

3.4 Le choix de la segmentation du signal de la parole

La segmentation ou le fenêtrage du signal de la parole analysée dans les méthodes observées doit être considéré avec précaution. Selon le choix de la méthode, plus particulièrement dans l'autocorrélation et la cepstrale, son importance est majeure. Ces deux méthodes font l'estimation moyenne de la période de f_0 sur un segment de longueur L échantillons et pour cela il leur faut au moins détecter deux principaux événements pour pouvoir définir une période sur une segmentation donnée. Si la segmentation est trop courte, l'algorithme ne pourra évaluer avec précision la période et si la segmentation est trop grande les non-stationnarités du signal seront camouflées par une moyenne des événements périodiques évolutifs.

Cependant dans le cas de la TOD la longueur de la segmentation est moins importante, car elle procède sur la reconnaissance de l'instant de la fermeture glottique uniquement. Avec un signal synthétisé dont la $1/f_0 = 10$ ms, différentes segmentations ont été testées où L valait 40, 25.6, et 12.8 ms, l'erreur relative de l'estimation de la période donnait moins de 2% pour les premières valeurs de L et moins de 6% pour la dernière. Cette dégradation de la précision avec la diminution de L est due aux effets de bord. Dans le cas d'une ondelette non causale, l'algorithme ne détient pas suffisamment de données pour adéquatement calculer la TOD lorsque le segment est plus petit que la moitié de la longueur de l'ondelette LW , il faut opter pour le recouvrement partiel des segmentations au moins l'équivalent à $L_w/2$ pour se libérer de cette limitation.

3.5 Étude de la parole non synthétisée

Deux voix, féminine et masculine, ont été analysées par les auteurs [11] à l'aide d'une fréquence d'échantillonnage $f_e = 16$ KHz, la période de la f_0 était estimée en segmentant la parole par une fenêtre rectangulaire de longueur $L = 32$ ms où la TOD est réalisée. Pour éviter les effets de coupure sur la convolution, le calcul se fait sur un bloc inclus

dans le segment dont les limites temporelles sont $[LW/2, L - LW/2]$. Le segment suivant pris en charge est décalé par un saut $S = LW/2$.

Les voix ont aussi été soumises aux deux autres méthodes avec les mêmes paramètres de saisie. Lorsqu'il s'agissait de la voix masculine, les trois méthodes donnaient une estimation de la période de f_0 similaire. Dans le cas de la voix féminine, sur les parties de signaux vocaux, l'autocorrélation et la TOD rendaient un résultat similaire cependant la méthode du cepstrum donnait une estimation plus changeante. Sur les parties de signaux non vocaux les méthodes d'autocorrélation et de la cepstrale comportaient des erreurs quand venait le temps de distinguer les parties vocales des non vocales surtout autour des 200 à 220 ms et 260 à 300 ms. La méthode de la cepstrale rencontre des difficultés de traitement lorsque les parties vocales de la voix féminine contiennent un contenu spectral riche en harmoniques de la f_0 , d'où son choix ardu à déterminer f_0 .

La TOD offre également une performance supérieure pour déterminer la f_0 indépendamment de la nature féminine ou masculine de la voix. La méthode de la cepstrale se comporte bien avec des voix masculines où la f_0 appartient au domaine spectral des basses fréquences de la voix. L'autocorrélation se comporte bien avec des voix féminines où la f_0 appartient au domaine spectral des hautes fréquences de la voix.

Par contre, l'autocorrélation comme la cepstrale n'arrivent pas à détecter les variations aussi facilement que la TOD qui contrairement aux autres méthodes n'assume pas à priori dans son traitement la stationnarité ou quasi-stationnarité à l'intérieur de la fenêtre d'analyse.

3.6 Conclusion sur les performances de la TOD

La TOD représente selon les auteurs Kadambe et Boudreaux-Bartels [11] et [12] un excellent choix comme méthode pour la détection de la fondamentale de la parole. Ses points forts par rapport aux autres méthodes approchées sont les suivants :

- Elle ne fait pas l'hypothèse de la stationnarité ou de la quasi-stationnarité dans le signal analysé
- Elle offre une estimation très précise dans un signal bruité (erreur relative maximale de 2 % pour un RSB de -18 dB en essai de simulation)
- Elle localise parfaitement le début de toute période de fondamentale f_0 et les successives dans un segment donné devenant ainsi un choix idéal pour des applications faisant intervenir des méthodes de synchronisation.
- Elle est idéale pour une grande gamme de période de fondamentale traitant également les voix féminines ou masculines.
- Sa mise en œuvre est relativement simple en terme de calcul et de traitement numérique car elle se contente de deux ou trois échelles d'ondelettes pour obtenir ses résultats.

3.7 Les variantes de l'algorithme de la TOD

La plupart des experts [11] et [12] qui adoptent la solution de la TOD sont d'avis que le type d'ondelette le plus propice à ce genre d'application est le B-spline. D'autres auteurs tels que Qiu, Koh et Yang de la référence [13] ont proposé une ondelette à leur avis plus facile à construire et cependant s'apparentant et conservant les caractéristiques de la famille des splines. Cette fonction est impaire et bien localisée dans le domaine fréquentiel et temporel. Elle permet donc d'avoir :

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0 \quad (3.5)$$

L'approche de ces derniers auteurs [13] est substantiellement différente de l'algorithme présenté au début [11]. De plus ils ne se contentent pas d'utiliser la TOD dans le domaine temporel uniquement mais aussi dans le fréquentiel.

Il faut remarquer que même l'emploi de la TOD dans le domaine temporel est modifié. Plutôt que de soumettre en analyse un segment de la parole sur trois différentes échelles

dyadiques par exemple, il débute l'analyse du segment à l'échelle 2^j où $j = 1$ et c'est le résultat obtenu du traitement de la TOD qui est alors soumis de nouveau à celle-ci mais cette fois à l'échelle incrémentée, soit $j + 1$, le résultat suivant est alors traité par la TOD à l'échelle suivante $j = 3$. Le résultat final est supérieur avec cet algorithme qui élimine rapidement les détails du signal analysé avec l'emploi récursif de la TOD sur les résultats successifs plutôt que sur le signal original.

La fondamentale peut aussi être déterminée dans le domaine fréquentiel à l'aide de la TOD. La fréquence fondamentale est mesurée par la distance séparant deux crêtes spectrales adjacentes. A partir d'un segment de voix on extrait son logarithme spectral qui est lui traité par la TOD cette fois sur une seule échelle $j = 3$.

Dans le cas temporel ou fréquentiel le facteur de translation de l'ondelette reste ici inchangé soit $\tau = 0$.

Les auteurs G.A. Shelby, C.M. Cooper et R.R. Adhami [14] ont suggéré de redresser le signal à analyser c'est-à-dire ne retenir que la partie positive du signal, cela réduit la quantité d'information à traiter sans pour autant nuire au traitement de la TOD pour déduire la période de f_0 .

3.8 Notre choix d'ondelettes

Le présent document a gardé les échelles de la première méthode [11]. Nous n'avons pas retenu l'approche des auteurs [10] sur la méthode choisie pour la détection de la f_0 car selon [15] la transformée de Hilbert pour trouver la fréquence instantanée dans un milieu bruité n'est pas une méthode robuste.

Cependant nous garderons la transformée de Hilbert pour chercher l'enveloppe d'un signal dans la partie ICA vue plus loin. Nous avons choisi l'algorithme de la deuxième

méthode dans le domaine temporel [13] pour trouver la f_0 , par contre les choix de l'ondelette de ces auteurs n'ont pas été retenus.

Bien que respectant le critère du choix de l'ondelette qui demande une ondelette s'apparentant au signal à analyser, nous avons préféré la première dérivée d'une gaussienne comme illustrée sur la figure 7 et recommandée par [13].

Elle s'apparente effectivement très bien au signal de la fermeture glottique de la voix. L'idée du redressement [14] du signal a été retenue dans l'évaluation des maxima locaux des fenêtres sur signal à analyser par la TOD.

Cette ondelette choisie va être contractée ou dilatée selon l'échelle retenue. Comme suggéré par les auteurs [11], [13] nous avons gardé trois échelles dyadiques de telle sorte qu'elles s'insèrent dans la bande spectrale de la fréquence fondamentale de la voix humaine.

La figure 8 donne l'occupation du domaine spectrale de l'ondelette pour les trois échelles choisies.

Nous avons repris l'exemple du signal utilisé par les méthodes classiques de détection de la fréquence fondamentale f_0 vues au chapitre 2. Les résultats sont montrés sur la figure 9, où nous pouvons voir comment la f_0 est déterminée sur le signal analysé.

La figure 9a est le signal analysé qui est traité par la transformée en ondelettes sur les trois échelles consécutives. Le résultat est montré sur la figure 9c où l'on remarque des maxima claires sans ambiguïté. L'algorithme par la suite relève la durée entre chaque maximum donnant la période de la f_0 . La figure 9b est le résultat de ce calcul.

