(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 429 316 B1

(12) **FASCICULE DE BREVET EUROPEEN**

(45) Date de publication et mention de la délivrance du brevet:
12.01.2005 Bulletin 2005/02

(51) Int Cl.[7]: **G10L 21/02**, H03G 5/16, H04B 3/04

(21) Numéro de dépôt: **03027552.3**

(22) Date de dépôt: **01.12.2003**

(54) **Procédé et système de correction multi-références des déformations spectrales de la voix introduites par un réseau de communication**

Verfahren und Vorrichtung zur multi-referenz Korrektur der durch ein Kommunikationsnetzwerk verursachten spektralen Sprachverzerrungen

System and method for multi-referenced correction of spectral voice distortions introduced by a communication network

(84) Etats contractants désignés:
DE ES GB IT
Etats d'extension désignés:
AL LT LV MK

(30) Priorité: **11.12.2002 FR 0215618**

(43) Date de publication de la demande:
16.06.2004 Bulletin 2004/25

(73) Titulaire: **FRANCE TELECOM**
75015 Paris (FR)

(72) Inventeurs:
- **Mahe, Gael**
  22300 Lannion (FR)
- **Gilloire, André**
  22300 Lannion (FR)

(74) Mandataire: **Vannini, Torquato**
A.P.I. Conseil,
Rue M. Dormoy
64000 Pau (FR)

(56) Documents cités:
FR-A- 2 822 999

- **MAH G. ET GILLOIRE A.: "Correction of the voice timbre distortions on the telephone network" EUROSPEECH 2001, 3 septembre 2001 (2001-09-03), - 7 septembre 2001 (2001-09-07) pages 1867-1870, XP002248934 Aalborg, Denmark**
- **MAH GA L: "Correction centralisée des distortions spectrales de la parole sur les réseaux téléphoniques" 19 décembre 2002 (2002-12-19), UNIVERSIT DE RENNES 1 , RENNES, FRANCE , XP002248935 * page 3 - page 7 * * page 87 - page 116 ***

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).

EP 1 429 316 B1

Description

[0001] L'invention concerne un procédé de correction multi références des déformations spectrales de la voix introduites par un réseau de communication. Elle concerne également un système pour la mise en oeuvre du procédé.

[0002] La présente invention a pour but l'amélioration de la qualité de la parole transmise sur les réseaux de communication, en offrant des moyens pour corriger les déformations spectrales du signal de parole, déformations provoquées par différents maillons de la chaîne de transmission du réseau.

[0003] La description qui en est donnée dans la suite fait explicitement référence à la transmission de la parole sur lignes téléphoniques "classiques" (c'est-à-dire filaires), mais elle s'applique également à tout type de réseau de communication (fixe, mobile ou autre) introduisant des déformations spectrales dans le signal, les paramètres pris comme référence pour qualifier le réseau devant être modifiés en fonction du réseau.

[0004] On va rappeler dans ce qui suit les différentes déformations rencontrées dans le cas du Réseau Téléphonique Commuté (RTC).

1.1. Les dégradations du timbre de la voix sur le Réseau RTC :

[0005] La figure 1 représente un schéma d'une liaison RTC. La parole émise par un locuteur est transmise par un terminal émetteur 10, transportée par la ligne d'abonné 20, subit une conversion 30 analogique-numérique (loi A), est transmise par le réseau numérique 40, subit une conversion numérique (loi A)- analogique 50, est transmise par la liaison d'abonné 60, passe par le terminal récepteur 70 pour être enfin reçue par le destinataire.

[0006] Chaque locuteur est relié par une ligne analogique (paire torsadée) au central téléphonique le plus proche. Il s'agit d'une transmission analogique en bande de base référence 1 et 3 sur la figure 1. La liaison entre les centraux, emprunte un réseau entièrement numérique 40. Le spectre de la voix est affecté par deux types de distorsions lors de la transmission analogique du signal en bande de base.

[0007] Le premier type de distorsion est le filtrage passe-bande des terminaux et des points d'accès à la partie numérique du réseau. Les caractéristiques typiques de ce filtrage sont décrites par l'UIT-T sous le nom de "système de référence intermédiaire" (SRI) [UIT-T, Recommandation P.48, 1988]. Ces caractéristiques fréquentielles, issues de mesures réalisées dans les années 70, tendent cependant à devenir obsolètes. C'est pourquoi l'UIT-T préconise depuis 1996 d'utiliser un SRI "modifié" [UIT-T, Recommandation P.830, 1996], dont la caractéristique nominale est représentée sur la figure 2 pour la partie émission, et sur la figure 3 pour la partie réception. Entre 200 et 3400 Hz, la tolérance est de $\pm$ 2,5 dB ; en dessous de 200 Hz, la décroissance de la caractéristique du système global doit être d'au moins 15 dB par octave. On appelle respectivement, selon la terminologie de l'UIT-T, "système émetteur" et "système récepteur" les parties émission et réception du SRI.

