



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) **EP 1 194 923 B1**

(12) **FASCICULE DE BREVET EUROPEEN**

(45) Date de publication et mention
de la délivrance du brevet:
18.01.2006 Bulletin 2006/03

(21) Numéro de dépôt: **00953223.5**

(22) Date de dépôt: **04.07.2000**

(51) Int Cl.:
G10L 19/02 (2006.01)

(86) Numéro de dépôt international:
PCT/FR2000/001904

(87) Numéro de publication internationale:
WO 2001/003116 (11.01.2001 Gazette 2001/02)

(54) **PROCEDES ET DISPOSITIFS D'ANALYSE ET DE SYNTHESE AUDIO**

VERFAHREN UND SYSTEM FÜR AUDIO ANALYSE UND SYNTHESE
METHODS AND DEVICE FOR AUDIO ANALYSIS AND SYNTHESIS

(84) Etats contractants désignés:
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
MC NL**

(30) Priorité: **05.07.1999 FR 9908638**

(43) Date de publication de la demande:
10.04.2002 Bulletin 2002/15

(73) Titulaire: **NORTEL NETWORKS France
78117 Châteaufort (FR)**

(72) Inventeurs:
• **CAPMAN, François
F-78000 Versailles (FR)**
• **MURGIA, Carlo
Irvine, CA 92618 (US)**

(74) Mandataire: **Loisel, Bertrand
Cabinet Plasseraud
65/67 rue de la Victoire
75440 Paris Cedex 09 (FR)**

(56) Documents cités:
**EP-A- 0 893 791 US-A- 5 878 388
US-A- 5 911 130**

- **DATABASE INSPEC [en ligne] INSTITUTE OF ELECTRICAL ENGINEERS, STEVENAGE, GB; LAGLER A ET AL: "Real-time fixed-point DSP-implementation of spectral subtraction algorithm for speech enhancement in noisy environment" Database accession no. 6604117 XP002139930 & 9TH INTERNATIONAL CZECH - SLOVAK SCIENTIFIC CONFERENCE. RADIOELEKTRONIKA 99. CONFERENCE PROCEEDINGS, PROCEEDINGS OF 9TH INTERNATIONAL CZECH - SLOVAK SCIENTIFIC CONFERENCE. RADIOELEKTRONIKA 99, BRNO, CZECH REPUBLIC, 27-28 APRIL 1999, pages 186-189, 1999, Brno, Czech Republic, Brno Univ. Technol, Czech Republic ISBN: 80-214-1327-1**
- **AHMADI S ET AL: "New techniques for sinusoidal coding of speech at 2400 bps" CONFERENCE RECORD OF THIRTIETH ASILOMAR CONFERENCE ON SIGNALS, SYSTEMS AND COMPUTERS (CAT. NO.96CB36004), CONFERENCE RECORD OF THE THIRTIETH ASILOMAR CONFERENCE ON SIGNALS, SYSTEMS AND COMPUTERS, PACIFIC GROVE, CA, USA, 3-6 NOV. 1996, pages 770-774 vol.1, XP002138769 1997, Los Alamitos, CA, USA, IEEE Comput. Soc. Press, USA ISBN: 0-8186-7646-9**

EP 1 194 923 B1

Il est rappelé que: Dans un délai de neuf mois à compter de la date de publication de la mention de la délivrance du brevet européen, toute personne peut faire opposition au brevet européen délivré, auprès de l'Office européen des brevets. L'opposition doit être formée par écrit et motivée. Elle n'est réputée formée qu'après paiement de la taxe d'opposition. (Art. 99(1) Convention sur le brevet européen).

Description

[0001] La présente invention concerne l'analyse et la synthèse des signaux audio, à partir de représentations de ces signaux dans le domaine spectral.

[0002] Elle s'applique en particulier, mais non exclusivement, au codage de la parole, en bande étroite ou en bande élargie, dans diverses gammes de débit de codage. Parmi les autres domaines d'application, on peut citer le débruitage par soustraction spectrale (voir EP-A-0 534 837 ou WO99/14739)

[0003] Dans les procédés d'analyse en question, le spectre du signal est obtenu en transformant des trames successives vers le domaine fréquentiel. La transformation employée est le plus souvent la transformée de Fourier rapide (TFR) ; mais d'autres transformées connues sont utilisables. Dans le cas fréquent d'un échantillonnage du signal à 8 kHz, le nombre N d'échantillons par trame est typiquement de l'ordre de 100 à 500, ce qui représente des trames de quelques dizaines de millisecondes. Pour bénéficier de la résolution maximale en fréquence, la TFR est effectuée sur 2N points, N échantillons à zéro étant ajoutés aux N échantillons de la trame.

[0004] Le spectre obtenu par transformée de Fourier de la trame de signal est la convolution du spectre réel du signal par la transformée de Fourier de la fenêtre d'analyse du signal. Cette fenêtre d'analyse, qui pondère les échantillons de chaque trame, est nécessaire à la prise en compte de la durée finie de la trame. Si la trame de signal est directement soumise à la TFR, c'est-à-dire si on utilise une fenêtre d'analyse rectangulaire, le spectre obtenu est perturbé par les pics secondaires de la TFR de la fenêtre d'analyse. Pour limiter cet inconvénient, qui est particulièrement sensible lorsque des paramètres représentant le signal ou le bruit doivent être extraits des spectres, on a recours à des fenêtres ayant de meilleures propriétés spectrales, c'est-à-dire des fonctions de pondération dont le support est limité à N échantillons et dont la transformée de Fourier a son énergie concentrée dans un pic étroit avec une forte atténuation des pics secondaires. Les plus courantes de ces fenêtres sont les fenêtres de Hamming, de Hanning et de Kaiser.

[0005] Dans la méthode d'analyse et de synthèse dite OLA (« Overlap-And-Add »), les trames successives présentent des recouvrements mutuels de 50 % (N/2 échantillons). Comme les fenêtres d'analyse couramment utilisées vérifient la propriété $f_A(i+N/2) + f_A(i) = 1$, la synthèse peut être effectuée simplement en effectuant la somme à recouvrement des trames de N échantillons successivement calculées par transformée de Fourier inverse des spectres.

[0006] US-A-5,878,388 se rapporte à un procédé d'analyse et de synthèse de la voix. L'article d'AHMADI S. et Al : "New techniques for sinusoidal coding of speech at 2400 bps", CONFERENCE RECORD OF THIRTIETH ASILOMAR CONFERENCE ON SIGNALS, SYSTEMS AND COMPUTERS, Pacific Grove, CA, USA, IEEE Comput. Soc. Press, USA, se rapporte à un procédé d'analyse et de synthèse de la voix. L'article de R.E. Crochiere, « A weighted overlap-add method for short-time Fourier analysis synthesis », IEEE Trans. on ASSP, Vol ASSP 28, février 1980, pages 99-102, se rapporte à un procédé d'analyse et de synthèse par transformation de Fourier à court-terme. US-A-5,911,130 se rapporte de la compression et décompression de signal audio.

[0007] Dans le but d'affiner la représentation spectrale, certaines méthodes dites WOLA (« Weighted OLA ») utilisent à l'analyse des trames dont les recouvrements mutuels sont de plus de 50 %. A la synthèse, il est nécessaire de repondérer les échantillons des trames avant de les sommer. Ces méthodes augmentent la complexité de l'analyse et de la synthèse. Dans des applications de codage, elles augmentent également le débit de transmission requis.

[0008] Un but de l'invention est de proposer un schéma d'analyse et synthèse de signaux audio qui permette de limiter la cadence des trames d'analyse, tout en utilisant des fenêtres d'analyse ayant de bonnes propriétés spectrales.

[0009] L'invention propose un procédé d'analyse d'un signal audio traité par trames successives de N échantillons, dans lequel on pondère les échantillons de chaque trame par une fenêtre d'analyse de type Hamming, Hanning, Kaiser ou analogue, on calcule un spectre du signal audio en transformant chaque trame d'échantillons pondérés dans le domaine fréquentiel, et on traite le spectre du signal audio pour délivrer des paramètres de synthèse d'un signal dérivé du signal audio analysé. Suivant l'invention, les trames successives comportent une alternance de trames pour lesquelles on délivre des jeux complets de paramètres de synthèse, qui présentent des recouvrements mutuels de moins de N/2 échantillons, soit moins de 50 %, et de trames pour lesquelles on délivre des jeux incomplets de paramètres de synthèse.

[0010] Les trames pour lesquelles ne sont pas délivrés des jeux complets de paramètres de synthèse peuvent ne faire l'objet d'aucune analyse spectrale. En variante, une analyse peut néanmoins être effectuée pour ces trames, afin de délivrer des jeux incomplets de paramètres de synthèse incluant des données représentant une erreur d'interpolation d'au moins un des paramètres de synthèse et/ou des données représentant un filtre d'interpolation d'au moins un des paramètres de synthèse.

[0011] Dans un premier domaine d'application du procédé, le traitement du spectre du signal audio comporte une extraction de paramètres de codage en vue de la transmission et/ou du stockage du signal audio codé. Dans un second domaine d'application du procédé, le traitement du spectre du signal audio comporte un débruitage par soustraction spectrale. D'autres domaines d'application peuvent encore être envisagés parmi les traitements audio.

[0012] Un second aspect de l'invention se rapporte à un procédé de synthèse d'un signal audio, dans lequel on évalue un ensemble de trames recouvrantes successives de N échantillons du signal audio pondérés par une fenêtre d'analyse, en transformant dans le domaine temporel des estimations spectrales correspondant respectivement aux dites trames,

et on combine les trames évaluées pour former le signal synthétisé. Pour un sous-ensemble des trames évaluées, les estimations spectrales sont obtenues en traitant des paramètres de synthèse respectivement associés aux trames dudit sous-ensemble tandis que, pour les trames ne faisant pas partie du sous-ensemble, les estimations spectrales sont obtenues avec une interpolation d'une partie au moins des paramètres de synthèse. Les trames successives dudit sous-ensemble présentent des décalages temporels mutuels de M échantillons, le nombre M étant plus grand que N/2, tandis que les trames successives dudit ensemble présentent des décalages temporels mutuels de M/p échantillons, p étant un entier plus grand que 1. On modifie chaque trame évaluée en lui appliquant un traitement correspondant à une division par ladite fenêtre d'analyse et à une multiplication par une fenêtre de synthèse, et on forme le signal synthétisé comme une somme à recouvrement des trames modifiées. Les échantillons d'une trame ayant des rangs i numérotés de 0 à N-1, la fenêtre de synthèse $f'_S(i)$ a un support limité aux rangs i allant de N/2 - M/p à N/2 + M/p et

vérifie $f'_S(i) + f'_S(i+M/p) = A$ pour $N/2 - M/p \leq i < N/2$, A étant une constante positive.

[0013] L'invention propose également des dispositifs de traitement audio comprenant des moyens pour la mise en oeuvre des procédés d'analyse et de synthèse ci-dessus.

[0014] D'autres particularités et avantages de la présente invention apparaîtront dans la description ci-après d'exemples de réalisation non limitatifs, en référence aux dessins annexés, dans lesquels :

- la figure 1 est un schéma synoptique d'un codeur audio selon l'invention ;
- les figures 2 et 3 sont des diagrammes illustrant la formation des trames de signal audio dans le codeur de la figure 1 ;
- les figures 4 et 5 sont des graphiques montrant un exemple de spectre du signal audio et illustrant l'extraction des enveloppes supérieure et inférieure de ce spectre ;
- la figure 6 est un schéma synoptique d'un exemple de moyens de quantification utilisables dans le codeur de la figure 1 ;
- la figure 7 est un schéma synoptique de moyens utilisables pour extraire des paramètres se rapportant à la phase de la composante non-harmonique dans une variante du codeur de la figure 1 ;
- la figure 8 est un schéma synoptique d'un décodeur audio correspondant au codeur de la figure 1 ;
- la figure 9 est un organigramme d'un exemple de procédure de lissage de coefficients spectraux et d'extraction de phases minimales mise en oeuvre dans le décodeur de la figure 8 ;
- la figure 10 est un schéma synoptique de modules d'analyse et de mixage spectral de composantes harmonique et non-harmonique du signal audio ;
- les figures 11 à 13 sont des graphiques montrant des exemples de fonctions non-linéaires utilisables dans le module d'analyse de la figure 10;
- les figures 14 et 15 sont des diagrammes illustrant une façon de procéder à la synthèse temporelle des trames de signal dans le décodeur de la figure 8 ;
- les figures 16 et 17 sont des graphiques montrant des fonctions de fenêtrage utilisables dans la synthèse des trames suivant les figures 14 et 15;
- les figures 18 et 19 sont des schémas synoptiques de moyens d'interpolation utilisables dans une variante de réalisation du codeur et du décodeur ;
- la figure 20 est un schéma synoptique de moyens d'interpolation utilisables dans une autre variante de réalisation du codeur ; et
- les figures 21 et 22 sont des diagrammes illustrant une autre façon de procéder à la synthèse temporelle des trames de signal dans le décodeur de la figure 8, à l'aide d'une interpolation de paramètres.

[0015] Le codeur et le décodeur décrits ci-après sont des circuits numériques qui peuvent, comme il est usuel dans le domaine du traitement des signaux audio, être réalisés par programmation d'un processeur de signal numérique (DSP) ou d'un circuit intégré d'application spécifique (ASIC).

[0016] Le codeur audio représenté sur la figure 1 traite un signal audio d'entrée x qui, dans l'exemple non-limitatif considéré ci-après, est un signal de parole. Le signal x est disponible sous forme numérique, par exemple à une fréquence d'échantillonnage F_e de 8 kHz. Il est par exemple délivré par un convertisseur analogique-numérique traitant le signal de sortie amplifié d'un microphone. Le signal d'entrée x peut également être formé à partir d'une autre version, analogique ou numérique, codée ou non, du signal de parole.

[0017] Le codeur comprend un module 1 qui forme des trames successives de signal audio pour les différents traitements effectués, et un multiplexeur de sortie 6 qui délivre un flux de sortie Φ contenant pour chaque trame des jeux de paramètres de quantification à partir desquels un décodeur sera capable de synthétiser une version décodée du signal audio.

[0018] La structure des trames est illustrée par les figures 2 et 3. Chaque trame 2 est composée d'un nombre N d'échantillons consécutifs du signal audio x. Les trames successives présentent des décalages temporels mutuels

correspondant à M échantillons, de sorte que leur recouvrement est de $L = N - M$ échantillons du signal. Dans l'exemple considéré, où $N = 256$, $M = 160$ et $L = 96$, la durée des trames 2 est de $N/F_e = 32$ ms, et une trame est formée toutes les $M/F_e = 20$ ms.

[0019] De façon classique, le module 1 multiplie les échantillons de chaque trame 2 par une fonction de fenêtrage f_A , de préférence choisie pour ses - bonnes propriétés spectrales. Les échantillons $x(i)$ de la trame étant numérotés de $i = 0$ à $i = N - 1$, la fenêtre d'analyse $f_A(i)$ peut ainsi être une fenêtre de Hamming, d'expression :

$$f_A(i) = 0,54 + 0,46 \cdot \cos\left(2\pi \frac{i - (N-1)/2}{N}\right) \quad (1)$$

ou une fenêtre de Hanning d'expression :

$$f_A(i) = \frac{1}{2} \left(1 + \cos\left(2\pi \frac{i - (N-1)/2}{N}\right) \right) \quad (2)$$

ou encore une fenêtre de Kaiser, d'expression :

$$f_A(i) = \frac{I_0\left(\alpha \sqrt{1 - \left(\frac{i - (N-1)/2}{N}\right)^2}\right)}{I_0(\alpha)} \quad (3)$$

où α est un coefficient par exemple égal à 6, et $I_0(\cdot)$ désigne la fonction de Bessel d'indice 0.

[0020] Le codeur de la figure 1 procède à une analyse du signal audio dans le domaine spectral. Il comporte un module 3 qui calcule la transformée de Fourier rapide (TFR) de chaque trame de signal. La trame de signal est mise en forme avant d'être soumise au module de TFR 3 : le module 1 lui adjoint $N = 256$ échantillons à zéro afin d'obtenir la résolution maximale de la transformée de Fourier, et il effectue d'autre part une permutation circulaire des $2N = 512$ échantillons afin de compenser les effets de phase résultant de la fenêtre d'analyse. Cette modification de la trame est illustrée par la figure 3. La trame dont on calcule la transformée de Fourier rapide sur $2N = 512$ points commence par les $N/2 = 128$ derniers échantillons pondérés de la trame, suivis par les $N = 256$ échantillons à zéro, et se termine par les $N/2 = 128$ premiers échantillons pondérés de la trame.

[0021] Le module de TFR 3 obtient le spectre du signal pour chaque trame, dont le module et la phase sont respectivement notés $|X|$ et φ_X , ou $|X(i)|$ et $\varphi_X(i)$ pour les index de fréquence $i = 0$ à $i = 2N - 1$ (grâce à la symétrie de la transformée de Fourier et des trames, on peut se limiter aux valeurs pour $0 \leq i < N$).