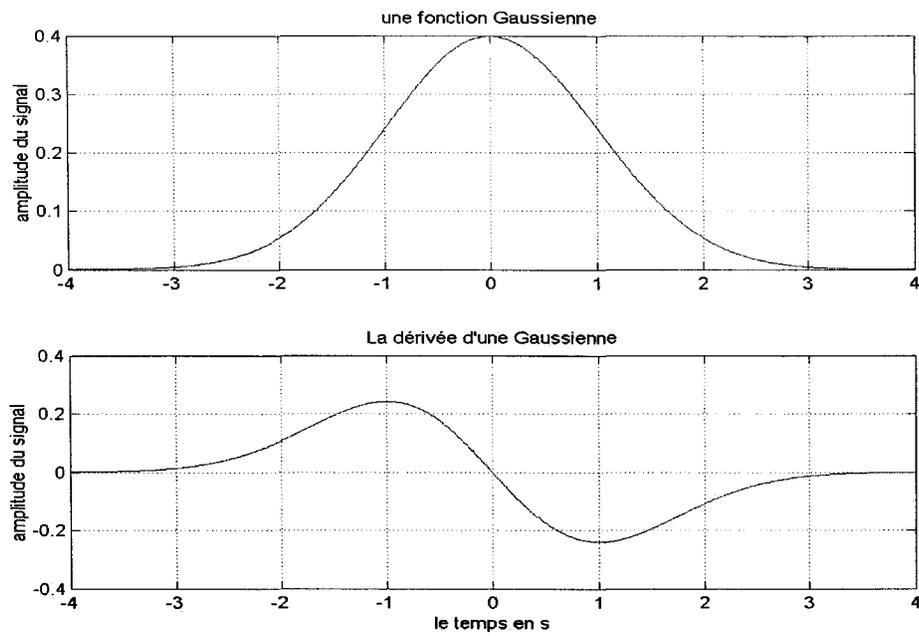


Figure 7 La Gaussienne et sa dérivée comme ondelette mère

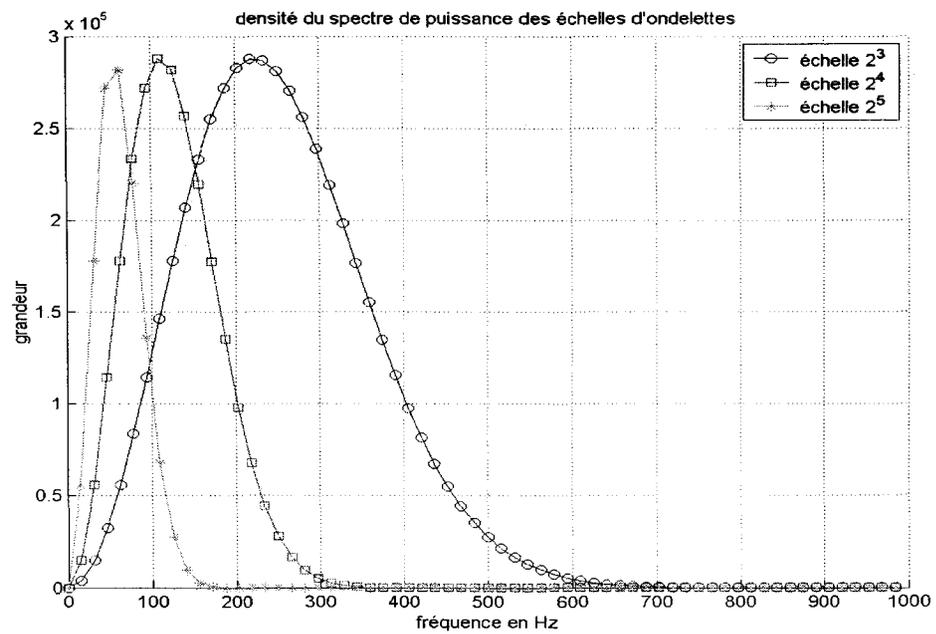


Figure 8 Domaine spectrale des trois échelles d'ondelette

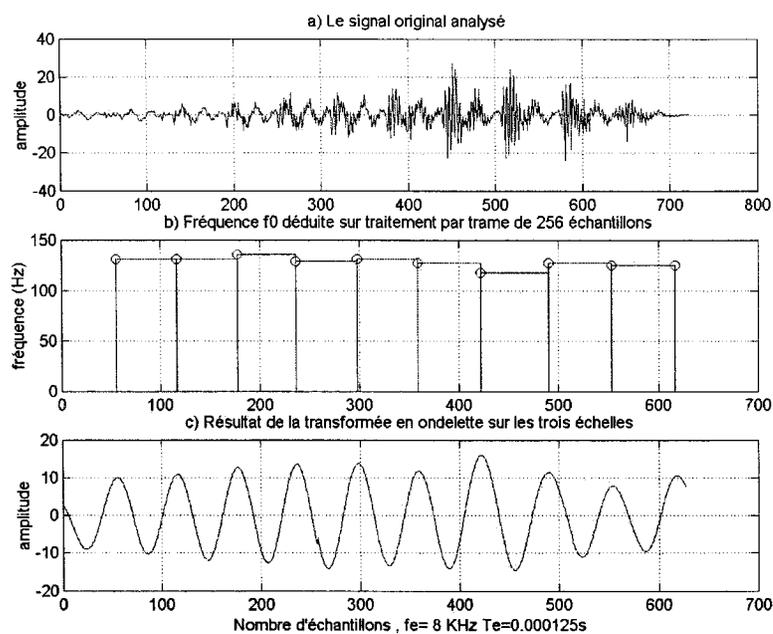


Figure 9 Détermination de la f_0 par transformée en ondelettes

Chaque trait vertical de la figure 9b correspond à un maximum de la figure 9c et entre deux maxima on les relie par un trait horizontal dont la projection sur l'ordonnée donne la fréquence de la f_0 qui varie légèrement entre 120 Hz et 135 Hz.

Les grandes lignes de l'algorithme sont montrées avec les principales opérations sur la figure 10. Cet algorithme est mis en œuvre plus en détail dans les programmes tant de simulation que d'application en langage C qu'on peut consulter respectivement dans les annexes I et II.

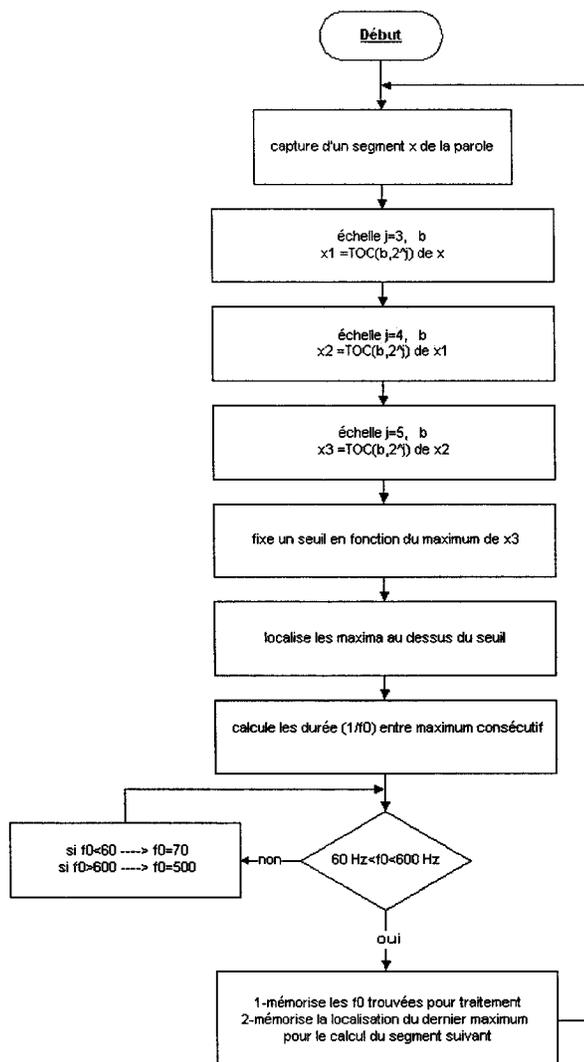


Figure 10 L'algorithme de détection de la f_0 avec TOC sur trois échelles

3.9 le banc de filtres de type passe-bande adaptatif

La démarche consiste à recourir au principe des liens harmoniques qu'on retrouve particulièrement dans tout son voisé dont la nature bien que non-stationnaire puisse être partitionnée pour relever les éléments stationnaires ou plus précisément quasi-périodiques. Pour se faire, on utilise un banc de filtres de type passe-bande centrés sur la fréquence fondamentale f_0 et ses harmoniques. Nous obtiendrons ainsi selon la f_0

mesurée, un nombre de sous-bandes du signal étudié reliées aux harmoniques de la f_0 et restant à l'intérieur de la fréquence de Nyquist.

C'est par ce contenu fréquentiel en harmoniques que le système auditif humain est selon la théorie retenue [10], le plus susceptible de distinguer les sons voisés issus d'un milieu bruité. Il est à remarquer que la largeur de bande des filtres du banc est dans l'ordre de grandeur de la fréquence centrale des passe-bandes. Cela confère à cette phase du traitement du signal un comportement similaire à celui qu'on retrouve dans le système auditif humain de la cochlée [10] pour ce qui est de la non-linéarité d'échelle de la réponse aux stimuli sonores.

3.10 Construction des passe-bandes du banc de filtres

Les filtres de type passe-bande sont élaborés selon l'information de la référence [10] qui veut simuler la réponse du système auditif humain où la bande passante de chaque filtre est proportionnelle à la f_0 . les filtres vont s'adapter à la fréquence fondamentale f_0 du moment de la façon suivante : le premier passe-bande va évaluer sa fréquence centrale à celle de la f_0 , qui est fixée approximativement sur la moyenne géométrique des fréquences de coupure du passe-bande. En définissant par α la fréquence de coupure basse et β la fréquence de coupure haute, nous pouvons résumer ainsi les particularités du premier passe-bande :

$$f_0 = \sqrt{\alpha \beta} \quad (3.6)$$

$$\beta = 2 \alpha \quad (3.7)$$

$$LB = \beta - \alpha = \alpha \quad (3.8)$$

$$\alpha = \frac{f_0}{\sqrt{2}} \quad \text{et} \quad \beta = \sqrt{2} f_0 \quad (3.9)$$

Le passe-bande suivant fera coïncider sa fréquence de coupure basse à celle de coupure haute du premier passe-bande autrement dit $\beta_0 = \alpha_1$. Dès lors la fréquence de coupure haute du passe-bande suivant est déduite de la relation (3.7), il en est de même de la fréquence centrale par la relation (3.6) qui est aussi une harmonique de la f_0 . Finalement la largeur de bande du passe-bande suivant sera le double de la largeur de bande du premier.

Ces relations sont conservées pour tous les passe-bandes qui se suivent. Le nombre de sous-bandes sera égal au nombre de passe-bandes permises par le domaine fréquentiel que la fréquence de Nyquist délimite.

