is equal to 96kb/s. The order of AR model for IA coding of each IMF in $IA-IP$ approach is fixed to 9 [48]. For IP coding, each extrema is characterized by two indices (time and amplitude) and each index is encoded by 8 bits. Values of NMR, BR and ODG obtained at BR equal to 96 kb/s with $IMF_{extrema}$ and $IA-IP$ coders are summarized in table V.1.

Table V.1: Compression results of audio signals (gspi, harp, quar, song, trpt and violin) by $IMF_{extrema}$, $IA-IP$, AAC, MP3 and the wavelet.

	Signal	gspi	harp	quar	song	trpt	violin
$IMF_{extrema}$	BR [kb/s]	96	96	96	96	96	96
	NMR	-5.87	-6.21	-6.23	-7.24	-6.47	-5.56
	ODG	-0.75	-0.67	-0.71	**-0.62**	-0.76	-0.71
$IA-IP$	BR [kb/s]	96	96	96	96	96	96
	NMR	-4.96	-3.1	-2.89	-4.12	-3.84	-3.17
	ODG	-0.78	-1.02	-1.05	-0.8	-0.78	-0.81
AAC	BR [kb/s]	96	96	96	96	96	96
	NMR	-6.12	-8.27	-6.36	-6.74	-8. 19	-6.49
	ODG	**-0.67**	**-0.59**	**-0.62**	-0.7	**-0.69**	**-0.66**
MP3	BR [kb/s]	96	96	96	96	96	96
	NMR	-2.14	-1.17	-1.29	-2.46	-2.23	-2.59
	ODG	-0.98	-1.04	-1.1	-0.89	-0.94	-0.96
Wavelet	BR [kb/s]	96	98	96	96	102	95
	NMR	-3.25	-2.73	-1.83	-3.52	-3.3	-2.97
	ODG	-0.79	-1.08	-1.19	-0.84	-0.81	-0.96

A careful examination of the results reported in table V.1, shows that both proposed approaches perform remarkably better than wavelets method in terms of BR and decoded listening quality. Compared to MP3 codec, at fixed BR to 96 kb/s, both approaches yield higher objective quality. Indeed, most of ODG index vary between -1 and 0 and have value greater than offered by MP3, which reflects the good quality of decoded signal. Compared to AAC codec, the performance of $IMF_{extrema}$

and $IA-IP$ varies with the signal type. Indeed, in /song/ signal, $IMF_{extrema}$ performs better than AAC codec, while in the other signals AAC coded perform better than the proposed approaches. Overall, the performance of the $IMF_{extrema}$ audio are not so far from those of the AAC coder. In the case of $IMF_{envelope}$ and $IA-IF$ coding approaches, we have succeeded fix the BR to 64 kb/s, since the bit allocation in both proposed approaches is variable. In the $IMF_{envelope}$ approach, the time index of maxima and scaled factor are encoded by 5 bits, while the amplitude index is encoded variably, in a way that the final BR is equal to 64kb/s. The offset value is coded over 8 bits. The order of AR model for IA coding of each IMF in $IA-IF$ approach is fixed to 9 [48]. For IF coding, the order of AR model of IF is detected, based on partial autocorrelation coefficient (Sec.IV.4.2.1). Finally, the coefficients and noise variance of both AR models are encoded by 6 bits. Results of $IMF_{envelope}$ and $IA-IF$ coders, wavelet approach, MP3 and AAC codecs, at BR 64kb/s, for different signals are shown in table V.3. According to this table, we

Table V.2: Compression results of audio signals (gspi, harp, quar, song, trpt and violin) by $IMF_{envelope}$, $IA-IF$, AAC, MP3 and wavelet methods.

	Signal	gspi	harp	quar	song	trpt	violin
$IMF_{envelope}$	BR [kb/s]	64	64	64	64	64	64
	NMR	-5.37	-5.65	-5.47	-5.13	-5.32	-5.04
	ODG	**-0.82**	**-0.73**	**-0.74**	**-0.79**	**-0.84**	**-0.83**
$IA-IF$	BR [kb/s]	64	64	64	64	64	64
	NMR	-3.65	-4.29	-5.47	-3.37	-5.73	-5.68
	ODG	**-0.84**	-0.74	-0.75	**-0.72**	-0.90	-0.92
AAC	BR [kb/s]	64	64	64	64	64	64
	NMR	-3.43	-6.46	-4.78	-4.23	-6.15	-4.59
	ODG	-0.85	**-0.73**	-0.75	-0.89	-0.88	-0.86
MP3	BR [kb/s]	64	64	64	64	64	64
	NMR	1.42	1.21	1.27	1.23	2.68	1.86
	ODG	-1.12	-1.87	-1.91	-1.09	-1.27	-1.34
Wavelet	BR [kb/s]	65	67	64	65	66	64
	NMR	-2.30	-3.67	1.64	-3.40	-1.35	-2.52
	ODG	-0.86	-1.27	-1.74	-0.98	-0.97	-1.08

conclude that $IMF_{envelope}$ coding, at BR 64 kb/s, performs remarkably better than MP3 and wavelet methods in terms of perceptual quality. Compared to AAC coder, we remark that $IMF_{envelope}$ gives also better results especially with the /gspi/ and /song/ audio signals. Even $IA-IF$ approach provides the best results compared to the wavelet approach and MP3 codec. However, improvement in perceptual quality

is achieved for only gspi and song signals compared to the AAC coder. For other signals, $IA-IF$ and AAC coder have comparable performances, where ODG varies between -1 and 0. This reflects the good decoded signal quality.

V.4 Conclusion

In this chapter, we have illustrated the EMD based coding on different audio signals and results compared to wavelet approach and to AAC and MP3 codecs. The obtained results in terms of BR and of ODG measure show that the proposed methods perform much better than MP3 codec and wavelet compression. The $IMF_{envelope}$ is the most efficient approach that performs better than the AAC codec. The effectiveness of this coding is observed especially for audio signals /gspi/, /song/, /trpt/ and /violin/. Further, the efficiency of $IMF_{envelope}$ is essentially due to use of a psychoacoustic model and the symmetry property of the IMF, which enable good audio quality at low BR. In addition, the decoding by spline interpolation is very easy. The proposed codings are adaptive and without any prior assumptions. Overall, the obtained results and the comparison to well established coding methods demonstrate the potential of the EMD as a promising audio coding tool. We show in the next chapter, how the EMD can also exploited for watermarking purpose.

CHAPTER VI

Audio watermarking based on the EMD

Contents

VI.1 Introduction 112

VI.2 Proposed watermarking algorithm 112

VI.2.1 Synchronization code 114

VI.2.2 Watermark embedding 115

VI.2.3 Watermark extraction 116

VI.3 Performance analysis 117

VI.4 Results 118

VI.5 Conclusion 124

This chapter introduces a new adaptive audio watermarking algorithm, based on EMD, dedicated to copyright protection. The audio signal is divided into frames and each one is decomposed adaptively into IMFs. The watermark and the synchronization codes are embedded into the extrema of the last IMF of each frame, a low frequency mode stable under different attacks and preserving an audio perceptual quality of the host signal. The watermarking technique is chosen in the category of Quantization Index Modulation (QIM) due to its good robustness and blind nature. Parameters of QIM are chosen to guarantee that the embedded watermark in the last IMF is inaudible. The data embedding rate of the proposed algorithm is 46.9-50.3b/s. Relying on exhaustive simulations,

we show the robustness of the hidden watermark for additive noise, MP3 compression, re-quantization, filtering, cropping and resampling. The comparison analysis shows that our method has better performance than watermarking schemes reported recently.