[0008] La seconde distorsion affectant le spectre de la voix est l'atténuation des lignes d'abonné. Dans un modèle simple de la ligne analogique locale [donné dans une Note Technique CNET NT/LAA/ELR/289 par Cadoret, 1983], on considère que celle-ci introduit un affaiblissement du signal dont la valeur en dB dépend de sa longueur et est proportionnelle à la racine carrée de la fréquence. L'affaiblissement est de 3 dB à 800 Hz pour une ligne moyenne (environ 2 km), de 9,5 dB à 800 Hz pour les lignes les plus longues (jusqu'à 10 km). Selon ce modèle, l'affaiblissement d'une ligne, représenté sur la figure 4, a pour expression :

$$A_{dB}(f) = A_{dB}(800Hz).\sqrt{\frac{f}{800}} \qquad (0.1)$$

[0009] A ces distorsions s'ajoute le filtrage anti-repliement du codeur MIC (réf 30). Celui-ci est typiquement un filtre passe-bande 200-3400 Hz avec une réponse presque plate sur la bande passante et une forte atténuation en dehors de la bande, selon le gabarit de la figure 5 par exemple [National Semiconductor, août 1994 : Documentation technique « TP3054 ,TP3057].

[0010] Au final, la voix subit une distorsion spectrale telle que représentée sur la figure 6 pour les différentes combinaisons de trois types de ligne analogique en émission et en réception (soit 6 distorsions), sous l'hypothèse d'équipements respectant la caractéristique nominale du SRI modifié. La voix apparaît ainsi étouffée si une des lignes analogiques est longue et souffre dans tous les cas d'un manque de "présence" dû à l'affaiblissement des composantes basse fréquence.

1.2. Les dégradations du timbre de la voix sur le Réseau Numérique à Intégration de Services (RNIS) et le réseau mobile GSM.

[0011] Dans le RNIS et le réseau GSM, le signal est numérisé dès le terminal. Les seules parties analogiques sont

les transducteurs en émission et en réception associés à leurs chaînes d'amplification et de conditionnement respectives. L'UIT-T a défini des gabarits d'efficacité en fréquence à l'émission représentés sur la figure 7, et à la réception représenté sur la figure 8, valables à la fois pour les téléphones numériques filaires [UIT-T, Recommandation P.310, mai 2000] et les terminaux numériques mobiles ou sans fil [UIT-T, Recommandation P.313, septembre 1999].

[0012] Par ailleurs, pour les réseaux GSM, il est reconnu que le codage et le décodage modifient légèrement l'enveloppe spectrale du signal. Cette altération est représentée sur la figure 9 pour un bruit rose codé puis décodé en mode EFR (Enhanced Full Rate).

[0013] L'effet de ces filtrages sur le timbre est principalement un affaiblissement des composantes basse fréquence, moins marqué cependant que dans le cas du RTC.

[0014] L'invention concerne la correction de ces distorsions spectrales par un traitement centralisé, c'est-à-dire un dispositif installé dans la partie numérique du réseau, comme indiqué sur la figure 10 pour le RTC.

[0015] L'objectif d'une correction du timbre de la voix est que le timbre de la voix en réception soit le plus proche possible de celui de la voix émise par le locuteur, que l'on appellera voix originale.

2. État de la technique

[0016] Une compensation des distorsions spectrales introduites dans le signal de parole par les divers éléments de la liaison téléphonique est permise à ce jour par des dispositifs à base d'égalisation. Celle-ci peut être fixe ou s'adapter en fonction des conditions de transmission.

2.1 L'égalisation fixe

[0017] Des dispositifs d'égalisation centralisée ont été proposés dans les brevets US 5333195 [Duane O. Bowker] et US 5471527 [Helena S. Ho]. Ces égaliseurs sont des filtres fixes qui restaurent le niveau des basses fréquences atténuées par l'émetteur. Bowker propose par exemple un gain de 10 à 15 dB sur la bande 100-300 Hz. Ces méthodes présentent deux inconvénients :

* L'égaliseur ne compense que le filtrage de l'émetteur, de sorte qu'à la réception, les composantes basse-fréquence restent fortement affaiblies par le filtrage SRI de réception.
* Cette égalisation fixe compense des conditions de transmission (ligne et système d'émission) moyennes. Si les conditions réelles sont trop différentes (par exemple si les lignes analogiques sont longues) le dispositif ne corrige pas suffisamment le timbre, voire l'altère plus que la liaison sans égalisation.

2.2 L'égalisation adaptative.

[0018] Le dispositif décrit dans le brevet US 5915235 [Andrew P De Jaco] vise à corriger la réponse fréquentielle non idéale d'un transducteur de téléphone mobile. L'égaliseur est décrit comme étant placé entre le convertisseur analogique-numérique et le codeur CELP, mais peut être aussi bien dans le terminal que dans le réseau. Le principe de l'égalisation est de rapprocher le spectre du signal reçu d'un spectre idéal. Deux méthodes sont proposées.