[0022] Un détecteur de fréquence fondamentale 4 estime pour chaque trame de signal une valeur de la fréquence fondamentale F_0 . Le détecteur 4 peut appliquer toute méthode connue d'analyse du signal de parole de la trame pour estimer la fréquence fondamentale F_0 , par exemple une méthode basée sur la fonction d'autocorrélation ou la fonction AMDF, éventuellement précédée d'un module de blanchiment par prédiction linéaire. L'estimation peut également être effectuée dans le domaine spectral ou dans le domaine cepstral. Une autre possibilité est d'évaluer les intervalles de temps entre les ruptures consécutives du signal de parole attribuables à des fermetures de la glotte du locuteur intervenant pendant la durée de la trame. Des méthodes bien connues utilisables pour détecter de telles micro-ruptures sont décrites dans les articles suivants : M. Basseville et al., « Sequential detection of abrupt changes in spectral characteristics of digital signals » (IEEE Trans. on Information Theory, 1983, Vol. IT-29, n° 5, pages 708-723) ; R. Andre-Obrecht, « A new statistical approach for the automatic segmentation of continuous speech signals » (IEEE Trans. on Acous., Speech and Sig. Proc., Vol. 36, N°1, janvier 1988) ; et C. MURGIA et al., « An algorithm for the estimation of glottal closure instants using the sequential detection of abrupt changes in speech signals » (Signal Processing VII, 1994, pages 1685-1688).

[0023] La fréquence fondamentale estimée F_0 fait l'objet d'une quantification, par exemple scalaire, par un module 5, qui fournit au multiplexeur de sortie 6 un index iF de quantification de la fréquence fondamentale pour chaque trame du signal.

[0024] Le codeur utilise des modélisations paramétriques cepstrales pour représenter une enveloppe supérieure et

une enveloppe inférieure du spectre du signal audio. La première étape de la transformation cepstrale consiste à appliquer au module du spectre du signal une fonction de compression spectrale, qui peut être une fonction logarithmique ou en racine. Le module 8 du codeur opère ainsi, pour chaque valeur $X(i)$ du spectre du signal ($0 \leq i < N$), la transformation suivante:

$$LX(i) = \text{Log}(|X(i)|) \quad (4)$$

dans le cas d'une compression logarithmique ou

$$LX(i) = |X(i)|^\gamma \quad (5)$$

dans le cas d'une compression en racine, γ étant un exposant compris entre 0 et 1.

[0025] Le spectre comprimé LX du signal audio est traité par un module 9 qui extrait des amplitudes spectrales associées aux harmoniques du signal correspondant aux multiples de la fréquence fondamentale estimée F_0 . Ces amplitudes sont ensuite interpolées par un module 10 afin d'obtenir une enveloppe supérieure comprimée notée LX_sup .

[0026] Il est à noter que la compression spectrale pourrait de façon équivalente être effectuée après la détermination des amplitudes associées aux harmoniques. Elle pourrait également être effectuée après l'interpolation, ce qui ne ferait que modifier la forme des fonctions d'interpolation.

[0027] Le module 9 d'extraction des maxima tient compte de l'éventuelle variation de la fréquence fondamentale sur la trame d'analyse, des erreurs que peut commettre le détecteur 4, ainsi que des imprécisions liées au caractère discret de l'échantillonnage en fréquence. Pour cela, la recherche des amplitudes des pics spectraux ne consiste pas simplement à prendre les valeurs $LX(i)$ correspondant aux index i tels que $i \cdot F_0 / 2N$ soit la fréquence la plus proche d'une harmonique de fréquence $k \cdot F_0$ ($k \geq 1$). L'amplitude spectrale retenue pour une harmonique d'ordre k est un maximum local du module du spectre au voisinage de la fréquence $k \cdot F_0$ (cette amplitude est obtenue directement sous forme comprimée lorsque la compression spectrale 8 est effectuée avant l'extraction des maxima 9).

[0028] Les figures 4 et 5 montrent un exemple de forme du spectre comprimé LX , où on voit que les amplitudes maximales des pics harmoniques ne coïncident pas nécessairement avec les amplitudes correspondant aux multiples entiers de la fréquence fondamentale estimée F_0 . Les flancs des pics étant assez raides, une petite erreur de positionnement de la fréquence fondamentale F_0 , amplifiée par l'indice d'harmonique k , peut distordre fortement l'enveloppe supérieure estimée du spectre et provoquer une mauvaise modélisation de la structure formantique du signal. Par exemple, prendre directement l'amplitude spectrale pour la fréquence $3 \cdot F_0$ dans le cas des figures 4 et 5 produirait une erreur importante dans l'extraction de l'enveloppe supérieure au voisinage de l'harmonique d'ordre $k = 3$, alors qu'il s'agit d'une zone énergétiquement importante dans l'exemple dessiné. En effectuant l'interpolation à partir du véritable maximum, on évite ce genre d'erreur d'estimation de l'enveloppe supérieure.

[0029] Dans l'exemple représenté sur la figure 4, l'interpolation est effectuée entre des points dont l'abscisse est la fréquence correspondant au maximum de l'amplitude d'un pic spectral, et dont l'ordonnée est ce maximum, avant ou après compression.

[0030] L'interpolation effectuée pour calculer l'enveloppe supérieure LX_sup est une simple interpolation linéaire. Bien entendu une autre forme d'interpolation pourrait être utilisée (par exemple polynomiale ou spline).

[0031] Dans la variante préférée représentée sur la figure 5, l'interpolation est effectuée entre des points dont l'abscisse est une fréquence $k \cdot F_0$ multiple de la fréquence fondamentale (en fait la fréquence la plus proche dans le spectre discret) et dont l'ordonnée est l'amplitude maximale, avant ou après compression, du spectre au voisinage de cette fréquence multiple.

[0032] En comparant les figures 4 et 5, on peut voir que le mode d'extraction selon la figure 5, qui repositionne les pics sur les fréquences harmoniques, conduit à une meilleure précision sur l'amplitude des pics que le décodeur attribuera aux fréquences multiples de la fréquence fondamentale. Il peut se produire un léger déplacement en fréquence de la position de ces pics, ce qui n'est pas perceptuellement très important et n'est d'ailleurs pas évité non plus dans le cas de la figure 4. Dans le cas de la figure 4, les points d'ancrage pour l'interpolation sont confondus avec les sommets des pics harmoniques. Dans le cas de la figure 5, on impose que ces points d'ancrage se trouvent précisément aux fréquences multiples de la fréquence fondamentale, leurs amplitudes correspondant à celles des pics.

[0033] L'intervalle de recherche du maximum d'amplitude associé à une harmonique de rang k est centré sur l'index

i de la fréquence de la TFR la plus proche de $k.F_0$, c'est-à-dire $i = \left\lfloor 2Nk \frac{F_0}{F_e} + \frac{1}{2} \right\rfloor$, où $[a]$ désigne l'entier égal ou

immédiatement inférieur au nombre a. La largeur de cet intervalle de recherche dépend de la fréquence d'échantillonnage F_e , de la taille 2N de la TFR et de la gamme de variation possible de la fréquence fondamentale. Cette largeur est typiquement de l'ordre d'une dizaine de fréquences avec les exemples de valeurs précédemment considérés. On peut la rendre réglable en fonction de la valeur F_0 de la fréquence fondamentale et du numéro k de l'harmonique.

[0034] Afin d'améliorer la résolution dans les basses fréquences et donc de représenter plus fidèlement les amplitudes des harmoniques dans cette zone, une distorsion non-linéaire de l'échelle des fréquences est opérée sur l'enveloppe supérieure comprimée par un module 12 avant que le module 13 effectue la transformée de Fourier rapide inverse (TFRI) fournissant les coefficients cepstraux cx_sup .

[0035] La distorsion non-linéaire permet de minimiser plus efficacement l'erreur de modélisation. Elle est par exemple effectuée selon une échelle de fréquences de type Mel ou Bark. Cette distorsion peut éventuellement dépendre de la fréquence fondamentale estimée F_0 . La figure 1 illustre le cas de l'échelle Mel. La relation entre les fréquences F du spectre linéaire, exprimées en hertz, et les fréquences F' de l'échelle Mel est la suivante :

$$F' = \frac{1000}{\log_{10}(2)} \times \log_{10} \left(1 + \frac{F}{1000} \right) \quad (6)$$

[0036] Afin de limiter le débit de transmission, une troncature des coefficients cepstraux cx_sup est effectuée. Le module de TFRI 13 a besoin de calculer seulement un vecteur cepstral de NCS coefficients cepstraux d'ordres 0 à NCS-1. A titre d'exemple, NCS peut être égal à 16.

[0037] Un post-filtrage dans le domaine cepstral, appelé post-liffrage, est appliqué par un module 15 à l'enveloppe supérieure comprimée LX_sup . Ce post-liffrage correspond à une manipulation des coefficients cepstraux cx_sup délivrés par le module de TRFI 13, qui correspond approximativement à un post-filtrage de la partie harmonique du signal par une fonction de transfert ayant la forme classique :

$$H(z) = \left(1 - \mu z^{-1} \right) \frac{A(z/\gamma_1)}{A(z/\gamma_2)} \quad (7)$$

où A(z) est la fonction de transfert d'un filtre de prédiction linéaire du signal audio, γ_1 et γ_2 sont des coefficients compris entre 0 et 1, et μ est un coefficient de préaccentuation éventuellement nul. La relation entre le coefficient post-liffré d'ordre i, noté $c_p(i)$, et le coefficient cepstral correspondant $c(i) = -cx_sup(i)$ délivré par le module 13 est alors :

$$c_p(0) = c(0) \quad \text{---} \\ c_p(i) = \left(1 + \gamma_2^i - \gamma_1^i \right) c(i) - \frac{\mu^i}{i} \quad \text{pour } i > 0 \quad (8)$$

[0038] Le coefficient de préaccentuation optionnel μ peut être contrôlé en posant comme contrainte de préserver la valeur du coefficient cepstral $cx_sup(1)$ relatif à la pente. En effet, la valeur $c(1) = cx_sup(1)$ d'un bruit blanc filtré par le filtre de préaccentuation correspond au coefficient de préaccentuation. On peut ainsi choisir ce dernier de la façon suivante :

$$\mu = (\gamma_2 - \gamma_1) \cdot c(1).$$

[0039] Après le post-liffrage 15, un module de normalisation 16 modifie encore les coefficients cepstraux en imposant la contrainte de modélisation exacte d'un point du spectre initial, qui est de préférence le point le plus énergétique parmi les maxima spectraux extraits par le module 9. En pratique, cette normalisation modifie seulement la valeur du coefficient $c_p(0)$.

[0040] Le module de normalisation 16 fonctionne de la façon suivante : il recalcule une valeur du spectre synthétisé à la fréquence du maximum indiqué par le module 9, par transformée de Fourier des coefficients cepstraux tronqués et post-liftrés, en tenant compte de la distorsion non-linéaire de l'axe des fréquences ; il détermine un gain de normalisation g_N par la différence logarithmique entre la valeur du maximum fournie par le module 9 et cette valeur recalculée ; et il ajoute le gain g_N au coefficient cepstral post-liftré $c_p(0)$. Cette normalisation peut être vue comme faisant partie du post-liftrage.

[0041] Les coefficients cepstraux post-liftrés et normalisés font l'objet d'une quantification par un module 18 qui transmet des index de quantification correspondants $icxs$ au multiplexeur de sortie 6 du codeur.

[0042] Le module 18 peut fonctionner par quantification vectorielle à partir de vecteurs cepstraux formés de coefficients post-liftrés et normalisés, notés ici $cx[n]$ pour la trame de signal de rang n . A titre d'exemple, le vecteur cepstral $cx[n]$ de NCS = 16 coefficients cepstraux $cx[n,0]$, $cx[n,1]$, ..., $cx[n,NCS-1]$ est distribué en quatre sous-vecteurs cepstraux contenant chacun quatre coefficients d'ordres consécutifs. Le vecteur cepstral $cx[n]$ peut être traité par les moyens représentés sur la figure 6, faisant partie du module de quantification 18. Ces moyens mettent en oeuvre, pour chaque composante $cx[n,i]$, un prédicteur de la forme :

$$cx_p[n,i] = (1 - \alpha(i)) \cdot rcx[n,i] + \alpha(i) \cdot rcx[n-1,i] \quad (9)$$

où $rcx[n]$ désigne un vecteur résiduel de prédiction pour la trame de rang n dont les composantes sont respectivement notées $rcx[n,0]$, $rcx[n,1]$, ..., $rcx[n,NCS-1]$, et $\alpha(i)$ désigne un coefficient de prédiction choisi pour être représentatif d'une corrélation inter-trame supposée. Après quantification des résidus, ce vecteur résiduel est défini par :

$$rcx[n,i] = \frac{cx[n,i] - \alpha(i) \cdot rcx_q[n-1,i]}{2 - \alpha(i)} \quad (10)$$

où $rcx_q[n-1]$ désigne le vecteur résiduel quantifié pour la trame de rang $n-1$, dont les composantes sont respectivement notées $rcx_q[n,0]$, $rcx_q[n,1]$, ..., $rcx_q[n,NCS-1]$.

[0043] Le numérateur de la relation (10) est obtenu par un soustracteur 20, dont les composantes du vecteur de sortie sont divisées par les quantités $2 - \alpha(i)$ en 21. Aux fins de la quantification, le vecteur résiduel $rcx[n]$ est subdivisé en quatre sous-vecteurs, correspondant à la subdivision en quatre sous-vecteurs cepstraux. Sur la base d'un dictionnaire obtenu par apprentissage préalable, l'unité 22 procède à la quantification vectorielle de chaque sous-vecteur du vecteur résiduel $rcx[n]$. Cette quantification peut consister, pour chaque sous-vecteur $srcx[n]$, à sélectionner dans le dictionnaire le sous-vecteur quantifié $srcx_q[n]$ qui minimise l'erreur quadratique $\|srcx[n] - srcx_q[n]\|^2$. L'ensemble $icxs$ des index de quantification icx , correspondant aux adresses dans le ou les dictionnaires des sous-vecteurs résiduels quantifiés $srcx_q[n]$, est fourni au multiplexeur de sortie 6.

[0044] L'unité 22 délivre également les valeurs des sous-vecteurs résiduels quantifiés, qui forment le vecteur $rcx_q[n]$. Celui-ci est retardé d'une trame en 23, et ses composantes sont multipliées par les coefficients $\alpha(i)$ en 24 pour fournir le vecteur à l'entrée négative du soustracteur 20. Ce dernier vecteur est d'autre part fourni à un additionneur 25, dont l'autre entrée reçoit un vecteur formé par les composantes du résidu quantifié $rcx_q[n]$ respectivement multipliées par les quantités $1 - \alpha(i)$ en 26. L'additionneur 25 délivre ainsi le vecteur cepstral quantifié $cx_q[n]$ que récupérera le décodeur.

[0045] Le coefficient de prédiction $\alpha(i)$ peut être optimisé séparément pour chacun des coefficients cepstraux. Les dictionnaires de quantification peuvent aussi être optimisés séparément pour chacun des quatre sous-vecteurs cepstraux. D'autre part, il est possible, de façon connue en soi, de normaliser les vecteurs cepstraux avant d'appliquer le schéma de prédiction/quantification, à partir de la variance des cepstres.

[0046] Il est à noter que le schéma ci-dessus de quantification des coefficients cepstraux peut n'être appliqué que pour certaines seulement des trames. Par exemple, on peut prévoir un second mode de quantification ainsi qu'un processus de sélection de celui des deux modes qui minimise un critère de moindres carrés avec les coefficients cepstraux à quantifier, et transmettre avec les index de quantification de la trame un bit indiquant lequel des deux modes a été sélectionné.

[0047] Les coefficients cepstraux quantifiés $cx_sup_q = cx_q[n]$ fournis par l'additionneur 25 sont adressés à un module 28 qui recalcule les amplitudes spectrales associées à une ou plusieurs des harmoniques de la fréquence fondamentale F_0 (figure 1). Ces amplitudes spectrales sont par exemple calculées sous forme comprimée, en appliquant la transformée de Fourier aux coefficients cepstraux quantifiés en tenant compte de la distorsion non-linéaire de l'échelle des fréquences utilisée dans la transformation cepstrale. Les amplitudes ainsi recalculées sont fournies à un module d'adaptation 29 qui les compare à des amplitudes de maxima déterminées par le module d'extraction 9.

[0048] Le module d'adaptation 29 contrôle le post-lifre 15 de façon à minimiser un écart de module entre le spectre du signal audio et les valeurs de module correspondantes calculées en 28. Cet écart de module peut être exprimé par une somme de valeurs absolues de différences d'amplitudes, comprimées ou non, correspondant à une ou plusieurs des fréquences harmoniques. Cette somme peut être pondérée en fonction des amplitudes spectrales associées à ces fréquences.

[0049] De façon optimale, l'écart de module pris en compte dans l'adaptation du post-lifrage tiendrait compte de toutes les harmoniques du spectre. Cependant, afin de réduire la complexité de l'optimisation, le module 28 peut ne resynthétiser les amplitudes spectrales que pour une ou plusieurs fréquences multiples de la fréquence fondamentale F_0 , sélectionnées sur la base de l'importance du module du spectre en valeur absolue. Le module d'adaptation 29 peut par exemple considérer les trois pics spectraux les plus intenses dans le calcul de l'écart de module à minimiser.

[0050] Dans une autre réalisation, le module d'adaptation 29 estime une courbe de masquage spectral du signal audio au moyen d'un modèle psychoacoustique, et les fréquences prises en compte dans le calcul de l'écart de module à minimiser sont sélectionnées sur la base de l'importance du module du spectre relativement à la courbe de masquage (on peut par exemple prendre les trois fréquences pour lesquelles le module du spectre dépasse le plus de la courbe de masquage). Différentes méthodes classiques sont utilisables pour calculer la courbe de masquage à partir du signal audio. On peut par exemple utiliser celle développée par J.D. Johnston (« Transform Coding of Audio Signals Using Perceptual Noise Criteria », IEEE Journal on Selected Area in Communications, Vol. 6, No. 2, février 1988).