VI.1 Introduction

We propose in this chapter an adaptive watermarking scheme based on the EMD. The IMFs are nearly orthogonal to each other, and all have nearly zero means. The number of extrema is decreased when going from one mode to the next, and the whole decomposition is guaranteed to be completed with a finite number of modes. The IMFs are fully described by their local extrema and thus can be recovered using these extrema [34],[47]. Low frequency components such as higher order IMFs are signal dominated [8],[44] and thus their alteration can lead to degradation of the signal. Thus, these modes can be considered to be good locations for watermark placement. Watermarks inserted into lower order IMFs (high frequency) are most vulnerable to attacks. The watermark can also be embedded into the trend (coarsest mode) of the host signal, but our experiments indicate that this mode is not highly robust to attacks. It has been argued that for watermarking robustness, the watermark bits are usually embedded in the perceptually components, mostly, the low frequency components of the host signal [37]. To simultaneously have better resistance against attacks and imperceptibility, we embed the watermark in the last IMF. We choose in our method a watermarking technique in the category of Quantization Index Modulation (QIM) due to its good robustness and blind nature [12]. Parameters of QIM are chosen to guarantee that the embedded watermark in the last IMF is inaudible. The watermark is associated with a synchronization code to facilitate its location. Audio signal is first segmented into frames where each one is decomposed adaptively into IMFs. Bits are inserted into the extrema of the last IMF such that the watermarked signal inaudibility is guaranteed.

VI.2 Proposed watermarking algorithm

The idea of the proposed watermarking method is to hid into the original audio signal a watermark together with a Synchronized Code (SC) in the time domain. The input signal is first segmented into frames and the EMD is conducted on every frame to extract the associated IMFs (Fig. VI.1). Then a binary data sequence consisted of SCs and informative watermark bits (Fig. VI.2) is embedded in the extrema of the last IMF (very low frequency component). The sequence is embedded P times. However, since the number of IMFs and the number of extrema depend on the amount of data of each frame, the number of binary sequence ($\leq P$) to be

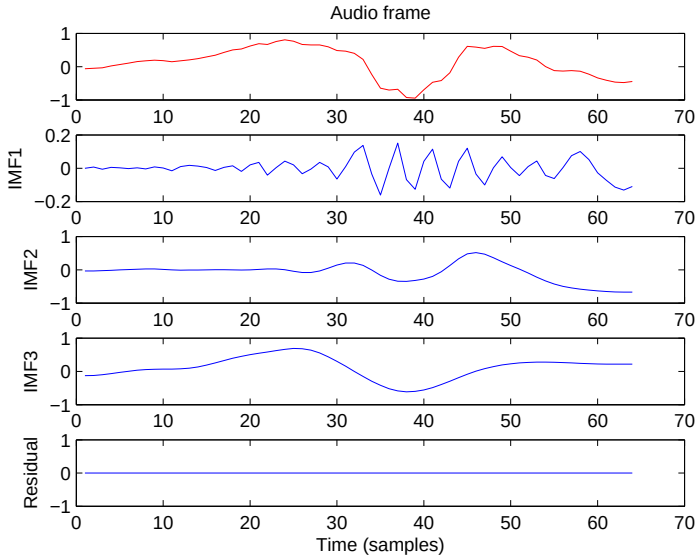


Figure VI.1: Decomposition of an audio frame into IMFs.

Sync-code	Watermark bits	Sync-code

Figure VI.2: Data structure $\{m_i\}$.

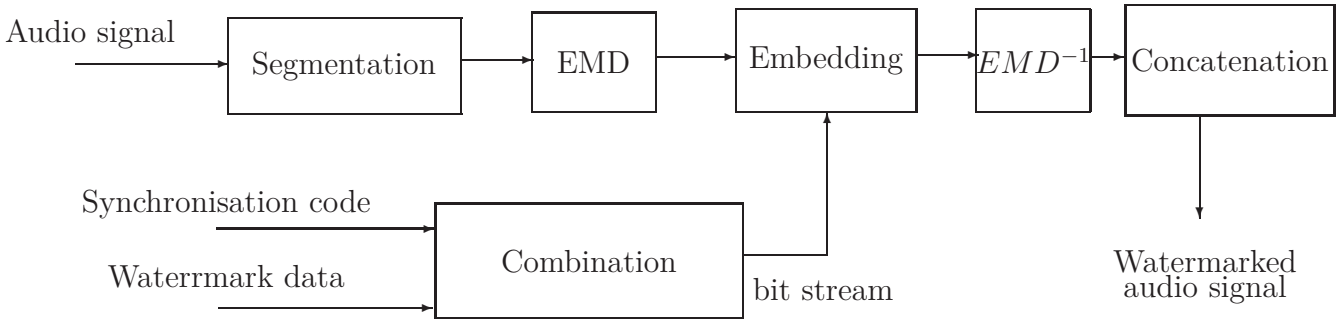


Figure VI.3: Watermark embedding.

embedded varies from one frame to the following. Finally, inverse transformation (EMD^{-1}) is applied to the modified extrema to recover the watermarked audio signal by superposition of the IMFs of each frame followed by the concatenation of the frames (Fig. VI.3). For data extraction, the watermarked audio signal is splitted into frames and EMD applied to each frame (Fig. VI.4). Binary data sequences are extracted from each last IMF by searching for SCs (Fig. VI.5). We show in figure VI.6 the last IMF before and after watermarking. This figure shows that there is little difference in terms of amplitudes between the two modes. EMD being fully data adaptive, thus its is important to guarantee that the number of IMFs will be same before and after embedding the watermark (Figs. VI.1,VI.4). In fact, if the numbers of IMFs are different, there is no guarantee that the IMF always contains the watermark information to extract. To overcome this problem the sifting of the watermarked signal is forced to extract the same number of IMFs as before watermarking. The proposed watermarking scheme is blind, that is, the host signal is not required for watermark extraction. Overview of the proposed method is detailed as follows:

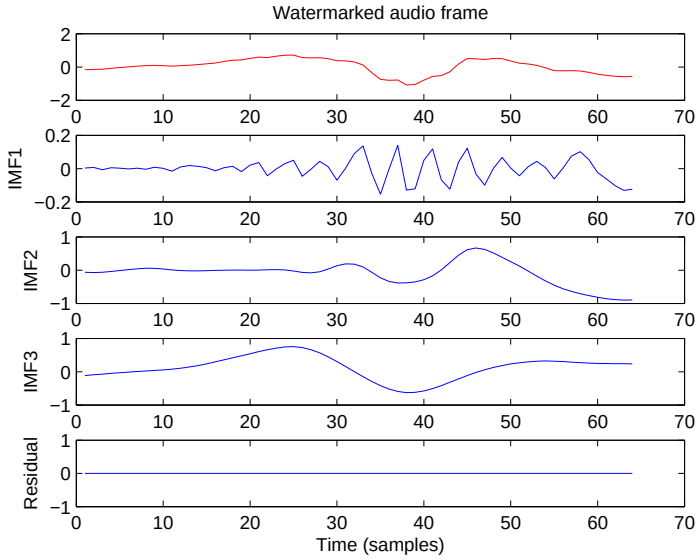


Figure VI.4: Decomposition of the watermarked audio frame by EMD.

VI.2.1 Synchronization code

To locate the embedding position of the hidden watermark bits in the host signal a SC is used. This code is unaffected by cropping and shifting attacks [86].

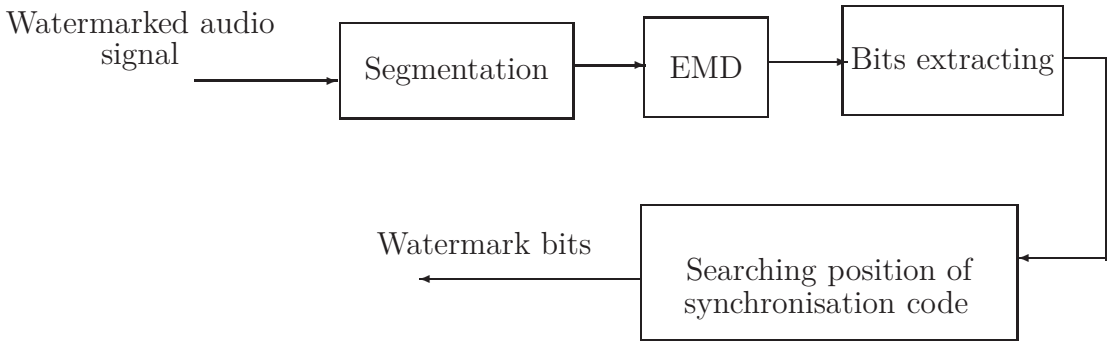


Figure VI.5: Watermark extraction.

