CHAPITRE 2

LA TRANSFORMÉE EN ONDELETTES CONTINUE

Il existe plusieurs méthodes basées sur les analyses spectrales pour étudier la nature évolutive sur le temps ou l'espace de différents phénomènes physiques, chimiques et biologiques.

La méthode la plus répandue est l'analyse de Fourier qui emploie les propriétés de la transformée de Fourier pour fournir l'information fréquentielle et le déphasage d'un signal variant dans l'espace ou le temps. Cette méthode bien que largement utilisée dans différents domaines scientifiques et industriels a montré certaines limitations dans un nombre d'applications où les signaux en cause s'écartaient plus ou moins de la nature du signal stationnaire et prédictif propice à l'analyse de Fourier.

Lorsque l'étude d'un phénomène implique la mesure de signaux quasi stationnaires ou non stationnaires dont les variations temporelles ou spatiales sont à fortes discontinuités, la transformation de Fourier est remplacée par celle à fenêtre glissante. S'il s'agit de signaux quasi stationnaires à discontinuités ponctuelles alors la transformée en ondelettes devient plus appropriée pour ces types de signaux.

La transformation en ondelettes, présentée par Mallat [8] , Goswami et Chan [9], permet un ensemble très diversifié de méthodes que ce rapport ne pourra aborder entièrement. Cependant il est utile dès à présent de mentionner la distinction fondamentale de principe qui existe entre les transformations en ondelettes et de Fourier.

La transformation de Fourier permet l'analyse et la synthèse d'un signal à partir de ses composantes élémentaires que sont les sinusoïdes modulées en fréquences et en phase.

La transformation en ondelettes permet d'analyser et de synthétiser un signal à partir de composantes élémentaires que sont les ondelettes. Ces dernières sont généralement des

signaux oscillatoires d'énergie finie et de durée également finie dont l'intégrale est nulle. Les ondelettes sont caractérisées par deux paramètres. L'un est associé à l'échelle ou à l'étendue temporelle de l'ondelette, l'autre est associé à la translation temporelle (ou spatiale) de l'ondelette par rapport à un moment donné.

2.1 La transformée de Fourier à fenêtre glissante (TFFG)

Lorsqu'on veut localiser ou détecter un changement dans un signal, on a recours à l'analyse de la transformée de Fourier à fenêtre temporelle glissante (TFFG). L'idée de base de la TFFG est d'utiliser une fonction fenêtre $\varphi(t)$ de telle sorte que les largeurs des fenêtres tant temporelles $\Delta\varphi(t)$ que fréquentielles $\Delta F(\varphi(t))$ seront bornées ou à support compact. Cela permet d'obtenir la réponse fréquentielle à un temps donné. Le principe d'incertitude d'Heisenberg limite la borne inférieure du produit de la largeur des fenêtres temporelles et fréquentielles; il doit être supérieur à ½. Le produit sera égal à ½ seulement si la fonction fenêtre est gaussienne. Un des exemples de la TFFG est la transformée de Gabor qui utilise une fonction gaussienne. La fenêtre d'analyse étant fixe, ce procédé ne peut localiser (dans le temps) toutes les gammes de fréquences de façon égale.

2.2 Transformée en ondelettes continue (TOC)

Comme dans le cas de la TFFG, une des caractéristiques maîtresses de la transformée en ondelettes continue est de pouvoir localiser un phénomène transitoire ou un changement brusque dans un signal. Contrairement à la TFFG, la fenêtre d'analyse s'allonge s'il s'agit de localiser les basses fréquences et se rétrécit pour résoudre les hautes fréquences. L'ondelette est une fonction localisée dans le temps et contenant des oscillations. Une ondelette ne peut être choisie arbitrairement, car après la transformation, il faut pouvoir reconstituer le signal. Il faut aussi que les fonctions d'intérêt puissent être représentées par une combinaison d'ondelettes dilatées et décalées. La condition suffisante d'admissibilité est que la moyenne temporelle d'une ondelette soit nulle et d'énergie finie (et supérieure à 0). Cela signifie que le spectre