[0051] Pour réaliser l'adaptation du post-lifrage, le module 29 peut utiliser un modèle d'identification de filtre. Une méthode plus simple consiste à prédéfinir un ensemble de jeux de paramètres de post-lifrage, c'est-à-dire un ensemble de couples γ_1, γ_2 dans le cas d'un post-lifrage selon les relations (8), à effectuer les opérations incombant aux modules 15, 16, 18 et 28 pour chacun de ces jeux de paramètres, et à retenir celui des jeux de paramètres qui conduit à l'écart de module minimal entre le spectre du signal et les valeurs recalculées. Les index de quantification fournis par le module 18 sont alors ceux qui se rapportent au meilleur jeu de paramètres.

[0052] Par un processus analogue à celui de l'extraction des coefficients cx_sup représentant l'enveloppe supérieure comprimée LX_sup du spectre du signal, le codeur détermine des coefficients cx_inf représentant une enveloppe inférieure comprimée LX_inf . Un module 30 extrait du spectre comprimé LX des amplitudes spectrales associées à des fréquences situées dans des zones du spectre intermédiaires par rapport aux fréquences multiples de la fréquence fondamentale estimée F_0 .

[0053] Dans l'exemple illustré par les figures 4 et 5, chaque amplitude associée à une fréquence située dans une zone intermédiaire entre deux harmoniques successives $k.F_0$ et $(k+1).F_0$ correspond simplement au module du spectre pour la fréquence $(k+1/2).F_0$ située au milieu de l'intervalle séparant les deux harmoniques. Dans une autre réalisation, cette amplitude pourrait être une moyenne du module du spectre sur une petite plage entourant cette fréquence $(k+1/2).F_0$.

[0054] Un module 31 procède à une interpolation, par exemple linéaire, des amplitudes spectrales associées aux fréquences situées dans les zones intermédiaires pour obtenir l'enveloppe inférieure comprimée LX_inf .

[0055] La transformation cepstrale appliquée à cette enveloppe inférieure comprimée LX_inf est effectuée suivant une échelle de fréquences résultant d'une distorsion non-linéaire appliquée par un module 32. Le module de TFRI 33 calcule un vecteur cepstral de NCI coefficients cepstraux cx_inf d'ordres 0 à NCI-1 représentant l'enveloppe inférieure. NCI est un nombre qui peut être sensiblement plus petit que NCS, par exemple NCI = 4.

[0056] La transformation non-linéaire de l'échelle des fréquences pour la transformation cepstrale de l'enveloppe inférieure peut être réalisée vers une échelle plus fine aux hautes fréquences qu'aux basses fréquences, ce qui permet avantageusement de bien modéliser les composantes non-voisées du signal aux hautes fréquences. Toutefois, pour assurer une homogénéité de représentation entre l'enveloppe supérieure et l'enveloppe inférieure, on pourra préférer adopter dans le module 32 la même échelle que dans le module 12 (Mel dans l'exemple considéré).

[0057] Les coefficients cepstraux cx_inf représentant l'enveloppe inférieure comprimée sont quantifiés par un module 34, qui peut fonctionner de la même manière que le module 18 de quantification des coefficients cepstraux représentant l'enveloppe supérieure comprimée. Dans le cas considéré, où on se limite à NCI = 4 coefficients cepstraux pour l'enveloppe inférieure, le vecteur ainsi formé est soumis à une quantification vectorielle de résidu de prédiction, effectuée par des moyens identiques à ceux représentés sur la figure 6 mais sans subdivision en sous-vecteurs. L'index de quantification $icx = icxi$ déterminé par le quantificateur vectoriel 22 pour chaque trame relativement aux coefficients cx_inf est fourni au multiplexeur de sortie 6 du codeur.

[0058] Le codeur représenté sur la figure 1 ne comporte aucun dispositif particulier pour coder les phases du spectre aux harmoniques du signal audio.

[0059] En revanche, il comporte des moyens 36-40 pour coder une information temporelle liée à la phase de la composante non-harmonique représentée par l'enveloppe inférieure.

[0060] Un module 36 de décompression spectrale et un module 37 de TFRI forment une estimation temporelle de la trame de la composante non-harmonique. Le module 36 applique une fonction de décompression réciproque de la fonction de compression appliquée par le module 8 (c'est-à-dire une exponentielle ou une fonction puissance $1/\gamma$) à

l'enveloppe inférieure comprimée LX_{inf} produite par le module d'interpolation 31. Ceci fournit le module de la trame estimée de la composante non-harmonique, dont la phase est prise égale à celle φ_X du spectre du signal X sur la trame. La transformée de Fourier inverse effectuée par le module 37 fournit la trame estimée de la composante non-harmonique.

[0061] Le module 38 subdivise cette trame estimée de la composante non-harmonique en plusieurs segments temporels. La trame délivrée par le module 37 se composant de $2N = 512$ échantillons pondérés comme illustré par la figure 3, le module 38 considère seulement les $N/2 = 128$ premiers échantillons et les $N/2 = 128$ derniers échantillons, et les subdivise par exemple en huit segments de 32 échantillons consécutifs représentant chacun 4 ms de signal.

[0062] Pour chaque segment, le module 38 calcule l'énergie égale à la somme des carrés des échantillons, et forme un vecteur E1 formé de huit composantes réelles positives égales aux huit énergies calculées. La plus grande de ces huit énergies, notée EM, est également déterminée pour être fournie, avec le vecteur E1, à un module de normalisation 39. Celui-ci divise chaque composante du vecteur E1 par EM, de sorte que le vecteur normalisé Emix est formé de huit composantes comprises entre 0 et 1. C'est ce vecteur normalisé Emix, ou vecteur de pondération, qui est soumis à la quantification par le module 40. Celui-ci peut opérer une quantification vectorielle avec un dictionnaire déterminé lors d'un apprentissage préalable. L'index de quantification iEm est fourni par le module 40 au multiplexeur de sortie 6 du codeur.

[0063] La figure 7 montre une variante de réalisation des moyens employés par le codeur de la figure 1 pour déterminer le vecteur Emix de pondération énergétique de la trame de la composante non-harmonique. Les modules 36, 37 de décompression spectrale et de TFRI fonctionnent comme ceux qui portent les mêmes références sur la figure 1. Un module de sélection 42 est ajouté pour déterminer la valeur du module du spectre soumis à la transformée de Fourier inverse 37. Sur la base de la fréquence fondamentale estimée F_0 , le module 42 identifie des régions harmoniques et des régions non-harmoniques du spectre du signal audio. Par exemple, une fréquence sera considérée comme appartenant à une région harmonique si elle se trouve dans un intervalle de fréquences centré sur une harmonique $k.F_0$ et de largeur correspondant à une largeur de raie spectrale synthétisée, et à une région non-harmonique sinon. Dans les régions non-harmoniques, le signal complexe soumis à la TFRI 37 est égal à la valeur du spectre, c'est-à-dire que son module et sa phase correspondent aux valeurs $|X|$ et φ_X fournies par le module de TFR 3. Dans les régions harmoniques, ce signal complexe a la même phase φ_X que le spectre et un module donné par l'enveloppe inférieure après décompression spectrale 36. Cette façon de procéder selon la figure 7 procure une modélisation plus précise des régions non-harmoniques.

[0064] Le décodeur représenté sur la figure 8 comprend un démultiplexeur d'entrée 45 qui extrait du flux binaire Φ , issu d'un codeur selon la figure 1, les index iF , $icxs$, $icxi$, iEm de quantification de la fréquence fondamentale F_0 , des coefficients cepstraux représentant l'enveloppe supérieure comprimée, des coefficients représentant l'enveloppe inférieure comprimée, et du vecteur de pondération Emix, et les distribue respectivement à des modules 46, 47, 48 et 49. Ces modules 46-49 comportent des dictionnaires de quantification semblables à ceux des modules 5, 18, 34 et 40 de la figure 1, afin de restituer les valeurs des paramètres quantifiés. Les modules 47 et 48 ont des dictionnaires pour former les résidus de prédiction quantifiés $rcx_q[n]$, et ils en déduisent les vecteurs cepstraux quantifiés $cx_q[n]$ avec des éléments identiques aux éléments 23-26 de la figure 6. Ces vecteurs cepstraux quantifiés $cx_q[n]$ fournissent les coefficients cepstraux cx_sup_q et cx_inf_q traités par le décodeur.

[0065] Un module 51 calcule la transformée de Fourier rapide des coefficients cepstraux cx_sup pour chaque trame de signal. L'échelle des fréquences du spectre comprimé qui en résulte est modifiée non-linéairement par un module 52 appliquant la transformation non-linéaire réciproque de celle du module 12 de la figure 1, et qui fournit l'estimation LX_sup de l'enveloppe supérieure comprimée. Une décompression spectrale de LX_sup , opérée par un module 53, fournit l'enveloppe supérieure X_sup comportant les valeurs estimées du module du spectre aux fréquences multiples de la fréquence fondamentale F_0 . Le module 54 synthétise l'estimation spectrale X_v de la composante harmonique du signal audio, par une somme de raies spectrales centrées sur les fréquences multiples de la fréquence fondamentale F_0 et dont les amplitudes (en module) sont celles données par l'enveloppe supérieure X_sup .

[0066] Bien que le flux numérique d'entrée Φ ne comporte pas d'informations spécifiques sur la phase du spectre du signal aux harmoniques de la fréquence fondamentale, le décodeur de la figure 8 est capable d'extraire de l'information sur cette phase à partir des coefficients cepstraux cx_sup_q représentant l'enveloppe supérieure comprimée. Cette information de phase est utilisée pour affecter une phase $\varphi(k)$ à chacune des raies spectrales déterminées par le module 54 dans l'estimation de la composante harmonique du signal.

[0067] En première approximation, le signal de parole peut être considéré comme étant à phase minimale. D'autre part, il est connu que l'information de phase minimale peut se déduire facilement d'une modélisation cepstrale. Cette information de phase minimale est donc calculée pour chaque fréquence harmonique. L'hypothèse de phase minimale signifie que l'énergie du signal synthétisé est localisée au début de chaque période de la fréquence fondamentale F_0 .

[0068] Pour être plus proche d'un signal de parole réel, on introduit un peu de dispersion au moyen d'un post-liffrage spécifique des cepstres lors de la synthèse de la phase. Avec ce post-liffrage, effectué par le module 55 de la figure 8, il est possible d'accentuer les résonances formantiques de l'enveloppe et donc de contrôler la dispersion des phases. Ce post-liffrage est par exemple de la forme (8).

[0069] Pour limiter les ruptures de phase, il est préférable de lisser les coefficients cepstraux post-liftrés, ce qui est effectué par le module 56. Le module 57 déduit des coefficients cepstraux post-liftrés et lissés la phase minimale affectée à chaque raie spectrale représentant un pic harmonique du spectre.

[0070] Les opérations effectuées par les modules 56, 57 de lissage et d'extraction de la phase minimale sont illustrées par l'organigramme de la figure 9. Le module 56 examine les variations des coefficients cepstraux pour appliquer un lissage moins important en présence de variations brusques qu'en présence de variations lentes. Pour cela, il effectue le lissage des coefficients cepstraux au moyen d'un facteur d'oubli λ_c choisi en fonction d'une comparaison entre un seuil d_{th} et une distance d'entre deux jeux successifs de coefficients cepstraux post-liftrés. Le seuil d_{th} est lui-même adapté en fonction des variations des coefficients cepstraux.

[0071] La première étape 60 consiste à calculer la distance d entre les deux vecteurs successifs relatifs aux trames $n-1$ et n . Ces vecteurs, notés ici $cxp[n-1]$ et $cxp[n]$, correspondent pour chaque trame à l'ensemble des NCS coefficients cepstraux post-liftrés représentant l'enveloppe supérieure comprimée. La distance utilisée peut notamment être la distance euclidienne entre les deux vecteurs ou encore une distance quadratique.

[0072] Deux lissages sont d'abord effectués, respectivement au moyen de facteurs d'oubli λ_{min} et λ_{max} , pour déterminer une distance minimale d_{min} et une distance maximale d_{max} . Le seuil d_{th} est ensuite déterminé à l'étape 70 comme étant situé entre les distances minimale et maximale d_{min} , d_{max} : $d_{th} = \beta \cdot d_{max} + (1-\beta) \cdot d_{min}$, le coefficient β étant par exemple égal à 0,5.

[0073] Dans l'exemple représenté, les facteurs d'oubli λ_{min} et λ_{max} sont eux-mêmes sélectionnés parmi deux valeurs distinctes, respectivement λ_{min1} , λ_{min2} et λ_{max1} , λ_{max2} comprises entre 0 et 1, les indices λ_{min1} , λ_{max1} étant chacun sensiblement plus près de 0 que les indices λ_{min2} , λ_{max2} . Si $d > d_{min}$ (test 61), le facteur d'oubli λ_{min} est égal à λ_{min1} (étape 62) ; sinon il est pris égal à λ_{min2} (étape 63). A l'étape 64, la distance minimale d_{min} est prise égale à $\lambda_{min} \cdot d_{min} + (1-\lambda_{min}) \cdot d$. Si $d > d_{max}$ (test 65), le facteur d'oubli λ_{max} est égal à λ_{max1} (étape 66) ; sinon il est pris égal à λ_{max2} (étape 67). A l'étape 68, la distance maximale d_{max} est prise égale à $\lambda_{max} \cdot d_{max} + (1-\lambda_{max}) \cdot d$.

[0074] Si la distance d entre les deux vecteurs cepstraux consécutifs est plus grande que le seuil d_{th} (test 71), on adopte pour le facteur d'oubli λ_c une valeur λ_{c1} relativement proche de 0 (étape 72). On considère dans ce cas que le signal correspondant est de type non stationnaire, de sorte qu'il n'y a pas lieu de conserver une grande mémoire des coefficients cepstraux antérieurs. Si $d \leq d_{th}$, on adopte à l'étape 73 pour le facteur d'oubli λ_c une valeur λ_{c2} moins proche de 0 afin de lisser davantage les coefficients cepstraux. Le lissage est effectué à l'étape 74, où le vecteur $cxl[n]$ de coefficients lissés pour la trame courante n est déterminé par :

$$cxl[n] = \lambda_c \cdot cxl[n-1] + (1 - \lambda_c) \cdot cxp[n] \quad (11)$$

[0075] Le module 57 calcule ensuite les phases minimales $\varphi(k)$ associées aux harmoniques $k \cdot F_0$. De façon connue, la phase minimale pour une harmonique d'ordre k est donnée par :

$$\varphi(k) = -2 \cdot \sum_{m=1}^{NCS-1} cxl[n,m] \cdot \sin(2\pi m k F_0 / F_e) \quad (12)$$

où $cxl[n,m]$ désigne le coefficient cepstral lissé d'ordre m pour la trame n .

[0076] A l'étape 75, l'index d'harmonique k est initialisé à 1. Pour initialiser le calcul de la phase minimale affectée à l'harmonique k , la phase $\varphi(k)$ et l'index cepstral m sont initialisés respectivement à 0 et 1 à l'étape 76. A l'étape 77, le module 57 ajoute à la phase $\varphi(k)$ la quantité $-2 \cdot cxl[n,m] \cdot \sin(2\pi m k F_0 / F_e)$. L'index cepstral m est incrémenté à l'étape 78 et comparé à NCS à l'étape 79. Les étapes 77 et 78 sont répétées tant que $m < NCS$. Quand $m = NCS$, le calcul de la phase minimale est terminé pour l'harmonique k , et l'index k est incrémenté à l'étape 80. Le calcul de phases minimales 76-79 est renouvelé pour l'harmonique suivante tant que $k \cdot F_0 < F_e/2$ (test 81).

[0077] Dans l'exemple de réalisation selon la figure 8, le module 54 tient compte d'une phase constante sur la largeur de chaque raie spectrale, égale à la phase minimale $\varphi(k)$ fournie pour l'harmonique correspondante k par le module 57.

[0078] L'estimation X_v de la composante harmonique est synthétisée par sommation de raies spectrales positionnées aux fréquences harmoniques de la fréquence fondamentale F_0 . Lors de cette synthèse, on peut positionner les raies spectrales sur l'axe des fréquences avec une résolution supérieure à la résolution de la transformée de Fourier. Pour cela, on précalcule une fois pour toutes une raie spectrale de référence selon la résolution supérieure. Ce calcul peut consister en une transformée de Fourier de la fenêtre d'analyse f_A avec une taille de transformée de 16384 points, procurant une résolution de 0,5 Hz par point. La synthèse de chaque raie harmonique est alors effectuée par le module

54 en positionnant sur l'axe des fréquences la raie de référence à haute résolution, et en sous-échantillonnant cette raie spectrale de référence pour se ramener à la résolution de 16,625 Hz de la transformée de Fourier sur 512 points. Ceci permet de positionner avec précision la raie spectrale.

[0079] Pour la détermination de l'enveloppe inférieure, le module de TFR 85 du décodeur de la figure 8 reçoit les NCI coefficients cepstraux quantifiés cx_inf_q d'ordres 0 à NCI - 1, et il les complète avantageusement par les NCS - NCI coefficients cepstraux cx_sup_q d'ordre NCI à NCS -1 représentant l'enveloppe supérieure. En effet, on peut estimer en première approximation que les variations rapides de l'enveloppe inférieure comprimée sont bien reproduites par celles de l'enveloppe supérieure comprimée. Dans une autre réalisation, le module de TFR 85 pourrait ne considérer que les NCI paramètres cepstraux cx_inf_q .

