



**REPRESENTATION TEMPS-FREQUENCE ET ACOUSTIQUE
DES SALLES**

N. Martin*, J. Mars*, J.C. Pierron**, J. Martin **

* CEPHAG/ENSIEG bp 46 - 38402 St-Martin-d'Hères Cédex - France

** CSTB Division acoustique des Salles 24, rue Joseph Fourier - 38400 St-Martin-d'Hères Cédex - France

RESUME

Nous présentons une comparaison des images temps-fréquence de réponses impulsionnelles de salles de spectacles obtenues par différentes méthodes d'analyse. Deux critères sont ensuite proposés pour permettre la distinction entre deux salles dans le cas où les critères habituels (critères objectifs) s'avèrent insuffisants.

I - INTRODUCTION

Depuis de nombreuses années les acousticiens recherchent à caractériser l'acoustique d'une salle de spectacles. L'approche la plus naturelle consiste à déterminer un questionnaire soumis à différents auditeurs écoutant un spectacle dans une salle puis à partir de leurs réponses d'analyser les impressions qu'ils ressentent : on établit donc un jeu de critères dits subjectifs (clarté, intimité, chaleur, brillance...) qui au travers d'études psychoacoustiques permettent de juger de la qualité de la salle. Le point fort de cette méthode qui est d'être lié à la perception réelle des phénomènes dans une vraie salle, devient caduque dès lors que l'on veut faire de la prédiction, la salle n'étant pas construite. C'est pourtant le problème le plus fréquemment rencontré par le concepteur d'une salle en projet et qui ne connaît que les plans de l'architecte.

La démarche suivante a consisté à essayer de trouver des relations entre les critères subjectifs et les phénomènes physiques engendrés au cours de la propagation du son dans un volume ; ceci a donné naissance à des critères dits objectifs [JORDAN V.L. 1981] qui se déterminent sur la réponse impulsionnelle de la salle.

Bien que ces critères soient largement utilisés comme outils de description, de nombreux auteurs s'intéressent à leur pertinence tant au niveau de leur corrélation avec les impressions subjectives qu'au niveau de la discrimination des salles entre elles [JULLIEN 1986] [BARRON 1988]. Ces travaux montrent la difficulté que présentent ces critères pour répondre complètement à ces questions. Les figures 1 et 2 illustrent le cas de deux salles très différentes et qui ne peuvent être distinguées simplement en comparant les valeurs des principaux critères calculées sur leurs réponses impulsionnelles respectives.

Il s'agit de l'Opéra Bastille à Paris (2700 places, volume = 21000 m³ et de la salle Olivier Messiaen à Grenoble (320 places, 2300 m³) désignées respectivement par les abréviations BST et OLM sur les figures. La figure 1 représente le temps de réverbération T_R en secondes (temps que met l'énergie pour décroître de 60 dB après extinction de la source) mesuré par bande d'octave pour les deux salles. Les mesures étant réalisées en de nombreux points, les traits verticaux traduisent les fluctuations statistiques sur cet ensemble de mesures. De manière similaire, la figure 2 représente le critère de clarté 80 ms C_{80} (rapport de l'énergie qui arrive dans les 80 premières ms sur l'énergie qui arrive après les 80 premières ms).

ABSTRACT

We present a comparison between time-frequency images of concert hall impulse response estimated by various analysis methods. Two criteria are then proposed to allow distinction between two halls in case usual criteria (objective criteria) are proving insufficient.

Les barres d'erreur se chevauchant largement, il est clair qu'il n'est pas possible de discriminer ces deux salles simplement par la connaissance du T_R ou de la clarté. Il en est de même pour les autres critères principalement utilisés (G , EDT , T_c ...). Cependant, l'oreille est capable de séparer ces deux salles sans ambiguïté. C'est pourquoi nous nous sommes intéressés aux représentations temps-fréquence puisque dans leur principe elles se rapprochent de l'analyse faite par l'oreille.

Dans cet article, il faut cependant préciser que seule une première étape est abordée ; elle tente de répondre à la question suivante : les images temps-fréquences fournies par les différentes méthodes connues et complétées par un traitement adéquat, permettent-elles d'apporter une information supplémentaire dans la discrimination des salles ? L'étape suivante qui est la signification psychoacoustique de cette information n'est pas du tout abordée.

