



MESURE DE L'INSTANT DE FERMETURE GLOTTALE
AVEC LA TRANSFORMATION DE WIGNER-VILLE

J.C. VALIERE, J.F. ALLARD, C. DEPOLLIER

Laboratoire d'Acoustique - UA 1101 - Université du Maine - B.P. 535 - Route de Laval
72017 LE MANS Cedex - France

Résumé

Pour montrer l'intérêt de la pseudo distribution de Wigner-Ville lissée dans le cas des signaux non stationnaires à structure complexe, un signal de parole voisée a été étudié simultanément avec cette transformation et la transformation de Fourier à court terme (TFCT). L'intérêt de la première transformation apparaît clairement, celle-ci permettant, contrairement à la TFCT, de mettre nettement en évidence et d'obtenir l'instant de fermeture glottale avec une bonne précision.

Abstract

In order to point out the interest of the smoothed pseudo Wigner-Ville distribution in the case of non stationary signals with a complex structure, a voiced speech signal has been studied simultaneously with this transformation and the STFT. The interest of the first transformation appears clearly, the Wigner-Ville distribution allowing, unlike the STFT, to point out and to get the instant of glottal closure with a good accuracy.

1. Introduction

La pseudo distribution de Wigner-Ville [1], [2] a été utilisée pour analyser des signaux à large bande non stationnaires, en particulier des réponses impulsionnelles de filtres et de haut-parleurs [3].

Un signal de parole voisée n'est pas stationnaire, l'excitation du conduit vocal créant des perturbations périodiques. C'est un signal assez complexe dans son contenu fréquentiel et les tentatives d'analyse par la pseudo distribution de Wigner-Ville sont d'interprétation difficile [4], à cause des interférences entre les différentes composantes du spectre [5]. Il est nécessaire d'effectuer un lissage temporel pour atténuer l'effet de ces interférences [6]. Nous montrerons qu'avec ce lissage temporel, la pseudo distribution de Wigner-Ville peut être utilisée pour l'étude de tels signaux.

2. La pseudo distribution de Wigner-Ville

La transformation de Wigner-Ville [7], [8] associée à un signal temporel $x(t)$ la fonction W_x dépendant du temps t et de la fréquence f ci-dessous :

$$W_x(t, f) = \int_{-\infty}^{+\infty} x(t + \frac{\tau}{2}) x^*(t - \frac{\tau}{2}) e^{-i2\pi f \tau} d\tau \quad (1)$$

Cette fonction est une représentation temps fréquence de l'énergie du signal $x(t)$. Elle a l'inconvénient de comporter, à cause de sa structure bilinéaire, des interférences entre les différentes composantes fréquentielles et/ou temporelles du signal. Les interférences entre composantes de fréquence positive et de fréquence négative peuvent être éliminées par remplacement du signal $x(t)$ par le signal analytique $z(t)$

associé. Cette opération permet aussi de placer la fréquence de repliement à la moitié de la fréquence d'échantillonnage alors qu'elle est au quart de la fréquence d'échantillonnage si $x(t)$ est utilisé [9]. Une partie des interférences subsistantes peut être supprimée en utilisant une fenêtre d'observation temporelle $h(t)$ à valeurs réelles qui isolera les composantes fréquentielles non simultanées [5]. La nouvelle représentation temps fréquence ainsi obtenue est la pseudo distribution de Wigner-Ville [1]. Si elle est associée au signal analytique, elle a pour expression :

$$PW_z(t, f) = \int_{-\infty}^{+\infty} h^2(\frac{\tau}{2}) z(t + \frac{\tau}{2}) z^*(t - \frac{\tau}{2}) e^{-2\pi i f \tau} d\tau \quad (2)$$

