TREIZIÈME COLLOQUE GRETSI - JUAN-LES-PINS DU 16 AU 20 SEPTEMBRE 1991

TRANSFORMATION EN ONDELETTES SUR UNE ECHELLE FREQUENTIELLE AUDITIVE

C. d'Alessandro & D. Beautemps †

†actuellement Institut de la Communication Parlée, CNRS UA368-/ENSERG - Université Stendhal.INPG: 46, Avenue Félix-Viallet 38031 Grenoble Cédex

LIMSI-CNRS, BP 133, 91403 Orsay Cédex

RÉSUMÉ

Ce papier présente une représentation par ondelettes sur une échelle fréquentielle auditive, pour le traitement de la parole. Les relations entre analyse par ondelettes et filtrage linéaire sont discutées. Un système d'analyse-synthèse basé sur ces principes est présenté, ainsi que des exemples d'analyse et de modifications de la parole.

ABSTRACT

This paper presents a wavelet representation for speech signal, using an auditory based frequency scale. The relationships between wavelet representation and linear filtering are discussed. An analysis-synthesis system based on these principles is presented, and some examples of analyses and modifications are shown. Stylization of the auditory spectrograms is also discussed, according to auditory-acoustic parameters.

1 INTRODUCTION

Les méthodes de représentation temps/fréquence non-paramétriques linéaires sont fondamentales traitement de la parole. Les outils les plus répandus sont le spectrographe acoustique et, pour la modification, le vocodeur de phase, liés à la transformée de Fourier à court terme. Cette méthode utilise une répartition linéaire en fréquence, et des largeurs de bandes constantes. La répartition des fréquences centrales sur une échelle logarithmique, avec des largeurs de bandes proportionnelles à la fréquence se comprend aujourd'hui dans le cadre de la transformée en ondelettes [4]. Un troisième type d'échelle spectro-temporelle provient des études sur la perception, pour mieux rendre compte des capacités d'analyse de l'oreille humaine [5]. La figure 1 confronte le trois types de répartition des filtres: lineaire (100 Hz et 300 Hz de largeur de bande, correspondant au spectrographe en bande étroite et en bande large), logarithmique (en 1/3 d'octave), et en échelle Bark. Il apparaît clairement que l'échelle auditive se comporte comme le spectrographe en bande étroite sous environ 800 Hz (figure 1.A), et comme l'analyse en 1/3 d'octave au delà (figure 1.B).

Le propos de cette communication est de développer une représentation en ondelettes du signal sur une échelle Bark, pour obtenir 1. un spectrographe auditif 2. un système d'analyse/synthèse qui permette de manipuler le signal dans le plan temps-fréquence, en accord avec la résolution spectro-temporelle auditive.

La deuxième section introduit l'interprétation en ondelettes de l'analyse/synthèse par banc de filtres. La troisième discute succintement des applications: spectrographe auditif, vocodeur.

2 Interprétation en ondelettes du filtrage linéaire

Le propos de cette section est de montrer sous quelles conditions générales un signal $s(\tau)$, $s \in L^2$ peut être représenté comme une somme discrète d'ondelettes indéxées par leurs positions spectro-temporelles (t_n, f_m) et pondérées par les coefficients (c_{nm}), suivant une expression de la forme:

$$s(\tau) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} c_{nm} o[t_n, f_m](\tau) \qquad (1)$$

L'interprétation de cette expression en terme de filtrage linéaire est le point-clef du développement qui va suivre [2].

Afin de pouvoir interpréter facilement la représentation, les ondelettes utilisées sont des fonctions localisées en temps/fréquence. Ces fonctions doivent posséder un maximum spectro-temporel principal, et être négligeables en dehors d'un pavé $[t_1, t_2] \times [f_1, f_2]$. On peut alors évaluer le comportement local de s par comparaison avec les fonctions $o[t_n, fm]$, et considérer l'expression (1) comme une décomposition de s sur un ensemble discret de points du plan temps-fréquence.