[0019] La première méthode (illustrée par la figure 4 du brevet précité de De Jaco), consiste à calculer des coefficients d'autocorrélation à long terme RLT :

$$R_{LT}(n,i) = \alpha\, R_{LT}(n-1,i) + (1-\alpha)R(n,i), \qquad (0.2)$$

avec RLT(n,i) $i^{ième}$ coefficient d'autocorrélation à long terme à la nième trame, R(n,i) $i^{ième}$ coefficient d'autocorrélation spécifique à la $n^{ième}$ trame, et α constante de lissage fixée par exemple à 0,995. De ces coefficients sont dérivés les coefficients LPC à long terme, qui sont les coefficients d'un filtre blanchisseur. À la sortie de ce filtre, le signal est filtré par un filtre fixe qui lui imprime les caractéristiques spectrales à long terme idéales, ie celles qu'il aurait à la sortie d'un transducteur ayant la réponse fréquentielle idéale. Ces deux filtres sont complétés par un gain multiplicatif égal au rapport entre les énergies à long terme de l'entrée du blanchisseur et de la sortie du deuxième filtre.

[0020] La deuxième méthode, illustrée par la figure 5 du brevet précité de De Jaco, consiste à diviser le signal en sous-bandes, et, pour chaque sous-bande, appliquer un gain multiplicatif de manière à atteindre une énergie cible, ce gain étant défini comme le rapport entre l'énergie cible de la sous-bande et l'énergie à long terme (obtenue par un lissage de l'énergie instantanée) du signal dans cette sous-bande.

[0021] Ces deux méthodes présentent l'inconvénient de ne corriger que la réponse non idéale du système d'émission, et pas celle du système de réception.

[0022] Le dispositif du brevet US 5905969 [Chafik Mokbel] a pour objet de compenser le filtrage du système d'émission et de la ligne d'abonné pour améliorer la reconnaissance centralisée de la parole et/ou la qualité de la parole transmise. Comme le présente la figure 3a de Mokbel, le spectre du signal est divisé en 24 sous-bandes, et chaque énergie de sous-bande est multipliée par un gain adaptatif. L'adaptation du gain est réalisée selon l'algorithme du gradient stochastique, par minimisation de l'erreur quadratique, l'erreur étant définie comme la différence entre l'énergie de sous-bande et une énergie de référence définie pour chaque sous-bande. L'énergie de référence est modulée à chaque trame par l'énergie de la trame courante, de manière à respecter les variations naturelles de niveau à court terme du signal de parole. La convergence de l'algorithme permet d'obtenir en sortie les 24 signaux de sous-bande égalisés.

[0023] Si l'application visée est l'amélioration de la qualité vocale, le signal de parole égalisé est obtenu par transformée de Fourier inverse des énergies de sous-bande égalisées.

[0024] Le brevet de Mokbel, ne mentionne pas de résultats en termes d'amélioration de la qualité vocale, et reconnaît que la méthode est sous-optimale, en ce qu'elle réalise une convolution circulaire. Par ailleurs, il est douteux qu'un signal de parole puisse être reconstruit correctement par transformée de Fourier inverse d'énergies de bandes distribuées selon l'échelle MEL. Enfin, le dispositif décrit ne corrige pas le filtrage du système de réception et de la ligne analogique de réception.

[0025] La compensation de l'effet de ligne est réalisée dans la méthode « Mokbel », de soustraction cepstrale, dans le but d'améliorer la robustesse de la reconnaissance de la parole. On montre que le cepstre du canal de transmission peut être estimé par le cepstre moyen du signal reçu, celui-ci étant préalablement blanchi par un filtre de préaccentuation. Cette méthode permet une nette amélioration des performances des systèmes de reconnaissance, mais est considérée comme une méthode "off-line", 2 à 4 s étant nécessaires pour estimer le cepstre moyen.

[0026] 2.3 Un autre état de la technique combine une pré-égalisation fixe et une égalisation adaptée et a fait l'objet du dépôt d'une demande de brevet FR 2822999 par le Déposant. Le dispositif décrit vise à corriger le timbre de la parole en combinant deux filtres.

[0027] Un filtre fixe, appelé pré-égaliseur, compense les distorsions d'une liaison téléphonique moyenne, définie comme étant constituée de deux lignes d'abonné moyennes et de systèmes d'émission et de réception respectant les réponses fréquentielles nominales définies dans [UIT-T, Recommandation P.48,App.I,1988]. Sa réponse fréquentielle, sur la bande [Fc-3150 Hz], est l'inverse de la réponse globale .de la partie analogique de cette liaison moyenne, Fc étant la fréquence basse limite d'égalisation.

[0028] Cette pré-égalisation est complétée par un égaliseur adapté, qui adapte la correction de manière plus précise aux conditions réelles de transmission. La réponse fréquentielle de l'égaliseur adapté est donnée par :

$$|EQ(f)| = \frac{1}{|S_RX(f).L_RX(f)|} \sqrt{\frac{\gamma_{ref}(f)}{\gamma_x(f)}}, \qquad (0.3)$$

avec L_RX la réponse fréquentielle de la ligne de réception, S_RX la réponse fréquentielle du système de réception et $\gamma_x(f)$ le spectre à long terme de la sortie x du pré-égaliseur.

[0029] Le spectre à long terme est défini comme la moyenne temporelle des spectres à court terme des trames successives de signal ; $\gamma_{ref(f)}$, appelé spectre de référence, est le spectre moyen de la parole défini par l'UIT [UIT-T/P. 50/App. I, 1998], pris comme approximation du spectre à long terme original du locuteur. Du fait de cette approximation, la réponse fréquentielle de l'égaliseur adapté est très irrégulière et seule sa forme générale est pertinente. C'est pourquoi elle doit être lissée. L'égaliseur adapté étant réalisé sous la forme d'un filtre temporel RIF, ce lissage dans le domaine fréquentiel est obtenu par un étroit fenêtrage (symétrique) de la réponse impulsionnelle.