d'une ondelette est similaire à un filtre passe haut puisque la moyenne temporelle est nulle. La transformée continue en ondelettes est donnée par:

$$W_{\psi}f(b,a) = \int_{-\infty}^{\infty} f(t)\overline{\psi_{b,a}(t)}dt \tag{2.1}$$

où l'ondelette s'exprime par :

$$\psi_{b,a}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{2.2}$$

Remarquons la présence du trait haut signifiant le complexe conjugué, car certains types d'ondelettes sont complexes. Si on substitue l'équation 2.2 dans l'équation 2.1, on constate qu'il s'agit d'une corrélation du signal $f(t)$ avec la fonction $\Psi(t)$, celle-ci étant faite à l'échelle a. On peut aussi dire que l'équation 2.1 représente la convolution de $f(t)$ avec la fonction $\Psi_{b,a}(t)$ involuée, c'est à dire $\Psi_{b,a}(-t)$.

Les paramètres b et a sont des paramètres de translation et de dilatation (ou contraction) respectivement. La transformation engendre donc une fonction à deux variables. En fait, on peut considérer que la transformation en ondelettes continue est constituée par les sorties d'une suite continue de filtres empilés les uns sur les autres sur l'axe vertical des échelles alors que le temps est sur l'axe horizontal du plan. Chacun des filtres est la transformée de Fourier de l'ondelette à l'échelle appropriée selon notre position dans l'axe.

Dans la pratique, la convolution du signal avec l'ondelette peut être effectuée à l'aide d'une transformée de Fourier discrète (TFD) du signal, suivie d'une multiplication par la TFD de l'ondelette discrétisée, et suivie finalement par une transformée de Fourier

inverse de ce produit. Ces opérations sont faites pour chaque échelle *a*. La transformée de Fourier inverse selon les auteurs Goswami et Chan [9] étant :

$$W_n(a) = \sum_{k=0}^{N-1} F_k \psi(aw_k) e^{i2\pi kn/N} \tag{2.3}$$

Où la DFT du signal est :

$$F_k = \frac{1}{N} \sum_{n=0}^{N-1} f_n e^{-i2\pi kn/N} \tag{2.4}$$

Pour assurer une comparaison des transformées entre les échelles, il y a une normalisation préalable de l'ondelette pour obtenir une énergie unitaire à chaque échelle.

Le résultat de la transformée en ondelettes est un coefficient d'ondelette à une échelle *a* et à une translation *n* données. Les valeurs des coefficients représentent la contribution relative d'une ondelette avec une contraction donnée et une translation donnée à la composition du signal.

La transformée en ondelettes inverse est mathématiquement donnée par :

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} db \int_{-\infty}^{\infty} \frac{1}{a^2} \left[W_\psi f(b,a) \right] \psi_{b,a}(t) da \tag{2.5}$$

Poursuivant l'analogie avec les bancs de filtres, on peut considérer la transformée inverse conne étant une banque de filtres superposés sur l'axe vertical et permettant de récupérer le signal d'origine si le produit du filtre de transformation avec celui de synthèse donne 1.

Plus l'ondelette est contractée dans le domaine du temps plus son spectre est dilaté dans celui de la fréquence. L'avantage principal de la transformée en ondelettes continue est d'obtenir un produit de largeur des fenêtres (temps * fréquence) constant. Cela permet de localiser un changement si, à un instant donné, on considère la valeur des coefficients d'ondelettes sur l'axe vertical. L'ondelette temporelle se contracte au fur et à mesure que l'on remonte l'axe vertical, ce qui permet une localisation adéquate de toutes les fréquences à tout instant.