II - METHODOLOGIE TEMPS-FREQUENCE ET RESULTATS

Les réponses impulsionnelles mesurées in situ sont analysées en temps-fréquence par deux types de méthodes :

-les méthodes de Fourier : Fourier à court terme et distribution de pseudo Wigner-Ville-Lissée;

-les méthodes non-Fourier : Capon glissant, AR glissant et Lagunas glissant, méthodes hybrides (ARCAP-LAGCAP). N'étant pas l'objet du présent article, le principe des méthodes ne sera pas décrit [SFA 1990] [AUGER 1991]. Dans ce choix de méthodes, seule la distribution de Wigner-Ville est une méthode à proprement dite non-stationnaire. Toutes les autres supposent une stationnarité locale. Etant donné qu'il s'agit d'une réponse impulsionnelle de salle, et pour se rapprocher le plus possible de la perception de l'oreille, l'échelle des fréquences choisie n'est pas une simple échelle linéaire ou logarithmique, mais une suite de gammes. Quelle que soit la méthode, le spectre obtenu à la fréquence f_i est moyenné sur une bande ΔB_i variant de

$$\left(\sqrt[12]{2}\right)^{-1/2} \cdot f_i \text{ à } \left(\sqrt[12]{2}\right)^{1/2} \cdot f_i$$

Cette largeur de bande correspond à un demi-ton de la gamme naturelle. Une succession de 12 bandes représentent une octave (soit 12 demi-tons) où la fréquence centrale finale est double de celle de la première bande $f_{12} = 2f_1$.

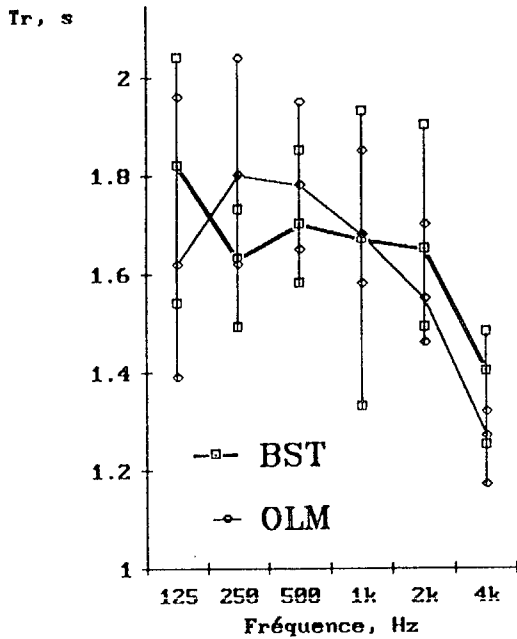


Figure 1 : Temps de réverbération

Donc, nous avons pour une date t et une fréquence f_i ; un spectre

$$\text{moyenné } \rho_m(t, f_i) = \frac{1}{N_f} \sum_{f \in \Delta B_i} \rho(t, f)$$

$$\text{avec } \Delta B_i = \left[\left(\frac{12\sqrt{2}}{\sqrt{2}} \right)^2 f_i, \left(\frac{12\sqrt{2}}{\sqrt{2}} \right)^2 f_i \right]$$

. N_f nombre de points de fréquence dans la bande ΔB_i

. $\rho(t, f)$ spectre à la fréquence f et à la date t estimé par une des méthodes citées.

Sur la figure 4 sont représentés les résultats obtenus sur les réponses impulsionnelles mesurées dans l'Opéra Bastille à Paris et dans la salle O. Messiaen à Grenoble, toutes deux sans spectateurs, et analysées par les méthodes de Fourier à court terme, Pseudo-Wigner-Ville-Lissée et Capon glissant. Les résultats par les autres méthodes n'ont pu être insérés mais seront commentés.

En observant la représentation de la transformée de Fourier glissante, nous observons la répartition de l'énergie sur les différentes notes émises par une source, soit une décroissance lente au cours de la propagation dans la salle (c'est-à-dire au cours du temps), décroissance variant suivant la fréquence. Des coupes à fréquence constante donnent immédiatement cette information. La durée de réverbération (T_r) de la salle en fonction de la fréquence est estimée par des coupes à date constante. Nous remarquons un T_r relativement plat quelle que soit la fréquence pour les deux salles et surtout la bonne présence des aïgues dans le cas de la Bastille : salle présentant une bonne brillance (octave 2000 Hz).