[0030] Cette méthode permet de restaurer un timbre proche de celui du signal original sur la bande d'égalisation [Fc-3150 Hz], mais :

- pour certains locuteurs, l'approximation de leur spectre à long terme original par le spectre de référence est très grossière, de sorte que l'égaliseur introduit une distorsion perceptible ;
- le fort lissage de la réponse fréquentielle de l'égaliseur, rendu nécessaire par l'erreur d'approximation, interdit de corriger des distorsions spectrales fines.

[0031] L'invention a pour but de remédier aux inconvénients de l'état de la technique. Elle a pour objet un procédé et un système pour améliorer la correction du timbre en réduisant l'erreur d'approximation du spectre à long terme original des locuteurs.

[0032] A cet effet, il est proposé de classer les locuteurs selon leur spectre à long terme et d'approcher celui-ci non plus par un spectre de référence unique mais par un spectre de référence par classe. Le procédé proposé permet de réaliser un traitement d'égalisation apte à déterminer la classe du locuteur et à égaliser suivant le spectre de référence

de la classe. Cette réduction de l'erreur d'approximation permet de lisser moins fortement la réponse fréquentielle de l'égaliseur adapté, le rendant apte à corriger des distorsions spectrales plus fines.

[0033] La présente invention a plus particulièrement pour objet un procédé de correction des déformations spectrales de la voix, introduites par un réseau de communication, comprenant une opération d'égalisation sur une bande de fréquence [F1-F2], adaptée à la distorsion réelle de la chaîne de transmission, cette opération étant réalisée au moyen d'un filtre numérique ayant une réponse fréquentielle fonction du rapport entre un spectre de référence et un spectre correspondant au spectre à long terme du signal voix des locuteurs, principalement caractérisé en ce qu'il comprend :

* Préalablement à l'opération d'égalisation du signal voix d'un locuteur en communication :
  - La constitution de classes de locuteurs avec une référence de voix par classe,
* Puis, pour un locuteur donné en communication :
  - Le classement de ce locuteur c'est-à-dire son affectation à une classe à partir de critères de classement prédéfinis pour lui faire correspondre une référence de voix qui est la plus proche de la sienne,
  - L'égalisation du signal numérisé de la voix du locuteur opérée avec comme spectre de référence, la référence de voix de la classe à laquelle ledit locuteur a été affecté.

[0034] Selon une autre caractéristique, la constitution de classes de locuteurs comprend :

- Le choix d'un corpus de N locuteurs enregistrés dans des conditions non dégradées et la détermination de leur spectre de fréquence à long terme,
- La classification des locuteurs du corpus selon leur cepstre partiel c'est à dire le cepstre calculé à partir du spectre à long terme restreint à la bande d'égalisation [F1-F2] et en appliquant un critère de classification prédéfini sur ces cepstres pour obtenir K classes,
- Le calcul du spectre de référence associé à classe de manière à obtenir une référence de voix correspondant à chacune des classes.

[0035] Selon une autre caractéristique, le spectre de référence sur la bande de fréquences d'égalisation [F1-F2], associé à chaque classe, est calculé par transformée de Fourier du centre de la classe défini par son cesptre partiel.

[0036] Selon une autre caractéristique, la classification d'un locuteur comprend :

- L'utilisation du pitch moyen du signal voix et du cepstre partiel de ce signal comme paramètres de classement,
- L'application d'une fonction discriminante à ces paramètres pour classer ledit locuteur.

[0037] Selon l'invention le procédé comprend en outre une étape de pré-égalisation du signal numérique par un filtre fixe ayant une réponse fréquentielle dans la bande de fréquence [F1-F2], correspondant à l'inverse d'une déformation spectrale de référence introduite par la liaison téléphonique.

[0038] Selon une autre caractéristique, l'égalisation du signal numérisé de la voix d'un locuteur comprend :

- la détection d'une activité vocale sur la ligne pour déclencher un enchaînement de traitements comprenant le calcul du spectre à long terme, la classification du locuteur, le calcul du module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] et le calcul des coefficients du filtre numérique différenciés selon la classe du locuteur, à partir de ce module,
- la commande du filtre avec les coefficients obtenus,
- le filtrage du signal sortant du pré-égaliseur par ledit filtre.

[0039] Selon une autre caractéristique, le calcul du module [EQ] de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] est réalisé par la mise en oeuvre de la relation suivante :

$$|EQ(f)| = \frac{1}{|S_RX(f).L_RX(f)|} \sqrt{\frac{\gamma_{ref}(f)}{\gamma_x(f)}}, \qquad (0.3)$$

dans laquelle $\gamma_{ref}(f)$ est le spectre de référence de la classe à laquelle appartient ledit locuteur,

et dans laquelle L_RX est la réponse fréquentielle de la ligne de réception, S_RX la réponse fréquentielle du système de réception et $\gamma x(f)$ le spectre à long terme du signal x d'entrée du filtre.