Cette représentation temps fréquence a été utilisée pour des réponses impulsionnelles de filtres et de haut parleurs [3]. Des interférences subsistent entre composantes de fréquence positive. Un lissage temporel [10] permet de diminuer leur amplitude. La distribution associée est la pseudo distribution de Wigner-Ville lissée (PWL). Son expression est :

$$PWL_z(t, f) = \int_{-\infty}^{+\infty} g(t-u) PW_z(u, f) du \quad (3)$$

Dans cette expression, $g(t)$ est la fenêtre de lissage temporel. Pour affaiblir sensiblement un terme d'interférence entre deux composantes fréquentielles distantes de Δf , la largeur T de la fenêtre de lissage doit vérifier la relation [6] :

$$T \geq 1/\Delta f \quad (4)$$

Pour un signal échantillonné, on peut faire choix pour la PWL de l'expression ci-dessous [11] :



$$P\hat{W}L_z(t, f) = 2 \sum_{\tau=-N+1}^{N-1} |h(\tau)|^2$$

$$\sum_{k=-M+1}^{M-1} g_M(k) z(t+k\tau) z^*(t+k-\tau) e^{-i4\pi f\tau} \quad (5)$$

Cette formulation permet l'utilisation de la transformée de Fourier rapide. En effet l'égalité (5) peut être réécrite sous la forme :

$$P\hat{W}L_z(t, \frac{n}{2N}) = 4\text{Re} \left[\sum_{\tau=0}^{N-1} e^{-2\pi i \frac{n\tau}{N}} |h_N(\tau)|^2 \right]$$

$$\sum_{k=-M+1}^{M-1} g_M(k) z(t+k\tau) z^*(t+k-\tau) -$$

$$- \frac{1}{2} \sum_{k=-M+1}^{M-1} g_M(k) |z(t+k)|^2 \quad (6)$$

Cette dernière expression légèrement modifiée a été utilisée pour l'analyse de signaux simples non stationnaires [2]. Nous l'utiliserons pour des signaux à structure plus complexe en choisissant comme exemples des échantillons de parole voisée.

3. La parole voisée et la localisation temporelle de l'excitation glottale

Une voyelle est générée par une excitation périodique du conduit vocal créée par la fermeture de la glotte [12]. Un tel signal est représenté en Figure |1| pour la voyelle [ɔ]. La fermeture glottale précède immédiatement l'instant où la dérivée temporelle du signal de parole à son module maximal. Une portion du signal temporel de la Figure |1| a été représentée en Figure |2| avec une forte dilatation temporelle.

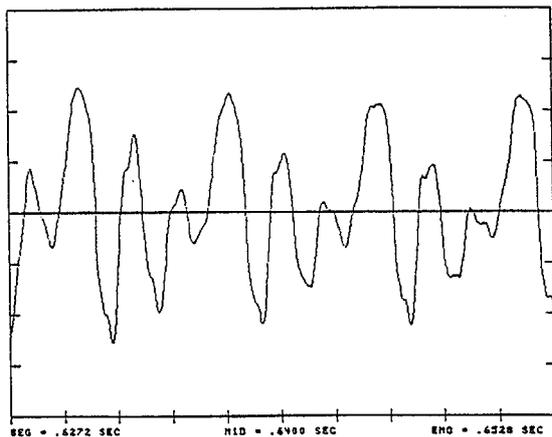


Fig.1: Représentation temporelle de la voyelle [ɔ]

Une fermeture glottale apparaît en Figure |2| pour t compris entre (0.6381 et 0.6383 ms). Différentes techniques ont été proposées pour localiser l'instant de fermeture de la glotte, par exemple celle de Sobakin [14], modifiée ultérieurement par Strube [13], qui utilise le déterminant de la matrice d'autocorrélation. Au moment où la glotte se ferme, le conduit vocal est soumis à une excitation et ne vibre pas librement : les interférences détruisent alors la structure des formants. On cherchera par examen des formants à déterminer l'instant de fermeture glottale.