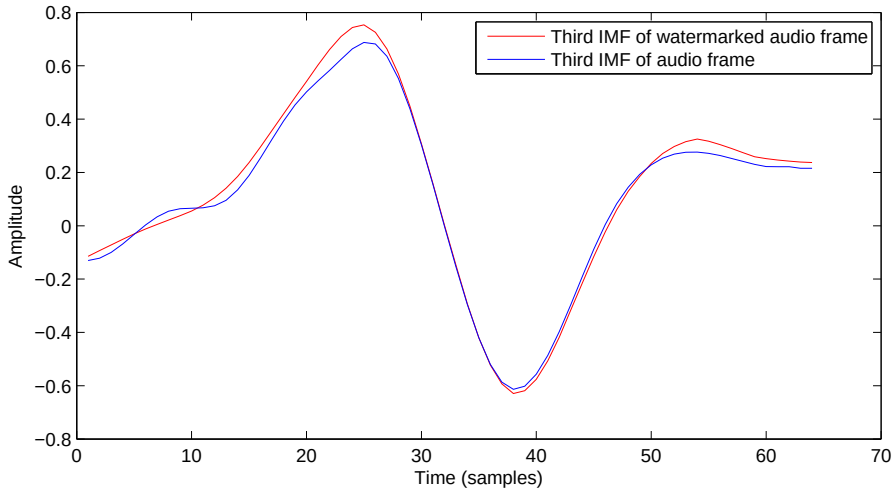


Figure VI.6: Illustration of the last IMF of an audio frame before and after watermarking.

Let U be the original SC and V be an unknown sequence of the same length. Sequence V is considered as a SC if only the number of different bits between U et V (bit by bit) is less or equal than to a predefined threshold τ [86].

VI.2.2 Watermark embedding

Before embedding, SCs are combined with watermark bits to form a binary sequence denoted by $m_i \in \{0, 1\}$, i-th bit of watermark (Fig. VI.2). Basics of our watermark embedding are shown in figure VI.3 and detailed are follows:

Step 1: Split original audio signal into frames.

Step 2: Decompose each frame into IMFs.

Step 3: Embed P times the binary sequence $\{m_i\}$ into extrema of the last IMF

(IMF_C) by QIM [12]:

$$e_i^* = \begin{cases} \lfloor e_i/\mathrm{S} \rfloor.\mathrm{S} + \mathrm{sgn}(3\mathrm{S}/4) & \text{if } m_i = 1 \\ \lfloor e_i/\mathrm{S} \rfloor.\mathrm{S} + \mathrm{sgn}(\mathrm{S}/4) & \text{if } m_i = 0 \end{cases} \quad \text{(VI.1)}$$

where e_i and e_i^* are the extrema of IMF_C of the host signal and the watermarked signal respectively. sgn function is equal to "+" if e_i is a maxima, and "-" if it is a minima. $\lfloor\ \rfloor$ denotes the floor function, and S denotes the embedding strength chosen to maintain the inaudibility constraint.

Step 4: Reconstruct the frame (EMD^{-1}) using modified IMF_C and concatenate the watermarked frames to retrieve the watermarked signal.

VI.2.3 Watermark extraction

For watermark extraction, host signal is splitted into frames and EMD is performed on each one as in embedding. Binary data is extracted using rule given by equation (VI.2). We then search for SCs in the extracted data. This procedure is repeated by shifting the selected segment (window) one sample at time until a SC is found. With the position of SC determined, we can then extract the hidden information bits, which follows the SC. Let $y = \{m_i^*\}$ denote the binary data to be extracted and U denote the original SC. To locate the embedded watermark we search the SCs in the sequence $\{m_i^*\}$ bit by bit. Let N_1 and N_2 be the numbers of bits of SC and watermark respectively. The extraction is performed without using the original audio signal. Basic steps involved in the watermarking extraction, shown in figure VI.5, are given as follows:

Step 1: Split the watermarked signal into frames.

Step 2: Decompose each frame into IMFs.

Step 3: Extract the extrema $\{e_i^*\}$ of IMF_C.

Step 4: Extract m_i^* from e_i^* using the following rule [86]:

$$m_i^* = \begin{cases} 1 & \text{if } e_i^* - \lfloor e_i^*/\mathrm{S} \rfloor.\mathrm{S} \geq \mathrm{sgn}(\mathrm{S}/2) \\ 0 & \text{if } e_i^* - \lfloor e_i^*/\mathrm{S} \rfloor.\mathrm{S} < \mathrm{sgn}(\mathrm{S}/2) \end{cases} \quad \text{(VI.2)}$$

Step 5: Set the start index of the extracted data, y, to $I = 1$ and select $L = N_1$ samples (sliding window size).

Step 6: Evaluate the similarity between the extracted segment $V = y(I : L)$ and U bit by bit. If the similarity value is $\geq \tau$, then V is taken as the SC and go to Step

8. Otherwise proceed to the next step.

Step 7: Increase I by 1 and slide the window to the next $L = N_1$ samples and repeat Step 6.

Step 8: Evaluate the similarity between the second extracted SC, $V' = y(I + N_1 + N_2 : I + 2N_1 + N_2)$. If the similarity value is $\geq \tau$, then V' is taken as the SC and the sequence $y(I + N_1 : I + N_1 + N_2 - 1)$ is taken as the mark, and go to Step 9. Otherwise repeat Step 7.

Step 9: $I \leftarrow I + N_1 + N_2$, if the new I value is equal to sequence length of bits, go to Step 10, else repeat Step 7.

Step 10: Extract the P watermarks and make comparison bit by bit between these marks, for correction, and finally extract the desired watermark.

VI.3 Performance analysis

We evaluate the performance of our method in terms of data payload, error probability of SC, SNR, Bit Error Rate (BER) and Normalized cross-Correlation (NC). The SNR is defined as

$$\texttt{SNR} = 10 \log_{10} \frac{\sum_{i=1}^{T} X^2(i)}{\sum_{i=1}^{T} (X(i) - \tilde{X}(i))^2} \tag{VI.3}$$

where X and $\tilde{X}$ denote the original and the watermarked audio signals respectively. According to International Federation of the Photographic Industry (IFPI) recommendations, a watermark audio signal should maintain more than 20 dB SNR. To evaluate the watermark detection accuracy after attacks, we used the BER and the NC defined as follows [41]:

$$\texttt{BER}(W, \tilde{W}) = \frac{\texttt{Number of error bits}}{\texttt{Number of total bits}} = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} W(i,j) \oplus \tilde{W}(i,j)}{M \times N} \tag{VI.4}$$

where $\oplus$ is the XOR operator. W and $\tilde{W}$ are the original and the recovered watermark respectively. `BER` is used to evaluate the watermark detection accuracy after

signal processing operations.

$$\mathtt{NC}(W,\tilde{W}) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} W(i,j)\tilde{W}(i,j)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} W^2(i,j)}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} \tilde{W}^2(i,j)}} \tag{VI.5}$$

NC is used to evaluate the similarity between the original watermark and the extracted watermark. A large NC indicates the presence of watermark while a low value suggests the lack of watermark. Two types of errors may occur while searching the SCs: the False Positive Error (FPE) and the False Negative Error (FNE). These errors are very harmful because they impair the credibility of the watermarking system. The associated probabilities of these errors are given by [41],[86]:

$$\begin{aligned} P_{FPE} &= \frac{1}{2^p}\sum_{k=p-\tau}^{p} C_p^k \\ P_{FNE} &= \frac{1}{2^p}\sum_{k=1+\tau}^{p} C_p^k (BER)^k (1-BER)^{p-k} \end{aligned} \tag{VI.6}$$

where p is the SC length and τ is the threshold. P_{FPE} is the probability that a SC is detected in false location while P_{FNE} is the probability that a watermarked signal is declared as unwatermarked by the decoder. We also use as performance measure the payload which quantifies the amount of information to be hidden. More precisely, the data payload refers to the number of bits that are embedded into that audio signal within a unit of time and is measured in the unit of bits per second (b/s). The data payload, D, is defined as follows:

$$D = \frac{L_h}{M_b} \tag{VI.7}$$

where L_h is the length in seconds of the host audio signal and M_b is the number of bits of the watermark data.