[0040] Selon une variante, le calcul du module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] est réalisé par la mise en oeuvre de la relation suivante :

$$C^{p}_{eq} = C^{p}_{ref} - C^{p}_{x} - C^{p}_{s_rx} - C^{p}_{l_rx}, \quad (0.13)$$

dans laquelle C^{p}_{eq}, C^{p}_{x}, $C^{p}_{s_rx}$ et $C^{p}_{l_rx}$ sont les cepstres partiels respectifs de l'égaliseur adapté, du signal d'entrée x du filtre égaliseur, du système de réception et de la ligne de réception, C^{p}_{ref} étant le cepstre partiel de référence, centre de la classe du locuteur. Le module [EQ] restreint à la bande F1-F2 est alors calculé par Transformée de Fourier discrète de C^{p}_{EQ}.

[0041] L'invention a également pour objet un système de correction des déformations spectrales de la voix, introduites par un réseau de communication, comprenant des moyens d'égalisation adaptée dans une bande de fréquence [F1-F2] qui comportent un filtre numérique dont la réponse fréquentielle est fonction du rapport entre un spectre de référence et un spectre correspondant au spectre à long terme d'un signal voix, principalement caractérisé en ce que ces moyens comprennent en outre :

- des moyens de traitements du signal pour le calcul des coefficients du filtre numérique munis :
  - d'un bloc de traitement du signal pour calculer le module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] selon la relation suivante :

$$|EQ(f)| = \frac{1}{|S_RX(f).L_RX(f)|} \sqrt{\frac{\gamma_{ref}(f)}{\gamma_{x}(f)}}, \quad (0.3)$$

dans laquelle $\gamma_{ref}(f)$ est le spectre de référence, qui peut être différent d'un locuteur à un autre et qui correspond à une référence de classe prédéterminée à laquelle appartient ledit locuteur, et dans laquelle L_RX est la réponse fréquentielle de la ligne de réception, S_RX la réponse fréquentielle du système de réception et $\gamma_{x}(f)$ le spectre à long terme du signal x d'entrée du filtre;

  - d'un deuxième bloc de traitements pour le calcul de la réponse impulsionnelle à partir du module de réponse fréquentielle ainsi calculé, afin de déterminer les coefficients du filtre différenciés selon la classe du locuteur.

[0042] Selon une autre caractéristique, le premier bloc de traitement comprend des moyens de calcul du cepstre partiel du filtre égaliseur selon la relation :

$$C^{p}_{eq} = C^{p}_{ref} - C^{p}_{x} - C^{p}_{s_rx} - C^{p}_{l_rx}, \quad (0.13)$$

dans laquelle C^{p}_{eq}, C^{p}_{x}, $C^{p}_{s_rx}$ et $C^{p}_{l_rx}$ sont les cepstres partiels respectifs de l'égaliseur adapté, du signal d'entrée x du filtre égaliseur, du système de réception et de la ligne de réception, C^{P}_{ref} étant le cepstre partiel de référence, centre de la classe du locuteur, le module de [EQ] restreint à la bande F1-F2 est alors calculé par Transformée de Fourier Discrète de C^{P}_{eq}.

[0043] Selon une autre caractéristique, le premier bloc de traitements comprend un sous-ensemble pour calculer les coefficients du cepstre partiel d'un locuteur en communication et un deuxième sous-ensemble pour opérer le classement de ce locuteur, ce deuxième sous-ensemble comprenant un bloc de calcul du pitch F_0, un bloc d'estimation du pitch moyen à partir du pitch calculé F_0, et un bloc de classement appliquant une fonction discriminante sur le vecteur x ayant pour composantes le pitch moyen et les coefficients du cepstre partiel pour classer ledit locuteur.

[0044] Selon l'invention, le système comprend en outre un pré-égaliseur, le signal égalisé à partir de spectres de référence différenciés selon la classe du locuteur étant le signal x de sortie du pré-égaliseur.

[0045] D'autres particularités et avantages de l'invention apparaîtront clairement dans la description suivante, qui est donnée à titre d'exemple illustratif et non limitatif et qui est faite en regard des figures annexées qui représentent :

- La figure 1, une liaison téléphonique schématisée pour un réseau téléphonique commuté (RTC),
- La figure 2, la courbe de réponse fréquentielle en émission du système de référence intermédiaire SRI modifié,
- La figure 3, la courbe de réponse fréquentielle en réception du système de référence intermédiaire SRI modifié,
- La figure 4, la réponse en fréquence des lignes d'abonnés selon leur longueur,
- La figure 5,le gabarit du filtre anti-repliement du codeur MIC,

- La figure 6, les distorsions spectrales subies par la parole sur le réseau téléphonique commuté avec SRI moyen et différentes combinaisons de lignes analogiques,
- La figure 7, le gabarit à l'émission pour les terminaux numériques,
- La figure 8, le gabarit à la réception pour les terminaux numériques,
- La figure 9, la distorsion spectrale introduite par le codage-décodage GSM en mode EFR (Enhanced Full Rate)
- La figure 10, le schéma d'un réseau de communication avec un système de correction des distorsions de la parole,
- La figure 11, les étapes de calcul du cepstre partiel,
- La figure 12, la classification des cepstres partiels selon le critère de la variance,
- Les figures 13a et 13b, les spectres à long terme correspondant aux centres des classes de locuteurs respectivement pour hommes et femmes,
- La figure 14, les caractéristiques fréquentielles des filtrages appliqués au corpus pour définir le corpus d'apprentissage,
- La figure 15, la réponse fréquentielle du pré-égaliseur pour différentes fréquences Fc,
- La figure 16, le schéma de réalisation du système de correction par égalisation différenciée par classe de locuteurs,
- La figure 17, une variante d'exécution du système selon la figure 16.