Le banc de filtre est lié à la fréquence fondamentale f_0 et sa structure se présente dans le tableau I :

Tableau I

Structure du banc de filtre

Les n bande-passantes	La fréquence centrale (Hz)	α coupure basse	β coupure haute	LB= β - α Largeur de bande
0	f_0	$\alpha_0 = f_0/2^{1/2}$	$\beta_0 = 2^{1/2} f_0$	$LB_0 = \alpha_0$
1	$2f_0$	$2\alpha_0$	$2\beta_0$	$2LB_0$
2	$4f_0$	$4\alpha_0$	$4\beta_0$	$4LB_0$
3	$8f_0$	$8\alpha_0$	$8\beta_0$	$8LB_0$
n	$2^n f_0$	$2^n \alpha_0$	$2^n \beta_0$	$2^n LB_0$

Le nombre n de passe-bande du banc de filtre, peut être déduit connaissant la fréquence fondamentale f_0 et la fréquence de Nyquist f_{Nyq} puisque la contrainte de Nyquist impose que :

$$2^n \beta_0 \leq f_{Nyq}$$

$$n \leq \frac{\ln\left(\frac{f_{Nyq}}{f_0\sqrt{2}}\right)}{\ln 2} \quad (3.10)$$

3.11 Le filtre numérique RIF pour les passe-bandes

Le filtre numérique RIF à réponse impulsionnelle finie (appelé également non récursif) est le type qui sera retenu pour notre application car malgré une demande substantielle en calcul comparativement aux autres filtres numériques, il offre deux qualités incontournables qui sont une stabilité intrinsèque alliée à la garantie d'absence de distorsion de phase lorsque ses coefficients sont agencés dans un ordre symétrique. Les lignes qui suivent présentent brièvement les principales propriétés du filtre RIF. La section actuelle traitant de l'élaboration des filtres en général et du banc de filtres en particulier s'est surtout appuyée sur la référence [16] du traitement numérique des signaux.

L'équation aux différences (EAD) d'un filtre RIF consiste à modéliser la sortie en terme des entrées présentes et passées, de là sa nature non récursive.

$$S(n) = b_0 E(n) + b_1 E(n-1) + \dots + b_N E(n-N) \quad (3.11)$$

La transformée en Z de l'EAD prend la forme suivante :

$$\frac{S(z)}{E(z)} = \sum_{k=0}^N b_k z^{-k} = H(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_N z^{-N} \quad (3.12)$$

Où les b_k sont les coefficients de la réponse impulsionnelle finie du filtre, et N l'ordre du filtre. Si on multiplie $H(z)$ par un facteur unitaire z^N/z^N on obtient la relation intéressante suivante :

$$\frac{S(z)}{E(z)} = \sum_{k=0}^N b_k z^{-k} = H(z) = \frac{b_0 z^N + b_1 z^{N-1} + b_2 z^{N-2} + \dots + b_N}{z^N} \quad (3.13)$$

Elle montre que tous les pôles sont nuls et appartiennent au cercle unitaire du domaine de z c'est-à-dire $|z| < 1$. Cette caractéristique suffit à confirmer la stabilité de la

fonction de transfert $H(z)$. Dans l'expression de $H(z)$ l'ordre de la fonction est spécifié par N .

La réponse en fréquence d'un filtre donne par sa transformée de Fourier continue inverse la réponse impulsionnelle idéale non causale à symétrie paire infinie et de déphasage nul (voir équation 3.15) cette dernière ne peut être concrétisée dans une mise en œuvre réelle. Toutefois une approximation de cette dernière permet de concrétiser sa réalisation en faisant intervenir deux opérations. La première consiste à la rendre finie par sa troncature ou fenêtrage, la seconde provoque une translation temporelle ou retard pour la rendre causale. C'est ainsi qu'après ces opérations nous obtenons la réponse impulsionnelle causale $H(z)$ de l'équation 3.27 qui introduit un déphasage. Cependant ce dernier est linéaire car la symétrie est préservée dans la translation temporelle.

Lorsque l'ordre N de $H(z)$ est paire le nombre de ses termes est impair et la symétrie se déplace autour d'un point central $N/2$. La transformée de Fourier continue de $H(z)$ devient alors dans ce cas $H(e^{j\omega})$ où son module est réel et son déphasage alors vaut $\theta(\omega) = -\omega N/2 \pm k\pi$.

Le déphasage est donc une fonction linéaire de la fréquence. Cette dernière caractéristique est essentielle pour le traitement des signaux vocaux, car le délai de groupe (voir équation 3.14) devient constant quelle que soit la fréquence. Ainsi on est assuré de l'absence de distorsion de phase.

$$\frac{d \theta(\omega)}{d \omega} = -\frac{N}{2} \quad (3.14)$$

3.12 La réponse impulsionnelle d'un passe-bande

En partant de l'exigence de la réponse fréquentielle idéale et par l'entremise de la transformée de Fourier inverse continue nous obtenons la réponse impulsionnelle idéale d'un passe-bande $h_{\text{INC}}(m)$ où l'abréviation inc signifie idéale et non causale.

$$h_{INC}(m) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} h_{INC}(e^{j\omega}) e^{jm\omega} d\omega \quad (3.15)$$

Où

$$h_{INC}(e^{j\omega}) = \begin{cases} 1 & \text{si } \omega_1 \leq |\omega| \leq \omega_2 \\ 0 & \text{si } \omega_1 > |\omega| > \omega_2 \end{cases} \quad (3.16)$$

Alors l'intégration se résume en deux parties

$$h_{INC}(m) = \frac{1}{2\pi} \left[\int_{-\omega_2}^{-\omega_1} e^{jm\omega} d\omega + \int_{\omega_1}^{\omega_2} e^{jm\omega} d\omega \right] \quad (3.17)$$

Son résultat est :

$$h_{INC}(m) = \begin{cases} \frac{\omega_2 - \omega_1}{\pi} & \text{si } m = 0 \\ \frac{\sin m\omega_2 - \sin m\omega_1}{m\pi} & \text{si } m \neq 0 \end{cases} \quad (3.18)$$

Sa transformée en Z s'exprime ainsi :

$$h_{INC}(z) = \sum_{m=-\infty}^{+\infty} h_{INC}(m) z^{-m} \quad (3.19)$$

Sa nature infinie ne peut aboutir à sa réalisation concrète. Il faut donc la rendre finie au moyen d'une troncature de son domaine autour de sa symétrie en la multipliant par une fonction de fenêtrage finie qui pour les besoins de la démonstration sera une fonction fenêtre rectangulaire. Une fenêtre plus performante sera retenue dans notre application et expliquée plus loin. Ainsi nous obtenons :

$$h_{NC}(m) = h_{INC}(m) \text{ fen}(m) \quad (3.20)$$

Où

$$fen(m) = \begin{cases} 1 & \text{si } -M \leq m \leq +M \\ 0 & \text{autrement} \end{cases} \quad (3.21)$$

La réponse impulsionnelle est devenue $h_{NC}(m)$ en remplaçant sa nature idéale infinie par sa forme tronquée mais encore non causale. Maintenant pour rendre la réponse impulsionnelle causale il faut introduire un délai ou retard de M .

La transformée en Z de $h_{NC}(m)$ s'exprime ainsi :

$$h_{NC}(z) = \sum_{m=-M}^{+M} h_{NC}(m) z^{-m} \quad (3.22)$$

En introduisant un retard de M la transformée en Z devient causale et s'exprime par :

$$h_C(z) = z^{-M} h_{NC}(z) = \sum_{m=-M}^{+M} h_{NC}(m) z^{-(m+M)} \quad (3.23)$$

Pour mieux formaliser $h_C(z)$ le changement de variable $n = m+M$ est souhaité, ce qui donne :

$$h_C(z) = \sum_{n=0}^{+2M} h_{NC}(n-M) z^{-n} \quad (3.24)$$

La borne supérieure de la sommation de $h_C(z)$ étant généralement admise comme l'ordre de la transformée alors si $N=2M$ l'expression devient :

$$h_C(z) = \sum_{n=0}^N h_{NC}\left(n - \frac{N}{2}\right) z^{-n} \quad (3.25)$$

Puisque

$$h_C(n) = h_{NC}\left(n - \frac{N}{2}\right) \quad (3.26)$$

La transformée en Z de la réponse impulsionnelle devenue concrète et réalisable est donc

$$h_C(z) = \sum_{n=0}^N h_C(n) z^{-n} \quad (3.27)$$

Les transformations nécessaires pour rendre la réponse impulsionnelle idéale et abstraite en une réponse impulsionnelle concrète et réalisable n'ont pas altéré la symétrie paire de la réponse impulsionnelle. Cependant la transformation pour aboutir à la causalité a déplacé le centre de symétrie avec les conséquences qui suivent.

La réponse impulsionnelle non causale $h_{NC}(m)$ définie par son domaine tronqué $-M \leq m \leq +M$ préserve la symétrie paire centré sur $h_{NC}(0)$ d'où la propriété suivante : $h_{NC}(m) = h_{NC}(-m)$. Lorsque le facteur en z du terme centré de sa transformée en Z est mis en évidence et que l'on fait la transformée de Fourier on constate comme pour la réponse idéale que la réponse impulsionnelle non-causale présente un déphasage nul.

Mais pour la réponse impulsionnelle causale $h_C(m)$ son domaine s'est vu décalé par l'ajout d'un retard M redéfinissant ainsi son domaine comme $0 \leq m \leq +2M$. Pour garder la cohérence des changements de variables vus plus haut la réponse impulsionnelle causale devient $h_C(m)$ définie par son domaine $0 \leq n \leq N$.