VI.4 Results

To evaluate the performance of our scheme, simulations are performed on audio signals including classic, jazz, rock and pop, sampled at 44.1 kHz. The embedded watermark, W, is a binary logo image of size M×N=34×48= 1632 bits (Fig. VI.7). We convert this 2D binary image into 1D sequence in order to embed it into the

audio signal. The SC used is a 16 bit Barker code 1111100110101110. Each au-

Figure VI.7: Binary watermark.

dio signal is divided into frames of size 64 samples and the threshold τ is set to 4. The S value is fixed to 0.98. These parameters have been chosen to have a good compromise between imperceptibility of the watermarked signal, payload and robustness. Figure VI.8 shows a portion of the pop signal and its watermarked version. Perceptual quality assessment can be performed using subjective listening

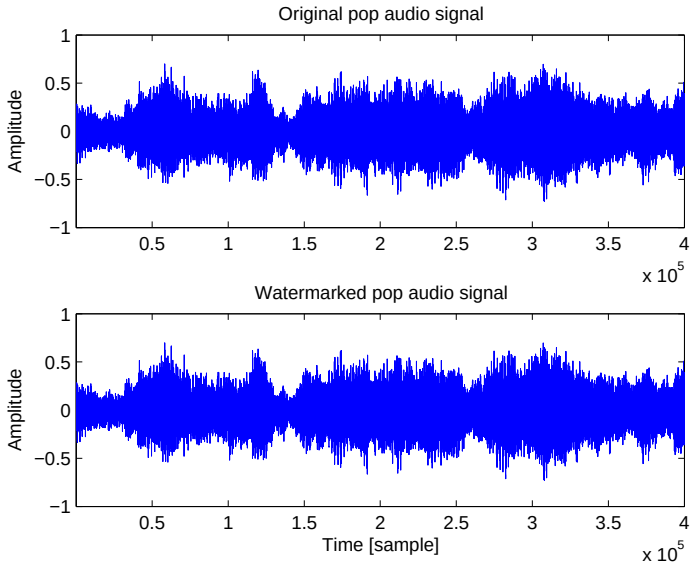


Figure VI.8: A portion of the pop audio signal and its watermarked version.

tests by human acoustic perception or using objective evaluation tests by measuring the SNR and Objective Difference Grade (ODG). In this work we use the second approach. ODG and SNR values of the four watermarked signals are reported in table VI.1. The SNR values are above 20 dB showing the good choice of S value and confirming to IFPI standard. All ODG values of the watermarked audio signal are between -1 and 0 which demonstrates their good quality.

Table VI.1: SNR and ODG between original and watermarked audio.

Audio file	SNR (dB)	ODG
classic	25.67	-0.5
jazz	26.38	-0.4
pop	24.12	-0.6
rock	25.49	-0.5

Robustness test

To asses the robustness of our approach, different attacks are performed:

- Noise: White Gaussian Noise (WGN) is added to the watermarked signal until the resulting signal has an SNR of 20 dB.
- Low pass filtering: A second order Butterworth filter, which eliminated frequency more than 11025 Hz, is used.
- Denoising: Filter the watermarked audio signal using Wiener filter.
- Cropping: Segments of 512 samples are removed from the watermarked signal at thirteen positions and subsequently replaced by segments of the watermarked signal contaminated with WGN.
- Resampling: The watermarked signal, originally sampled at 44.1 kHz, is resampled at 22.05 kHz and restored back by sampling again at 44.1 kHz.
- MP3 compression 64 kb/s and 32 kb/s: Using MP3, the watermarked signal is compressed and then decompressed.
- Requantization: The watermarked signal is re-quantized down to 8 bits/sample and then back to 16 bits/sample.

Table VI.2 shows the extracted watermarks with the associated NC and BER values for different attacks on pop audio signal. NC values are all above 0.9682 and BER values are all below 4%. The extracted watermark are visually similar to the original watermark. These results shows the robustness of watermarking method for pop audio signal. Even in the case of WGN attack with SNR of 20dB, our approach does not detects any error. This is mainly due to the insertion of the watermark into IMF_C extrema. In fact low frequency subband has high robustness against noise addition [41],[86].

Table VI.2: BER and NC of extracted watermark for pop audio signal by proposed approach.

Attack type	BER %	NC	Extracted watermark
No attack	0	1	
AWGN (20dB)	0	1	
Low pass filtering	0	0.9994	
Denoising	6	0.9482	
Cropping	0	1	
Resampling	3	0.9783	
MP3 (64 Kb/s)	0	0.9996	
MP3 (32 Kb/s)	1	0.9876	
Requantization	0	1	

Table VI.3 reports similar results for classic, jazz and rock audio files. NC values are all above 0.9964 and BER values are all below 3%, demonstrating the good performance robustness of our method on these audio files. This is robustness is due to the fact that even the perceptual characteristics of individual audio files vary, the EMD decomposition adapts to each one. Table VI.4 shows comparison results in

Table VI.3: BER and NC of extracted watermark for different audio signals (Classical, Jazz, Rock) by proposed approach.

Audio signal	Attack type	BER %	NC
Classical	No attack	0	1
	AWGN	0	1
	Low pass filtering	0	1
	Denoising	0	1
	Cropping	0	1
	Resampling	2	0.9986
	MP3(64 kb/s)	0	1
	MP3 (32 kb/s)	0	1
	Requantization	0	1
Jazz	No attack	0	1
	AWGN	0	1
	Low pass filtering	1	0.9989
	Denoising	6	0.9964
	Cropping	0	1
	Resampling	2	0.9973
	MP3(64 kb/s)	0	1
	MP3 (32 kb/s)	1	0.9983
	Requantization	0	1
Rock	No attack	0	1
	AWGN	0	1
	Low pass filtering	0	1
	Denoising	0	1
	Cropping	0	1
	Resampling	1	0.9989
	MP3(64 kb/s)	0	1
	MP3 (32 kb/s)	0	1
	Requantization	0	1

terms of payload and robustness to MP3 compression attack of our method to nine recent watermarking schemes. Due to diversity of these embedding approaches, the comparison is sorted by attempted data payload. It can be seen that our method achieves the highest payload for the three audio files. Also, for these signals our scheme has a good performance against MP3 (32kb/s) compression, where the max-

imum of BER against this last is of 1%.

Table VI.4: BER and NC of extracted watermark for different audio signals (Classical, Jazz, Rock) by proposed approach.

Reference	payload (b/s)	Robustness to MP3 (kb/s)
Proposed algorithm	46.9-50.3	32
Bhat K[41]	45.9	32
Lie[54]	43	80
Cvejic[14]	27.1	32
Yeo[89]	10	96
Tachibana[80]	8.5	96
Li[53]	4.2	32
Mansour[57]	2.3	56
Xiang[88]	2	64
Kirovski[49]	0.5-1	32

Figure VI.9 plots the P_{FPE} versus p. We see that P_{FPE} tends to 0 when $p \geq 16$. So, this confirms the choice of SC length. Figure VI.10 shows that the P_{FNE} is dependent on the length of watermark bits. So, we note that for the embedding bits length ≥ 30, the P_{FNE} tends to 0. Since the watermark bits used is of 1632 bits (≥ 30), the obtained P_{FNE} is very low.