[0080] Le module 86 convertit l'échelle de fréquences de manière réciproque de la conversion opérée par le module 32 du codeur, afin de restituer l'estimation LX_inf de l'enveloppe inférieure comprimée, soumise au module de décompression spectrale 87. En sortie du module 87, le décodeur dispose d'une enveloppe inférieure X_inf comportant les valeurs du module du spectre dans les vallées situées entre les pics harmoniques.

[0081] Cette enveloppe X_inf va moduler le spectre d'une trame de bruit dont la phase est traitée en fonction du vecteur de pondération quantifié $Emix$ extrait par le module 49. Un générateur 88 délivre une trame de bruit normalisé dont les segments de 4 ms sont pondérés dans un module 89 conformément aux composantes normalisées du vecteur fourni par le module 49 pour la trame courante. Ce bruit est un bruit blanc filtré passe-haut pour tenir compte du faible niveau qu'a en principe la composante non-voisée aux basses fréquences. A partir du bruit pondéré en énergie, le module 90 forme des trames de $2N = 512$ échantillons en appliquant la fenêtre d'analyse f_A , l'insertion de 256 échantillons à zéro et la permutation circulaire pour la compensation de phase conformément à ce qui a été expliqué en référence à la figure 3. La transformée de Fourier de la trame résultante est calculée par le module TFR 91.

[0082] L'estimation spectrale X_{uv} de la composante non-harmonique est déterminée par le module de synthèse spectrale 92 qui effectue une pondération fréquence par fréquence. Cette pondération consiste à multiplier chaque valeur spectrale complexe fournie par le module de TFR 91 par la valeur de l'enveloppe inférieure X_inf obtenue pour la même fréquence par le module de décompression spectrale 87.

[0083] Les estimations spectrales X_v , X_{uv} des composantes harmonique (voisée dans le cas d'un signal de parole) et non-harmonique (ou non-voisée) sont combinées par un module de mixage 95 contrôlé par un module 96 d'analyse du degré d'harmonicité (ou de voisement) du signal.

[0084] L'organisation de ces modules 95, 96 est illustrée par la figure 10. Le module d'analyse 96 comporte une unité 97 d'estimation d'un degré de voisement W dépendant de la fréquence, à partir duquel sont calculés quatre gains dépendant de la fréquence, à savoir deux gains g_v , g_{uv} contrôlant l'importance relative des composantes harmonique et non-harmonique dans le signal synthétisé, et deux gains g_{v_ϕ} , g_{uv_ϕ} utilisés pour bruiteur la phase de la composante harmonique.

[0085] Le degré de voisement $W(i)$ est une valeur à variation continue comprise entre 0 et 1 déterminée pour chaque index de fréquence i ($0 \leq i < N$) en fonction de l'enveloppe supérieure $X_sup(i)$ et de l'enveloppe inférieure $X_inf(i)$ obtenues pour cette fréquence i par les modules de décompression 53, 87. Le degré de voisement $W(i)$ est estimé par l'unité 97 pour chaque index de fréquence i correspondant à une harmonique de la fréquence fondamentale F_0 , à savoir

$$i = \left\lfloor 2Nk \frac{F_0}{F_e} + \frac{1}{2} \right\rfloor \text{ pour } k = 1, 2, \dots, \text{ par une fonction croissante du rapport entre l'enveloppe supérieure } X_sup \text{ et}$$

l'enveloppe inférieure X_inf à cette fréquence, par exemple selon la formule :

$$W(i) = \min \left\{ 1, \frac{10 \cdot \log_{10} [X_sup(i)/X_inf(i)]}{Vth(F_0)} \right\} \quad (13)$$

Le seuil $Vth(F_0)$ correspond à la dynamique moyenne calculée sur un spectre synthétique purement voisé à la fréquence fondamentale. Il est avantageusement choisi dépendant de la fréquence fondamentale F_0 .

[0086] Le degré de voisement $W(i)$ pour une fréquence autre que les fréquences harmoniques est obtenu simplement comme étant égal à celui estimé pour l'harmonique la plus proche.

[0087] Le gain $g_v(i)$, qui dépend de la fréquence, est obtenu en appliquant une fonction non-linéaire au degré de voisement $W(i)$ (bloc 98). Cette fonction non-linéaire a par exemple la forme représentée sur la figure 11 :

$$g_v(i) = 0 \quad \text{si } 0 \leq W(i) \leq W1$$

$$g_v(i) = \frac{W(i) - W1}{W2 - W1} \quad \text{si } W1 < W(i) < W2 \quad (14)$$

$$g_v(i) = 1 \quad \text{si } W2 \leq W(i) \leq 1$$

les seuils W1, W2 étant tels que $0 < W1 < W2 < 1$. Le gain g_{uv} peut être calculé de manière semblable au gain g_v (la somme des deux gains g_v , g_{uv} étant constante, par exemple égale à 1), ou déduit simplement de celui-ci par la relation $g_{uv}(i) = 1 - g_v(i)$, comme schématisé par le soustracteur 99 sur la figure 10.

[0088] Il est intéressant de pouvoir bruite la phase de la composante harmonique du signal à une fréquence donnée si l'analyse du degré de voisement montre que le signal est plutôt de type non-harmonique à cette fréquence.

la phase ϕ_v de la composante harmonique mixée est le résultat d'une combinaison linéaire des phases (ϕ_v , ϕ_{uv} des composantes harmonique et non-harmonique X_v , X_{uv} synthétisées par les modules 54, 92. Les gains g_{v_ϕ} , g_{uv_ϕ} respectivement appliqués à ces phases sont calculés à partir du degré de voisement W et pondérés également en fonction de l'index de fréquence i, étant donné que le bruitage de la phase n'est véritablement utile qu'au-delà d'une certaine fréquence.

[0089] Un premier gain g_{v1_ϕ} est calculé en appliquant une fonction non-linéaire au degré de voisement W(i), comme schématisé par le bloc 100 sur la figure 10. Cette fonction non-linéaire peut avoir la forme représentée sur la figure 12 :

$$g_{v1_\phi}(i) = G1 \quad \text{si } 0 \leq W(i) \leq W3$$

$$g_{v1_\phi}(i) = G1 + (1 - G1) \frac{W(i) - W3}{W4 - W3} \quad \text{si } W3 < W(i) < W4 \quad (15)$$

$$g_{v1_\phi}(i) = 1 \quad \text{si } W4 \leq W(i) \leq 1$$

les seuils W3 et W4 étant tels que $0 < W3 < W4 < 1$, et le gain minimal G1 étant compris entre 0 et 1.

[0090] Un multiplieur 101 multiplie pour chaque fréquence d'index i le gain g_{v1_ϕ} par un autre gain g_{v2_ϕ} dépendant seulement de l'index de fréquence i, pour former le gain $g_{v_\phi}(i)$. Le gain $g_{v2_\phi}(i)$ dépend non-linéairement de l'index de fréquence i, par exemple comme indiqué sur la figure 13 :

$$g_{v2_\phi}(i) = 1 \quad \text{si } 0 \leq i \leq i1$$

$$g_{v2_\phi}(i) = 1 - (1 - G2) \frac{i - i1}{i2 - i1} \quad \text{si } i1 < i < i2 \quad (16)$$

$$g_{v2_\phi}(i) = G2 \quad \text{si } i2 \leq i \leq 1$$

les index i1 et i2 étant tels que $0 < i1 < i2 \leq N$, et le gain minimal G2 étant compris entre 0 et 1. Le gain $g_{uv_\phi}(i)$ peut être calculé simplement comme étant égal à $1 - g_{v_\phi}(i) = 1 - g_{v1_\phi}(i) \cdot g_{v2_\phi}(i)$ (soustracteur 102 de la figure 10).

[0091] Le spectre complexe Y du signal synthétisé est produit par le module de mixage 95, qui réalise la relation de mixage suivante, pour $0 \leq i < N$:

$$Y(i) = g_v(i) \cdot |X_v(i)| \cdot \exp[j \phi_v(i)] + g_{uv}(i) \cdot X_{uv}(i) \quad (17)$$

avec

$$\varphi_v(i) = g_{v_\varphi}(i) \cdot \varphi_v(i) + g_{uv_\varphi}(i) \cdot \varphi_{uv}(i) \quad (18)$$

5 où $\varphi_v(i)$ désigne l'argument du nombre complexe $X_v(i)$ fourni par le module 54 pour la fréquence d'index i (bloc 104 de la figure 1.0), et $\varphi_{uv}(i)$ désigne l'argument du nombre complexe $X_{uv}(i)$ fourni par le module 92 (bloc 105 de la figure 10). Cette combinaison est réalisée par les multiplieurs 106-110 et les additionneurs 111-112 représentés sur la figure 10.

10 **[0092]** Le spectre mixé $Y(i)$ pour $0 \leq i < 2N$ (avec $Y(2N-1-i) = Y(i)$) est ensuite transformé dans le domaine temporel par le module de TFRI 115 (figure 8). On ne retient que les $N/2 = 128$ premiers et les $N/2 = 128$ derniers échantillons de la trame de $2N = 512$ échantillons produite par le module 115, et on applique la permutation circulaire inverse de celle illustrée par la figure 3 pour obtenir la trame synthétisée de $N = 256$ échantillons pondérés par la fenêtre d'analyse f_A .

[0093] Les trames successivement obtenues de cette manière sont finalement traitées par le module de synthèse temporelle 116 qui forme le signal audio décodé x .

15 **[0094]** Le module de synthèse temporelle 116 effectue une somme à recouvrement de trames modifiées par rapport à celles successivement évaluées en sortie du module 115. La modification peut être vue en deux étapes illustrées respectivement par les figures 14 et 15.

[0095] La première étape (figure 14) consiste à multiplier chaque trame 2' délivrée par le module de TFRI 115 par une fenêtre $1/f_A$ inverse de la fenêtre d'analyse f_A employée par le module 1 du codeur. Les échantillons de la trame 2" qui en résultent sont donc pondérés uniformément.

20 **[0096]** La seconde étape (figure 15) consiste à multiplier les échantillons de cette trame 2" par une fenêtre de synthèse f_S vérifiant les propriétés suivantes :

$$f_S(N-L+i) + f_S(i) = A \quad \text{pour } 0 \leq i < L \quad (19)$$

$$f_S(i) = A \quad \text{pour } L \leq i < N-L \quad (20)$$

où A désigne une constante positive arbitraire, par exemple $A = 1$. La fenêtre de synthèse $f_S(i)$ croît progressivement de 0 à A pour i allant de 0 à L . C'est par exemple une demi-sinusoïde surélevée :

$$f_S(i) = \frac{A}{2} \cdot (1 - \cos[(i + 1/2)\pi/L]) \quad \text{pour } 0 \leq i < L \quad (21)$$

40 **[0097]** Après avoir repondéré chaque trame 2" par la fenêtre de synthèse f_S , le module 116 positionne les trames successives avec leurs décalages temporels de $M = 160$ échantillons et leurs recouvrements temporels de $L = 96$ échantillons, puis il effectue la somme des trames ainsi positionnées dans le temps. Du fait des propriétés (19) et (20) de la fenêtre de synthèse f_S , chaque échantillon du signal audio décodé x ainsi obtenu est affecté d'un poids global uniforme égal à A . Ce poids global provient de la contribution d'une trame unique si l'échantillon a dans cette trame un rang i tel que $L \leq i < N - L$, et comporte les contributions sommées de deux trames successives si $0 \leq i < L$ ou $N-L \leq i < N$.

45 **[0098]** On peut ainsi effectuer la synthèse temporelle de façon simple même si, comme dans le cas considéré, le recouvrement L entre deux trames successives est plus petit que la moitié de la taille N de ces trames.

[0099] Les deux étapes exposées ci-dessus pour la modification des trames de signal peuvent être fusionnées en une seule étape. Il suffit de précalculer une fenêtre composée $f_C(i) = f_S(i)/f_A(i)$, et de multiplier simplement les trames 2' de $N = 256$ échantillons délivrées par le module 115 par la fenêtre composée f_C avant d'effectuer la sommation à recouvrement.

50 **[0100]** La figure 16 montre l'allure de la fenêtre composée f_C dans le cas où la fenêtre d'analyse f_A est une fenêtre de Hamming et la fenêtre de synthèse f_S a la forme donnée par les relations (19) à (21).

[0101] D'autres formes de la fenêtre de synthèse f_S vérifiant les relations (19) et (20) peuvent être employées. Dans la variante de la figure 17, c'est une fonction affine par morceaux définie par :

$$f_S(i) = A \cdot i/L \quad \text{pour } 0 \leq i < L \quad (22)$$

[0102] Afin d'améliorer la qualité de codage du-signal audio, le codeur de la figure 1 peut augmenter la cadence de formation et d'analyse des trames, afin de transmettre davantage de paramètres de quantification au décodeur. Dans la structure de trame représentée sur la figure 2, une trame de $N = 256$ échantillons (32 ms) est formée toutes les 20 ms. Ces trames de 256 échantillons pourraient être formées à une cadence supérieure, par exemple de 10 ms, deux trames successives ayant alors un décalage de $M/2 = 80$ échantillons et un recouvrement de 176 échantillons.

[0103] Dans ces conditions, on peut transmettre les jeux complets de paramètres de quantification iF , $icxs$, $icxi$, iEm pour seulement un sous-ensemble des trames, et transmettre pour les autres trames des paramètres permettant d'effectuer une interpolation adéquate au niveau du décodeur. Dans l'exemple envisagé ci-dessus, le sous-ensemble pour lequel des jeux de paramètres complets sont transmis peut être constitué par les trames de rang entier n , dont la périodicité est de $M/F_e = 20$ ms, et les trames pour lesquelles une interpolation est effectuée peuvent être celles de rang demi-entier $n + 1/2$ qui sont décalées de 10 ms par rapport aux trames du sous-ensemble.

[0104] Dans la réalisation illustrée par la figure 18, les notations $cx_q[n-1]$ et $cx_q[n]$ désignent des vecteurs cepstraux quantifiés déterminés, pour deux trames successives de rang entier, par le module de quantification 18 et/ou par le module de quantification 34. Ces vecteurs comprennent par exemple quatre coefficients cepstraux consécutifs chacun. Ils pourraient également comprendre davantage de coefficients cepstraux.

[0105] Un module 120 effectue une interpolation de ces deux vecteurs cepstraux $cx_q[n-1]$ et $cx_q[n]$, afin d'estimer une valeur intermédiaire $cx_i[n-1/2]$. L'interpolation effectuée par le module 120 peut être une simple moyenne arithmétique des vecteurs $cx_q[n-1]$ et $cx_q[n]$. En variante, le module 120 pourrait appliquer une formule d'interpolation plus sophistiquée, par exemple polynomiale, en se fondant également sur les vecteurs cepstraux obtenus pour des trames antérieures à la trame $n-1$. D'autre part, si plus d'une trame interpolée est intercalée entre deux trames consécutives de rang entier, l'interpolation tient compte de la position relative de chaque trame interpolée.

[0106] A l'aide des moyens précédemment décrits, le codeur calcule également les coefficients cepstraux $cx[n-1/2]$ relatifs à la trame de rang demi-entier. Dans le cas de l'enveloppe supérieure, ces coefficients cepstraux sont ceux fournis par le module de TFRI 13 après post-lifrage 15 (par exemple avec les mêmes coefficients de post-lifrage que pour la trame précédente $n-1$) et normalisation 16. Dans le cas de l'enveloppe inférieure, les coefficients cepstraux $cx[n-1/2]$ sont ceux délivrés par le module de TFRI 33.

[0107] Un soustracteur 121 forme la différence $ecx[n-1/2]$ entre les coefficients cepstraux $cx[n-1/2]$ calculés pour la trame de rang demi-entier et les coefficients $cx_i[n-1/2]$ estimés par interpolation. Cette différence est fournie à un module de quantification 122 qui adresse des index de quantification $icx[n-1/2]$ au multiplexeur de sortie 6 du codeur. Le module 122 fonctionne par exemple par quantification vectorielle des erreurs d'interpolation $ecx[n-1/2]$ successivement déterminées pour les trames de rang demi-entier.

[0108] Cette quantification de l'erreur d'interpolation peut être effectuée par le codeur pour chacun des NCS + NCI coefficients cepstraux utilisés par le décodeur, ou seulement pour certains d'entre eux, typiquement ceux d'ordres les plus petits.

[0109] Les moyens correspondants du décodeur sont illustrés par la figure 19. Le décodeur fonctionne essentiellement comme celui décrit en référence à la figure 8 pour déterminer les trames de signal de rang entier. Un module d'interpolation 124 identique au module 120 du codeur estime les coefficients intermédiaires $cx_i[n-1/2]$ à partir des coefficients quantifiés $cx_q[n-1]$ et $cx_q[n]$ fournis par le module 47 et/ou le module 48 à partir des index $icxs$, $icxi$ extraits du flux Φ . Un module d'extraction de paramètres 125 reçoit l'index de quantification $icx[n-1/2]$ depuis le démultiplexeur d'entrée 45 du décodeur, et en déduit l'erreur d'interpolation quantifiée $ecx_q[n-1/2]$ à partir du même dictionnaire de quantification que celui utilisé par le module 122 du codeur. Un additionneur 126 fait la somme des vecteurs cepstraux $cx_i[n-1/2]$ et $ecx_q[n-1/2]$ afin de fournir les coefficients cepstraux $cx[n-1/2]$ qui seront utilisés par le décodeur (modules 51-57, 95, 96, 115 et/ou modules 85-87, 92, 95, 96, 115) pour former la trame interpolée de rang $n-1/2$.