2.1 Coefficients d'ondelettes et filtres d'analyse

Soit $w[f](t)$ la réponse impulsionnelle d'un filtre passe-bande, de fréquence centrale f, $b[f](t)$ le signal obtenu en filtrant s par $w[f]$. Soit *l'ondelette analysante* $o[t, f]$ centrée au point spectro-temporel (t, f), et obtenue dans le domaine temporel à partir de $w[f]$ par inversion du sens du temps, décalage à l'instant t et conjugaison complexe (notée par *): $o[t, f](\tau) = w^*[f](t - \tau)$

On peut interpréter la réponse temporelle à l'instant t du filtre $w[f]$ comme la corrélation du signal s et de l'ondelette o.

$$\begin{aligned} b[f](t) = \int_{-\infty}^{+\infty} s(\tau) w[f](t-\tau) d\tau &= (s * w[f])(t) \\ = \int_{-\infty}^{+\infty} s(\tau) o^*[t, f](\tau) d\tau &= < s, o[t, f] > \end{aligned}$$

2.2 Echantillonnage spectro-temporel

Pour aboutir à une décomposition discrète comme 1, considérons maintenant $w[f_m]$ comme le m^{ieme} filtre passe-bande d'un banc de filtres. Il est possible de reconstruire le signal initial si cet échantillonnage fréquentiel permet de retrouver le spectre de s, ce qui aboutira à la condition (4). Le signal $b[f_m](t)$, interprété comme ensemble des coefficients d'ondelettes centrées en (t, f_m), est un signal passe-bande, à cause du filtre $w[f_m]$. Le signal filtré peut être échantillonné à la période Δt_m, ce qui donne l'expression des coefficients d'ondelettes c_{nm}:

$$\begin{aligned} b[f_m](n\Delta t_m) &= (s * w[f_m])(n\Delta t_m) \\ =< s, o[n\Delta t_m, f_m] > &= c_{nm} \end{aligned}$$

Le signal continu peut être reconstruit exactement à condition d'être échantillonné avec un pas Δt_m dont la limite supérieure est $1/bw(f_m)$, fixée par la largeur de bande du filtre. Les filtres utilisés n'étant pas des filtres passe-bandes idéaux, cette largeur de bande est en pratique une largeur de bande effective, suivant un critère d'atténuation. Le choix adopté est précisé à la section suivante.

Un nouveau signal $\hat{b}[f_m](t)$ est formé à partir de $b[f_m](t)$, par produit avec un peigne de Dirac:

$$\hat{b}[f_m](t) = b[f_m](t) \sum_{n=-\infty}^{+\infty} \delta(t - n\Delta t_m)$$
$$= \sum_{n=-\infty}^{+\infty} b[f_m](n\Delta t_m)\delta(t - n\Delta t_m)$$

Le spectre (noté par un ˜ de $\hat{b}[f_m]$) se déduit de celui de $b[f_m]$ par périodisation, de période $1/\Delta t_m$:

$$\tilde{\hat{b}}[f_m](\nu) = \tilde{b}[f_m](\nu) * \{\frac{1}{\Delta t_m} \sum_{n=-\infty}^{+\infty} \delta(\nu - \frac{n}{\Delta t_m})\}$$
$$= \frac{\tilde{b}[f_m](\nu)}{\Delta t_m} + \{\frac{1}{\Delta t_m} \sum_{n\neq 0} \tilde{b}[f_m](\nu - \frac{n}{\Delta t_m})\}$$

2.3 Synthèse

Soit $r[f_m](t)$ le signal continu obtenu à partir du signal échantilloné $\hat{b}[f_m](t)$ filtré par le filtre de synthèse $u[f_m]$. Dans le domaine spectral:

$$\tilde{r}[f_m](\nu) = \tilde{u}[f_m](\nu)\{\frac{\tilde{\hat{b}}[f_m](\nu)}{\Delta t_m} + \frac{1}{\Delta t_m} \sum_{n\neq 0} \tilde{b}[f_m](\nu - \frac{n}{\Delta t_m})\} \quad (2)$$

une condition suffisante pour reconstituer un signal $r[f_m]$ qui soit un signal passe-bande en phase avec s dans sa bande est de choisir, dans le domaine temporel, des filtres d'analyse et de synthèse en miroir et en quadrature (au facteur de pondération $1/\Delta t_m$ près, qui rend compte de la cadence d'échantillonnage): $\tilde{u}[f_m](\nu) = \Delta t_m \tilde{w}[f_m]^*(\nu)$, en effet, on peut alors écrire:

$$\tilde{r}[f_m](\nu) = |\tilde{w}^*[f_m]|^2(\nu)\tilde{s}(\nu) + \overbrace{\tilde{w}^*[f_m](\nu) \sum_{n\neq 0} \tilde{b}[f_m](\nu - \frac{n}{\Delta t_m})}^{(A)} \quad (3)$$