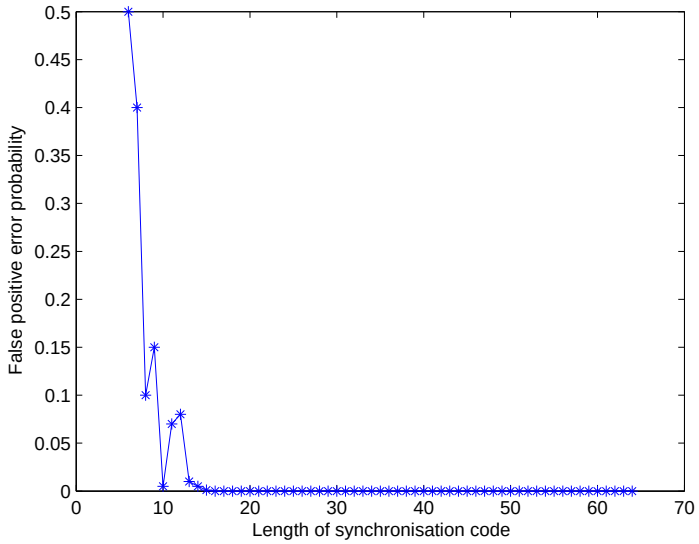


Figure VI.9: P_{FPE} versus synchronization code length.

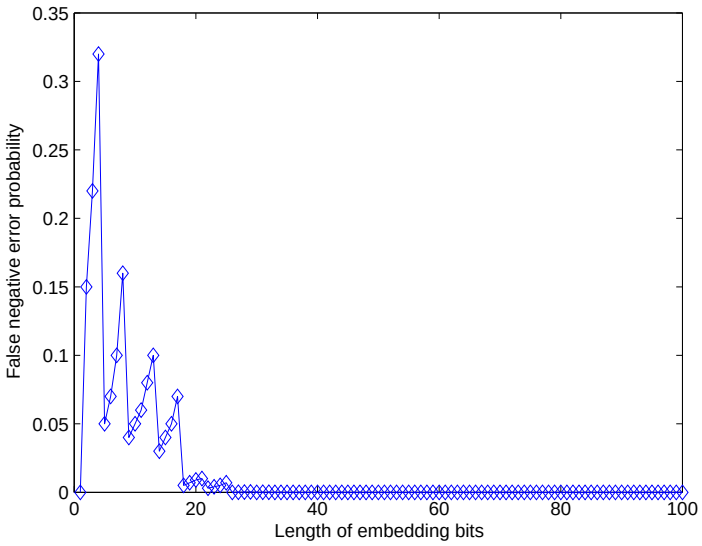


Figure VI.10: P_{FNE} versus the length of embedding bits.

VI.5 Conclusion

In this chapter a new adaptive watermarking scheme based on the EMD is proposed. Watermark is embedded in very low frequency mode (last IMF), thus achieving good performance against various attacks. Watermark is associated with synchronization codes and thus the synchronized watermark has the ability to resist shifting and cropping. Data bits of the synchronized watermark are embedded in the extrema of the last IMF of the audio signal based on QIM. Extensive simulations over different audio signals indicate that the proposed watermarking scheme has greater robustness against common attacks than nine recently proposed algorithms. This scheme has higher payload and better performance against MP3 compression compared to these earlier audio watermarking methods. In all audio test signals, the watermark introduced no audible distortion. Experiments demonstrate that the watermarked audio signals are indistinguishable from original ones. These performances take advantage of the self-adaptive decomposition of the audio signal being marked provided by the EMD. The proposed scheme achieves very low false positive and false negative error probability rates. Our watermarking method involves easy calculations and does not use the original audio signal. In the conducted experiments the embedding strength S is kept constant for all audio files. To further improve the performance of the method, the S parameter should be adapted to the type and magnitudes of the original audio signal.

Conclusions and perspectives

The purpose of this thesis was to investigate the potential of EMD (Huang transform) as analyzing tool for audio and speech processing. Main contributions, around EMD, of this dissertation are: speech denoising, audio coding and audio watermarking for copyright protection.

In chapter **I**, EMD is presented. This expansion into IMFs is performed in adaptive way. Unlike FT or WT, basis functions of EMD are derived from the signal itself and hence, the decomposition is adaptive in contrast to FT or WT where the basis functions are fixed. This is one reason that motivated our choice for the EMD. Further an other interest of the EMD is that no assumptions concerning the linearity or the stationarity are made about the signal to be analyzed. IMFs are orthogonal [34] and their extraction is nonlinear, but their linear recombination is accurate. We have shown that the decomposition results by EMD are conditioned by sampling and signals interpolation. Based on simulations, we noted that this decomposition is organized in a filter bank structure toward a Gaussian white noise. Both orthogonality of modes and filter bank nature of EMD are important properties exploited for denoising, decoding and watermarking purposes. In general the EMD results are conditioned by the sampling rate and interpolation used. The reported results are obtained with signals oversampled and using cubic splines interpolation which is commonly used to approximate upper and lower envelopes in EMD. To improve the obtained results, it would be useful to test other interpolation approaches such as cubic Hermite spline or regularized interpolation. Although EMD has advantages in signal decomposition, it still has limitations such as end effect and the mode mixing (caused by signal intermittency). These shortcomings

must be resolved to improve the performance of the signal decomposition. As its name, EMD is still an empirical technique. Our results show that EMD is very effective for denoising, coding and watermarking, but it still needs theoretical support. Even, EMD based processing methods have shown good performances compared to MMSE filter or wavelet approach it is interesting, as future work, to extend the comparison to other approaches such as methods based on matching pursuit [15],[32].

In chapter **II**, three denoising approaches based on the EMD are proposed. Two approaches were dedicated to white noise and the third one has focused on a large class of noises including correlated case. For approaches dedicated to white noise, EMD-Shrinkage is used especially in the case where the estimated noise level is not reliable. However, when the estimation of noise level is accurate, EMD combined with MMSE filter (EMD-MMSE) improves the denoising results. Furthermore, the obtained results also show that it is more efficient to apply the thresholding or the filtering to the extracted modes (IMFs) of the signal than to the signal itself. In the case of colored noise, the EMD-ACWA gives better results compared to ACWA filter and to wavelet approach. Indeed, the effectiveness of the ACWA filter is improved when it is associated with the EMD. In particular, we have also shown that it is more efficient in term of performance to combine EMD with the ACWA filter than with other classical filters such as MMSE filter. This is essentially due to the fact that as the EMD, the ACWA filter operates in time domain and exploits the local statistics of the signal. Furthermore, the assumptions of signal stationarity and white are not required. Since ACWA filter performs a sliding window analysis, performances of EMD-ACWA are partly dependent on proper choice of window length. The optimal size of the window is in general not known (depends on SNR and signal) and is determined only through experimentation. As future research we plan to work on a strategy to choosing optimal length value of the window. Ongoing research work is also to apply the proposed denoising to a large class of real signals to confirm the obtained results.

Chapter **III** is dedicated to speech denoising. As in chapter **II**, EMD is used in conjunction with ACWA filter. The aim was to improve the previously obtained denoising performances. This was achieved by taking into account the class of speech frame (voiced/unvoiced). The obtained results have shown that the number of denoised IMFs depends on whether the noisy frame is voiced or unvoiced. Thus,

an energy criterion is used to detect voiced frames while a stationarity index is used to distinguish between unvoiced and transient sequences. Obtained results for clean speech signals corrupted with additive white Gaussian noise with varying SNR values show that the proposed method performs better than the ACWA filtering of all IMFs (EMD-ACWA), wavelet denoising approach (db4) and ACWA filtering of the noisy signal. As shown from the reported results, taking into account the statistical properties over the time of signal (voiced/unvoiced) in the filtering process improves noticeably the performances of the speech denoising in terms of both SNR and PESQ. To capture the stationarity of the speech frame, the index used is based on spectrogram. This TFR is chosen due to its simple use. However, the spectrogram performs less better in term of temporal and frequency resolution than other TFRs such Wigner-Ville distribution. As result, other TFRs than spectrogram should be tested to see if there is enough stationarity in the data.