Lorsque nous examinons les figures obtenues dans une même salle pour différentes positions des récepteurs (non représentées ici), il apparaît nettement des images distinctes en fonction de la position pour des temps correspondant à des distances inférieures au rayon de réverbération (premières réflexions fonction de la position) mais semblables au delà. L'image obtenue au-delà du rayon de réverbération fournit une caractéristique invariante pour une même salle.

Les images obtenues sur différentes salles sont toutes très distinctes excepté le cas critique présenté ici.

Si nous observons la transformée de Pseudo-Wigner-Ville Lissée, les résultats sont visuellement sensiblement équivalents. Des traitements ultérieurs pourraient éventuellement les distinguer.

La méthode de CAPON, estimateur de puissance, fournit une représentation plus lissée qui est équivalente à celle obtenue par

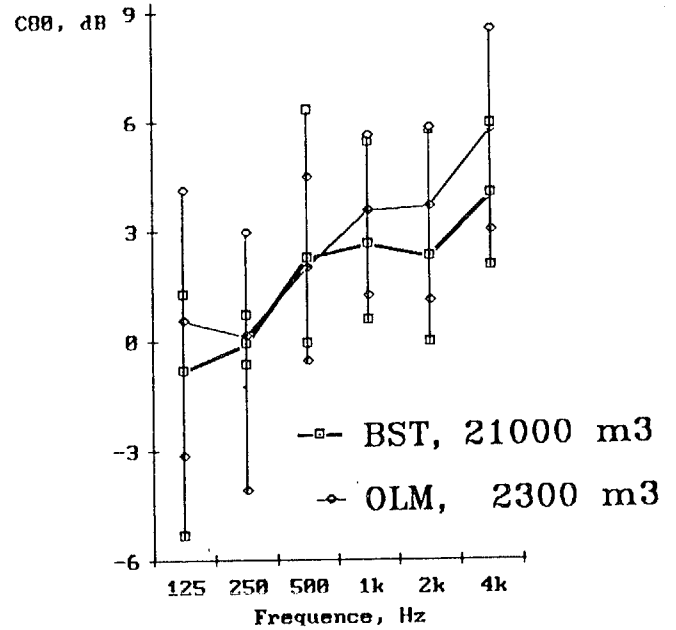


Figure 2 : Clarté 80 ms

la méthode de Lagunas, estimateur de densité spectrale de puissance. En effet, les signaux traités n'ont pas besoin de traitement haute-résolution. Ce sont des spectres à variation lente. Ainsi, le densité spectrale de puissance et la puissance, qui est l'intégrale de la densité spectrale, ont une représentation de même structure. Par contre, un modèle AR n'est pas adapté à ce type de signaux correspondant à un nombre de modes très important [KUTTRUFF 1979]. Il n'a pas été possible de trouver un ordre permettant une modélisation acceptable. A chaque mémoire temporelle sur laquelle la stationnarité est supposée locale, nous observons un signal à très large bande pour lequel la méthode de Lagunas-Capon est plus performante par rapport au modèle AR. Un modèle AR ne peut s'employer que sur des bandes "moyennement larges" sur lequel un ensemble de pôles très proches les uns des autres permet la modélisation. Ce résultat avait déjà été pressenti lors de l'étude de signaux vibratoires de machines tournantes [BONDANT 1990].

III - CRITERES DE DIFFERENCIATION

L'image temps-fréquence de la densité spectrale d'énergie rassemble un ensemble d'informations sur la qualité acoustique de la salle. On est donc amené à chercher des méthodes pour extraire automatiquement ces informations. Nous proposons deux critères issus des images temps-fréquence tout en précisant que ces critères ont été testés sur un petit nombre de salles, une validation est encore nécessaire. Ces critères sont des critères de distinction entre salles, l'étape suivante serait de les relier aux impressions subjectives.