[0110] Si certains seulement des coefficients cepstraux ont fait l'objet d'une quantification d'erreur d'interpolation, les autres sont déterminés par le décodeur par une interpolation simple, sans correction.

[0111] Le décodeur peut également interpoler les autres paramètres F_0 , $Emix$ utilisés pour synthétiser les trames de signal. La fréquence fondamentale F_0 peut être interpolée linéairement, soit dans le domaine temporel, soit (de préférence) directement dans le domaine fréquentiel. Pour l'interpolation éventuelle du vecteur de pondération énergétique $Emix$, il convient d'effectuer l'interpolation après dénormalisation et en tenant compte bien entendu des décalages temporels entre trames.

[0112] Il est à noter qu'il est particulièrement avantageux, pour interpoler la représentation des enveloppes spectrales, d'effectuer cette interpolation dans le domaine cepstral. Contrairement à une interpolation effectuée sur d'autres paramètres, tels que les coefficients LSP (« Line Spectrum Pairs »), l'interpolation linéaire des coefficients cepstraux correspond à l'interpolation linéaire des amplitudes spectrales comprimées.

[0113] Dans la variante représentée sur la figure 20, le codeur utilise les vecteurs cepstraux $cx_q[n]$, $cx_q[n-1]$, ..., $cx_q[n-r]$ et $cx_q[n-1/2]$ calculés pour les dernières trames passées ($r \geq 1$) pour identifier un filtre interpolateur optimal qui, lorsqu'on lui soumet les vecteurs cepstraux quantifiés $cx_q[n-r]$, ..., $cx_q[n]$ relatifs aux trames de rang entier, délivre un vecteur cepstral interpolé $cx_i[n-1/2]$ qui présente une distance minimale avec le vecteur $cx[n-1/2]$ calculé pour la

dernière trame de rang demi-entier.

[0114] Dans l'exemple représenté sur la figure 20, ce filtre interpolateur 128 est présent dans le codeur, et un soustracteur 129 retranche sa sortie $cx_i[n-1/2]$ du vecteur cepstral calculé $cx[n-1/2]$. Un module de minimisation 130 détermine le jeu de paramètres {P} du filtre interpolateur 128, pour lequel l'erreur d'interpolation $ecx[n-1/2]$ délivrée par le soustracteur 129 présente une norme minimale. Ce jeu de paramètres {P} est adressé à un module de quantification 131 qui fournit un index de quantification correspondant iP au multiplexeur de sortie 6 du codeur.

[0115] En fonction du débit alloué dans le flux Φ aux index de quantification des paramètres {P} définissant le filtre interpolateur optimal 128, on pourra adopter une quantification plus ou moins fine de ces paramètres, ou une forme plus ou moins élaborée du filtre interpolateur, ou encore prévoir plusieurs filtres interpolateurs quantifiés de manière distincte pour différents vecteurs de coefficients cepstraux.

[0116] Dans une réalisation simple, le filtre interpolateur 128 est linéaire, avec $r=1$:

$$cx_i[n-1/2] = \rho \cdot cx_q[n-1] + (1-\rho) \cdot cx_q[n] \quad (23)$$

et le jeu de paramètres {P} se limite au coefficient ρ compris entre 0 et 1.

[0117] A partir des index iP de quantification des paramètres {P} obtenus dans le flux binaire ϕ , le décodeur reconstruit le filtre interpolateur 128 (aux erreurs de quantification près), et traite les vecteurs spectraux $cx_q[n-r], \dots, cx_q[n]$ afin d'estimer les coefficients cepstraux $cx[n-1/2]$ utilisés pour synthétiser les trames de rang demi-entier.

[0118] De façon générale, le décodeur peut utiliser une méthode d'interpolation simple (sans transmission de paramètres de la part du codeur pour les trames de rang demi-entier), une méthode d'interpolation avec prise en compte d'une erreur d'interpolation quantifiée (selon les figures 17 et 18), ou une méthode d'interpolation avec un filtre interpolateur optimal (selon la figure 19) pour évaluer les trames de rang demi-entier en plus des trames de rang entier évaluées directement comme expliqué en référence aux figures 8 à 13. Le module 116 de synthèse temporelle peut alors combiner l'ensemble de ces trames évaluées pour former le signal synthétisé x de la manière expliquée ci-après en référence aux figures 14, 21 et 22.

[0119] Comme dans la méthode de synthèse temporelle précédemment décrite, le module 116 effectue une somme à recouvrement de trames modifiées par rapport à celles successivement évaluées en sortie du module 115, et cette modification peut être vue en deux étapes dont la première est identique à celle précédemment décrite en référence à la figure 14 (diviser les échantillons de la trame 2' par la fenêtre d'analyse f_A).

[0120] La seconde étape (figure 21) consiste à multiplier les échantillons de la trame renormalisée 2" par une fenêtre de synthèse f'_S vérifiant les propriétés suivantes :

$$f'_S(i) = 0 \quad \text{pour } 0 \leq i < N/2 - M/p \quad \text{et} \quad N/2 + M/p \leq i < N \quad (24)$$

$$f'_S(i) + f'_S(i + M/p) = A \quad \text{pour } N/2 - M/p \leq i < N/2 \quad (25)$$

où A désigne une constante positive arbitraire, par exemple $A = 1$, et p est l'entier tel que le décalage temporel entre les trames successives (calculées directement et interpolées) soit de M/p échantillons, soit $p = 2$ dans l'exemple décrit.

La fenêtre de synthèse $f'_S(i)$ croît progressivement pour i allant de $N/2 - M/p$ à $N/2$. C'est par exemple une sinusoïde

surélevée sur l'intervalle $N/2 - M/p \leq i < N/2 + M/p$. En particulier, la fenêtre de synthèse f'_S peut être, sur cet intervalle, une fenêtre de Hamming (comme représenté sur la figure 21) ou une fenêtre de Hanning.

[0121] La figure 21 montre les trames successives 2" repositionnées dans le temps par le module 116. Les hachures indiquent les portions éliminées des trames (fenêtre de synthèse à 0). On voit qu'en effectuant la somme à recouvrement des échantillons des trames successives, la propriété (25) assure une pondération homogène des échantillons du signal synthétisé.

[0122] Comme dans la méthode de synthèse illustrée par les figures 14 et 15, la procédure de pondération des trames obtenues par transformée de Fourier inverse des spectres Y peut être effectuée en une seule étape, avec une fenêtre

composée $f'_C(i) = f'_S(i)/f_A(i)$. La figure 22 montre la forme de la fenêtre composée f'_C dans le cas où les fenêtres f_A et f'_S sont de type Hamming.

[0123] Comme la méthode de synthèse temporelle illustrée par les figures 14 à 17, celle illustrée par les figures 14, 21 et 22 permet de prendre en compte un recouvrement L entre deux trames d'analyse (pour lesquelles l'analyse est effectuée de façon complète) plus petit que la moitié que la taille N de ces trames. De façon générale, cette dernière méthode est applicable lorsque les trames d'analyse successives présentent des décalages temporels mutuels M de plus de N/2 échantillons (même éventuellement de plus de N échantillons si un très bas débit est requis), l'interpolation conduisant à un ensemble de trames dont les décalages temporels mutuels sont de moins de N/2 échantillons.

[0124] Les trames interpolées peuvent faire l'objet d'une transmission réduite de paramètres de codage, comme décrit précédemment, mais cela n'est pas obligatoire. Ce mode de réalisation permet de conserver un intervalle M relativement grand entre deux trames d'analyse, et donc de limiter le débit de transmission requis, tout en limitant les discontinuités susceptibles d'apparaître en raison de la taille de cet intervalle par rapport aux échelles de temps typiques des variations des paramètres du signal audio, notamment les coefficients cepstraux et la fréquence fondamentale.

Revendications

1. Procédé d'analyse d'un signal audio (x) traité par trames successives de N échantillons, dans lequel on pondère les échantillons de chaque trame par une fenêtre d'analyse (f_A) de type Hamming, Hanning, Kaiser ou analogue, on calcule un spectre du signal audio en transformant chaque trame d'échantillons pondérés dans le domaine fréquentiel, et on traite le spectre du signal audio pour délivrer des paramètres (cx_sup, cx_inf, Emix) de synthèse d'un signal dérivé du signal audio analysé, **caractérisé en ce que** les trames successives comportent une alternance de trames pour lesquelles on délivre des jeux complets de paramètres de synthèse et de trames pour lesquelles on délivre des jeux incomplets de paramètres de synthèse, et **en ce que** les trames successives pour lesquelles on délivre des jeux complets de paramètres de synthèse présentent des recouvrements mutuels de moins de N/2 échantillons.
2. Procédé selon la revendication 1, dans lequel les jeux incomplets de paramètres de synthèse incluent des données (icx[n-1/2]) représentant une erreur (ecx[n-1/2]) d'interpolation d'au moins un des paramètres de synthèse.
3. Procédé selon la revendication 1, dans lequel les jeux incomplets de paramètres de synthèse incluent des données (iP) représentant un filtre (128) d'interpolation d'au moins un des paramètres de synthèse.
4. Procédé selon l'une quelconque des revendications 1 à 3, dans lequel le traitement du spectre du signal audio (x) comporte une extraction de paramètres de codage (cx_sup, cx_inf, Emix) en vue de la transmission et/ou du stockage du signal audio codé.
5. Procédé selon l'une quelconque des revendications 1 à 3, dans lequel le traitement du spectre du signal audio (x) comporte un débruitage par soustraction spectrale.
6. Dispositif de traitement audio, comprenant des moyens d'analyse qui réalisent un procédé selon l'une quelconque des revendications 1 à 5.
7. Procédé de synthèse d'un signal audio, dans lequel on évalue un ensemble de trames recouvrantes successives de N échantillons du signal audio pondérés par une fenêtre d'analyse (f_A), en transformant dans le domaine temporel des estimations spectrales (Y) correspondant respectivement aux dites trames, et on combine les trames évaluées pour former le signal synthétisé (x), **caractérisé en ce que**, pour un sous-ensemble des trames évaluées, les estimations spectrales sont obtenues en traitant des paramètres de synthèse (cx_sup_q, cx_inf_q, Emix) respectivement associés aux trames dudit sous-ensemble tandis que, pour les trames ne faisant pas partie du sous-ensemble, les estimations spectrales sont obtenues avec une interpolation d'une partie au moins des paramètres de synthèse, **en ce que** les trames successives dudit sous-ensemble présentent des décalages temporels mutuels de M échantillons, le nombre M étant plus grand que N/2, tandis que les trames successives dudit ensemble présentent des décalages temporels mutuels de M/p échantillons, p étant un entier plus grand que 1, **en ce qu'on** modifie chaque trame évaluée en lui appliquant un traitement correspondant à une division par ladite fenêtre d'analyse (f_A) et à une multiplication par une fenêtre de synthèse (f'_S), et on forme le signal synthétisé comme une

somme à recouvrement des trames modifiées, et **en ce que**, les échantillons d'une trame ayant des rangs i numérotés de 0 à $N-1$, la fenêtre de synthèse $f'_s(i)$ a un support limité aux rangs i allant de $N/2 - M/p$ à $N/2 + M/p$ et vérifie

$$f'_s(i) + f'_s(i+M/p) = A \text{ pour } N/2 - M/p \leq i < N/2, A \text{ étant une constante positive.}$$

8. Procédé selon la revendication 7, dans lequel la fenêtre de synthèse $f'_s(i)$ croit pour i allant de $N/2 - M/p$ à $N/2$.

9. Procédé selon la revendication 8, dans lequel la fenêtre de synthèse $f'_s(i)$ pour $N/2 - M/p \leq i < N/2 + M/p$ est une sinusoïde surélevée.

10. Procédé selon l'une quelconque des revendications 7 à 9, dans lequel des données ($icx_q[n-1/2]$) représentant une erreur d'interpolation ($ecx_q[n-1/2]$) sont associées aux trames ne faisant pas partie dudit sous-ensemble, et sont utilisées pour corriger au moins un des paramètres de synthèse interpolés ($cx_i[n-1/2]$).

11. Procédé selon l'une quelconque des revendications 7 à 9, dans lequel des données (iP) représentant un filtre interpolateur (128) sont associées aux trames ne faisant pas partie dudit sous-ensemble, et sont utilisées pour interpoler au moins un des paramètres de synthèse.

12. Procédé selon l'une quelconque des revendications 7 à 11, dans lequel les paramètres de synthèse comprennent des coefficients cepstraux ($cx[n]$) soumis à l'interpolation.

13. Dispositif de traitement audio, comprenant des moyens de synthèse qui réalisent un procédé selon l'une quelconque des revendications 7 à 12.

Patentansprüche

1. Verfahren zum Analysieren eines Audiosignals (x), das in Form von aufeinander folgenden Rahmen mit N Abtastwerten verarbeitet wird, wobei die Abtastwerte jedes Rahmens mittels eines Analysefensters (f_A) vom Typ Hamming, Hanning, Kaiser oder dergleichen gewichtet werden, wobei ein Spektrum des Audiosignals berechnet wird, indem jeder Rahmen aus gewichteten Abtastwerten im Frequenzbereich transformiert wird, und wobei das Spektrum des Audiosignals verarbeitet wird, um Syntheseparameter (cx_sup , cx_inf , $Emix$) eines von dem analysierten Audiosignal abgeleiteten Signals auszugeben, **dadurch gekennzeichnet, dass** die aufeinander folgenden Rahmen eine abwechselnde Folge von Rahmen aufweisen, für die vollständige Sätze von Syntheseparametern ausgegeben werden, und von Rahmen, für die unvollständige Sätze von Syntheseparametern ausgegeben werden, und dass diejenigen aufeinander folgenden Rahmen, für die vollständige Sätze von Syntheseparametern ausgegeben werden, eine gegenseitige Überdeckung von mindestens $N/2$ -Abtastwerten aufweisen.

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** die unvollständigen Sätze von Syntheseparametern Werte ($icx[n-1/2]$) beinhalten, die einen Interpolationsfehler ($ecx[n-1/2]$) wenigstens eines der Syntheseparameter darstellen.

3. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** die unvollständigen Sätze von Syntheseparametern Werte (iP) beinhalten, die einen Interpolationsfilter (128) wenigstens eines der Syntheseparameter darstellen.

4. Verfahren nach einem der Ansprüche 1 bis 3, **dadurch gekennzeichnet, dass** die Verarbeitung des Spektrums des Audiosignals (x) ein Extrahieren von Codierungsparametern (cx_sup , cx_inf , $Emix$) für die Übertragung und / oder Speicherung des codierten Audiosignals beinhaltet.

5. Verfahren nach einem der Ansprüche 1 bis 3, **dadurch gekennzeichnet, dass** die Verarbeitung des Spektrums des Audiosignals (x) ein Entrauschen durch spektrale Subtraktion beinhaltet.

6. Audioverarbeitungsrichtung mit Analysemitteln, die zur Durchführung eines Verfahrens nach einem der Ansprüche 1 bis 5 ausgebildet sind.

7. Verfahren zum Synthetisieren eines Audiosignals, wobei eine Gesamtheit von überlappenden, aufeinander folgenden Rahmen mit N mittels eines Analysefensters (f_A) gewichteten Abtastwerten des Audiosignals ausgewertet werden, indem den jeweiligen Rahmen entsprechende spektrale Schätzwerte (Y) im Zeitbereich transformiert werden, und wobei die ausgewerteten Rahmen zum Erzeugen des synthetisierten Signals (x) kombiniert werden, **dadurch gekennzeichnet, dass** für eine Untermenge der ausgewerteten Rahmen die spektralen Schätzwerte erhalten werden, indem den jeweiligen Rahmen der Untergruppe zugeordnete Syntheseparameter (cx_{sup_q} , cx_{inf_q} , Emix) verarbeitet werden, wohingegen für die nicht zu der Untermenge gehörigen Rahmen die spektralen Schätzwerte mit einer Interpolation wenigstens einer Teilmenge der Syntheseparameter erhalten werden, dass die aufeinander folgenden Rahmen der Untermenge zueinander zeitliche Verschiebungen von M Abtastwerten aufweisen, wobei die Zahl M größer ist als N/2, wohingegen die aufeinander folgenden Rahmen der Gesamtheit zueinander zeitliche Verschiebungen von M/p Abtastwerten aufweisen, wobei p eine ganze Zahl größer als 1 ist, dass jeder ausgewertete Rahmen modifiziert wird, indem auf ihn eine Verarbeitung entsprechend einer Division durch das Analysefenster (f_A) und entsprechend einer Multiplikation mit einem Synthesefenster (f_S') angewendet wird und wobei das synthetisierte Signal als ein Overlap-Add der modifizierten Rahmen gebildet wird, und dass bei Anordnung der Abtastwerte eines Rahmens in von 0 bis N-1 nummerierten Rängen das Synthesefenster $f_S'(i)$ eine auf die Ränge i von $N/2 - M/p$ bis $N/2 + M/p$ begrenzte Breite hat und der Beziehung $f_S'(i) + f_S'(i+M/p) = A$ für $N/2 - M/p \leq i < N/2$ genügt, wobei A eine positive Konstante ist.
8. Verfahren nach Anspruch 7, **dadurch gekennzeichnet, dass** das Synthesefenster $f_S'(i)$ für i zwischen $N/2 - M/p$ bis $N/2$ ansteigt.
9. Verfahren nach Anspruch 8, **dadurch gekennzeichnet, dass** das Synthesefenster $f_S'(i)$ für $N/2 - M/p \leq i < N/2 + M/p$ eine überhöhte Sinuskurve ist.
10. Verfahren nach einem der Ansprüche 7 bis 9, **dadurch gekennzeichnet, dass** einen Interpolationsfehler ($ecx_q[n-1/2]$) darstellenden Werte ($icx_q[n-1/2]$) denjenigen Rahmen zugeordnet sind, die nicht zu der Untermenge gehören, und dazu verwendet werden, wenigstens einen der interpolierten Syntheseparameter ($cx_i[n-1/2]$) zu korrigieren.
11. Verfahren nach einem der Ansprüche 7 bis 9, **dadurch gekennzeichnet, dass** einen Interpolationsfilter (128) darstellende Werte (iP) denjenigen Rahmen zugeordnet sind, die nicht zu der Untermenge gehören, und dazu verwendet werden, wenigstens einen der Syntheseparameter zu interpolieren.
12. Verfahren nach einem der Ansprüche 7 bis 11, **dadurch gekennzeichnet, dass** die Syntheseparameter Cepstral-koeffizienten ($cx[n]$) beinhalten, die in der Interpolation unterworfen sind.
13. Audioverarbeitungsrichtung mit Synthesemitteln, die zur Durchführung eines Verfahrens nach einem der Ansprüche 7 bis 12 ausgebildet sind.