La symétrie paire se voit ainsi décalée pour être centré sur $h_C(N/2)$ ce qui implique alors la propriété suivante : $h_C(n+N/2) = h_C(-n+N/2)$, cette dernière propriété sera mise à profit lors de la mise en œuvre des filtres dans le DSP, effectivement le calcul sera réduit de moitié puisqu'il suffit de calculer uniquement la moitié des termes de la réponse impulsionnelle. Maintenant si on fait la transformée de Fourier de la transformée en Z de $h_C(n)$ tout en mettant en évidence le facteur en z du terme central de symétrie, on obtient après simplifications l'expression suivante :

$$H_C(e^{j\omega}) = \sum_{n=0}^N h_C(n) e^{-jn\omega} = \pm P(\omega) e^{-j\frac{N}{2}\omega} \quad (3.28)$$

où $\pm P(\omega)$ est une fonction réelle de forme polynomiale donnant le module de l'amplitude de $h_C(e^{j\omega})$ et s'exprime comme suit :

$$P(\omega) = \sum_{n=0}^{N/2} C_n \cos(n\omega) \quad (3.29)$$

En tenant compte du signe du polynôme combiné avec l'argument de $h_C(e^{j\omega})$ on constate l'introduction d'un déphasage dans la réponse impulsionnelle causale $h_C(e^{j\omega})$. Son expression générale montre bien sa linéarité en fréquence.

$$\theta(\omega) = -\frac{N}{2}\omega \pm k\pi \quad k = \{0, 1, 2, \dots\} \quad (3.30)$$

3.13 La fonction de fenêtrage

La troncature de la réponse impulsionnelle idéale se fait telle que vue précédemment par un fenêtrage, en multipliant la réponse impulsionnelle idéale par une fonction fenêtre. Dans la démonstration plus haut, on utilise la fenêtre rectangulaire définie en (3.21). Cette dernière est cependant rarement utilisée dans la pratique car contrairement aux autres types de fenêtres elle ne minimise aucunement les effets déformateurs de la troncature sur la réponse impulsionnelle appelés phénomènes de Gibbs. Ces derniers sont des ondulations superposées à la synthèse de la fonction $h_C(e^{j\omega})$ et ceci, indépendamment du nombre de coefficients employés.

Le fait d'augmenter les coefficients diminue l'amplitude des oscillations devenues plus rapides sur les zones de continuité de la fonction sans toutefois réduire l'amplitude des oscillations qui convergent vers les points de discontinuité de la fonction.

C'est précisément pour réduire ces effets que nous cherchons des fenêtres plus performantes. Il existe plusieurs types de fenêtres connues et communément utilisées. Le choix d'une fenêtre va dépendre de l'atténuation maximale recherchée, de l'ordre requis et en relation avec lui de la largeur de la fenêtre. Les plus utilisées sont les fenêtres de Hamming, de Von Hann, de Blackman et de Kaiser.

3.14 La fenêtre de Kaiser

La mise en œuvre des filtres RIF emploie la fenêtre de Kaiser qui contrairement aux autres fenêtres s'ajuste avec une plus grande flexibilité à l'aide de deux paramètres qui sont sa longueur L ou nombre de termes de la fenêtre et β le second paramètre, nécessaire pour évaluer la fonction modifiée de Bessel d'ordre zéro du premier genre $I_0(x)$. Cette fonction va donner principalement la silhouette de la fenêtre et l'ajustement de la raideur de la bande de transition séparant la bande passante de l'atténuée et fixant l'atténuation maximale voulue.

Selon la valeur de β , on peut retrouver les autres fenêtres, ainsi lorsque $\beta = 0$ nous obtenons la fenêtre rectangulaire, quand $\beta = 5.44$ la fenêtre s'apparente très bien à celle de Hamming.

La simplicité de mise en œuvre des autres fenêtres leur permet de s'affranchir de la complexité et du coût élevé en calcul itératif qu'exige la souplesse de la fenêtre de Kaiser. Cet aspect ne sera pas à négliger lors de la mise en œuvre du DSP dans l'éventualité de contraintes en temps réel importantes, à moins de placer au préalable dans une table le résultat des calculs itératifs qui accélère le procédé ce qui a été notre choix.

La démarche pour construire la fenêtre de Kaiser est prise de la référence [16], il s'agit d'une approche d'optimisation alliée à un calcul basé sur des relations empiriques. Les étapes sont les suivantes :

- On débute en choisissant le coefficient d'ondulation minimale requise qu'il soit dans la bande passante δ_P ou atténuée δ_A autrement dit $\delta = \min(\delta_P, \delta_A)$ pour trouver son évaluation en dB telle que $A = -20 \log(\delta)$.
- On fixe la largeur de la bande de transition désirée séparant la bande passante de la bande atténuée de telle sorte que $\Delta\omega = \omega_{LA} - \omega_{LP}$

L'ordre N du filtre RIF sera déduit de la relation suivante :

$$N \geq \frac{A - 8}{2.285 \Delta\omega} \quad (3.31)$$

Il est à remarquer que suite à ce calcul approché, N retenu sera toujours un nombre pair afin que le nombre de termes de la fenêtre comme ceux de la réponse impulsionnelle tronquée soit impair $L = N + 1$ pour assurer la symétrie centrale des coefficients et garantir un déphasage linéaire.

La valeur de β est déterminée par les relations fixées selon la plage impliquée de l'atténuation A exprimée en dB.

$$\beta = \begin{cases} 0 & \text{si } A < 21 \\ 0.5842(A - 21)^{0.4} + 0.07886(A - 21) & \text{si } 21 \leq A \leq 50 \\ 0.1102(A - 8.7) & \text{si } A > 50 \end{cases} \quad (3.32)$$

Une fois β et N connus, le calcul des coefficients de la fenêtre de Kaiser causale $W_{kc}(n)$ sont déterminés par la relation suivante :

$$W_{kc}(n) = \begin{cases} \frac{I_0 \left[\beta \sqrt{1 - \left(\frac{n - N/2}{N/2} \right)^2} \right]}{I_0(\beta)} & \text{si } 0 \leq n \leq N \\ 0 & \text{autrement} \end{cases} \quad (3.33)$$

$I_0(x)$ est la fonction modifiée de Bessel d'ordre zéro du premier genre, elle est évaluée de la façon suivante :

$$I_0(x) = 1 + \sum_{k=1}^{\infty} \left[\frac{(x/2)^k}{k!} \right]^2 \quad (3.34)$$

En pratique une précision satisfaisante est atteinte en ne requérant tout au plus que 20 à 25 termes de la sommation de cette expression.

La figure 11 illustre l'allure d'une fenêtre de Kaiser, dont les paramètres sont à titre d'exemple ; $\delta_{\min} = 0.01$, $A = 40$ dB, et $\Delta\omega = 0.0897$, l'ordre ainsi obtenu est $N=158$

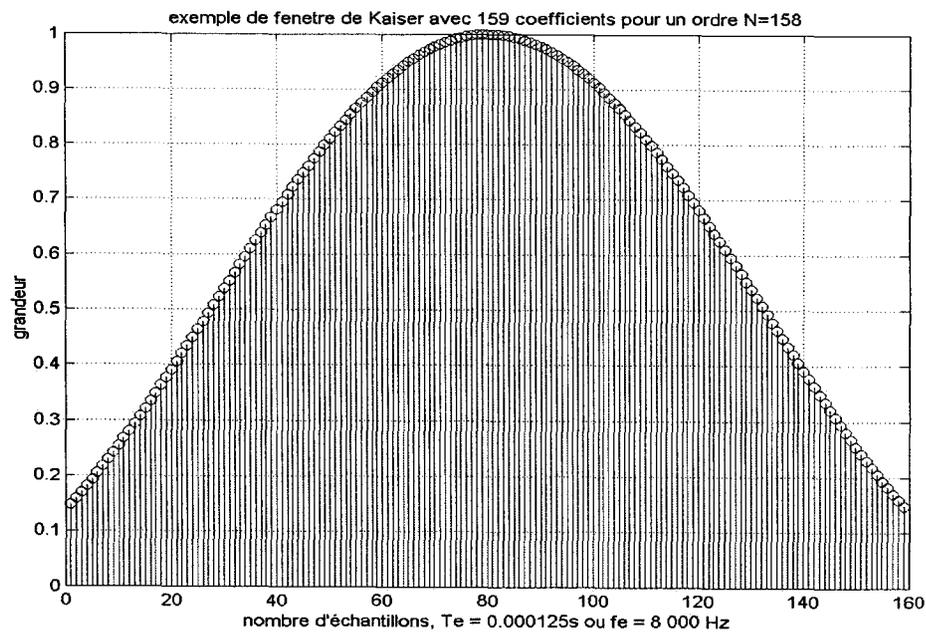


Figure 11 Fenêtre de Kaiser

3.15 Les signaux issus du banc de filtres

Le banc de filtres de type passe-bande centrés sur la fréquence fondamentale f_0 et ses harmoniques donne en sortie de chaque filtre une sous-bande du signal observé c'est-à-dire $r_{i,k}(t)$. Puisque le traitement ne veut retenir que l'harmonicit   propre de la voix choisie. Il va chercher l'enveloppe de chaque sous-bande de signal $\hat{A}_{i,k}(t)$ pour s'en servir comme signal modulant une porteuse    la fr  quence centrale du filtre passe-bande concern  . Selon la sous-bande en question cette fr  quence centrale est soit la fondamentale f_0 ou une de ses harmoniques.

Pour se faire, ce dernier traitement est réalisé à l'aide d'une transformée de Hilbert dont le principe est vu en détail plus loin et qui à partir de chaque sous-bande en entrée, extrait en sortie l'enveloppe de chaque sous-bande qui va moduler sa porteuse respective. Se faisant, le signal obtenu au cours du traitement $Z_{i,k}(t)$ a perdu l'information de la phase originale. Le processus de recouvrer l'information de la phase fait appel à la théorie de Wiener qui suit, pour donner le signal synthétisé avec le déphasage corrigé $x_{i,k}(t)$.

3.16 Les filtres de Wiener

Développés à partir de concepts temporels plutôt que fréquentiels, ces filtres sont conçus pour minimiser l'erreur quadratique moyenne entre leur signal en sortie et un signal désiré. Ils sont optimaux au sens du critère de l'erreur quadratique moyenne et nous verrons que dans ce cas les coefficients du filtre sont liés à la fonction d'autocorrélation du signal d'entrée et à l'intercorrélacion entre les signaux d'entrée et de sortie désirée. Cette partie du document a souvent puisé dans les références [17], [18], [19], [20], et [21], [22], [23], [24].