III-1 Opérations Morphologiques Binaires

Quel que soit d'ailleurs le type de signal considéré, l'interprétation automatique d'une représentation temps-fréquence conduit tout naturellement aux techniques de traitement d'images. Le traitement de ces images particulières obtenues par analyse temps-fréquence reste un domaine très ouvert [DOLIGEZ T. 1990].

Il est certain que la méthode utilisée pour estimer le spectre va influencer de manière forte sur la structure de l'image. Ainsi, l'interprétation du traitement d'image ne peut être dissociée du choix de la méthode temps-fréquence.

Notre première approche a consisté à simplifier l'image en ne gardant que les contours de zones correspondant à une énergie concentrée dans une bande, contours estimés à partir d'opérations morphologiques binaires sur l'image [SERRA 1982].

Cette détection de contours a été réalisée sur une image binarisée. Il est clair que cette opération de binarisation est sévère et élimine une grande partie de l'information. Notre objectif est de

définir un critère parmi un ensemble de critères permettant de caractériser le signal à partir d'une image temps-fréquence.

Les étapes de la méthode sont :

-binarisation de l'image à partir d'un seuil : les valeurs supérieures à ce seuil sont forcées à 1, les valeurs inférieures sont annulées ;

-ouverture de l'image par un élément structurant approché par une croix de taille r_1 (rayon de la boule dans laquelle s'insère la croix) correspondant à l'enchaînement d'une érosion et d'une dilatation de taille r_1 . Cette opération, dans le cas présent, effectuée un lissage de l'image ;

-détection de contours correspondant à l'image résultante d'une soustraction entre l'image initiale (résultat de l'opération précédente) et une érosion de l'image par une croix de taille r_2 ;

-suivi de contours.

Dans un premier temps, ces opérations morphologiques ont été réalisées sur les images obtenues par la méthode de Fourier (non représentée ici). Le noyau binaire obtenu et ce, pour différents seuils, est trop granuleux. Afin de pouvoir comparer deux salles, il est impératif d'appliquer une ouverture de taille supérieure à 1 qui conduit à "écraser" la partie intéressante. Aucune comparaison n'est alors possible.

Dans un deuxième temps, ces opérations morphologiques ont été réalisées sur les images d'aspect plus lissé obtenues par la méthode de CAPON. Le résultat présenté (figure 3) a été calculé avec un seuil de 62% par rapport au maximum, une ouverture de taille $r_1 = 1$ et une détection de contours à partir d'une érosion de taille $r_2 = 1$.

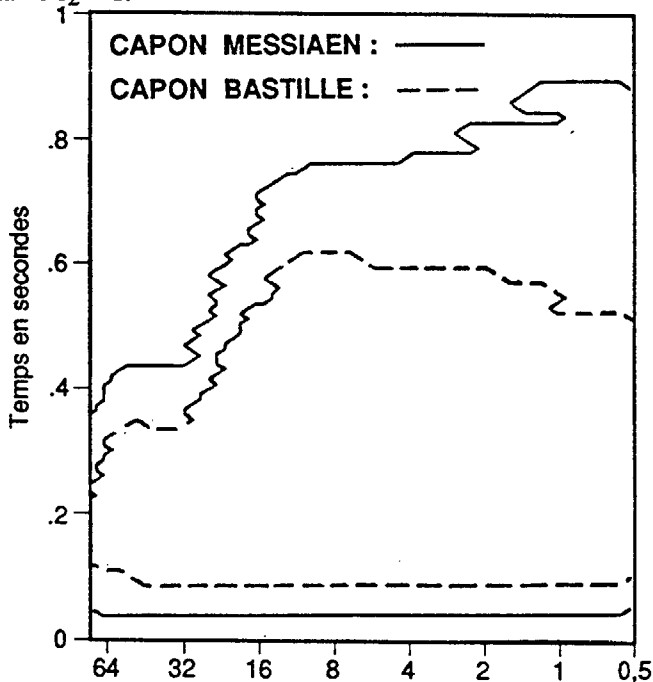


Figure 3 : Détection de contours (seuil 62%)

Contrairement aux images de Fourier, la superposition des contours obtenus permet la distinction entre ces deux salles critiques O. MESSIAEN et BASTILLE. Ce premier résultat est encourageant mais une étude plus approfondie est indispensable pour valider la méthode.