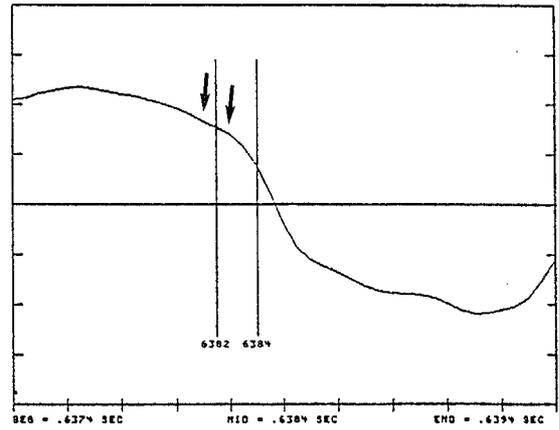


Fig.2 : Mise en évidence d'une fermeture glottale sur la voyelle [ɔ] de la Fig.1. La fermeture glottale apparaît entre les 2 flèches

4. Analyse d'un signal de parole par PWL

L'intérêt de la PWL est la bonne résolution en temps et en fréquence de cette représentation. L'utilisation du signal analytique à la place du signal réel, et un lissage temporel trop long, peuvent nuire à cette résolution. Pour mettre en évidence l'effet du filtrage de Hilbert sur la résolution temporelle, on a représenté en Figures |3| et |4| des PWL, calculées avec le signal temporel de la Figure |1| échantillonnée à 20 kHz, et avec le signal analytique obtenu par filtrage du signal temporel échantillonné à 10kHz.

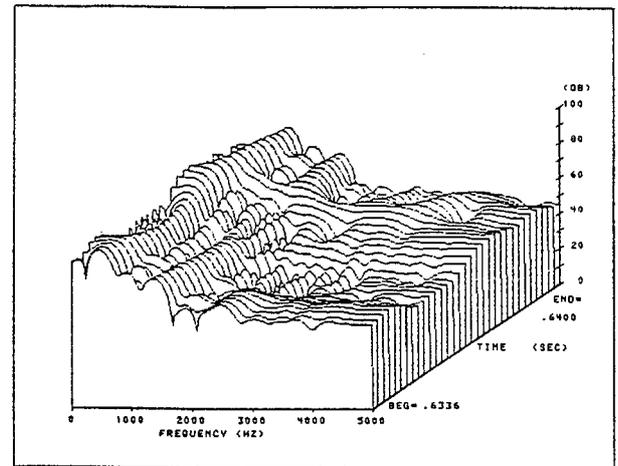


Fig.3 : Représentation par PWL (valeur absolue) du signal de la Fig.1. Durée de H = 6.4 ms. Durée de G = 2,1 ms. H et G sont des fenêtres de Kaiser Bessel.

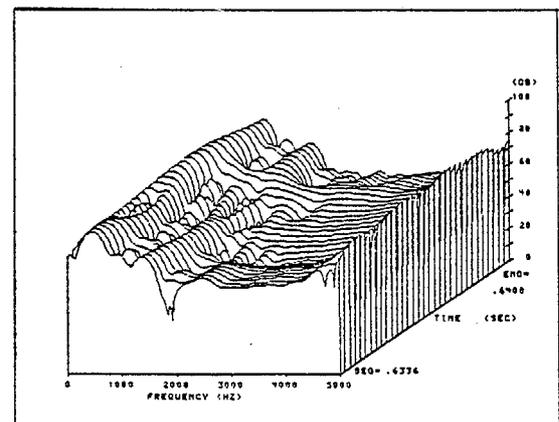


Fig.4 : Représentation par PWL (valeur absolue) du signal analytique associé au signal de la Fig.1. Durée de H = 6.4 ms. Durée de G = 2.1 ms. H et G sont des fenêtres de Kaiser Bessel.