[0046] Dans toute la suite les mêmes références portées sur les dessins correspondent aux mêmes éléments.

[0047] La description qui va suivre va d'abord présenter l'étape préalable de classification d'un corpus de locuteurs selon leur spectre à long terme. Cette étape permet de définir K classes et une référence par classe.

[0048] Un enchaînement de traitements permet de traiter le signal de parole (dès la détection d'une activité vocale par le système) de chaque locuteur pour d'une part classer les locuteurs c'est à dire les affecter à une classe selon des critères prédéterminés et pour d'autre part corriger la voix en utilisant la référence de la classe du locuteur.

[0049] Etape préalable de classification des locuteurs.

* Choix du corpus de définition des classes.

[0050] Le spectre de référence étant une approximation du spectre à long terme original des locuteurs, la définition des classes de locuteurs et de leurs spectres de référence respectifs nécessite de disposer d'un corpus de locuteurs enregistrés dans des conditions non dégradées. En particulier, le spectre à long terme d'un locuteur mesuré sur cet enregistrement doit pouvoir être considéré comme son spectre original, i.e. celui de sa voix à l'extrémité d'émission d'une liaison téléphonique.

Définition de l'individu : le cepstre partiel.

[0051] Le traitement proposé permet de disposer, dans chaque classe, d'un spectre de référence le plus proche possible du spectre à long terme de chaque membre de la classe. Cependant, seule la partie du spectre comprise dans la bande d'égalisation F1-F2 est prise en compte dans le traitement d'égalisation adaptée. Les classes sont donc constituées selon le spectre à long terme restreint à cette bande.

[0052] D'autre part, la comparaison entre deux spectres est effectuée à un faible niveau de résolution spectrale, de manière à ne refléter que l'enveloppe spectrale. C'est pourquoi, on se place de préférence, dans l'espace des premiers coefficients cepstraux d'ordre supérieur à 0 (le coefficient d'ordre 0 représentant l'énergie), le choix du nombre de coefficients dépendant de la résolution spectrale souhaitée.

[0053] On détermine donc dans le traitement, le "cepstre partiel à long terme", que l'on note Cp, comme la représentation cepstrale du spectre à long terme restreint à une bande de fréquence. Si l'on note k1 et k2 les indices de fréquence correspondant respectivement aux fréquences F1 et F2, et γ le spectre à long terme de la parole, le cepstre partiel est défini par la relation :

$$\tilde{C^P} = \mathrm{TFD}^{-1}\left(10\log\left(\gamma\left(k_1 \ldots k_2\right) \circ \gamma\left(k_2 - 1 \ldots k_1 + 1\right)\right)\right) \qquad (0.4)$$

où $\circ$ désigne l'opération de concaténation.

[0054] La (TFD) Transformée de Fourier Discrète Inverse est calculée par exemple par IFFT après interpolation des échantillons du spectre tronqué de manière à atteindre un nombre d'échantillons puissance de 2. Par exemple, en choisissant la bande d'égalisation 187-3187 Hz, correspondant aux indices fréquentiels 5 à 101 pour une représentation du spectre (symétrisé) sur 256 points (de 0 à 255), l'interpolation se fait simplement en intercalant une raie fréquentielle (interpolée linéairement) toutes les trois raies dans le spectre restreint à 187-3187 Hz.

[0055] Les étapes du calcul du cepstre partiel sont représentées sur la figure 11.

[0056] Pour que les coefficients cepstraux reflètent l'enveloppe spectrale mais pas l'influence de la structure harmonique du spectre de la parole sur les spectres à long terme, on ne conserve pas les coefficients d'ordre élevé. Les locuteurs à classer sont donc représentés par les coefficients d'ordres 1 à L de leur cepstre partiel à long terme, L valant typiquement 20.

* La Classification.

[0057] Les classes sont constituées par exemple de manière non-supervisée, selon une classification hiérarchique ascendante.

[0058] Celle-ci consiste à créer, à partir de N individus disjoints, une hiérarchie de partitions selon le processus suivant : à chaque étape, on agrège les deux éléments les plus proches, un élément étant soit un individu non agrégé, soit un agrégat d'individus constitué lors d'une précédente étape. La proximité entre deux éléments est déterminée par une mesure de dissimilarité que l'on appelle distance. Le processus se poursuit jusqu'à l'agrégation de toute la population. La hiérarchie de partitions ainsi créée peut se représenter sous la forme d'un arbre comme celui de la figure 12, contenant N-1 partitions imbriquées. Chaque coupure de l'arbre fournit une partition, d'autant plus fine que l'on coupe bas.