III-2 Utilisation de la surface temps-fréquence

Ce critère est basé sur la surface temps-fréquence des différentes zones de couleur de la représentation temps-fréquence. Cette surface est représentée en pourcentage par rapport à la surface totale. Pour que ce critère soit valable, il est impératif qu'il n'y ait pas de maximum maximorum "isolé" important qui modifierait notablement la représentation couleur et empêcherait toute comparaison de salles entre elles. Sur plus de 20 images analysées ; ce cas n'a jamais été rencontré. Afin d'éviter que les bruits basses fréquence ne polluent les résultats, la fréquence la

plus basse n'est plus 50 Hz (sol 0), mais 100 Hz (sol 1). Les résultats présentés dans le tableau suivant ont été obtenus après moyennage sur les pourcentages mesurés pour différentes places dans la salle pour une source fixe [PIERRON 1991].

Bastille	MOYENNE	Min	Max	Min %	Max %
	-16 dB à -24 dB	19.1	18.3	19.7	-4.2
-24 dB à -32 dB	20.5	19.0	21.3	-7.3	+3.9
-32 dB à -40 dB	20.5	19.8	21.3	-3.4	+3.9
Messiaen					
Messiaen	MOYENNE	Min	Max	Min %	Max %
	-16 dB à -24 dB	17.9	17.4	18.8	-2.8
-24 dB à -32 dB	18.4	18.3	18.5	-0.5	+0.5
-32 dB à -40 dB	20.3	19.7	20.8	-3	+2.5

Tableau : Pourcentage de surface temps-fréquence moyenne sur différentes places et écart-types.

L'observation de ces résultats, en particulier les écarts-types sur la moyenne, nous a permis de fixer le choix d'un intervalle de confiance à 5%. Cet intervalle de confiance appliqué sur la moyenne nous autorise à conclure sur le fait que les deux salles sont discriminés par ce critère. Il en est de même pour distinguer la Bastille vide de la Bastille pleine (tableaux non représentés ici).

Conclusion

Cette première étude nous a permis de comparer quelques techniques temps-fréquence sur des réponses impulsionnelles de salles. La transformée de Fourier à court-terme et la distribution de Wigner-Ville apportent des résultats sensiblement équivalents. En non-stationnaire, la méthode autorégressive semble peu adaptée à cause d'un nombre de modes trop important. Les méthodes de CAPON et LAGUNAS fournissent des contours de zone d'énergie plus lisses, ce qui permet une détection de contour par des opérations morphologiques binaires. Les courbes ainsi obtenues sont discriminantes dans le cas de deux salles critiques non séparées par les critères objectifs habituels de même que pour un deuxième critère basé sur le pourcentage de surface couleur. Ces deux critères ne peuvent encore pas être utilisés comme instruments de mesure ou de comparaison entre salles. Il faudrait les rattacher aux impressions subjectives de l'écoute dans les salles en complétant cette étude par une étude psychoacoustique des salles mesurées.

Références

- AUGER F., C. DEMARS, T. DOLIGÉZ, P. FLANDRIN, C. LAMBERT-NEBOUT, J. MARS, N. MARTIN, F. MOLINARO, J.P. OUARLEZ, O. RIOUL (1991) : Rapport de synthèse. Opération "Signaux-test", Méthodes Temps-fréquence - fiches synthétiques. Assemblée générale du GDR 134, Traitement du Signal et Images, Batz sur Mer, 8-9 Avril 1991.
- BARRON M., L.J. LEE (1988) : Energy relations in concert auditoriums I. J. Acoust. Soc. Am. 84, 1988, p.618.
- BONDANT L., J. MARS, N. MARTIN (1990) : Analyse spectrale de signaux non-stationnaires par des méthodes temps-fréquence. SFA, Lyon, 10-13 Avril 1990.
- DOLIGÉZ T. (1990) : Traitement d'images temps-fréquence. Dea Signal-Image-Parole - INP Grenoble, Sept. 1990.
- JORDAN V.L. (1981) : A Group of objective acoustical criteria for concert hall. Applied Acoustics 14, 1981, pp. 253-266.
- JULLIEN J.P. (1986) : Correlation among objective criteria of room acoustic quality. Proceedings 12th ICA Toronto, 1986.
- KUTTRUFF H (1979) : Room Acoustics. Applied Science Publishers LTD. London 1979.
- PIERRON J.C. (1991) : Utilisation des images temps-fréquence obtenues par transformation de Wigner-Ville en acoustique des salles. Rapport CSTB n° 2.90.506, Fév. 91.
- SERRA J. (1982) : Mathematical Morphologie and image analysis. Academic Press, Londres 1982.