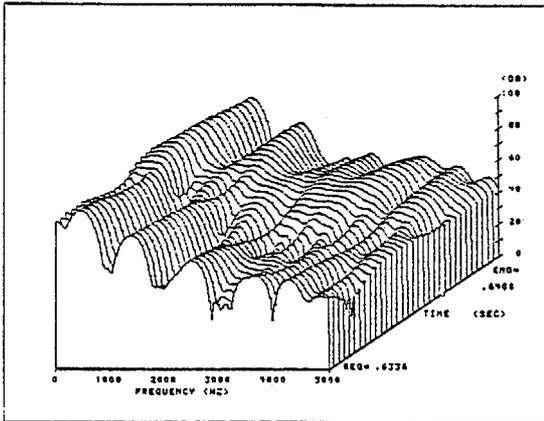


Fig. 5 : Représentation par TFCT du signal de la Fig.1. Durée de H = 6.4 ms. H est une fenêtre de Kaiser Bessel.

En Figure [3], le signal temporel a été utilisé dans les conditions suivantes ; la fenêtre h est une fenêtre de Kaiser-Bessel de durée 6,4 ms, la fenêtre g est une fenêtre de Kaiser-Bessel de durée 2,1 ms. En Figure [4], le signal analytique a été utilisé avec des fenêtres de même nature et de même durée. Le signal analytique a été obtenu avec un filtre de Hilbert de 43 éléments, de bande passante comprise entre 0,1 kHz et 4,9 kHz, obtenu par l'algorithme de Remez [15],[16]. Pour comparaison, les distributions temps-fréquence obtenues par TFCT avec la même fenêtre h sont représentées en Figure [5]. Les Figures [3] et [4] sont très similaires et il n'y a pas, pour le signal considéré, de diminution sensible de la résolution quand on utilise le signal analytique.

Deux distributions pratiquement sans formant apparaissent en Figure [3] et en Figure [4] quand l'élément central de la fenêtre h se trouve dans la région de la fermeture glottale. Cet instant peut ainsi être connu avec une précision de l'ordre de 0,2 ms, légèrement inférieure à celle obtenue par l'examen de la Figure [2]. La fenêtre g est alors trop courte pour supprimer les interférences quand les formants réapparaissent. Néanmoins l'intérêt d'une analyse par PWL avec une fenêtre de lissage courte apparaît évident quand on compare les Figures [3] et [4] avec la Figure [5], sur laquelle l'analyse par TFCT n'indique pas avec précision l'instant de fermeture glottale.

On peut affaiblir sensiblement les interférences en prenant une fenêtre de lissage plus large. L'analyse d'un [a] par PWL avec une fenêtre h identique aux précédentes et une fenêtre g de Kaiser-Bessel de 3,4 ms est représentée en Figure [6], ainsi que l'analyse par TFCT (Figure [7]).

Les deux représentations temps-fréquence sont alors très proches dans la bande passante du filtre anti-repliement qui s'étend jusqu'à 3,5 kHz. Il n'est plus possible de déterminer avec précision l'instant de fermeture de la glotte par utilisation de la PWL.

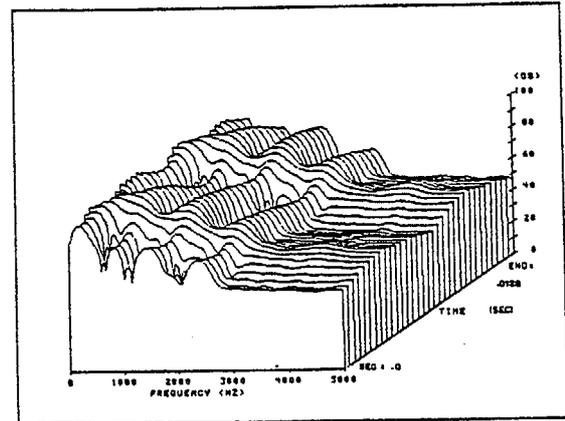


Fig. 6 : Représentation par PWL (valeur absolue) de la voyelle [a]. Durée de H = 6.4 ms. Durée de G = 3.1 ms. H et G sont des fenêtres de Kaiser Bessel.