In chapter **IV**, a new signal coding strategy is introduced. A salient property of the IMF is that it can be fully described by its extrema. This property is the core of the proposed signal coding. Firstly, two waveform coding schemes are introduced. These two codings are non-parametric approaches. The first scheme ($IMF_{extrema}$) consist in encoding the IMFs extrema. Motivated by quasi-symmetrical property of the IMF and in order to further reduce the bit rate a second scheme was proposed. Thus, one out of two IMF envelopes is coded ($IMF_{envelope}$). Secondly, two parametric approaches combining HHT and AR modeling are proposed. AR modeling is supported by the correlation of IA and IF values of the IMFs. So, this model is useful to exploit this correlation. In the first parametric approach ($IA-IF$), coefficients of the AR model of both IA and IF components are encoded. In the second method ($IA-IP$), we keep the same encoding for IA i.e., by linear prediction, and the IP extrema are coded by scalar quantization. On the whole, coding of IMF extrema provides a general framework for signal coding in adaptive way and potentially can be useful for a large class of signals. Even the proposed coding is illustrated on only audio signals (Chap. **V**), the developed algorithms can be easily extended, for example, to biomedical signals (ECG, EEG, MEG,...).

Results of the coding framework in audio context are reported in chapter **V**. To reduce the BR, $IMF_{extrema}$ and $IMF_{envelope}$ audio coders are associated with psychoacoustic model. We have also shown in $IA-IF$ approach, that the order

of the AR model for IF varies from one IMF to another. Therefore, for each IMF the AR order of associated IF function is determined using partial autocorrelation coefficient. Obtained results in terms of BR and of ODG show that the proposed methods perform much better than MP3 codec and wavelet compression. The $IMF_{envelope}$ is the most efficient approach, which provides better results compared to the AAC codec. Efficiency of the $IMF_{envelope}$ is essentially due to the use of a psychoacoustic model and the symmetry property of the IMF, which enable good audio quality at low BR. Results of the proposed empirical coding are not prejudiced by predetermined basis and/or subband filtering. This coding does not require any user parameters setting, except the stopping criterion of the EMD. Decoding by spline interpolation is very easy and the computational time of the method is much lower. Although different practical experiments have already been carried out on different kinds of audio sources, future works should consider large classes of audio signals as well as varied experimental conditions such as different sampling rates or frame size for improving the tuning of the method.

In chapter **VI**, a new adaptive audio watermarking algorithm based on EMD and dedicated for copyright protection is introduced. The principle of the proposed watermarking consists in embedding the watermark into extrema of the low frequency IMF. Low frequency components such as higher order IMFs are signal dominated and thus their alteration can lead to degradation of the signal. As result, these modes can be considered to be good locations for watermark placement. To simultaneously have better resistance against attacks and imperceptibility, we embed the watermark in the last IMF. We choose in our method a watermarking technique in the category of QIM due to its good robustness and blind nature. Parameters of QIM are chosen to guarantee that the embedded watermark in the last IMF is inaudible. Obtained results for audio signals demonstrate that the hidden data are robust against attacks such as additive noise, MP3 compression, requantization, cropping and filtering. Our method has hight data payload and performance against MP3 compression compared to nine audio watermarking approaches reported recently. Furthermore our approach has higher payload, where the data payload of the proposed algorithm, varies between 46.9 and 50.3 b/s, and better performance against MP3 compression compared to other watermarking approaches. Our watermarking method involves easy calculations and does not use the original audio signal. In the conducted experiments the embedding strength

S is kept constant for all audio files. To further improve the performance of the method, the S parameter should be adapted to the type and magnitudes of the original audio signal.

Even based on extensive simulations (synthetic and real data), the obtained results of denoising, encoding and watermarking compared to well established methods such as MMSE filter, wavelets approach, MP3 and AAC codecs illustrate the real potential of the EMD as analyzing tool (in adaptive way) in speech and audio processing. Although the developed tools are illustrated on 1D signals, they can be easily extended to image processing. On the whole, the obtained results can be further improved through a modification of the conventional sifting. More specifically, instead of interpolation (exact B-splines fitting) to construct the upper and lower envelopes of the signal to be decomposed we can use a smoothing (regularized B-splines). Advantage of this sifting is to give EMD more robustness against noise and to reduce the number of unwanted IMFs of conventional EMD. As result the number of IMFs to be denoised or encoded will be reduced. Also as future work we plan to explore some theoretical aspects of the EMD such interpolation, mode mixing or orthogonality of the modes. The formalism of the EMD remains an exciting challenge for the signal processing community.

Bibliography

[1] A. Ayenu-Prah and N. Attoh-Okine. A criterion for selecting relevant intrinsic mode functions in empirical mode decomposition. *J. on Advances in Adaptive Data Analysis*, 2(1):1–24, 2010.

[2] W. Bender, D. Gruhl, N. Morimoto, and A. Lu. Techniques for data hiding. *IBM Systems Journal*, 35:313–336, 1996.

[3] S. Benramdane, J.C. Cexus, A.O. Boudraa, and J.A. Astolfi. Transient turbulent pressure signal processing using empirical mode decomposition. *Proc. Physics in Signal and Image Processing*, Mhoulouse, 2007.

[4] R. Benzid. *Ondelettes et statistiques d'ordre supérieur appliquées aux signaux uni et bidimentionnels*. PhD thesis, Université de Batna, 2005.

[5] S. F. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. ASSP.*, 27:113–120, 1979.

[6] A. Bouchiki. *Détection et classification d'échos de cibles Sonar par THT (Transformation de Huang-Teager)*. PhD thesis, Université de Rennes I, 2010.

[7] A.O. Boudraa and J.C. Cexus. Denoising via empirical mode decomposition. In *Proc. IEEE ISCCSP*, Marrakech, Morocco, 2006.

[8] A.O. Boudraa and J.C. Cexus. EMD-based signal filtering. *IEEE Trans. Instrum. Measur.*, 56(6):2196–2202, December 2007.

[9] A.O. Boudraa, J.C. Cexus, and Z. Saidi. EMD-based signal noise reduction. *Int. J. Sig. Process.*, 1(1):33–37, 2004. ISSN: 1304-4494.

[10] K. Brandenburg and G. Stoll. Iso-mpeg-1 audio: A generic standard for coding of high-quality digital audio. *J. Audio Eng. Soc.*, 42(10):780–792, 1994.

[11] J.C. Cexus and A.O. Boudraa. Non-stationary signals analysis by teager-huang transform (THT). In *Proc. IEEE EUSIPCO*, Florence, Italy, 2006.

[12] B. Chen and G. W. Wornell. Quantization index modulation methods for digital watermarking and information embedding of multimedia. *J. VLSI Signal Processing Systems*, 27:7–33, 2001.

[13] N. Cvejic and T. Seppanen. Robust audio watermarking in wavelet domain using frequency hopping and patchwork method. In *Proc IEEE ISPA*, Rome, Italy, 2003.

[14] N. Cvejic and T. Seppanen. Spread spectrum audio watermarking using frequency hopping and attack characterisation. *Signal Processing*, 84(1):207–213, 2004.

[15] L. Daudet. Sparse and structured decompositions of signals with the molecular matching pursuit. *IEEE Trans. Speech Audio Processing*, 14(5):1808–1816, 2006.

[16] R. Deering and J.F. Kaiser. The use of a masking signal to improve empirical mode decomposition. *Proc. IEEE ICASSP*, 4:485–488, Philadelphia, 2005.

[17] E. Deger, K. ISlam Molla, K. Hirose, N. Minemastu, and K. Hasan. Speech enhancement using soft thresholding with DCT-EMD based hybrid algorithm. In *Proc. IEEE EUSIPCO*, 2007.

[18] E. Delechelle, J. Lemoine, and O. Niang. Empirical mode decomposition: An analytical approach for sifting process. *IEEE Signal Proc. Lett.*, 12(11):764–767, 2005.

[19] P. Depalle, G. Garcia, and X. Rodet. Analysis of sound for additive synthesis: tracking of partials using hidden Markov model. *Proc. ICMC*, pages 94–97, 1993.