Le terme (A) est nul, puisque le support fréquentiel de $\tilde{b}[f_m]$ est contenu dans celui de $\tilde{w}$, les supports fréquentiels de $\tilde{w}$ et de $\tilde{w}^*$ sont identiques, et donc les deux termes du produit (A) possèdent des supports disjoints.

L'équation 2, dans le domaine temporel se réduit à:

$$r[f_m](t) = \int_{-\infty}^{+\infty} \hat{b}[f_m](\tau)u[f_m](t - \tau)d\tau$$
$$= \int_{-\infty}^{+\infty} (\sum_{n=-\infty}^{+\infty} b[f_m](n\Delta t_m)\delta(\tau - n\Delta t_m))u[f_m](t - \tau)d\tau$$
$$= \sum_{n=-\infty}^{+\infty} b[f_m](n\Delta t_m) \int_{-\infty}^{+\infty} \delta(\tau - n\Delta t_m)u[f_m](t - \tau)d\tau$$
$$= \sum_{n=-\infty}^{+\infty} b[f_m](n\Delta t_m)u[f_m](t - n\Delta t_m)$$

Dans une bande, l'interprétation de la synthèse comme une somme pondérée d'ondelettes s'écrit:

$$r[f_m](t) = \sum_{n=-\infty}^{+\infty} c_{nm}p[n\Delta t_m, f_m](t)$$
$$= \sum_{n=-\infty}^{+\infty} b[f_m](n\Delta t_m)u[f_m](t - n\Delta t_m)$$

où les $p[t, f]$ sont les ondelettes de synthèse, qui doivent satisfaire à la condition: $p[t, f](\tau) = u[f](\tau - t) = \Delta t_m o[t, f](\tau)$

Un signal de synthèse complet $r(t)$ (que l'on veut proche de $s(t)$), est formé par sommation des signaux $r[f_m]$ passe-bande, à toutes les fréquences f_m, ce qui s'écrit dans le domaine fréquentiel:

$$\tilde{r}(\nu) = \sum_{m=-\infty}^{+\infty} K_m\tilde{r}[f_m](\nu)$$
$$= \sum_{m=-\infty}^{+\infty} K_m|\tilde{w}[f_m](\nu)|^2\tilde{s}(\nu)$$

Les K_m sont des constantes de normalisation des filtres afin de satisfaire à la condition pour une reconstruction exacte de s:

$$\sum_{m=-\infty}^{+\infty} K_m|\tilde{w}[f_m](\nu)|^2 = 1 \quad (4)$$

La formule générale d'analyse/synthèse par ondelettes s'écrit donc:

$$r(t) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \overbrace{\int_{-\infty}^{+\infty} o^*[n\Delta t_m, f_m](\tau)s(\tau)d\tau}^{(A)} \; C_m o[n\Delta t_m, f_m](t)$$

L'étape (A) est l'étape d'analyse, les constantes C_m sont égales à $K_m/\Delta t_m$, et compensent le recouvrement fréquentiel (par K_m) et le recouvrement temporel (par Δt_m) des ondelettes.

3 Applications

Un système d'analyse-synthèse utilisant une échelle Bark a été implémenté. Pour obtenir un banc de filtres non-déphaseurs, des F.I.R. ont été réalisés par fenêtrage [3].

L'ondelette prototype est une fenêtre de Hamming. Les fréquences centrales des ondelettes sont régulièrement réparties en échelle Bark. Leurs largeurs de bandes, dépendantes de la fréquence centrale, sont fixées à 1 Bark à -3 dB du sommet. Uue bande passante de 5 kHz correspond à environ 17 Bark (16.9), et une bande passante de 8 kHz à environ 20 Bark (19.7).

Les ondelettes du banc de filtre dans les différentes bandes de fréquence se déduisent par contraction/dilatation de l'ondelette prototype, afin d'obtenir la largeur de bande voulue, et par modulation (multiplication par une exponentielle complexe) à la fréquence centrale du filtre.