[0059] Dans ce type de classification, on choisit comme mesure de distance entre deux éléments, la variation d'inertie intra-classes résultant de leur agrégation. Une partition est en effet d'autant meilleure que les classes créées sont homogènes, c'est-à-dire que l'inertie intra-classes est faible. Dans le cas d'un nuage de points xi de masses respectives mi, répartis en classes q de centres de gravité respectifs gq, l'inertie intra-classes est définie par :

$$I_{\text{intra}} = \sum_{q} \sum_{i \in q} m_i \left\| x_i - g_q \right\|^2 . \qquad (0.5)$$

[0060] L'inertie intra-classes, nulle à l'étape initiale de l'algorithme de calcul, s'accroît inévitablement à chaque agrégation.

[0061] On utilise de préférence le principe connu de l'agrégation selon la variance. Selon ce principe on recherche, à chaque étape de l'algorithme mis en oeuvre, les deux éléments dont l'agrégation produit l'augmentation d'inertie intra-classes la plus faible.

[0062] La partition ainsi obtenue est améliorée par une procédure d'agrégation autour des centres mobiles, qui permet de réduire la variance intra-classes.

[0063] Le spectre de référence, sur la bande F1-F2, associé à chaque classe est calculé par transformée de Fourier du centre de la classe.

* Exemple de classification

[0064] Le traitement décrit ci-dessus est appliqué à un corpus de 63 locuteurs. L'arbre de classification du corpus est représenté sur la figure 12. Dans cette représentation, la hauteur d'un segment horizontal agrégeant deux éléments est choisie proportionnelle à leur distance, ce qui permet de visualiser la proximité des éléments regroupés dans une même classe. Cette représentation facilite le choix du niveau de coupure de l'arbre, et donc des classes retenues. La coupure doit être faite au-dessus des agrégations de niveau faible, qui regroupent des individus proches, et en dessous des agrégations de niveau élevé, qui associent des groupes d'individus bien distincts.

[0065] De cette manière, on obtient de façon nette quatre classes (K= 4). Ces classes sont assez homogènes du point de vue du sexe des locuteurs, et une coupure de l'arbre en deux classes fait apparaître à peu près une classe hommes et une classe femmes.

[0066] La consolidation de cette partition par une procédure d'agrégation autour des centres mobiles aboutit à quatre classes de cardinaux 11, 18, 18 et 16, plus homogènes que précédemment du point de vue du sexe : seuls un homme et deux femmes sont affectés à des classes ne correspondant par à leur sexe.

[0067] Les spectres restreints à la bande 187-3187 Hz correspondant aux centres de ces classes sont représentés sur les figures 13a et 13b pour les classes hommes et femmes ainsi que pour leurs sous-classes respectives. Ces spectres, résultats de la classification, sont utilisés comme référence multiple par l'égaliseur adapté.

* Utilisation de critères de classement des locuteurs.

[0068] Les classes de locuteurs étant définies, le traitement prévoit l'utilisation de paramètres et des critères pour affecter un locuteur à l'une ou l'autre des classes.

[0069] Cette affectation n'est pas réalisée simplement selon la proximité du cepstre partiel avec un des centres de classes, puisque ce cepstre est dévié par la partie de la liaison téléphonique en amont de l'égaliseur.

[0070] Il est proposé avantageusement d'utiliser des critères de classement robustes à cette déviation. Cette robustesse est assurée à la fois par le choix des paramètres de classement et par celui du corpus d'apprentissage des critères de classement.

* De préférence on utilise les paramètres de classement : pitch moyen et cepstre partiel.

[0071] Les classes précédemment définies sont homogènes du point de vue du sexe. Le pitch moyen étant à la fois assez discriminant pour un classement homme / femme et insensible aux distorsions spectrales induites par une liaison téléphonique, il est donc utilisé comme paramètre de classement, conjointement avec le cepstre partiel.

* Choix du corpus d'apprentissage des critères de classement.

[0072] On applique à ces paramètres une technique de discrimination par exemple la technique usuelle de l'analyse linéaire discriminante.

[0073] D'autres techniques connues peuvent être utilisées telles qu'une technique non linéaire utilisant un réseau de neurones.

[0074] Si l'on dispose de N individus décrits par des vecteurs de dimension p et répartis a priori en K classes, l'analyse linéaire discriminante consiste :

- dans un premier temps, à chercher les K-1 fonctions linéaires indépendantes qui séparent au mieux les K classes. Il s'agit de déterminer quelles sont les combinaisons linéaires des p composantes des vecteurs qui minimisent la variance intra-classes et maximisent la variance interclasses.
- dans un deuxième temps, à déterminer la classe d'un nouvel individu par application des fonctions linéaires discriminantes au vecteur le représentant.

[0075] Dans le cas présent, les vecteurs représentatifs des individus ont pour composantes le pitch et les coefficients 1 à L (typiquement, L = 20) du cepstre partiel. La robustesse des fonctions discriminantes à la déviation des coefficients cepstraux est assurée à la fois par la présence du pitch dans les paramètres et par le choix du corpus d'apprentissage. Celui-ci est composé d'individus dont la voix originale a subi une grande diversité de filtrages représentatifs des distorsions occasionnées par les liaisons téléphoniques.