[20] P.R. Deshmukh. Multiwavelet decomposition for audio coding. *IE(I) Journal-ET*, 87:38–41, 2006.

[21] E.S. Diop, R. Alexandre, and A.O. Boudraa. A pde charecterization of the intrinsic mode functions. In *Proc IEEE ICASSP*, Taipei, Taiwan, 2009.

[22] D.L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inform. Theory*, 41(3):613–627, 1995.

[23] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrica*, 81:425–455, 1994.

[24] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptopia with discussion. *Proc. Royal Stat. Soc., Series B,*, 57:301–396, 1995.

[25] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean square error short-time spectral estimator. *IEEE Trans. on Acoustic Speech and Signal Processing*, 32:1109–1121, 1984.

[26] P. Flandrin and P. Gonçalvès. Empirical mode decompositions as data-driven wavelet like expansions. *Int. J. of Wavelets, Multires. and Info. Proc.*, 2(4):477–496, 2004.

[27] P. Flandrin, P. Gonçalvès, and G. Rilling. Emd equivalent filter banks, from interpretation to applications. In *Hilbert-Huang Transform and its applications*, Wrld Scientific, New Jersey, 2005.

[28] P. Flandrin, G. Rilling, and P. Gonçalvès. Empirical mode decomposition as a filter bank. *IEEE Sig. Proc. Lett.*, 11(2):112–114, 2004.

[29] P. Flandrin, G. Rilling, and P. Gonçalvès. Sur la décomposition modale empirique. Paris, Septembre 2003.

[30] G. Gonon, S. Montresor, and M. Baudry. Improved entropic gain and adaptive time-frequency segmentation. application to audio coding. In *Proc IEEE EUROSPEECH*, September 2001.

[31] M.M. Goodwin and M. Vetterli. Matching pursuit and atomic signal models based on recursive filter banks. *IEEE Trans. Sig. Process.*, 47(7):1890–1901, 1999.

[32] R. Gribonval and E. Bacry. Harmonic decompositions of audio signals with matching pursuit. *IEEE Trans. Signal Processing*, 51(1):101111, 2003.

[33] N.E. Huang, M.J. Brenner, and L. Salvino. Hilbert-Huang transform stability spectral analysis applied to flutter flight test data. *AIAA Journal*, 44(4):772–786, 2006.

[34] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, and H.H. Liu. The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London*, 454(1971):903–995, 1998.

[35] N.E. Huang, M.L.C. Wu, S.R. Long, S.S.P. Shen, W. Qu, P. Gloersen, and K.L. Fan. A confidence limit for the empirical mode decomposition and hilbert spectral analysis. *the Royal Society of London Mathematical Physical and Engineering Sciences*, 459(2037):2317–2345, 2003.

[36] R. Huber and B. Kollmeie. Pemo-qa new method for objective audio quality assessment using a model of auditory perception. *IEEE Transactions on audio, speech and language processing*, 14(6), 2006.

[37] I.J.Cox, J. Kilian, T. Leighton, and T. Shamoon. A secure, robust watermark for multimedia. *Lecture Notes in Computer-Science*, 1174:185–206, 1996.

[38] ITU-R BS.1387-1 ITU Recommendation. Method for objective measurements of perceived audio quality. 2001.

[39] N. Jayant. Signal compression. *Int. J. High Speed Electron. Syst.*, 8(1):1–12, 1997.

[40] J.D. Johnston. Transfom coding of audiosignals using perceptual criteria. *IEEE. Select Areas Commun.*, 6:314–323, 1988.

[41] V. Bhat K, I. Sengupta, and A. das. An adaptive audio watermarking based on the singular value decomposition in the wavelet domain. *Digital Signal Processing*, 2010(20):1547–1558, 2010.

[42] K. Khaldi, M. Turki-Hadj Alouane, and A.O. Boudraa. A new EMD denoising approach dedicated to voiced speech signals. In *Proc. IEEE SCS*, Hammamet Tunisia, Movember 2008.

[43] K. Khaldi, M. Turki-Hadj Alouane, and A.O. Boudraa. Speech enhancement by adaptive weighted average filtering in the EMD framework. In *Proc. IEEE SCS*, Hammamet, Tunisia, November 2008.

[44] K. Khaldi, M. Turki-Hadj Alouane, and A.O. Boudraa. Voiced speech enhancement based on adaptive filtering of selected intrinsic mode functions. *J. on Advances in Adaptive Data Analysis*, 2(1):65–80, 2010.

[45] K. Khaldi, A.O. Boudraa, A. Bouchiki, and M. Turki-Hadj Alouane. Speech enhancement via EMD. *EURASIP Journal on Advances in Signal Processing*, 2008, 2008.

[46] K. Khaldi, A.O. Boudraa, M.Turki, and T. Chonavel. Codage audio perceptuel à bas débit par décomposition en modes empiriques. In *GRETSI*, Dijon, France, Septembre 2009.

[47] K. Khaldi, A.O. Boudraa, M.Turki, T. Chonavel, and I. Samaali. Audio encoding based on the empirical mode decomposition. In *Proc IEEE EUSIPCO*, Glasgow, Scotland, August 2009.

[48] K. Khaldi, A.O. Boudraa, B. Torrésani, T. Chonavel, and M. Turki. Audio encoding using huang and hilbert transforms. In *Proc IEEE ISCCSP*, Limassol, Cyprus, March 2010.

[49] D. Kiroveski and H.S. Malvar. Spread-spectrum watermarking of audio signals. *IEEE Trans. Signal Processing*, 51(4):1020–1033, 2003.

[50] D. Kiroveski and S. Malvar. Robust spread-spectrum audio watermarking. In *Proc IEEE ICASSP*, Utah, 2001.

[51] H. Laurent and C. Doncarli. Stationarity index for abrupt changes detection in the time frequency plane. *IEEE Signal Proc. Lett.*, 5(2):43–45, 1998.

[52] J.S. Lee. Digital image enhancement and noise filtering by using local statistics. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2(4):165–168, March 1980.

[53] W. Li, X. Xue, and P. Lu. Localised audio watermarking technique robust against time-scale modification. *IEEE Trans. Multimedia*, 8(1):60–69, 2006.

[54] W.N. Lie and L.C. Chang. Robust high-quality time-domain audio watermarking based on low frequency amplitude modification. *IEEE Trans. Multimedia*, 8(1):46–59, 2006.

[55] S. Mallat. *Une exploration des signaux en ondelettes*. Ellipses, 2000.

[56] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Trans. Sig. Process.*, 41:3397–3415, 1993.

[57] M.F. Mansour and A.H. Tewfik. Data embedding in audio using time-scale modification. *IEEE Trans. Speech Audio Processing*, 13(3):432–440, 2005.

[58] R.J. McAulay and T.F. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Trans. Acoustics, Speech and Signal Processing*, 34(4):744–754, 1986.

[59] M. Meguro, A. Taguchi, and N. Hamada. Data-dependent weighted average filterig for image sequence restoration. *Electroics and Communications in Japan, Part III: Fundamental Electronics Science*, 84(4):1–10, 2000.

[60] S. Meigen and V. Perrier. A new formulation for emperical mode decomposition based on constrined optimization. *IEEE Signal Proc. Lett.*, 14(12):932–935, 2007.

[61] A. Mertins. Signal analysis: Wavelets, filter banks, time-frequency transforms and applications. 1999.

[62] M. Misiti, Y. Misiti, G. Oppenheim, and J. M. Poggi. Matlab wavelet toolbox. *Math Works Inc.*, 1996.

[63] P. Noll. Mpeg digital audio coding. *IEEE Sig. Process. Magazine*, 14(5):59–81, 1997.

[64] P. Noll. Mpeg digital audio coding. *IEEE Sig. Proc. Mag.*, 14(5):59–81, 1997.

[65] ITU-T P.835. Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm. *ITU-T Recommendation P.835*, 2003.

[66] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*, volume 1. Cambridge University Press, 2nd edition, 1992.