L'interprétation d'une transformée en ondelettes continue est différente de celle donnée par la TFFG. Puisque la transformée de Fourier d'une ondelette est équivalente à un filtre passe-haut, une sinusoïde pure sera représentée par une bande plus ou moins large selon la largeur de la fenêtre fréquentielle, au lieu d'une seule ligne comme dans la TFFG. De plus l'axe vertical n'est pas un axe de fréquence mais plutôt un axe d'échelle. Il est possible de relier l'échelle et la fréquence si on connaît le type d'ondelette utilisée, donc son spectre. En outre, la localisation d'une discontinuité temporelle, qui contient toutes les fréquences, sera moins bonne aux basses fréquences, la fenêtre temporelle étant plus large.

La caractérisation de signaux avec cette méthode peut se faire de deux manières. L'expérience de l'usager lors de l'examen du plan de la transformation, ou des tests statistiques sur la puissance des coefficients par rapport à ceux obtenus avec du bruit blanc. Bien que cette approche soit possible, nous centrons notre propos sur la décomposition du signal dans le cadre de ce projet.

CHAPITRE 3

LA MÉTHODE DU REHAUSEMENT DE LA PAROLE

La recherche sur le projet s'est basée sur le document en référence [10] des auteurs Barros, Rutkowski, Itakura et Ohnishi dont le titre est : « Estimation of speech embedded in a reverberant and noisy environment ». Dans ce document on montre un système pour rehausser la parole représentant la plus grande énergie baignée dans un environnement de bruits additifs et convolutifs où s'ajoutent d'autres sources vocales ou d'autres perturbations. Le système illustré à la figure 6, est décrit par 4 blocs de processus dont les fonctions principales sont les suivantes :

- Deux microphones distancés l'un de l'autre, captent l'ensemble des signaux présents. Il y a donc deux canaux dont les signaux à traiter sont $X_1(t)$ et $X_2(t)$.
- Le bloc SIF (speech instantaneous frequency) permet à partir des signaux $X_1(t)$ et $X_2(t)$ de déterminer le ton de la voix (pitch) ou encore la fréquence fondamentale f0 de la parole qui sera nécessaire pour les autres blocs en aval. La méthode se base sur un algorithme utilisant une analyse spectrale et un filtre avec une ondelette de type Gabor pour estimer la fréquence f0 en employant la transformée de Hilbert pour déduire la fréquence instantanée de la parole.
- Un banc de filtres adaptatifs de type passe-bande centrés sur un nombre fini d'harmoniques de la f0 décomposent en sous-bandes les signaux venant des deux canaux. Le traitement du signal travaillé ainsi en bandes étroites permet de réduire voir supprimer les effets convolutifs inhérents à ce genre d'application. Chaque sous-bande en sortie d'un passe-bande aura un signal $r_{i,k}(t)$ où l'indice i et k sont respectivement le canal et la sous-bande associée à la fréquence fondamentale f0 ou à une de ses harmoniques kf0.
- Le bloc ACI (Analyse en Composantes Indépendantes) est constitué de plusieurs fonctions. À partir de chaque $r_{i,k}(t)$ une transformée de Hilbert est faite pour obtenir l'enveloppe du signal $\hat{A}_{i,k}(t)$ qui va permettre de moduler une pure harmonique à la fréquence de f0 ou kf0 selon la sous-bande. Ce signal synthétisé

$Z_{i,k}(t)$ doit se substituer à $r_{i,k}(t)$ et pour se faire, il doit recouvrer la phase de ce dernier au moyen d'un filtre adaptatif de Wiener. Nous obtenons alors $x_{i,k}(t)$ un signal synthétisé avec la correction de phase remplaçant $r_{i,k}(t)$. Pour chaque sous-bande, un traitement statistique et adaptatif va chercher les éléments indépendants pour ne garder que ce qui sont liés à f0 ou à kf0 donc les plus susceptibles d'être les constituants de la parole représentée par f0 ou kf0. On reconstitue de la sorte la parole recherchée épurée des éléments étrangers en sommant la contributions de toutes les sous-bandes des deux canaux.

- Le dernier bloc agit principalement en tant qu'algorithme de coordination entre les différents blocs mentionnés en maintenant à jour les éléments psycho-accoustiques.