Dans notre système, nous avons choisi la fenêtre de Hamming. Si w dénote la fenêtre de Hamming de support temporel $[-\frac{1}{2}, +\frac{1}{2}]$ secondes, l'ondelette centrée à l'instant t, à la fréquence f, et dont le lobe principal possède une largeur de bande b se déduit de w par:

$$o[t, f](\tau) = w(\frac{b(\tau - t)}{4})e^{2i\pi f(\tau - t)}$$

dont le support est de longueur $4/b$ secondes. La limite inférieure de la fréquence d'échantillonnage dans chaque bande d'analyse est b, si l'on considère la distortion due au repliement de spectre des lobes secondaires comme négligeable, ce qui est acceptable en pratique.

3.1 Spectrographe auditif

Le spectographe auditif visualise les coefficients d'ondelettes aux différents points spectro-temporels. Les informations spectrales et temporelles, liées à l'évolution des phases,et des du modules sont représentées. L'échelle d'amplitude est logarithmique.

Le spectrographe auditif illustre le comportement fréquentiel de l'échelle auditive: proche de la bande étroite en basse fréquence, et de l'analyse en 1/3 d'octave en haute fréquence. Les paramètres acoustiques descriptifs, ou indices acoustiques, observés sur les spectrogrammes auditifs sont assez différents de ceux observés avec le spectrographe en bande large, qui est utilisé d'ordinaire en phonétique acoustique, comme il est discuté dans [1] (voir figure 2).

3.2 Modification du signal

La visualisation spectrographique illustre la résolution spectro-temporelle du vocodeur correspondant. Ainsi, avec un vocodeur de phase en bande étroite, il est impossible de manipuler des transitoires, qui sont de toute façon invisibles sur un spectrogramme en bande étroite. De même, un vocodeur de phase en bande large est inadapté pour la modification individuelle des harmoniques, et ces harmoniques n'apparaissent pas individuellement sur un spectrogramme en bande large. Le vocodeur en échelle Bark est donc adapté pour la modification des paramètres acoustiques apparents sur les spectrogrammes en échelle Bark. Du grave vers l'aigu la portion du plan temps-fréquence représentée par une ondelette conserve une aire spectro-temporelle constante. Si cette aire est définie par le produit du support temporel par le support fréquentiel du lobe principal de l'ondelette, elle vaut 4 dans la cas de la fenêtre de Hamming.

4 Conclusion

Dans cet article le cadre du filtrage linéaire a été employé pour définir une représentation par ondelettes, interprétation tout à fait générale qui permet l'utilisation d'échelles fréquentielles d'analyse arbitraires. Ainsi, un système d'analyse/synthèse par ondelettes sur une échelle auditive a été construit. La visualisation spectrographique issue de cette analyse a été sommairement discutée. La modification localisée ou globale des caractéristiques spectro-temporelles du signal est aisée, dans les limites de la résolution spectro-temporelle imposée par l'échelle fréquentielle.

Références

1. d'ALESSANDRO, C., BEAUTEMPS, D., "Justification perceptive du spectrographe auditif" Proc. of XII^{th} Int. Cong. Phon. Sci., ICPhS'91, Aix-en-Provence, Aout 1991.
2. CROCHIERE, R. E. & RABINER, L. R., "Interpolation and Decimation of Digital Signals - A Tutorial Review", Proc. of the IEEE, Vol. 69, No. 3, March. 1981, pp 300-331.
3. HARRIS, F. J., "On the use of windows for harmonic analysis with the discrete Fourier transform" Proc. of the IEEE, Vol. 66 No 1., Jan. 1978, pp 51-83.
4. MALLAT, S. G., "A theory for multiresolution signal decomposition: the wavelet representation." IEEE trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-11, No. 7, Jul. 1989, pp 674-693.
5. TRAUNMULLER, H., "Analytical expressions for the tonotopic sensory scale", J. Acoust. Soc. Am., Vol. 88 No. 1, Jul. 1990, pp 97-100.