Le problème d'estimation linéaire est le suivant : $x(n)$ correspond au signal qui nous intéresse mais n'est pas directement accessible. Seul $y(n)$ l'est suite au résultat du passage de $x(n)$ dans un système linéaire.

On veut retrouver $x(n)$ à partir de $y(n)$. Une solution consiste à filtrer $y(n)$ de telle sorte que la sortie du filtre soit la plus proche possible de $x(n)$. On peut mesurer l'erreur de l'estimation par $e(n)$ définie par :

$$e(n) = x(n) - \hat{x}(n) \quad (3.35)$$

Évidemment, plus l'erreur $e(n)$ sera faible, plus l'estimation sera bonne. On cherche donc un filtre qui minimisera l'erreur. Minimiser $e^2(n)$ revient à minimiser une fonction quadratique dérivable. Par ailleurs, étant donné que les signaux intéressants sont

aléatoires, la fonction coût qui sera à minimiser est l'erreur quadratique moyenne souvent connue sous son appellation anglaise MSE (mean square error) définie par :

$$\xi(n) = E[e^2(n)] \quad (3.36)$$

Le filtre optimal de Wiener est un filtre qui minimisera la MSE.

3.17 Filtre de Wiener de type RIF

Le filtre adaptatif se construit autour du calcul d'un filtre RIF. Soit h , le filtre que nous recherchons et N la longueur de sa réponse impulsionnelle dont la forme matricielle est :

$$h = [h_0 \quad h_1 \quad \dots \quad h_{N-1}]^T \quad (3.37)$$

Le signal estimé est alors :

$$\hat{x}(n) = \sum_{k=0}^{N-1} h_k y(n-k) \quad (3.38)$$

ou en introduisant la notation matricielle pour $y(n)$ on a :

$$\hat{x}(n) = h^T y(n) \Leftrightarrow \hat{x}(n) = y^T(n) h \quad (3.39)$$

avec

$$y = [y(n) \quad y(n-1) \quad \dots \quad y(n-(N-1))]^T \quad (3.40)$$

Avec l'hypothèse que les signaux $x(n)$ et $y(n)$ sont stationnaires, on arrive à la fonction coût suivante :

$$\xi = E\left[(x(n) - \mathbf{h}^T \mathbf{y}(n))^2\right] \Leftrightarrow \xi = E\left[x^2(n) - 2\mathbf{h}^T \mathbf{y}(n)x(n) + \mathbf{h}^T \mathbf{y}(n)\mathbf{y}^T(n)\mathbf{h}\right] \quad (3.41)$$

$$\xi = E[x^2(n)] - 2\mathbf{h}^T \Phi_{yx} + \mathbf{h}^T \Phi_{yy} \mathbf{h} \quad (3.42)$$

où Φ_{yy} est la matrice d'autocorrélation de taille $N \times N$ est :

$$\Phi_{yy} = E[\mathbf{y}(n)\mathbf{y}^T(n)] \quad (3.43)$$

le vecteur d'intercorrélation Φ_{yx} de taille N est défini par :

$$\Phi_{yx} = E[\mathbf{y}(n)x(n)] \quad (3.44)$$

La fonction coût MSE $\xi(n)$ montre qu'elle dépend de la réponse impulsionnelle du filtre RIF \mathbf{h} . Pour en obtenir le minimum, on cherche les conditions d'annulation de la dérivée de la fonction coût par rapport aux variables qui sont les N points de la réponse impulsionnelle du filtre. On a donc une dérivée partielle totale c'est-à-dire un gradient à annuler.

La dérivée de la fonction coût par rapport au $j^{\text{ième}}$ point de la réponse impulsionnelle donne :

$$\frac{\partial \xi}{\partial h_j} = E\left[\frac{\partial}{\partial h_j}\{e^2(n)\}\right] = E\left[2e(n)\frac{\partial e(n)}{\partial h_j}\right] \quad (3.45)$$

En explicitant l'équation $e(n)$, on obtient :

$$\frac{\partial \xi}{\partial h_j} = E \left[2e(n) \frac{\partial}{\partial h_j} \{x(n) - h^T y(n)\} \right] \quad (3.46)$$

Le fait que la sortie du filtre $h^T y(n)$ peut s'écrire comme une somme de N produits dont un seul contient le terme h_j , nous permet d'écrire l'expression suivante :

$$\frac{\partial \xi}{\partial h_j} = E \left[2e(n) \frac{\partial}{\partial h_j} \{h_j y(n-j)\} \right] \Leftrightarrow \frac{\partial \xi}{\partial h_j} = E[-2e(n)y(n-j)] \quad (3.47)$$

Pour avoir les conditions d'annulation de cette équation sur tous les $j=\{0, \dots, N-1\}$, on utilise un ensemble de N équations qui s'écrit sous forme matricielle par la notion du vecteur gradient comme suit:

$$\nabla \xi = \begin{bmatrix} \partial \xi / \partial h_0 \\ \partial \xi / \partial h_1 \\ \vdots \\ \partial \xi / \partial h_j \\ \vdots \\ \partial \xi / \partial h_{N-1} \end{bmatrix} = -2E \begin{bmatrix} y(n)e(n) \\ y(n-1)e(n) \\ \vdots \\ y(n-j)e(n) \\ \vdots \\ y(n-N+1)e(n) \end{bmatrix} = -2E \left\{ \begin{bmatrix} y(n) \\ y(n-1) \\ \vdots \\ y(n-j) \\ \vdots \\ y(n-N+1) \end{bmatrix} e(n) \right\} = -2E[y(n)e(n)] \quad (3.48)$$

En utilisant les équations de l'erreur et de la fonction de transfert du filtre pour remplacer $e(n)$ on obtient :

$$\nabla \xi = -2E[y(n)(x(n) - y^T(n)h)] = -2E[y(n)x(n)] + 2E[y(n)y^T(n)]h \quad (3.49)$$

Cette équation devient en introduisant le vecteur d'intercorrélation et la matrice d'autocorrélation:

$$\nabla \xi = -2\Phi_{yx} + 2\Phi_{yy} h = 0 \quad (3.50)$$

La réponse impulsionnelle optimale h_{opt} annule cette équation d'où :

$$h_{opt} = \Phi_{yy}^{-1} \Phi_{yx} \quad (3.51)$$

Ce filtre de Wiener de type RIF permet d'obtenir une erreur quadratique minimale entre $x(n)$ désirée et son estimé donnée par :

$$\xi_{min} = E[x^2(n)] - h_{opt}^T \Phi_{yx} \quad (3.52)$$

3.18 Algorithme LMS pour le filtrage adaptatif

La mise en œuvre d'un filtre (estimateur) optimal de Wiener demande la connaissance des caractéristiques du signal, de la matrice d'autocorrélation et du vecteur d'intercorrélacion, du bruit et de la fonction de transfert du canal. Cela implique également que ces caractéristiques soient stables au cours du temps, ce qui n'est pas le cas en pratique. La matrice d'autocorrélation et le vecteur d'intercorrélacion comme la fonction coût sont inconnus généralement dans les cas pratiques.

On va approcher le filtre optimal de Wiener en utilisant une boucle de retour et un algorithme de minimisation : c'est ce que l'on appelle le filtrage adaptatif illustré sur la figure 12. Dans ce cas, on remplacera la connaissance des fonctions de corrélation par une phase d'apprentissage permettant de modifier itérativement la réponse impulsionnelle du filtre, en affectant les coefficients du filtre.

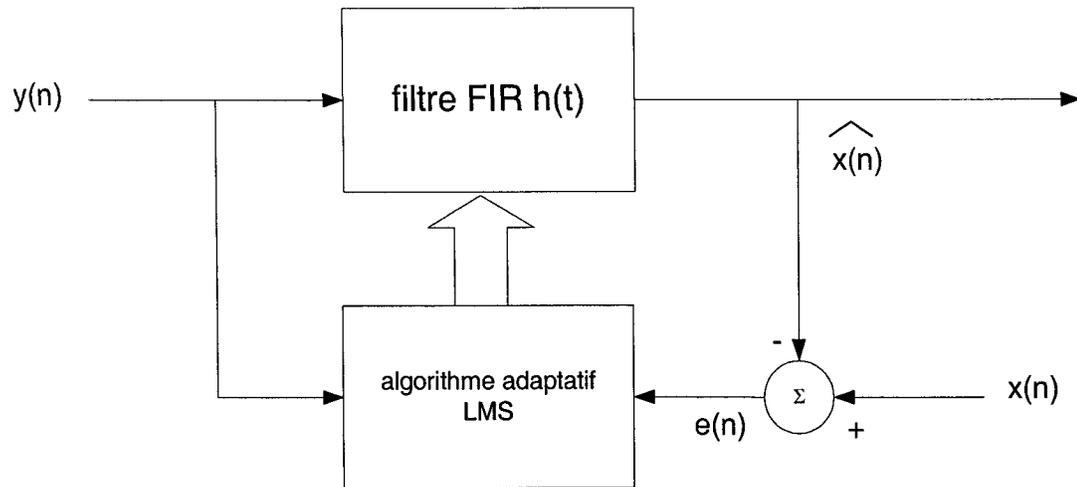


Figure 12 Principe du filtrage adaptatif

Le filtrage adaptatif a pour objet d'approcher ces filtres optimaux. Pour cela, les coefficients de la réponse impulsionnelle du filtre sont adaptés en fonction de l'erreur par une boucle de retour comme le montre la figure 12.

Cette adaptation nécessite une séquence d'apprentissage et une stratégie de mise à jour des coefficients du filtre dont l'objectif est la minimisation d'une erreur. Pour cela, on utilisera des algorithmes d'optimisation. L'algorithme choisi ici pour sa simplicité est le LMS largement utilisé en filtrage adaptatif.