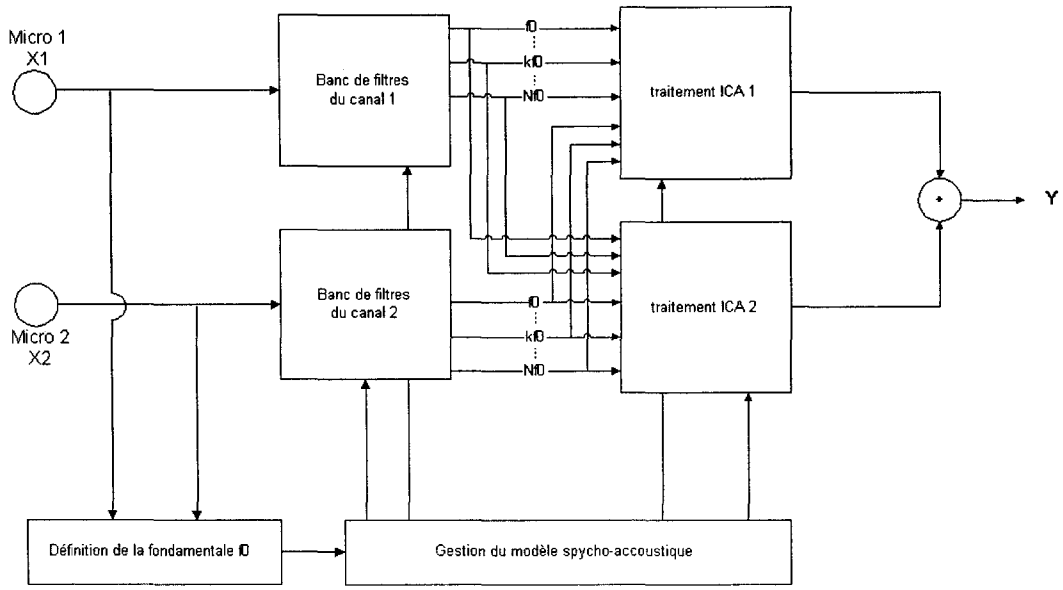

Figure 6 Schéma synoptique du rehaussement de la voix selon [10]

Le document [10] bien que présentant les grandes lignes du projet, reste toutefois très succinct sur les détails de sa réalisation et quelquefois certains éléments sont totalement absents des publications. Nous avons donc choisi de partir sur les bases du système

suggéré tout en adoptant des solutions originales aux niveaux des différents blocs fonctionnels lorsque la situation l'exigeait.

3.1 La détermination de la fréquence fondamentale par ondelettes

L'algorithme de détection et de détermination de la fréquence fondamentale f0 de la parole que nous avons privilégié, suite aux études précédentes, est basé sur l'utilisation des ondelettes.

Les méthodes faisant appel aux propriétés des ondelettes s'appuient sur le fait que la parole est un signal non stationnaire quasi-périodique avec d'importantes discontinuités. Ce genre de signal est réputé plus propice à l'analyse aux moyens d'ondelettes qui possèdent un comportement plus robuste vis-à-vis du bruit additif et convolutif. Parmi les méthodes présentées ici, nous avons retenu certains de leurs éléments pour mettre en œuvre notre méthode qui détermine la f0. Nous allons maintenant détailler leurs principaux éléments.

Nous sommes partis de la référence [11] des auteurs Kadambe et Boudreaux-Bartels qui emploient une transformée en ondelettes de type dyadique (TOD) car son facteur d'échelle est une grandeur à la base deux qu'on peut ajuster en intervenant sur sa puissance appartenant aux entiers relatifs (voir équation 3.1). Cette ondelette est selon les auteurs [11] propice pour les f0 hautes et basses et présente une robustesse au bruit. Elle est proposée avec une performance considérée supérieure aux méthodes classiques basées sur les méthodes de l'autocorrélation et de la cepstrale.