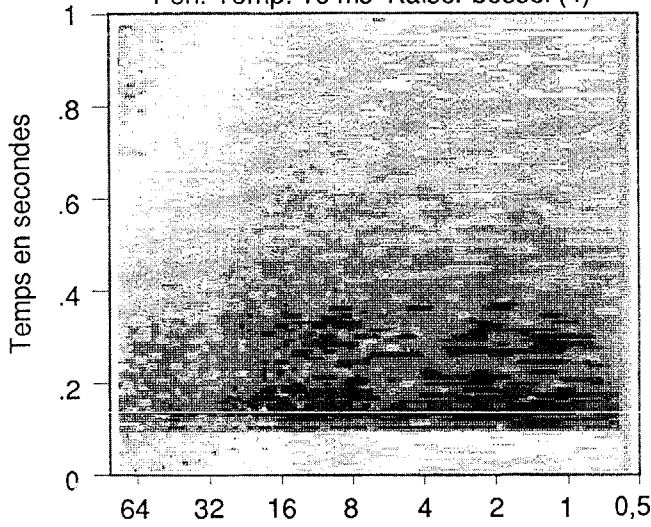
BASTILLE VIDE

O. MESSIAEN VIDE

Amplitude logarithmique - Décalage temporel : 8 ms

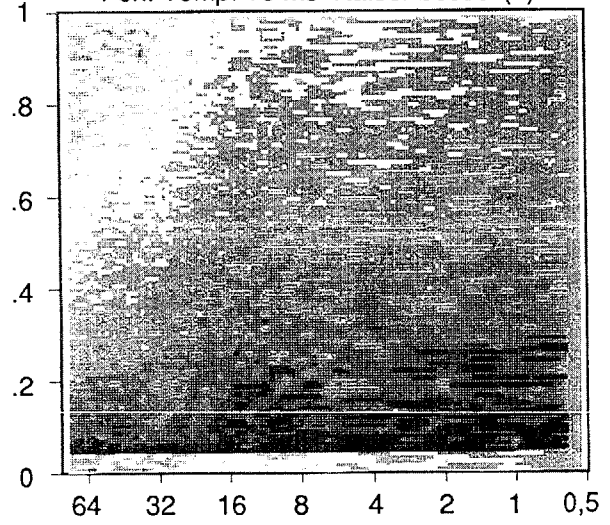
TRANSFORMEE DE FOURIER

Fen. Temp: 16 ms Kaiser.bessel (4)



TRANSFORMEE DE FOURIER

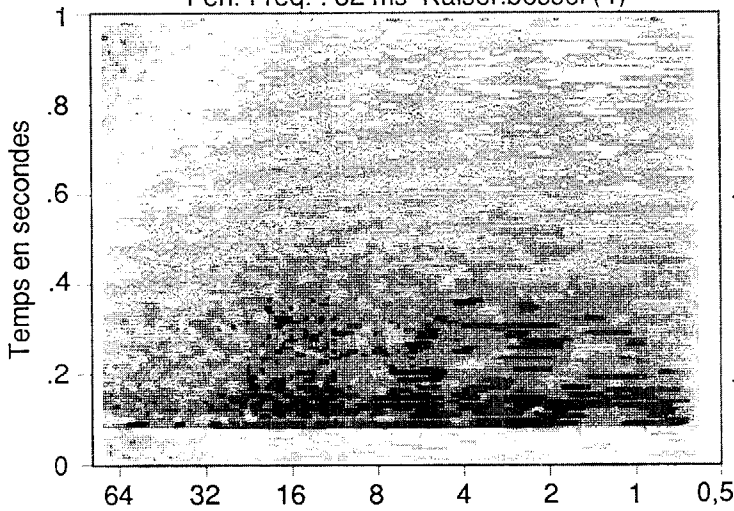
Fen. Temp: 16 ms Kaiser.bessel (4)