La réponse impulsionnelle d'un filtre adaptatif est donc variable dans le temps. Elle dépend du signal reçu, de la séquence d'apprentissage et de l'algorithme d'optimisation utilisé.

Le signal estimé est défini de la façon suivante :

$$\hat{x}(n) = \sum_{k=0}^{N-1} h_k(n-1)y(n-k) \Leftrightarrow \hat{x}(n) = \mathbf{h}^T(n-1)\mathbf{y} \quad (3.53)$$

Il est calculé à l'instant (n) en utilisant la réponse impulsionnelle du filtre calculée précédemment au moment $(n-1)$ par l'algorithme d'optimisation.

Il existe plusieurs techniques d'optimisation basées sur la descente en gradient qui peuvent être mise en œuvre. L'algorithme LMS (Least Mean Squares) dont le détail est donné ci-dessous est souvent utilisé dans les systèmes de filtrage adaptatif et c'est cette approche que nous retenons pour les besoins de notre application. On s'y réfère souvent comme l'algorithme du gradient stochastique.

3.19 Détail de l'algorithme LMS

Initialisation du filtre RIF $h(0) = 0$. À chaque échantillon n , les opérations suivantes sont accomplies :

$$\begin{aligned}\hat{x}(n) &= h^T(n-1) y(n) \\ e(n) &= x(n) - \hat{x}(n) \\ h(n) &= h(n-1) + 2\mu y(n) e(n)\end{aligned}\tag{3.54}$$

Dans notre application, $y(n)$ est notre signal d'entrée représenté par la fréquence kf_0 modulée par l'enveloppe $\hat{A}_{i,k}(t)$ obtenue par la transformée de Hilbert du signal $r_{i,k}(t)$ à la sortie du banc des filtres. Ces derniers signaux représentent selon la sous-bande k le signal désiré $x(n)$ de notre algorithme par lequel se définit la séquence d'apprentissage qui permettra de retrouver la phase originale en sortie du filtre adaptatif.

L'emploi de l'algorithme du LMS se justifie par le fait qu'il est essentiel en traitement de signal de traiter autant faire se peut, les données observées directement, autrement dit en ligne. Ainsi, on évite d'avoir recours dans les calculs, à l'espérance mathématique comme cela a été vu plus haut pour les besoins d'analyse.

On va donc travailler sur h pour minimiser la fonction coût :

$$\xi(n) = E\{e^2(n)\} \quad (3.55)$$

en mettant à jour h tel que :

$$h(n+1) = h(n) - \mu \nabla \xi(n) \quad (3.56)$$

Il faut pour cela évaluer le gradient et on déduit ci-dessous son expression :

$$\begin{aligned} \nabla \xi(n) &= \nabla E\{e^2(n)\} = E\{\nabla e^2(n)\} = E\{2e(n) \nabla e^*(n)\} \\ \text{et } \nabla e^*(n) &= \nabla(x(n) - \hat{x}(n)) = \nabla(x(n) - h^T y^*(n)) = -y^*(n) \end{aligned} \quad (3.57)$$

donc nous avons :

$$\nabla \xi(n) = \frac{\partial \xi(n)}{\partial h} = -2E\{e(n) y^*(n)\} \quad (3.58)$$

Ainsi la mise à jour de l'algorithme du LMS devient :

$$h(n+1) = h(n) - \mu \nabla \xi(n) = h(n) + 2\mu E\{e(n) y^*(n)\} \quad (3.59)$$

on élimine l'espérance mathématique en l'approximant par son estimée instantanée tout en se limitant au domaine réel, l'algorithme du gradient stochastique s'exprime finalement par :

$$h(n+1) = h(n) + 2\mu e(n) y(n) \quad (3.60)$$

Le facteur μ est le pas d'adaptation de l'algorithme et agit à titre de gain sur le gradient. Sa grandeur est cruciale dans l'atteinte de la convergence de l'algorithme. Dans le cas de signaux stationnaires, on démontre que la convergence est atteinte lorsque :

$$\lim_{n \rightarrow \infty} h_n = \Phi_{yy}^{-1} \Phi_{yx} \Leftrightarrow 0 < \mu < \frac{2}{\lambda_{\max}} \quad (3.61)$$

La borne supérieure de μ fait intervenir λ_{\max} qui est la valeur propre maximale de la matrice d'autocorrélation Φ_{yy} qui comme on le sait est rarement connue en pratique. De plus, la grandeur de cette borne supérieure n'est jamais approchée. On lui préfère à la place de λ_{\max} la trace de la matrice d'autocorrélation Φ_{yy} pour une borne plus sûre.

Comme la matrice d'autocorrélation Φ_{yy} se réfère généralement à des signaux stationnaires elle se présente sous le type Toeplitz. Alors sa trace s'évalue ainsi :

$$\text{tr}(\Phi_{yy}) = (p+1)\Phi_{yy}(0) = (p+1)E\{|y(n)|^2\} \quad (3.62)$$

où p est l'ordre du filtre, ou encore le nombre des éléments de la diagonale principale de la matrice d'autocorrélation Φ_{yy} . Encore ici l'espérance mathématique est substituée par son estimée instantanée.

Jusqu'à présent le pas d'adaptation est considéré constant bien que cette particularité ne soit pas idéale dans un contexte où la matrice d'autocorrélation Φ_{yy} est rarement connue et que les signaux ne sont pas toujours garantis stationnaires et peuvent varier en nature, on préfère alors un pas d'adaptation qui varie selon la puissance du processus traité. L'algorithme LMS est alors dit normalisé et on obtient :

$$\mu(n) = \frac{\beta}{L\|y(n)\|^2} \quad (3.63)$$

Où β du pas normalisé est un facteur qui se situe dans les limites telles que $0 < \beta < 2$. L représente l'ordre du filtre, ou encore le nombre des éléments de la diagonale principale de la matrice d'autocorrélation. La mise à jour de l'algorithme du gradient stochastique prend finalement la forme suivante:

$$h(n+1) = h(n) + \beta \frac{y(n)}{\varepsilon + L\|y(n)\|^2} e(n) \quad (3.64)$$

où ε est une petite valeur constante qui assure que le quotient ne s'emballe pas et maintienne la convergence advenant que $y(n)$ du dénominateur tend vers zéro.

Dans notre application [10], $y(n)$ est le signal d'entrée synthétisé par la fréquence kf_0 modulée par l'enveloppe $\hat{A}_{i,k}(t)$ qui est obtenue par la transformée de Hilbert du signal $r_{i,k}(t)$ filtré à la sortie par le k passe-bande du banc des filtres. Ces derniers signaux représentent selon la sous-bande k le signal désiré $x(n)$ de notre algorithme par lequel se définit la séquence d'apprentissage qui permettra de retrouver la phase originale en sortie du filtre adaptatif.

Donc la phase est retrouvée à partir du signal synthétisé en s'appuyant sur un filtre adaptatif de type Wiener [10]. Le processus est le suivant ;

En premier, le signal synthétisé se définit ainsi :

$$Z_{i,k}(t) = [Z_{1,i,k} \quad Z_{-1,i,k}]^T \quad (3.65)$$

Où i est le canal propre au capteur (microphone) considéré, k la sous-bande du banc de filtres associée à la fondamentale f_0 ou à une de ses harmoniques kf_0 sachant que le signal synthétisé se construit plus particulièrement comme :

$$Z_{q,i,k}(t) = \hat{A}_{i,k}(t) e^{qj2\pi f_0(t)t} \quad \text{où } q = (1, -1) \quad (3.66)$$

Ce signal est appliqué à l'entrée du filtre de Wiener dont le signal estimé présent à la sortie devient :

$$\hat{x}_{i,k} = h_{i,k}^T Z_{i,k} \quad (3.67)$$

De façon plus détaillée et d'un point de vue matriciel, cette dernière relation s'écrit :

$$X = [C_p \ C_m] \begin{bmatrix} Z_p \\ Z_m \end{bmatrix} = C_p Z_p + C_m Z_m \quad (3.68)$$

$$Z_p = A e^{j\omega} \quad \text{et} \quad Z_m = A e^{-j\omega}$$

Le signal en sortie aura la forme suivante :

$$x = C_p A e^{j\omega} + C_m A e^{-j\omega} \quad (3.69)$$

Les coefficients du filtre de Wiener dans $h(t)$ sont C_p et C_m et se trouvent à être l'un par rapport à l'autre son conjugué complexe. Ce sont eux, qui ajoutent la phase au signal synthétisé pour obtenir le signal estimé $x(t)$ identique au signal désiré $r_{i,k}(t)$ par l'adaptation permise selon l'algorithme de Wiener.

La phase du signal synthétisé peut être déduite en fonction du temps en posant l'expression complexe des coefficients du filtre de Wiener $h(t)$ ainsi :

$$C_p = a + jb \quad \& \quad C_m = a - jb \quad (3.70)$$

Alors le signal en sortie du filtre de Wiener devient :

$$\begin{aligned} x &= C_p A (\cos \omega + j \sin \omega) + C_m A (\cos \omega - j \sin \omega) \\ x &= C_p A \cos \omega + C_m A \cos \omega + j(C_p A \sin \omega - C_m A \sin \omega) \\ x &= (C_p + C_m) A \cos \omega + j(C_p - C_m) A \sin \omega \\ x &= 2a A \cos \omega + j 2jb A \sin \omega \\ x &= 2a A \cos \omega - 2b A \sin \omega \\ x &= 2A (a \cos \omega - b \sin \omega) \\ x &= 2A \sqrt{a^2 + b^2} \cos(\omega - \Phi) \quad \text{où} \quad \Phi = \text{tg}^{-1} \left(\frac{b}{a} \right) \end{aligned} \quad (3.71)$$

Ce signal représenté par le filtre adaptatif de Wiener le signal filtré $r_{i,k}(t)$ désiré mais cette fois-ci reconstruit à partir d'un signal synthétisé par l'enveloppe $\hat{A}_{i,k}(t)$ modulant en

amplitude une pure harmonique basée sur la fréquence fondamentale kf_0 corrigée par la contribution de la phase $\Phi(t)$.