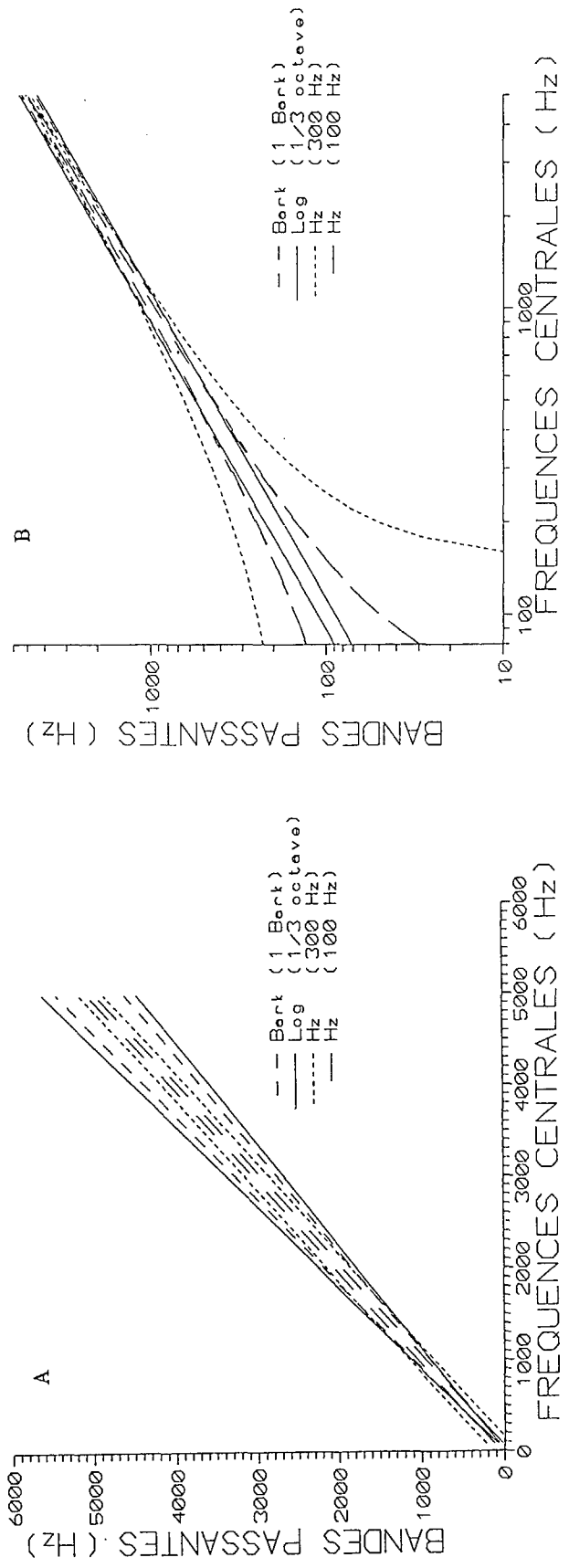

Figure 1: Comparaison des résolutions constantes en échelles Hz, Bark et Octaves. [A] Hautes fréquences; [B] Basses fréquences.