[67] J.G. Proakis and D.G. Manolakis. *Digital Signal Processing: Principles, Algorithms, and Applications*, volume 1. Prentice-Hall, 3rd edition, 1996.

[68] G. Rilling and P. Flandrin. Décomposition modale empirique : échantillonnage et résolution. In *GdR ISIS, Thème*, Marseille, France, Décembre 2004.Journée : Décompositions adaptatives II.

[69] G. Rilling, P. Flandrin, and P. Gonçalvès. On empirical mode decomposition and its algorithms. In *IEEE-EURASIP, Workshop on Nonlinear Signal and Image Processing, NSIP*, Grado(I), June 2003.

[70] A. Rix, J. Beerends, M. Hollier, and A. Hekstra. Perceptual evaluation of speech quality (pesq) - a new method for speech quality assessment of telephone networks and codecs. *Proc. IEEE ICASSP*, pages 749–752, 2001.

[71] F. Russo. Nonlinear fuzzy filters: An overview. *Proc. IEEE EUSIPCO*, pages 257–260, 1996.

[72] P. Scalart and J.V. Filho. Speech enhancement based on a priori signal to noise estimation. *Proc. IEEE ICASSP*, 2:629–632, 1996.

[73] R.C. Sharpely and V. Vatchev. Analysis of the intrinsic mode functions. *Constructive Approximation*, 24:17–47, 2006.

[74] D. Sinha and A. Tewfik. Low bit rate transparent audio compression using adapted wavelets. *IEEE Trans. ASSP*, 41(12):3463–3479, 1993.

[75] I.Y. Soon, S.N. Koh, and C.K. Yeo. Noisy speech enhancement using discrete cosine transform. *Speech Communication*, 24(3):249–257, 1998.

[76] G. Steidl, J. Weickert, T. Brox, Pavel Mrazek, and M. Welk. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs, 2004.

[77] N. Stevenson, M. Mesbah, and B. Boashash. A sampling limit for the empirical mode decomposition. In *IEEE ISSPA*, Sydney, australia, August 2005.

[78] G. Stoll, S. Nielsen, and L. Van de Kerkhof. Generic architecture of the iso/mpeg audio layer i and ii-compatible developments to improve quality and addition of new features. *Conv. Aud. Eng. Soc.*, 1993.

[79] M.D. Swanson, B. Zhu, and A.H. Tewfik. Robust audio watermarking using perceptual masking. *Signal Processing*, 66(3):337–355, 1998.

[80] R. Tachibana, S. Shimizu, S. Kobayashi, and T. Nakamura. An audio watermarking method using a two-dimensional pseudo-random array. *Signal Processing*, 82(10):1455–1469, 2002.

[81] V. Vatchev and R.C. Sharpely. Decomposition of functions into pairs of intrinsic mode functions. *The Royal Society London A*, 464(2097), 2008.

[82] R.N.J. Veldhuis, M. Breeuwer, and R.G. Van Der Waal. Subband coding of digital audio signals. *Phillips J. Res.*, 44:329–343, 1989.

[83] T. Welch. A technique for high-performance data compression. *Computer*, 17:8–19, 1984.

[84] B. Weng and K.E. Barner. Optimal and bidirectional optimal empirical mode decomposition. *Proc. IEEE ICASSP*, 3:1501–1504, Toulouse, 2007.

[85] B. Weng, M. Blanco-Velasco, and K.E. Barner. Ecg denoising based on the Empirical Mode Decomposition. New York City, USA, August 2006.

[86] Z. Wu and N. E. Huang. Statistical significance test of intrinsic mode functions. In *Hilbert-Huang Transform and its applications*, Wrld Scientific, New Jersey, 2005.

[87] Z. Wu and N.E. Huang. A study of the characteristics of white noise using the empirical mode decomposition method. *Proc. Roy. Soc. London A*, 460:1579–1611, 2004.

[88] S. Xiang, H.J. Kim, and J. Huang. Audio watermarking robust against time-scale modification and mp3 compression. *Signal Processing*, 88(10):2372–2387, 2008.

[89] I.K. Yeo and H.J. Kim. Modified patchwork algorithm: A novel audio watermarking scheme. *IEEE Trans. Speech Audio Processing*, 11(4):381–386, 2003.

APPENDIX A

Chapter III

In this appendix, we give brief descriptions of the quality measures used.

Input Signal-to-Noise Ratio (SNR_{in}): The input Signal to Noise Ratio (SNR_{in}) is given by:

$$\mathrm{SNR_{in}} = 10\log_{10}\frac{\sum_{t=1}^{T}(x(t))^2}{\sum_{t=1}^{T}(y(t)-x(t))^2} \tag{A.1}$$

where x and y are respectively the clean and the noisy signals.

Output Signal-to-Noise Ratio (SNR_{out}): The SNR_{out} is very sensitive to the time alignment of the original and distorted signal. The SNR_{out} is measured as

$$\mathrm{SNR_{out}} = 10\log_{10}\frac{\sum_{t=1}^{T}(\tilde{x}(t))^2}{\sum_{t=1}^{T}(x(t)-\tilde{x}(t))^2} \tag{A.2}$$

where $\tilde{x}$ is the reconstructed signal.

Perceptual Evaluation of Speech Quality (PESQ): The PESQ measure is the most complex to compute, and it is recommended by ITU-T for speech quality assessment of 3.2 kHz (narrow-band) handset telephony and narrow-band speech codec [65]. The note refers PESQ values type Mean Opinion Score (MOS), in the form of a scalar between -0.5 and 4.5.

APPENDIX B

Chapter V

Objective Difference Grade (ODG): The ODG is a perceptual criterion [38], which is located by 5 impairment grade (table B.1).

Table B.1: Impairment grade.

ODG	Impairment	Quality
0	Imperceptible	Excellent
-1	Perceptible, but not annoying	Good
-2	Slightly annoying	Fair
-3	Annoying	Poor
-4	Very annoying	Bad

Résumé - Abstract

Traitement et analyse des signaux sonores par transformée de Huang (EMD)

Résumé:

Dans cette thèse on a exploré l'apport de l'EMD en traitement et en analyse des signaux audio et de parole. Cette décomposition du signal en IMF est adaptative et ne fait pas d'hypothèses (stationnarité et linéarité) sur le signal à analyser. Le comportement en banc de filtre dyadique de l'EMD ainsi que la quasi-symétrie des modes et leur représentation via leurs extrema sont les propriétés qui sont l'origine des outils qu'on a développés: débruitage, codage et tatouage. Ces contributions sont illustrées sur des données synthétiques et réelles et les résultats comparés à ceux de méthodes éprouvées telles que le filtre MMSE, l'approche ondelettes et les codecs AAC et MP3 montrent les bonnes performances des outils développés autour de l'EMD. Ces résultats montrent les capacités de l'EMD comme outils de traitement et d'analyse de façon adaptative des signaux audio et de parole. Même si les outils développés ont été illustrés uniquement sur des signaux 1D, ils peuvent être étendus au cas du traitement des images avec des applications à des domaines variés telsque la biomédecine ou l'imagerie satellitaire.

Mots clés: EMD, Débruitage, Codage, Tatouage, Transformée de Hilbert.

Abstract:

This dissertation explores the potential of EMD as analyzing tool for audio and speech processing. This signal expansion into IMFs is adaptive and without any prior assumptions (stationarity and linearity) on the signal to be analyzed. Salient properties of EMD such as dyadic filter bank structure, quasi-symmetry of IMF and fully description of IMF by its extrema, are exploited for denoising, coding and watermarking purposes. These contributions are illustrated on synthetic and real data and results compared to well established methods such as MMSE filter, wavelets approach, MP3 and AAC codecs showing the good performances of EMD based signal processes. These findings demonstrate the real potential of EMD as analyzing tool (in adaptive way) in speech and audio processing. Although the developed tools are illustrated on 1D signals, they can be easily extended to image processing and find applications in areas such as Biomedicine or Satellite imaging.

Key words: EMD, Denoising, Encoding, Watermarking, Hilbert Transform.