Nous avons vu dans notre application [10] que l'emploi de la transformée de Hilbert du signal $r_{i,k}(t)$ filtré à la sortie d'un des k passe-bandes du banc de filtre était nécessaire pour élaborer l'enveloppe $\hat{A}_{i,k}(t)$. Cette enveloppe va moduler la fondamentale ou une de ses harmoniques pour synthétiser et se substituer aux signaux réels filtrés.

La transformée de Hilbert possède quelques propriétés intéressantes largement utilisées en traitement de signal et en communication, comme dans notre cas, elle est très souvent employée pour extraire l'enveloppe de signaux à bande étroite, c'est-à-dire des signaux dont la majorité de l'énergie est centrée autour d'une fréquence qui est dite porteuse dans le cas de la modulation d'amplitude. Avant d'aborder cet emploi, expliquons brièvement en quoi consiste la transformée de Hilbert.

Elle permet de transformer tout signal réel $x(t)$ en un signal analytique $z(t)$ sous sa forme complexe à fréquence uniquement positive. Ainsi :

$$z(t) = x(t) + j y(t) \quad (3.72)$$

Où

$$y(t) = TH \{x(t)\} \quad (3.73)$$

La partie imaginaire du signal analytique, $y(t)$ est la transformée de Hilbert du signal réel autrement dit de la partie réelle du signal analytique qui a subi un décalage temporel d'un quart de cycle.

Ce décalage temporel ou déphasage peut être réalisé par un filtre passe-tout provoquant un déphasage constant de $\pi/2$ aux fréquences négatives et de $-\pi/2$ pour les fréquences

positives, à cet égard, il est appelé filtre de transformée de Hilbert puisque sa sortie est la transformée de Hilbert du signal appliqué à son entrée. En notation complexe, un décalage temporel d'un quart de cycle ou si l'on veut, un déphasage de $\pm \pi/2$ se traduit ainsi :

$$e^{\pm j\pi/2} = \cos(\pi/2) \pm j \sin(\pi/2) = 0 \pm j = \pm j \quad (3.74)$$

Pour les signaux stationnaires à fréquences positives ou négatives, cette catégorie de signaux peut se traduire également comme :

$$x_+(t) = e^{+j\omega t} \text{ et } x_-(t) = e^{-j\omega t} \quad (3.75)$$

Si on produit un déphasage de $+\pi/2$ aux fréquences négatives et de $-\pi/2$ aux fréquences positives comme le fait la transformée de Hilbert, nous obtenons :

$$y_+(t) = -j e^{+j\omega t} \text{ et } y_-(t) = +j e^{-j\omega t} \quad (3.76)$$

Les signaux analytiques respectifs deviennent :

$$z_+(t) = x_+(t) + j y_+(t) \text{ et } z_-(t) = x_-(t) + j y_-(t) \quad (3.77)$$

Une fois ces expressions simplifiées nous avons :

$$\begin{aligned} z_+(t) &= e^{+j\omega t} + j(-j e^{+j\omega t}) \text{ et } z_-(t) = e^{-j\omega t} + j(+j e^{-j\omega t}) \\ z_+(t) &= e^{+j\omega t} - j^2 e^{+j\omega t} \text{ et } z_-(t) = e^{-j\omega t} + j^2 e^{-j\omega t} \\ z_+(t) &= 2e^{+j\omega t} \text{ et } z_-(t) = 0 \end{aligned} \quad (3.78)$$

Nous remarquons ainsi que les fréquences négatives sont annulées lorsque le signal devient analytique.

Jusqu'à maintenant, nous avons abordé la transformée de Hilbert dans le domaine temporel, cependant sa mise en œuvre dans notre application sur le processeur de traitement de signal préfère l'approche fréquentielle de la transformée de Hilbert qui s'avère plus aisée avec l'emploi de l'outil mathématique de la FFT (fast fourier transform) qui est la transformée de Fourier rapide.

Aussi la transformée de Hilbert dans le domaine fréquentiel est présentée ci-après. Si nous revenons à la représentation d'un signal analytique $z(t)$ qui est défini selon:

$$z(t) = f(t) + j \hat{f}(t) \quad (3.79)$$

Sa transformée de Fourier devient :

$$Z(\omega) = F(\omega) + j \hat{F}(\omega) \quad (3.80)$$

Pour être analytique $Z(\omega)$ ne peut se situer que sur le côté positif du spectre où seules les fréquences positives se retrouvent. Par conséquent pour toutes fréquences négatives, nous avons :

$$Z(\omega) = 0 \quad \text{si } \omega < 0 \quad (3.81)$$

Cela veut donc dire que :

$$F(\omega) = -j \hat{F}(\omega) \quad \text{ou encore que } \hat{F}(\omega) = j F(\omega) \quad \text{quand } \omega < 0 \quad (3.82)$$

Comme la fonction en fréquence dans la partie imaginaire est à symétrie impaire :

$$\hat{F}(\omega) = -j F(\omega) \quad \text{quand } \omega > 0 \quad (3.83)$$

En résumé la transformée de Hilbert dans le domaine fréquentiel est définie de la façon suivante :

$$\hat{F}(\omega) = -j F(\omega) \quad \text{quand } \omega > 0 \quad \text{ou } \hat{F}(\omega) = j F(\omega) \quad \text{quand } \omega < 0 \quad (3.84)$$

Sous sa forme concise la transformée de Hilbert s'exprime ainsi :

$$\hat{F}(\omega) = -j F(\omega) \operatorname{sgn}(\omega) \quad (3.85)$$

Rappelons ici les propriétés de la fonction signum $\operatorname{sgn}(\omega)$ qui se définit ainsi :

$$\operatorname{sgn}(\omega) = \frac{|\omega|}{\omega} = \begin{cases} 1 & \text{si } \omega > 0 \\ -1 & \text{si } \omega < 0 \end{cases} \quad (3.86)$$

En conclusion un signal analytique dans le domaine fréquentiel vaut :

si $\omega < 0$ alors

$$\begin{aligned} Z(\omega) &= F(\omega) + j \hat{F}(\omega) \\ Z(\omega) &= F(\omega) + j [jF(\omega)] \\ Z(\omega) &= F(\omega) + j^2 F(\omega) \\ Z(\omega) &= F(\omega) - F(\omega) = 0 \end{aligned} \tag{3.87}$$

Et

si $\omega > 0$ alors

$$\begin{aligned} Z(\omega) &= F(\omega) + j \hat{F}(\omega) \\ Z(\omega) &= F(\omega) + j [-jF(\omega)] \\ Z(\omega) &= F(\omega) - j^2 F(\omega) \\ Z(\omega) &= F(\omega) + F(\omega) = 2 F(\omega) \end{aligned} \tag{3.88}$$

Une fois le calcul dans le domaine fréquentiel obtenu, il suffit de faire la transformée de Fourier inverse ou une FFT inverse pour retrouver le signal analytique dans le domaine temporel. Le calcul du module de sa forme complexe nous donne généralement l'enveloppe lorsque le signal analytique est un signal de type à bande étroite ou de modulation d'amplitude.

3.20 Introduction sur L'ACI

La dernière partie du projet s'inspire de la démarche reposant sur la théorie de l'analyse en composantes indépendantes (ACI) [4], sans toutefois y adhérer totalement puisqu'elle a recourt également à d'autres aspects qui s'en dissocient à certains égards. Cependant la démarche appartient plus globalement aux méthodes dites de séparation et d'extraction aveugle de sources (SAS et EAS) communément connues sous l'acronyme anglo-saxon de BSS ou BSE (respectivement blind sources separation ou extraction).

L'ACI comme les SAS et EAS proviennent du besoin de retrouver les sources originales qui, après avoir été mélangées par le milieu les portant aboutissent sur des capteurs ou antennes donnant des signaux observés qui résultent de la combinaison des sources altérées par les propriétés naturelles du milieu qui agissent souvent comme le fait un

filtre mais variant dans le temps et tributaire de plusieurs facteurs connus comme inconnus.

C'est à partir des signaux observés que ces méthodes vont chercher les différentes sources originales sans connaissances à priori des particularités des sources ni des propriétés propres de mélange ou d'altération du milieu, de là, le terme aveugle employé par ces méthodes. Quand il s'agit de séparation aveugle des sources comme c'est généralement le cas de l'ACI, on cherche à retrouver les différentes sources existantes alors que lorsque l'extraction aveugle de sources est considérée, la recherche se limite à un nombre limité de sources parmi celles existantes dans le milieu observé.

La recherche des sources originales se précise en posant le problème dans son contexte mathématique et matriciel de la façon suivante :

$$X = AS \quad (3.89)$$

Le vecteur X représente les différents capteurs fournissant les signaux observés résultant de la combinaison des sources par les propriétés de mixtion du milieu définies par la matrice de mélange A. Enfin le vecteur S contient les différentes sources recherchées.

Par exemple pour un système composé de trois sources j du vecteur Sj mélangées par la matrice de mélange [a_{ij}] et observées par trois capteurs i donnant les signaux Xi la relation est présentée ainsi :

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} \quad (3.90)$$

$$x_1 = a_{11}s_1 + a_{12}s_2 + a_{13}s_3$$

$$x_2 = a_{21}s_1 + a_{22}s_2 + a_{23}s_3$$

$$x_3 = a_{31}s_1 + a_{32}s_2 + a_{33}s_3$$

Les méthodes comme l'ACI (ou SAS et EAS) cherchent par différents algorithmes à élaborer une matrice de séparation W qui à partir du vecteur observé va retrouver une estimation des sources recherchées. La démarche est la suivante :

$$\begin{aligned} Y &= WX \\ Y &= WAS \end{aligned} \quad (3.91)$$

L'objectif des différents algorithmes existants est de trouver par calculs statistiques, la matrice de séparation qui tend vers l'inverse de la matrice de mélange inconnue, de telle sorte que :

$$W = A^{-1} \quad \text{pour que} \quad \hat{Y} = S \quad (3.92)$$

L'enjeu consiste donc à définir cette matrice de séparation W par des méthodes empruntant différents domaines touchant à la théorie de l'information et du traitement du signal où les théories d'estimation et d'optimisation sont mises à contribution [4].