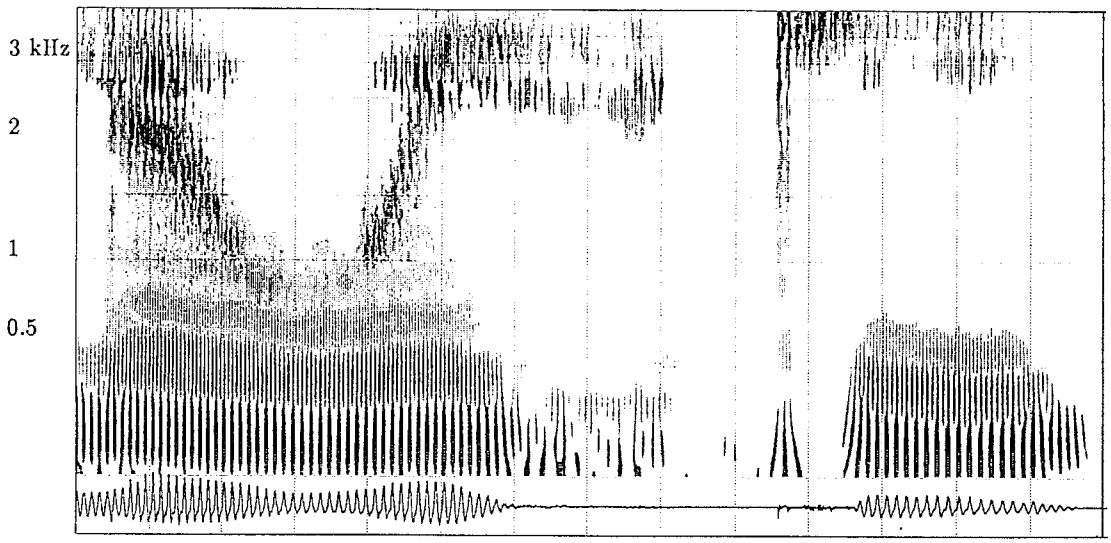

Figure 2. Spectrogramme auditif. Voix féminine: "le whiski".

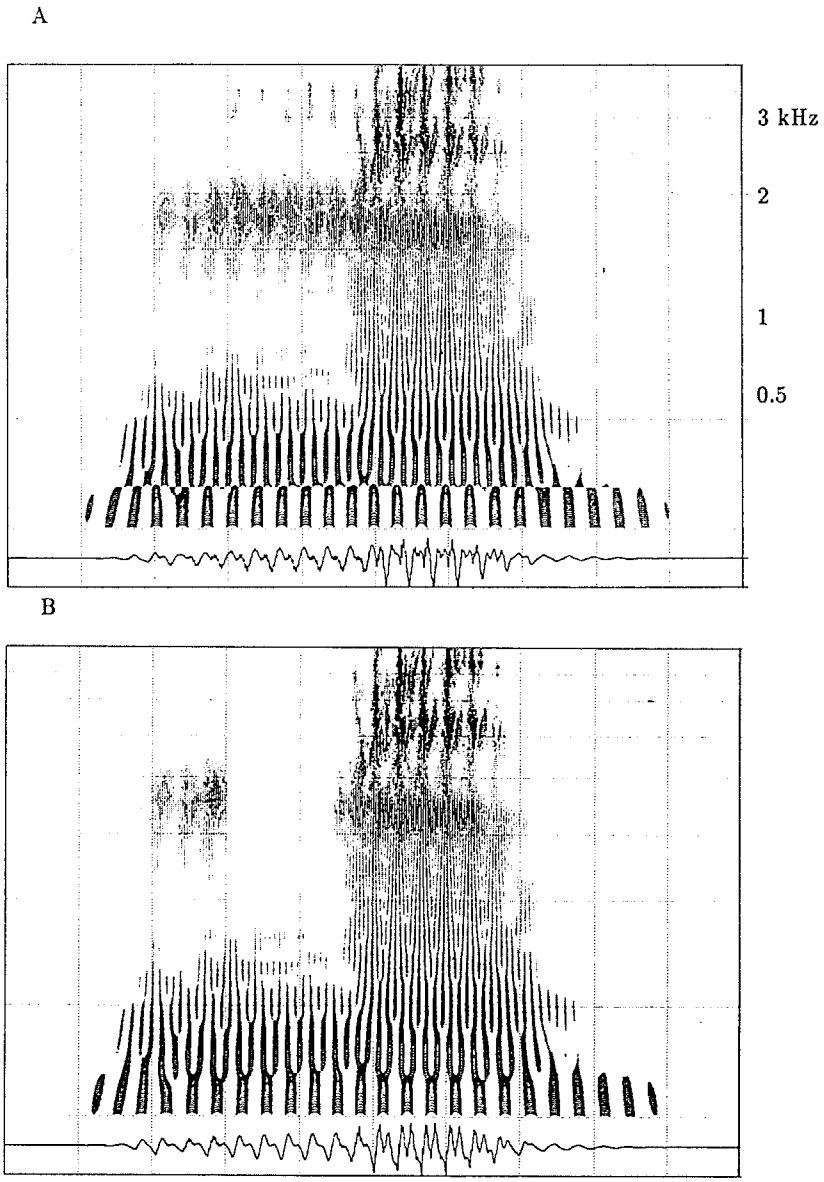

Figure 3. Modifications du signal et spectrogrammes auditif. Voix masculine: "le peu". [A] modification de la phase de l'harmonique 1; [B] filtrage de quatre périodes de voisement autour du second formant.