PSEUDO - WIGNER - VILLE LISSEE

Fen. Temp: 16 ms Kaiser.bessel (4)

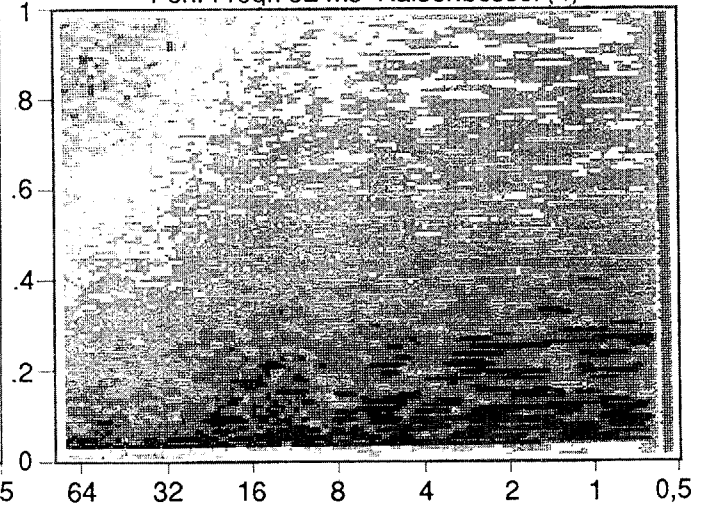
Fen. Fréq. : 32 ms Kaiser.bessel (4)



PSEUDO - WIGNER - VILLE LISSEE

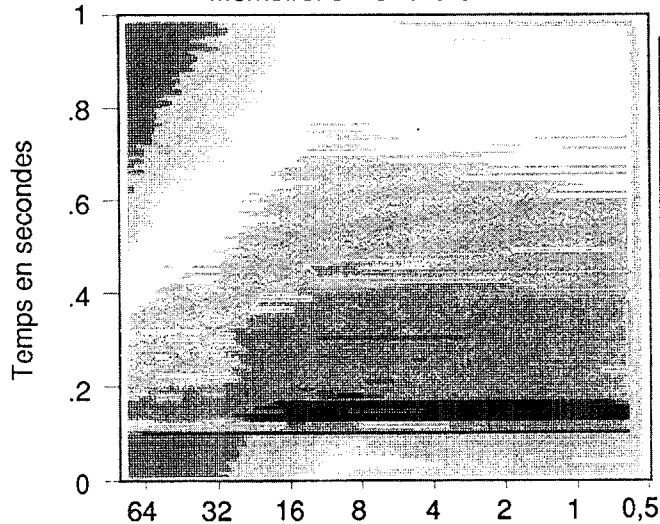
Fen. Temp: 16 ms Kaiser.bessel (4)

Fen. Fréq. : 32 ms Kaiser.bessel (4)



CAPON

Mémoire: 8 ms Ordre 16



CAPON

Mémoire: 8 ms Ordre 16

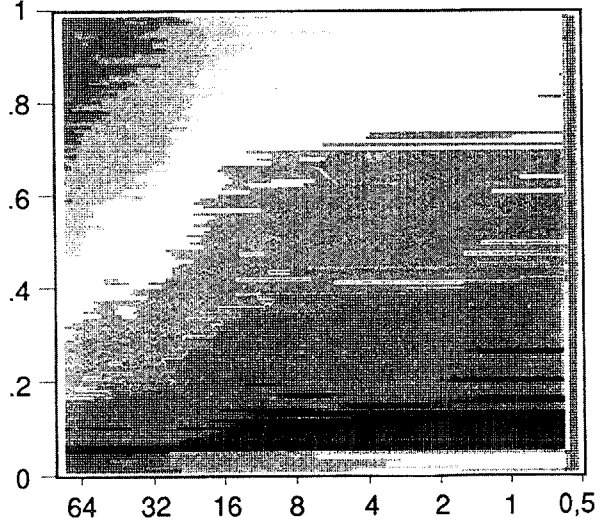


Figure 4 : Représentations Temps-Fréquence par Fourier glissant, Pseudo-Wigner-Ville Li ssée et Capon