Claims

1. A method of analyzing an audio signal (x) processed by successive frames of N samples, wherein the samples of each frame are weighted by an analysis window (f_A) of Hamming, Hanning, Kaiser or similar type, a spectrum of the audio signal is calculated by transforming each frame of weighted samples in the frequency domain, and the spectrum of the audio signal is processed so as to deliver parameters (cx_{sup} , cx_{inf} , Emix) for synthesizing a signal derived from the analyzed audio signal, **characterized in that** the successive frames comprise an alternation of frames for which are delivered complete sets of synthesis parameters and of frames for which are delivered incomplete sets of synthesis parameters, and **in that** the successive frames for which complete sets of synthesis parameters are delivered exhibit mutual overlaps of less than N/2 samples.
2. The method as claimed in claim 1, wherein the incomplete sets of synthesis parameters include data ($icx[n-1/2]$) representing an error ($ecx[n-1/2]$) of interpolation of at least one of the synthesis parameters.
3. The method as claimed in claim 1, wherein the incomplete sets of synthesis parameters include data (iP) representing a filter (128) for interpolating at least one of the synthesis parameters.
4. The method as claimed in any one of claims 1 to 3, wherein the processing of the spectrum of the audio signal (x)

comprises an extraction of coding parameters (cx_sup, cx_inf, Emix) with a view to the transmission and/or the storage of the coded audio signal.

- 5
5. The method as claimed in any one of claims 1 to 3, wherein the processing of the spectrum of the audio signal (x) comprises a denoising by spectral subtraction.
6. An audio processing device, comprising analysis means which carry out a method as claimed in any one of claims 1 to 5.
- 10
7. A method of synthesizing an audio signal, wherein a set of successive overlapping frames of N samples of the audio signal which are weighted by an analysis window (f_A) is evaluated, by transforming in the time domain spectral estimates (Y) corresponding respectively to said frames, and the evaluated frames are combined to form the synthesized signal (x), **characterized in that**, for a subset of the evaluated frames, the spectral estimates are obtained by processing synthesis parameters (cx_sup_q, cx_inf_q, Emix) respectively associated with the frames of said subset while, for the frames which do not form part of the subset, the spectral estimates are obtained with an interpolation of a part at least of the synthesis parameters, **in that** the successive frames of said subset exhibit mutual time shifts of M samples, the number M being larger than N/2, while the successive frames of said set exhibit mutual time shifts of M/p samples, p being an integer larger than 1, **in that** each evaluated frame is modified by applying thereto a processing corresponding to a division by said analysis window (f_A) and to a multiplication by a synthesis window (f'_S), and the synthesized signal is formed as an overlap sum of the modified frames, and **in that**, the samples of a frame having ranks i numbered from 0 to N-1, the synthesis window $f'_S(i)$ has a support limited to the ranks i ranging from N/2 - M/p to N/2 + M/p and satisfies $f'_S(i) + f'_S(i + M/p) = A$ for $N/2 - M/p \leq i < N/2$, A being a positive constant.
- 15
8. The method as claimed in claim 7, wherein the synthesis window $f'_S(i)$ increases for i ranging from N/2 - M/p to N/2.
9. The method as claimed in claim 9, wherein the synthesis window $f'_S(i)$ for $N/2 - M/p \leq i < N/2 + M/p$ is a raised sinusoid.
- 20
10. The method as claimed in any one of claims 7 to 9, wherein data (icx_q[n-1/2]) representing an interpolation error (ecx_q[n-1/2]) are associated with the frames which do not form part of said subset, and are used to correct at least one of the interpolated synthesis parameters (cx_i[n-1/2]).
- 25
11. The method as claimed in any one of claims 7 to 9, wherein data (iP) representing an interpolator filter (128) are associated with the frames which do not form part of said subset, and are used to interpolate at least one of the synthesis parameters.
- 30
12. The method as claimed in any one of claims 7 to 11, wherein the synthesis parameters comprise cepstral coefficients (cx[n]) subjected to the interpolation.
- 35
13. An audio processing device, comprising synthesis means which carry out a method as claimed in any one of claims 7 to 12.
- 40
- 45
- 50
- 55

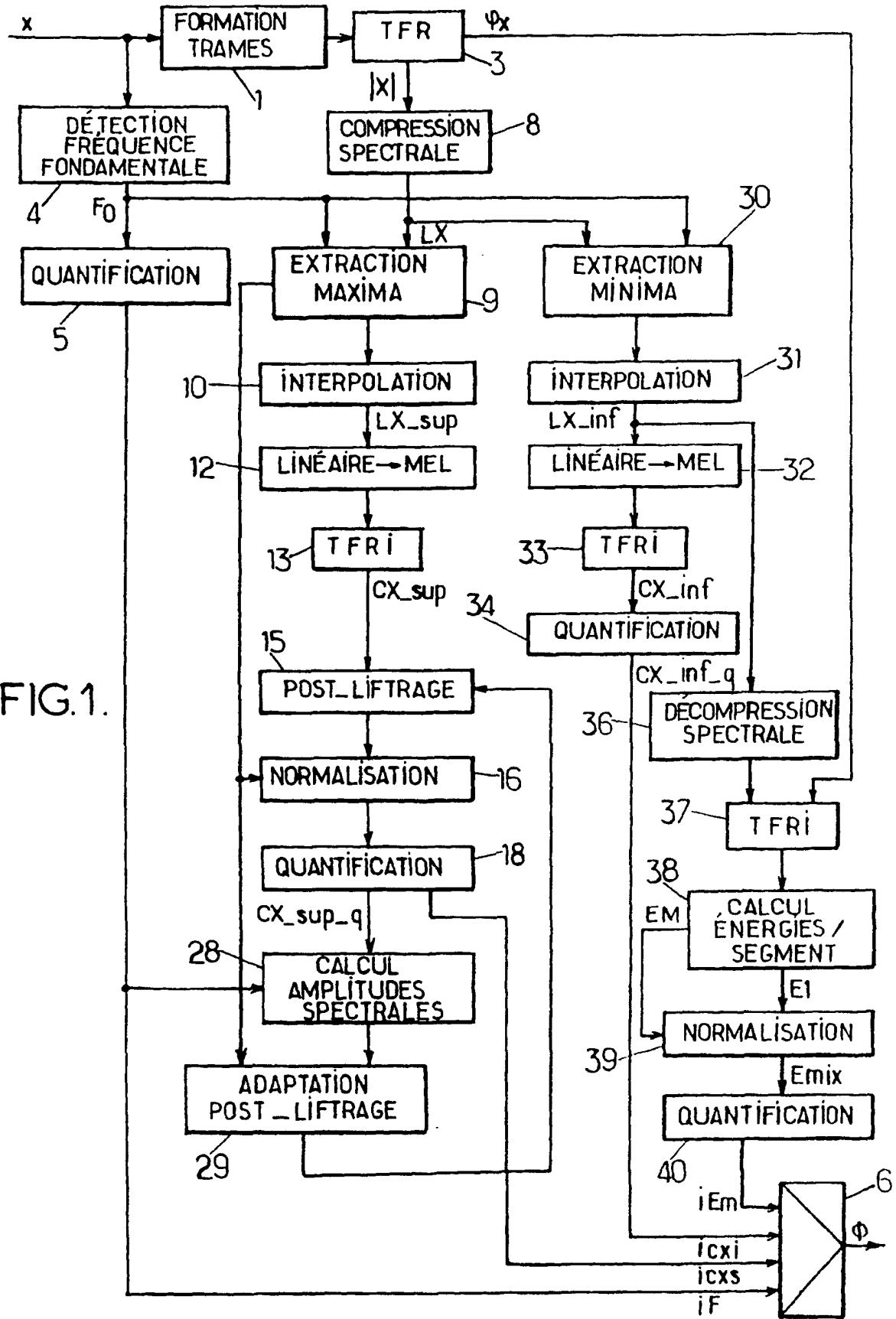
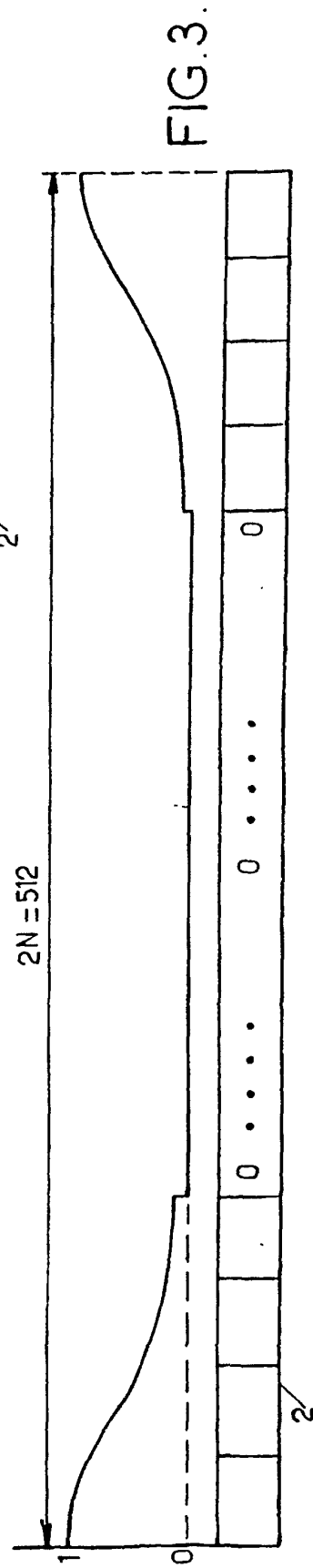
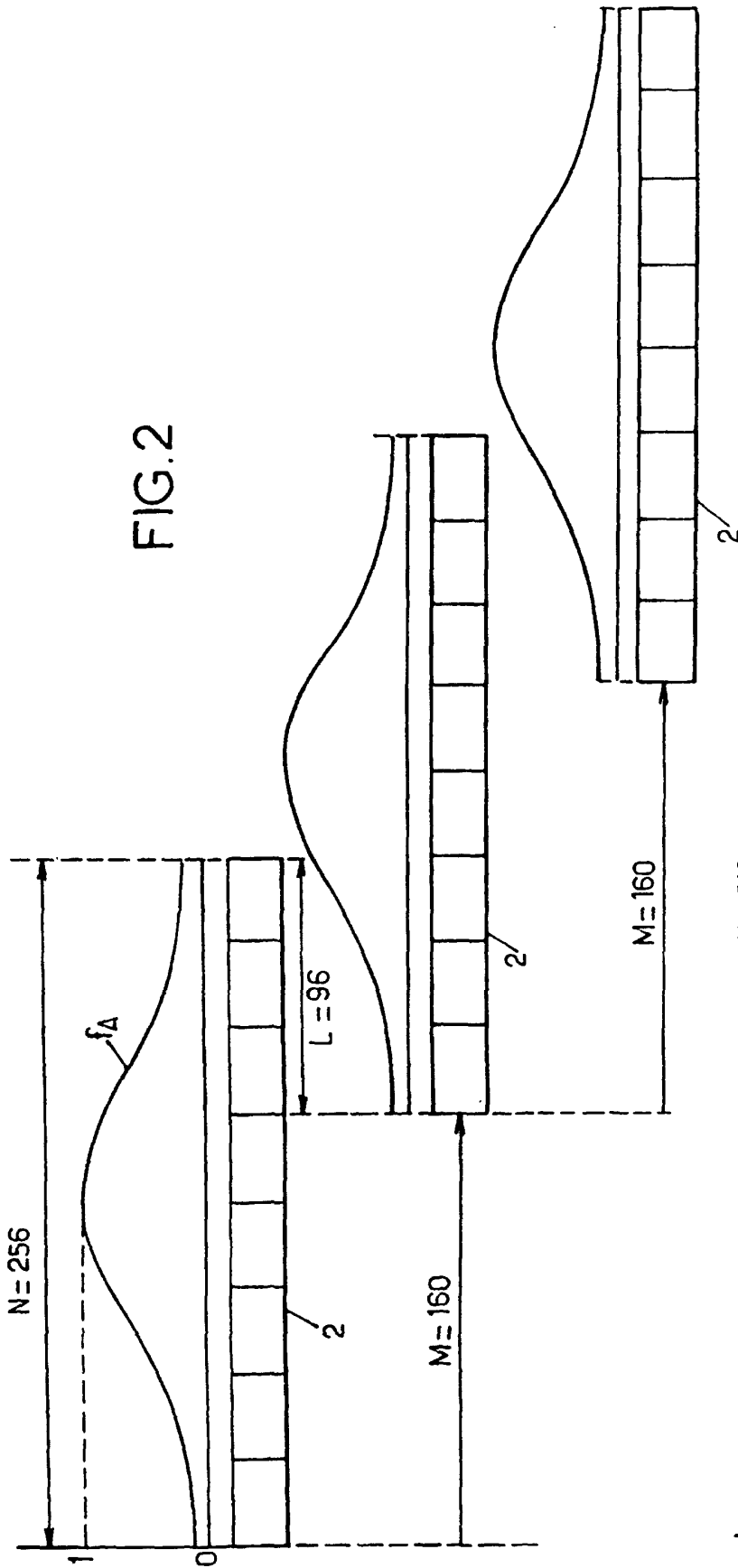


FIG. 1.



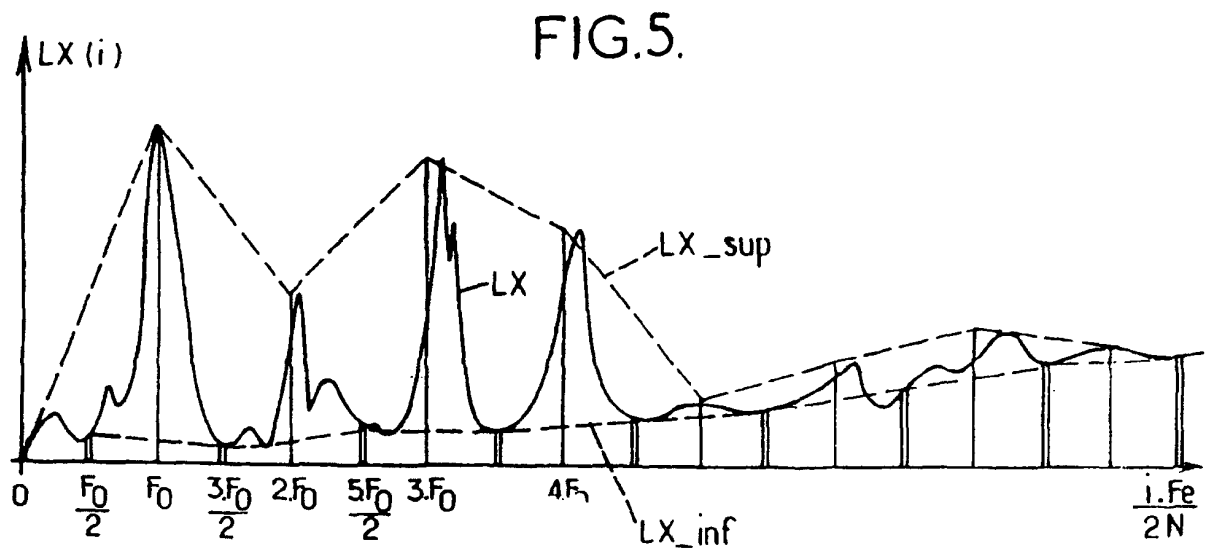
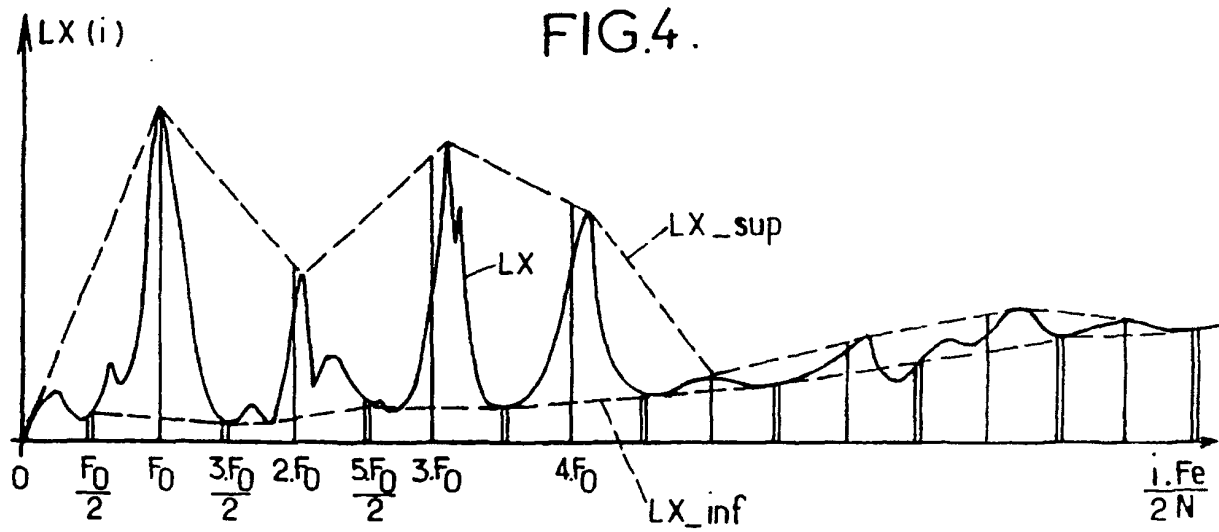


FIG. 6.