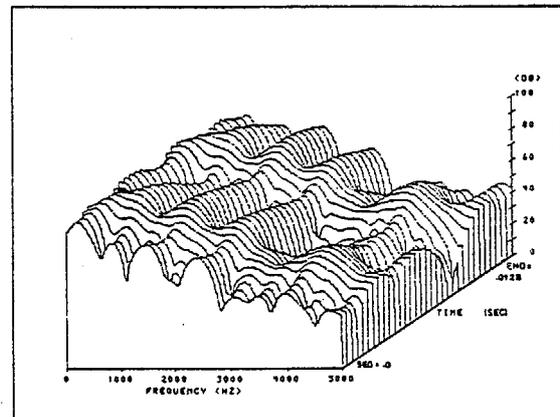


Fig. 7 : Représentation par TFCT de la voyelle [a] sur le même intervalle de temps que pour la Fig.5 et la même fenêtre H.

5. Conclusion

La PWL peut être plus performante que la transformation de Fourier à court terme pour l'étude des signaux non stationnaires à structure complexe. Néanmoins, l'affaiblissement des interférences et une bonne résolution dans les domaines fréquentiel et temporel sont des objectifs qui ne peuvent pas être atteints simultanément. La durée de la fenêtre de lissage temporel sera fonction d'un compromis dépendant de l'importance relative accordée à ces deux objectifs.

REFERENCES

- [1] T.A.C.M. Claasen and W.F.G. Mecklenbräuer
Philips J. Res. 35, 1980, p. 217-250.
- [2] P. Flandrin, B. Escudié
Traitement du Signal 2, 1985, p. 143-151.
- [3] Cornelis P. Janse and Arie J.M. Kaizer
J. Audio Eng. Soc. 31, 1983, p. 198-223.
- [4] D. Chester, F. Taylor, M. Doyle
IEEE ASSP Spectrum Estimation Workshop II,
Tampa, 1983, p. 98-102.



- |5| Franz Hlawatsch
Digital Signal Processing-84, V. Cappellini
et A.G. Constantinides eds., North-Holland,
Amsterdam, 1984, p. 363-367.
- |6| P. Flandrin
IEEE Int. Conf. on Acoustics, Speech and
Signal Processin, ICASSP-84, San Diego, 1984.
- |7| E.P. Wigner
Phys. Rev. 40, 1932, p. 749-759.
- |8| J. Ville
Cables et Transm., 2A, (1), 1948, p. 61-74.
- |9| Theo A.C.M. Clasen, and W.F.G. Mecklenbräuker
IEEE Trans. on Acoustics, Speech and Signal
Processing, ASSP-31, (5), 1983, p. 1067-1072.
- |10| W. Martin, P. Flandrin
Signal Processing : Theories And Applications
II, H. Schüssler ed., North-Holland,
Amsterdam, 1983, p. 455-458.
- |11| P. Flandrin, W. Martin et M. Zakharia
Digital Signal Processing-84, V. Cappellini
et A.G. Constantinides, eds., North-Holland,
Amsterdam, 1984, p. 262-266.
- |12| R.L. Miller
J. Acoust. Soc. Am. 31, 1959, p. 667-677.
- |13| Hans Werner Strube
J. Acoust. Soc. Am. 56, 1974, p. 1625-1629.
- |14| A.N. Sobakin
Soviet Physics-Acoustics 18, 1972, p. 84-90.
- |15| J.H. Mc Clellan, T.W. Parks, and L.R. Rabiner
IEEE Trans. on Audio and Electro-acoustics,
AU-21, N° 6, 1973, p. 506-526.
- |16| L.R. Rabiner and R.W. Schafer
The Bell System Technical Journal 53, 1974,
p. 363-390.