Pour alléger les calculs complexes très exigeants, on a recourt généralement à un prétraitement des données observées afin de travailler avec des signaux dont la moyenne est réduite et la variance unitaire.

De plus les hypothèses de bases concernant les signaux observés exigent pour garantir la performance des méthodes de l'ACI de s'assurer que les sources ou composantes indépendantes sont statistiquement indépendantes et que leur distribution statistique soit non-Gaussienne.

Ces caractéristiques sont d'ailleurs souvent rencontrées dans les signaux vocaux qui nous intéressent. Pour aller dans le sens de ces dernières considérations les données observées seront généralement blanchies ou encore décorréliées.

3.21 La séparation des sources dans notre cas d'étude

Maintenant que la présentation de l'ACI a été brièvement expliquée pour les besoins de compréhension, il sera possible de décrire les similitudes mais surtout les différences dans l'approche préconisée par les auteurs [10].

En ce qui a trait aux similitudes, c'est surtout l'aspect du prétraitement qui est directement appuyé sur l'approche de la méthode de l'ACI.

Pour les différences, il faut remarquer au départ que le nombre de capteur n'égale pas celui des sources ou composantes indépendantes comme vu précédemment. Cette égalité est souvent souhaitée, puisque la matrice de séparation estimée W doit tendre vers la matrice inverse de mélange soit A^{-1} . Hors généralement une matrice doit être carrée pour espérer s'inverser. Lorsque l'inégalité découle de la supériorité du nombre des sources sur celui des signaux observés, l'approche l'ACI est dite à bases surcomplètes et son traitement est ardu, on fait souvent appel aux calculs matriciels employant une matrice pseudo-inverse mais le coût en complexité de calcul est prohibitif pour l'appliquer dans un environnement d'un DSP comme le nôtre.

Mais la différence majeure par rapport à l'approche ACI est qu'on se dissocie du contexte aveugle par le fait que l'information sur les sources ou plus particulièrement sur la source vocale qui nous intéresse est disponible même si la matrice de mélange reste méconnue. On désire extraire seulement une source vocale dont l'énergie est la plus importante. Bien que cette source possède des aspects similaires aux signaux aléatoires elle conserve des portions dont les propriétés sont quasi stationnaires et où l'on peut retrouver une structure caractérisée par une fréquence fondamentale affectée d'un déphasage connu.

C'est précisément ces dernières informations qui permettent d'éviter de recourir à l'approche ACI à bases surcomplètes, elles évitent aussi les écueils souvent rencontrés

dans les résultats donnés par les méthodes générales d'ACI qui consiste dans l'ambiguïté sur l'ordre réel des sources retrouvées et la détermination de leur énergie respective.

3.22 La méthode préconisée pour l'extraction de la source désirée

Dans l'approche des auteurs [10], la méthode du traitement du calcul pour la minimisation de la moyenne des carrés de l'erreur s'effectue principalement en lot ou hors ligne. Sachant que :

$$y_k = W^T x \quad \text{et} \quad \varepsilon_k = y_k - e^{-jk\phi_0} \quad (3.93)$$

Et la fonction coût ou l'erreur est :

$$\xi(W) = E[\varepsilon_k^2] \quad (3.94)$$

atteint son minimum lorsque le gradient s'annule alors on obtient pour W l'expression :

$$W = E[xx^T]^{-1} E[xe^{-jk\phi_0}] = E[xe^{-jk\phi_0}] \quad \text{quand} \quad \nabla \xi(w) = 0 \quad \text{et} \quad E[xx^T]^{-1} = I \quad (3.95)$$

Φ_0 étant associée à la fondamentale ou une de ses harmoniques avec son déphasage, le vecteur entrée x ait été préalablement blanchi comme l'indique l'équation 3.95.

Nous suggérons plutôt un traitement en ligne des auteurs D.P. Mandic et A.Cichoki [25]. Il s'appuie essentiellement sur la nature des signaux vocaux ayant une structure dynamique bien connue, aussi l'algorithme procède sur la propriété prédictive des sources et vise également la minimisation de l'erreur de prédiction instantanée de la source à extraire présentant l'énergie principale comme hypothèse des auteurs [10].

L'algorithme doit rechercher et extraire parmi un ensemble de sources celle qui a l'énergie la plus importante. Pour se faire, il doit se référer à un ou des critères de séparation qui relèvent des particularités ou de la structure du signal recherchée. Dans le cas présent c'est évidemment l'information sur la fréquence fondamentale f_0 et son déphasage qui va être utile.

Le modèle de l'algorithme peut être interprété par la figure suivante où est illustré uniquement un canal pour fin de clarté, autrement dit, on ne considère qu'une rangée de la matrice de séparation ici w_{11} à w_{1M} , où M est le nombre de capteur ou microphone. Les coefficients du filtre prédictif du même canal étant de b_{11} à b_{1N} .

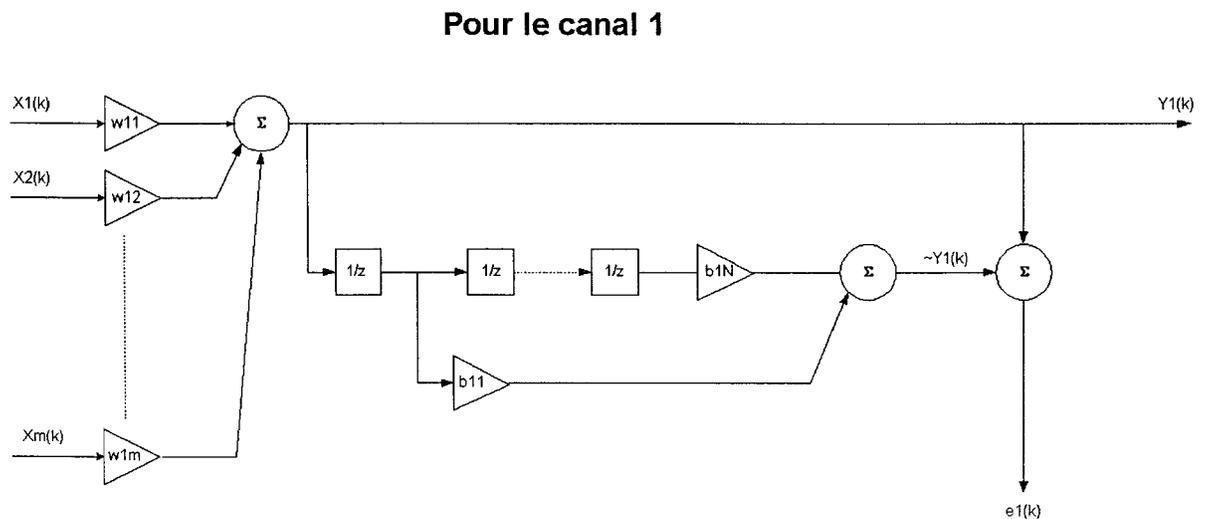


Figure 13 Matrice de séparation et prédicteur rehausseur selon [25]

L'erreur $e_1(k)$ sur la figure 13 va influencer la mise à jour des coefficients des filtres prédictifs et de la matrice de séparation W . Le nombre N de coefficient du filtre prédictif b_{11} à b_{1N} est variable, la figure 13 n'illustre que le premier et dernier coefficient du filtre prédictif.

La mise à jour des coefficients s'effectue de la façon suivante :

$$b_{1j}(k+1) = b_{1j}(k) - \mu_b \nabla_{b_{1j}} J(w_1(k), b_1(k)) \quad (3.96)$$

$$W_{1i}(k+1) = W_{1i}(k) - \mu_w \nabla_{w_{1i}} J(w_1(k), b_1(k)) \quad (3.97)$$

Sachant que l'expression de l'erreur se définit comme étant :

$$J(w_1(k), b_1(k)) = \frac{1}{2} e_1^2(k) \quad \text{et que} \quad e_1(k) = y_1(k) - \hat{y}_1(k) \quad (3.98)$$

Où

$$y(k) = W^T X \quad \text{et} \quad \hat{y}(k) = b^T y(k-1) \quad (3.99)$$

Après avoir effectué le calcul détaillé du gradient la mise à jour de la règle d'apprentissage devient :

$$b_{1j}(k+1) = b_{1j}(k) + \mu_b e_1(k) y(k-j) \quad (3.100)$$

Et

$$W_{1i}(k+1) = W_{1i}(k) - \mu_w e_1(k) x_i(k) \quad (3.101)$$

Les taux d'apprentissage ou pas d'adaptation μ_b et μ_w sont également adaptatifs et traités de la même façon que ceux vus plus haut dans les algorithmes du LMS, c'est-à-dire qu'ils sont normalisés par la puissance instantanée de leur entrée respective, mais le présent algorithme introduit un paramètre de combinaison convexe λ défini dans l'intervalle $[0,1]$ pour produire une pondération entre les deux pas d'adaptation. On a ainsi :

$$\mu_b = \frac{\lambda}{\|y(k)\|^2} \quad \text{et} \quad \mu_w = \frac{1-\lambda}{\|x(k)\|^2} \quad (3.102)$$

Cette approche sur les pas d'adaptation selon [25] est plus propice pour les signaux à dynamique élevée et non stationnaires et où les propriétés de mixtion sont variables dans le temps.

C'est le nombre de coefficients N du filtre prédictif qui va discriminer la source recherchée. Généralement les sources aux variations lentes exigent des prédicteurs avec des séries de coefficients plus longues que celles aux variations rapides. Connaissant la fréquence fondamentale il suffit de fixer en conséquence la valeur de N valant ainsi $1/f_0 = nT_e$, c'est-à-dire un multiple de la période d'échantillonnage.