La durée variable de la période de f0, soit 1/f0 peut se située entre 1,35 ms et 40 ms et dépend largement de facteurs psychologiques et physiologiques du locuteur. Une représentation temps-échelle de la transformée en ondelettes de type dyadique (TOD) devra localiser la fermeture glottique et déduire l'intervalle de cet événement répétitif qui définie la période de f0. La TOD du signal x(t) est définie par :

$$TOD_x\left(b,2^j\right)=\frac{1}{2^j}\int_{-\infty}^{+\infty}x(t)\,g^*\left(\frac{t-b}{2^j}\right)dt=x(t)*g^*_{2^j}(t) \tag{3.1}$$

et se calcule en employant un facteur d'échelle a = 2^j qui est discrétisé sur la séquence dyadique. Ensuite, la fonction d'ondelette complexe conjuguée *g*(t)* doit satisfaire les conditions mentionnées ci-après. Cette fonction s'exprime comme :

$$g_{2^j}(t) = \frac{1}{2^j}\, g\left(\frac{t}{2^j}\right) \tag{3.2}$$

D'un point de vue du traitement de signal, la TOD peut être vue comme le résultat de la sortie de bancs de filtres multibandes composés de passe-bandes en octave dont la réponse impulsionnelle est la fonction d'ondelette. La largeur de bande et la fréquence centrale de chaque passe bande sont proportionnelles à $1/2^j$. Chaque échelle ainsi correspond à des bancs de fréquences qui permet de ressortir le contenu fréquentiel du signal aussi les fonctions d'ondelettes peuvent être assimiler à des filtres passe-bandes dans le domaine fréquentiel.

La TOD a comme propriétés la linéarité et l'invariance sur sa translation temporelle que partagent également les signaux vocaux dont la modélisation est souvent une combinaison linéaire avec une translation temporelle de sinusoïdes amorties. Si le signal x(t) ou ses dérivées possèdent des discontinuités alors la TOD de x(t) affichera en module des maxima locaux autour de ces discontinuités. Cette particularité prouvera son utilité dans la détection de la f0 puisque la fermeture glottique correspond à une variation brusque du débit de l'air et fait ressortir le caractère transitoire du signal de la parole.

Mallat qui est cité par les auteurs de la référence [11] a montré que dans notre genre d'application le choix du type d'ondelette est important. Si l'on choisit une fonction

d'ondelette g(t) qui est la première dérivée d'une fonction lisse c'est à dire dont la transformée de Fourier possède une énergie principalement concentrée sur la région des basses fréquences alors la TOD montrera des maxima locaux sur les variations brusques du signal et des minima locaux pour des variations lentes du même signal. De plus, on remarque lorsque survient une très forte variation du signal au moment précis t0, la TOD donne un maximum local et ce pour plusieurs échelles dyadiques consécutives à ce même t0. Cette dernière observation permet d'élaborer des algorithmes qui mettront en corrélation les maxima locaux de la TOD sur plusieurs échelles consécutives.

3.2 L'algorithme de la TOD pour la détection de la fondamentale f0

La TOD sur un segment du signal de la parole étudiée d'une longueur L ms est calculée à l'échelle $a = 2^i$, où i = ib, ib +1, ..., ih. Ici ib et ih sont respectivement les limites basse et haute que peut prendre i.

Pour chaque échelle 2^i, on trouve les maxima locaux par rapport à b de la TOD $(b, 2^i)$ qui dépasse un seuil donné T, ici son niveau correspond à 80% du maximum global de la TOD. Si la localisation des maxima locaux coïncide sur au moins deux échelles alors on suppose qu'il s'agit d'un moment de fermeture glottique. Par la suite nous estimons la période de la fréquence fondamentale f0 de la parole en mesurant l'intervalle temporel entre de tels maxima locaux consécutifs.