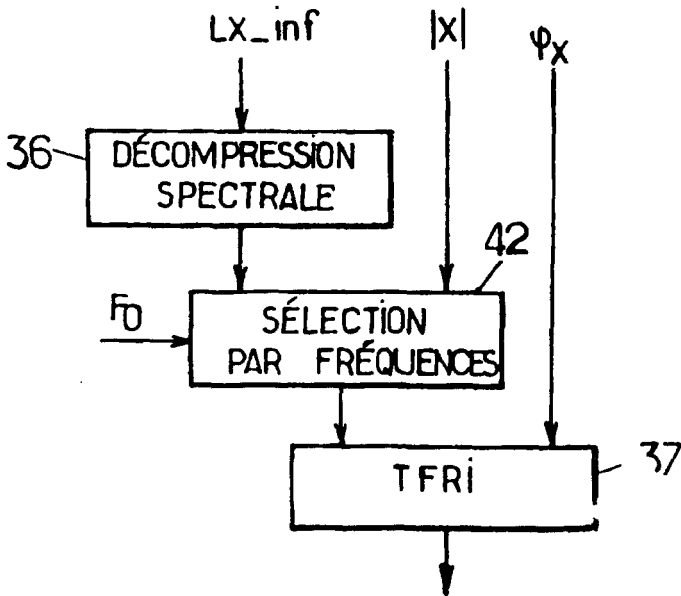
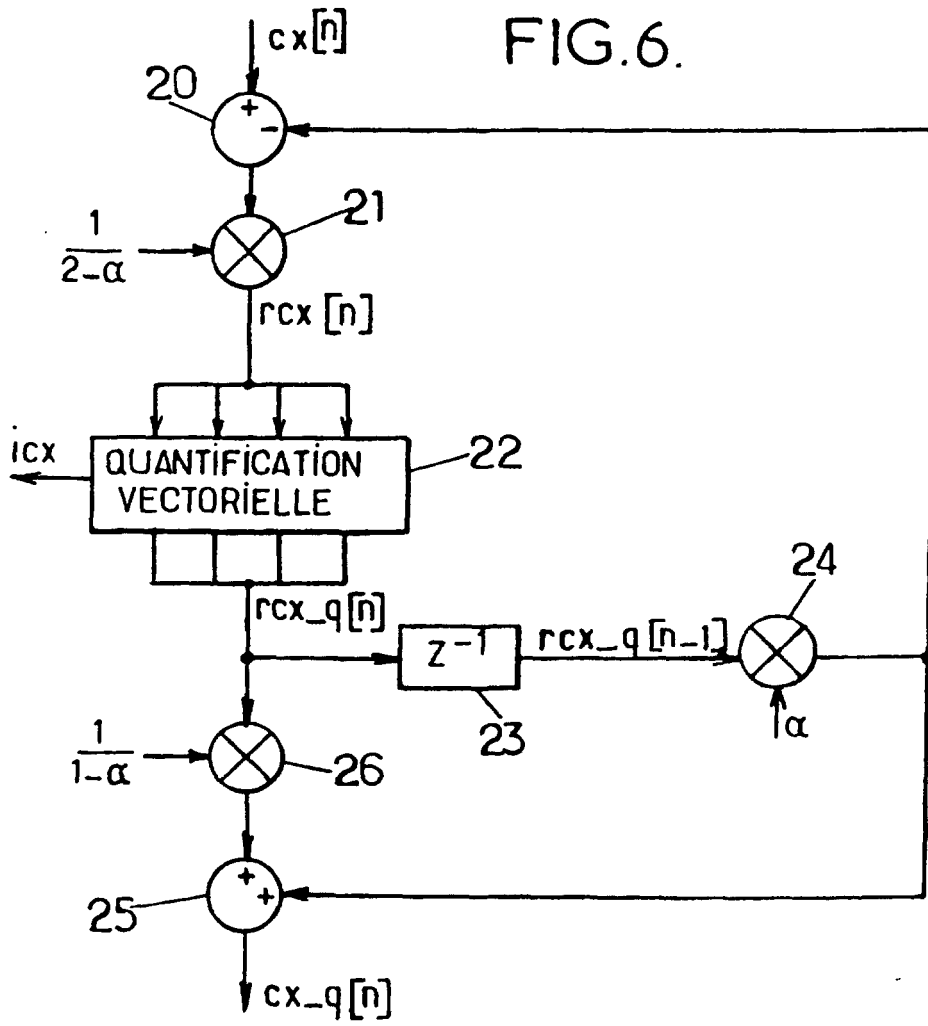


FIG. 7.

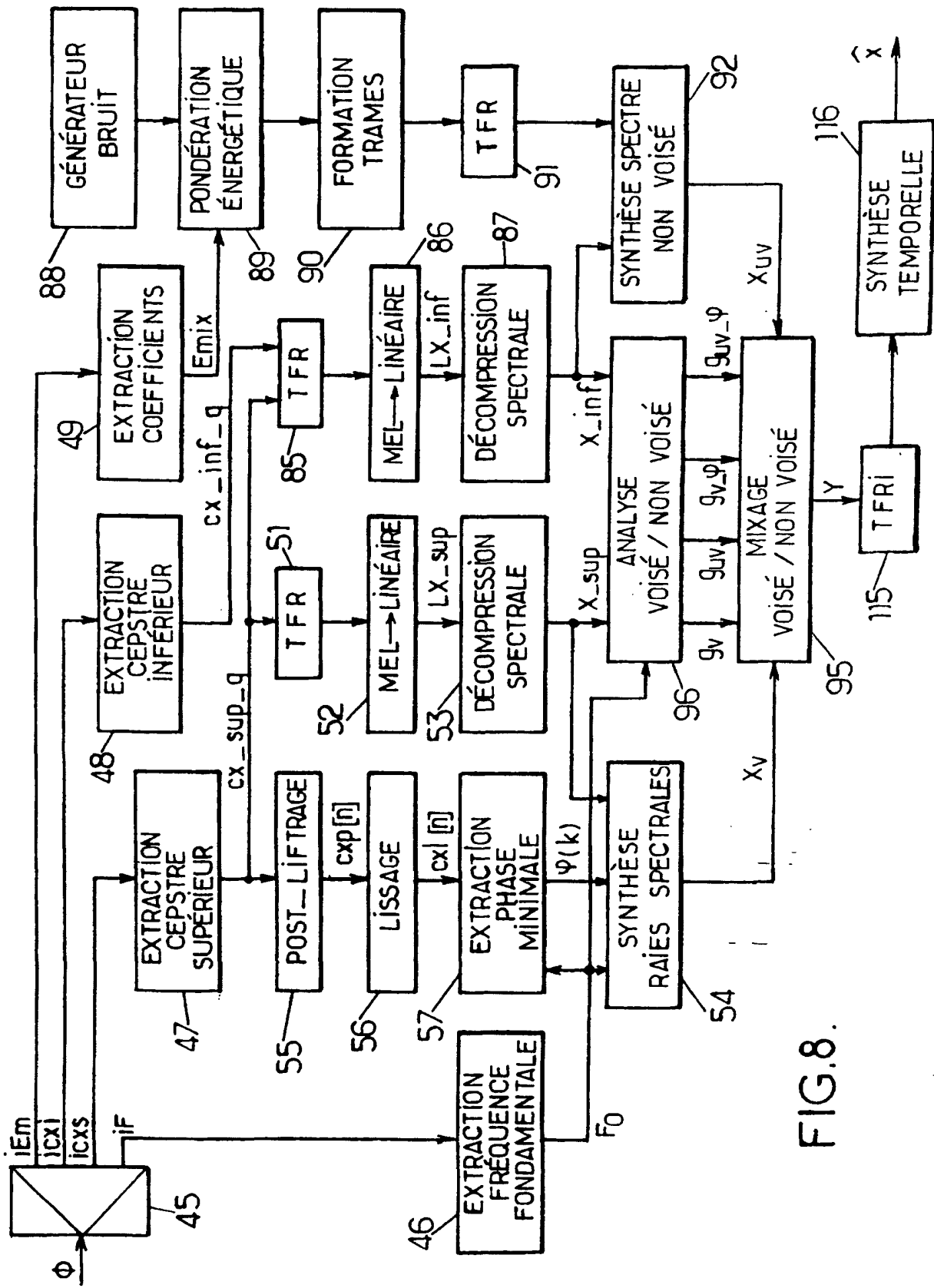


FIG. 8.

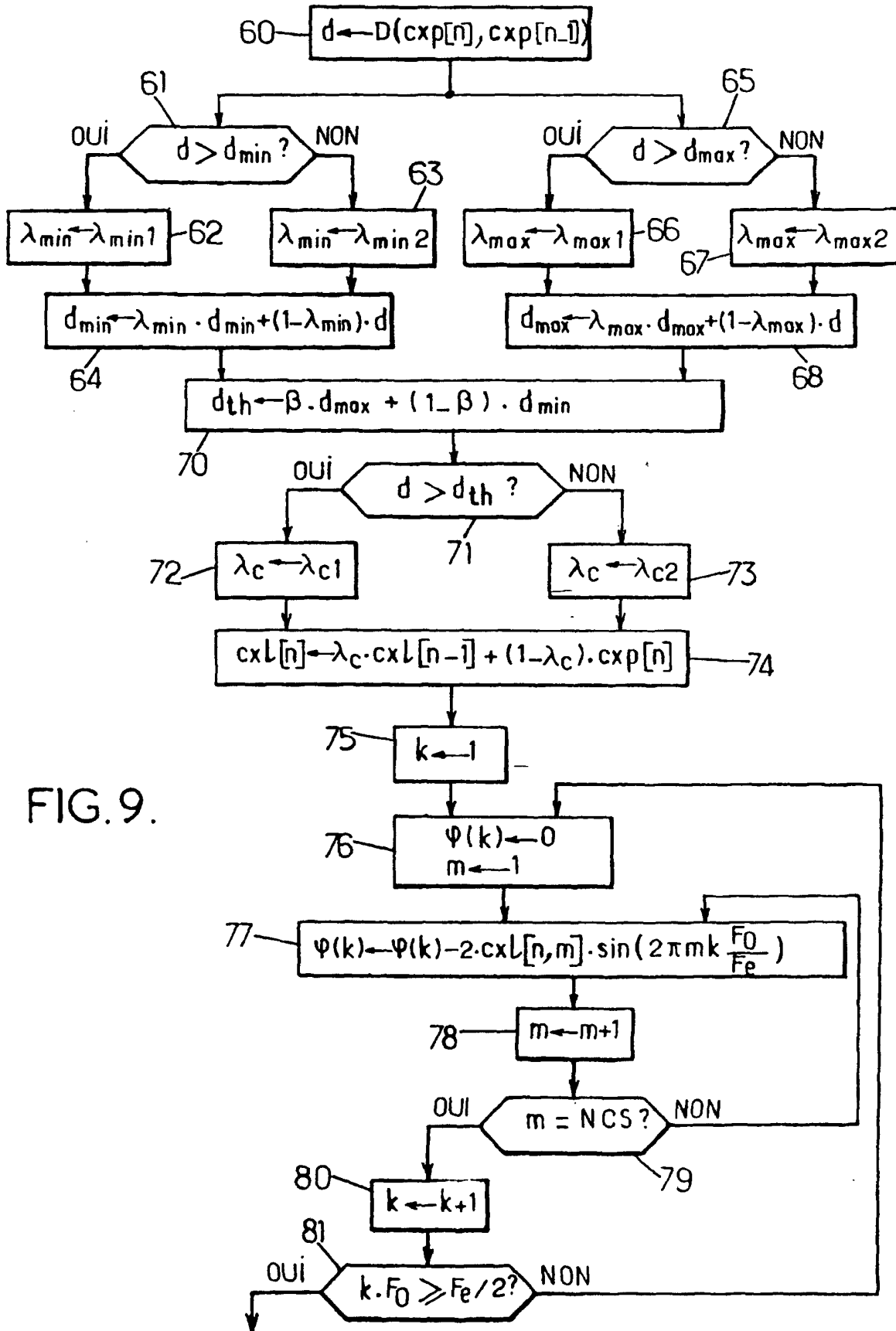


FIG. 9.

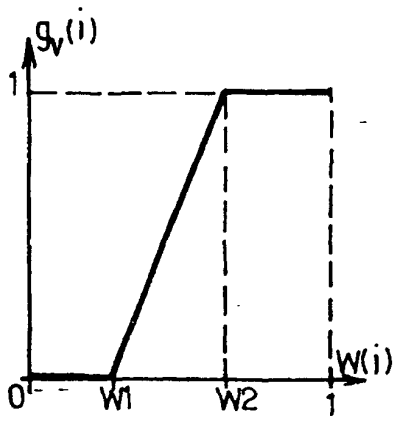
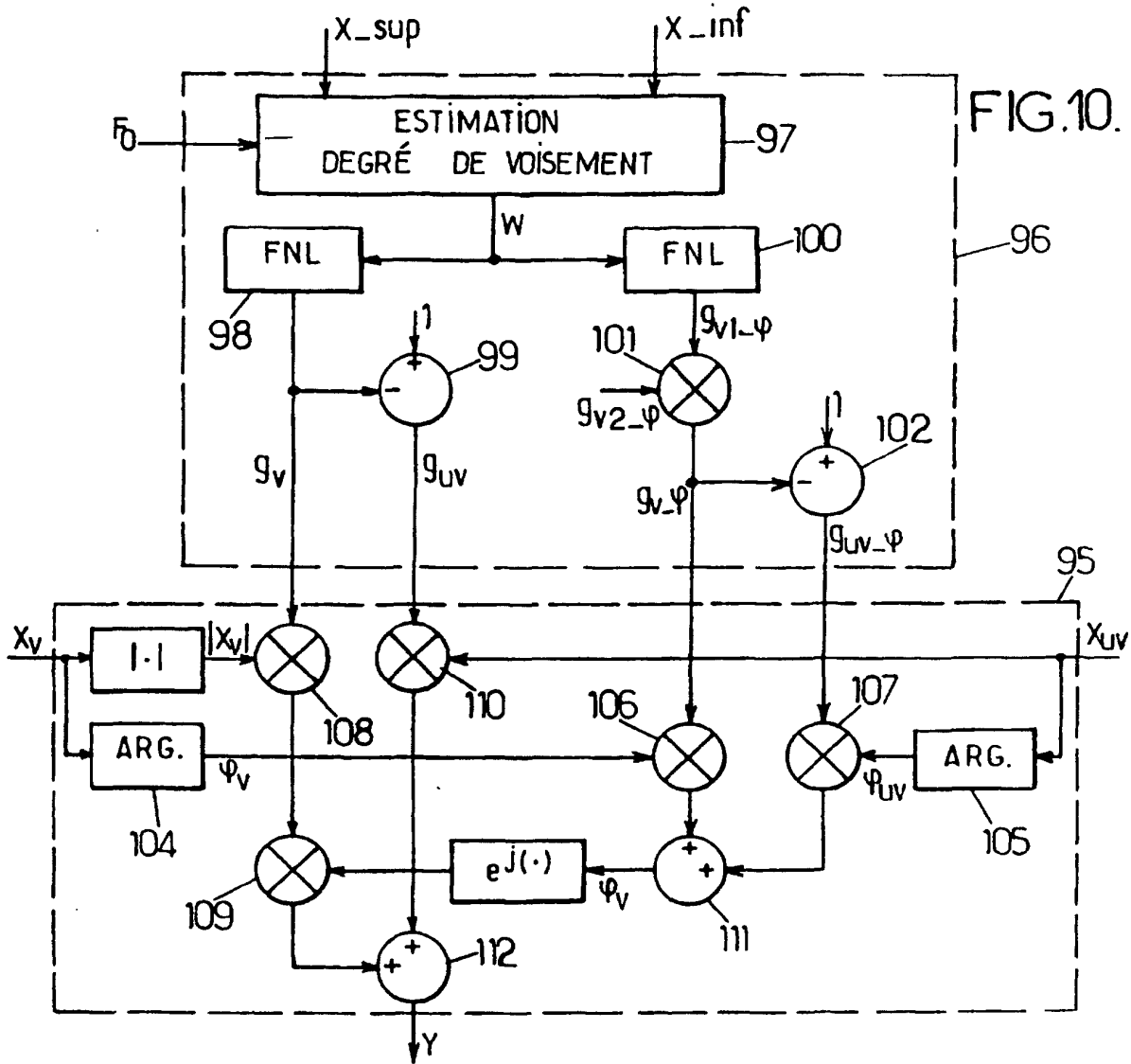


FIG. 11

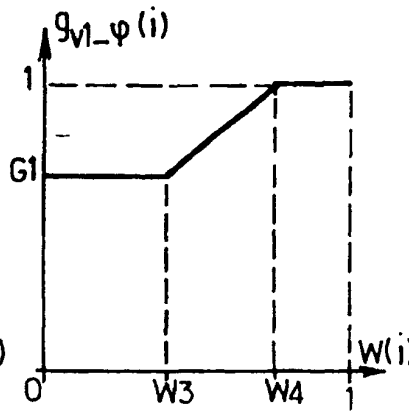


FIG. 12.