En théorie le nombre d'échelle est infini, par contre pour réduire le traitement du calcul nous limiterons ce nombre à une grandeur adéquate pour l'obtention de la TOD en se basant essentiellement sur la nature de la voix, également de durée finie. Dans ce contexte le nombre d'échelle dyadique suffisant [11] pour estimer la période de f0 à partir de la TOD est fixé à trois. Les échelles vont de 2^3 à 2^5 et sachant que l'échelle est inversement proportionnelle à la fréquence, les échelles choisies des ondelettes doivent couvrir le domaine spectral contenant la plage des fréquences fondamentales possibles.

La plus grande échelle tient compte de la volonté de retenir la bande des basses fréquences où se situe la f0. Ainsi nous calculons la TOD en commençant par l'échelle inférieure et nous doublons à chaque itération jusqu'à l'obtention de l'échelle supérieure. Évidemment le fait d'avoir uniquement trois échelles réduit la complexité du calcul de la TOD.

Les auteurs Kadambe et Boudreaux-Bartels [11] ont comparé les performances entre différentes méthodes d'estimation de la f0. La méthode d'estimation de la période de la f0 employant la TOD a été comparée aux méthodes les plus courantes employant l'analyse cepstrale et l'autocorrélation du signal.
La méthode d'analyse cepstrale d'un signal périodique aboutit à la même périodicité que le signal considéré. Le cepstral d'un signal x(t) est défini selon [11] comme :

$$c_x(t)=\left[\int_0^{\infty} \log\left|X(\omega)\right|^2 \cos(\omega t)\, d\omega\right]^2 \tag{3.3}$$

Quant à l'autocorrélation Rx(τ) du signal x(t), elle est définie par :

$$R_x(\tau)=\int_{-\infty}^{+\infty} x^*(t)\, x(t+\tau)dt \tag{3.4}$$

L'autocorrélation d'un signal périodique donne également une périodicité identique à celle du signal considéré.

Les deux méthodes déduisent la f0 en mesurant la durée entre les maxima consécutifs qui donne la période de la f0. C'est en général le premier maximum après celui placé à l'origine de l'abscisse qui donne la valeur prépondérante de la f0.

La comparaison s'est fait sur la précision de l'estimation de la période de la f0, sur la robustesse du traitement pour un signal dans un environnement bruité, sur la complexité

du calcul exigée par l'algorithme et finalement sur la facilité de choisir différentes segmentations du signal de la parole pour accomplir le traitement.

Lorsque f0 appartient au domaine des basses fréquences, les trois méthodes s'équivalent dans la précision de l'estimation de la période du ton de la voix. On constate lors des variations en sauts discontinus de la période de la fondamentale que la TOD donne un excellent résultat alors que les autres méthodes présentent leur pire performance. Cela est dû au fait que ces dernières méthodes font l'hypothèse de la stationnarité du signal dans la fenêtre d'analyse pour fixer la valeur de f0 et par conséquent donne une moyenne de la période estimée dans un segment du signal où les non-stationnarités sont noyées.

Toujours selon les auteurs Kadambe et Boudreaux-Bartels [11], lorsque le signal analysé est bruité avec des rapports signal/bruit (RSB) allant de 0 dB à –18 dB, c'est la TOD qui présente la meilleure précision (erreur relative de $\leq$ 2% pour un RSB de –18 dB) sur l'estimation de la période. Les autres méthodes affichent leur meilleure performance pour un RSB à 0 dB avec des erreurs relatives de 17% et 52% respectivement pour la méthode du cepstrum et de l'autocorrélation, lorsque le RSB atteint –18 dB les résultats se dégradent allant jusqu'à 77 % et 80 %.

3.3 La complexité du calcul dans les méthodes abordées

La méthode la moins exigeante en terme d'opérations et par conséquent la plus rapide d'exécution est l'autocorrélation, ensuite vient après la TOD. Ces deux méthodes impliquent principalement des sommes de produits. La plus complexe et exigeante en terme d'opérations, est la méthode de la cepstrale car elle demande une série d'opérations élaborées dont une transformée de Fourier rapide suivie du calcul du logarithme de la puissance spectrale pour finalement se terminer par une transformée de Fourier inverse.