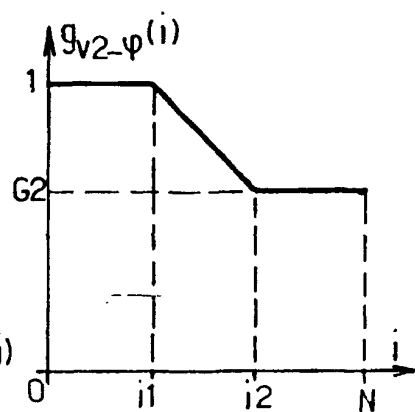


FIG. 13.

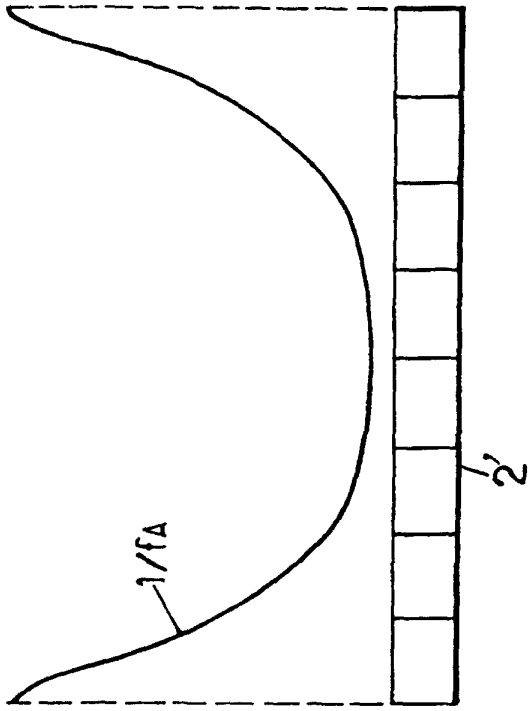


FIG. 14.

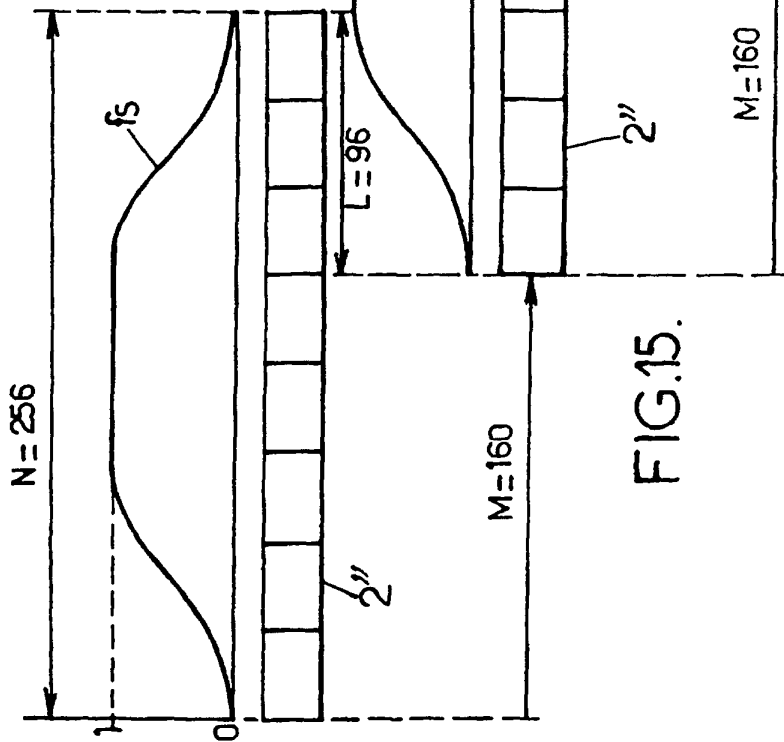


FIG. 15.

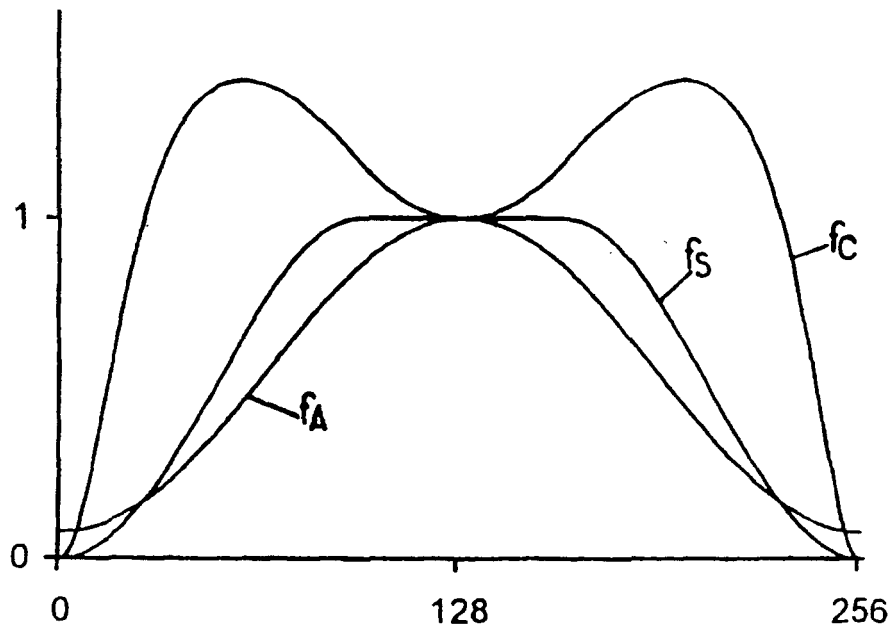


FIG.16.

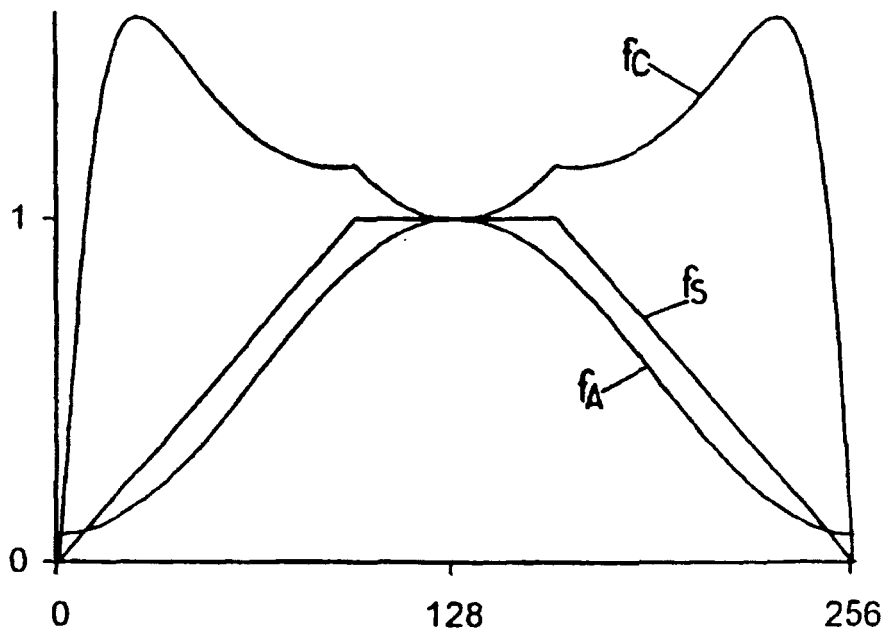


FIG.17.

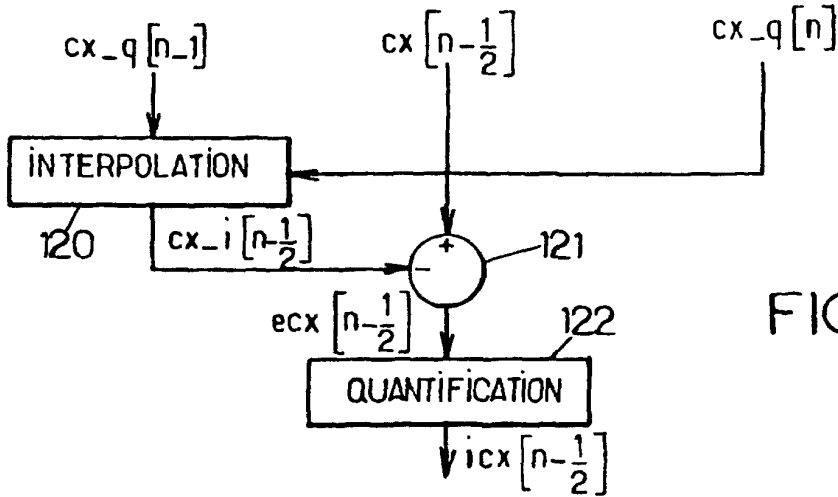


FIG. 18.

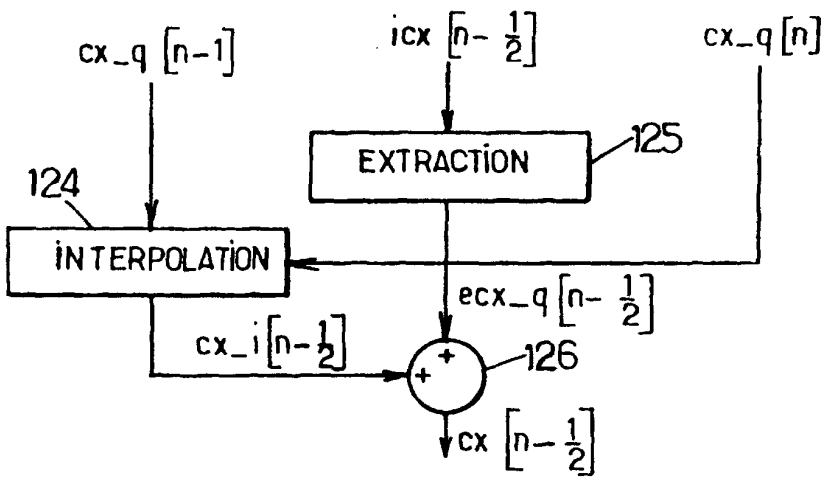


FIG. 19.

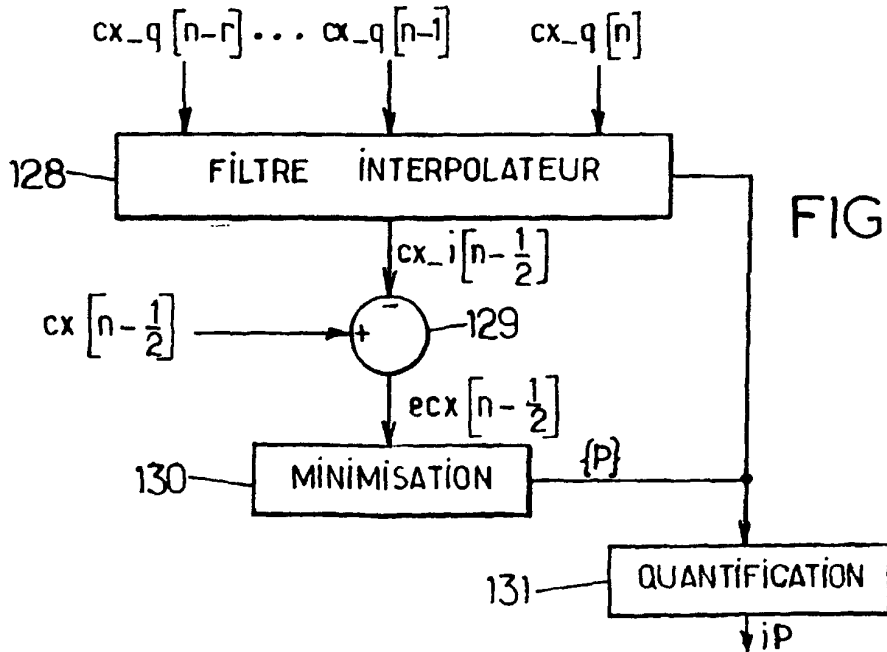


FIG. 20.

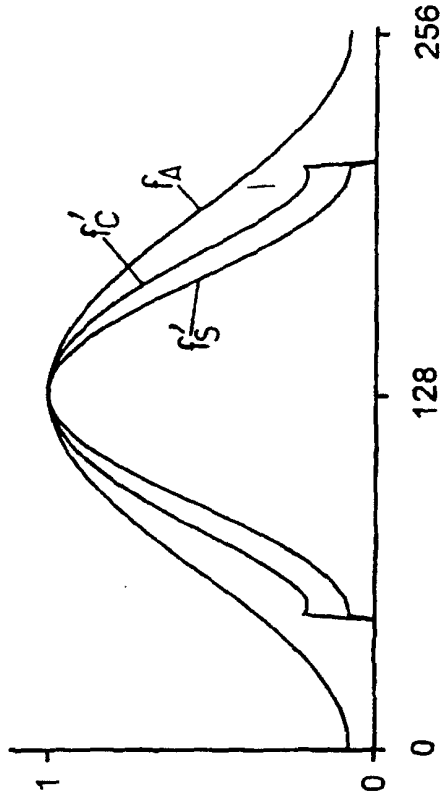


FIG.22.

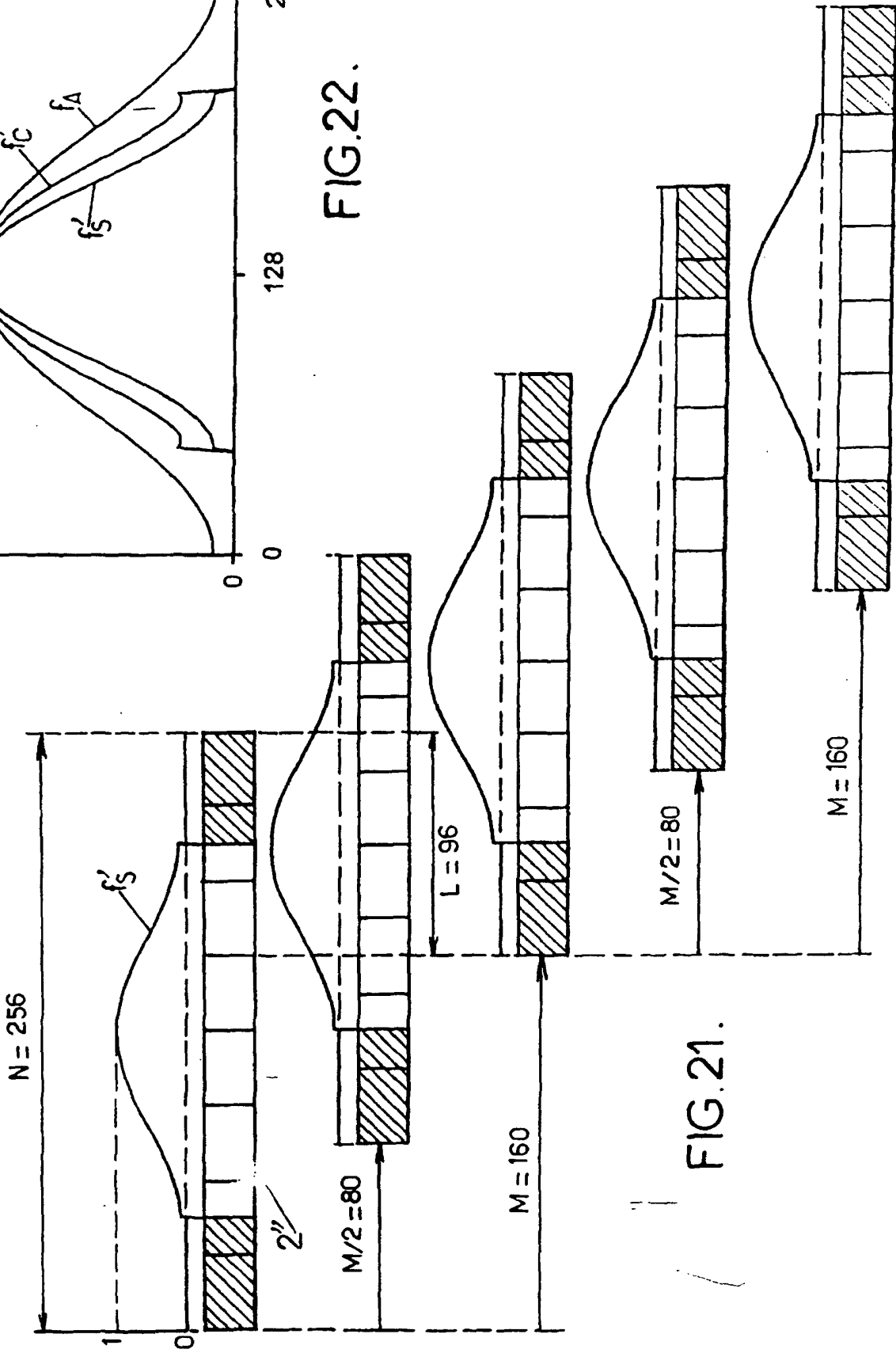


FIG.21.