[0076] Plus précisément, à partir d'un corpus de voix originales (non dégradées) de N locuteurs, on définit un corpus de N vecteurs de composantes $[\bar{F}_0 ; C^p(1); ...; C^p(L)]$, avec $\overline{F_0}$ le pitch moyen et C^p le cepstre partiel. La construction du corpus d'apprentissage desdites fonctions consiste à définir un ensemble de M biais cepstraux qui s'ajouteront chacun à chaque cepstre partiel représentatif d'un locuteur du corpus original, ce qui permet d'obtenir un nouveau corpus de NM individus.

[0077] Ces biais dans le domaine du cepstre partiel correspondent à une large gamme de distorsions spectrales sur la bande F1-F2, proches de celles pouvant résulter de la liaison téléphonique.

[0078] A titre d'exemple, on propose l'ensemble de réponses fréquentielles représentées sur la figure 14 pour la bande 187-3187 Hz : chaque réponse fréquentielle correspond à un chemin de gauche à droite dans le treillis. L'amplitude de leurs variations sur cette bande n'excède pas 20 dB, à l'instar des caractéristiques extrémales des systèmes d'émission et lignes.

[0079] A partir de ces 81 caractéristiques fréquentielles sont calculés les 81 biais correspondants dans le domaine du cepstre partiel, selon le traitement décrit pour la mise en oeuvre de la relation (0.4). Par addition de ces biais au corpus de 63 locuteurs précédemment utilisé, on obtient un corpus d'apprentissage comptant 5103 individus représentatifs de diverses conditions (locuteur, filtrage de la liaison).

[0080] Dans le cas de classement par analyse linéaire discriminante :

* Application des critères de classement.

[0081] Soit (a^k) $1 \leq k \leq K-1$ la famille de fonctions linéaires discriminantes définies à partir du corpus d'apprentissage. Un locuteur représenté par le vecteur $x = [\overline{F_0} ; C^p(1) ; ... ; C^p(L)]$ est affecté à la classe q si la probabilité conditionnelle de q sachant a(x), notée P(q|a(x)), est maximale, a(x) désignant le vecteur de composantes $(a^k(x))1 \leq k \leq K-1$. Selon le théorème de Bayes,

$$P(q|a(x)) = \frac{P(a(x)|q)P(q)}{P(a(x))} . \qquad (0.6)$$

[0082] Par conséquent, P(q|a(x)) est proportionnelle à P (a (x) | q) P (q). Dans le sous-espace engendré par les K-1 fonctions discriminantes, sous l'hypothèse d'une distribution multi-gaussienne des individus dans chaque classe, la densité de probabilité de a(x) à l'intérieur de la classe q a pour expression :

$$f_q(x) = \frac{1}{(2\pi)^{\frac{K-1}{2}}\sqrt{|S_q|}} \exp\left(-\frac{1}{2}\left(a(x)-a(\bar{x}^q)\right)' S_q^{-1}\left(a(x)-a(\bar{x}^q)\right)\right), \quad (0.7)$$

où $\bar{x}^q$ est le centre de la classe q, |Sq| désigne le déterminant de la matrice Sq, et Sq est la matrice des covariances de a à l'intérieur de la classe q, d'élément qénérique $\sigma^q jk$ que l'on peut estimer par :

$$\sigma_{jk}^q = \frac{1}{N_q}\sum_{i=1}^{N_q}\left(a^j(x^i)-a^j(\bar{x}^q)\right)\left(a^k(x^i)-a^k(\bar{x}^q)\right). \quad (0.8)$$

[0083] L'individu x sera affecté à la classe q qui maximise fq(x)P(q), ce qui revient à minimiser sur q la fonction sq (x) appelée score discriminant :

$$s_q(x) = \left(a(x)-a(\bar{x}^q)\right)' S_q^{-1}\left(a(x)-a(\bar{x}^q)\right) + \log\left(|S_q|\right) - 2\log\left(P(q)\right), \quad (0.9)$$

[0084] Le procédé de correction proposé est mis en oeuvre par le système de correction (égaliseur) implanté dans le réseau numérique 40 comme l'illustre la figure 10.

[0085] La figure 16 illustre le système de correction apte à mettre en oeuvre le procédé. La figure 17 illustre ce système selon une variante de réalisation comme cela va être détaillé dans la suite. Ces variantes portent sur le mode de calcul du module de la réponse fréquentielle de l'égaliseur adapté restreinte à la bande F1-F2.

[0086] Le pré-égaliseur 200 est un filtre fixe, dont la réponse fréquentielle, sur la bande F1-F2, est l'inverse de la réponse globale de la partie analogique d'une liaison moyenne telle que définie précédemment [UIT-T/P.830, 1996].

[0087] La raideur de la réponse fréquentielle de ce filtre implique une réponse impulsionnelle longue ; c'est pourquoi, de manière à limiter le retard introduit par le traitement, le pré-égaliseur est réalisé typiquement sous forme d'un filtre RII, d'ordre 20 par exemple.

[0088] La figure 15 représente les réponses fréquentielles typiques du pré-égaliseur pour trois valeurs de F1. La dispersion des retards de groupe est inférieure à 2 ms, de sorte que la distorsion de phase résultante n'est pas perceptible.

[0089] La chaîne de traitement 400 qui suit permet le classement du locuteur et l'égalisation adaptée différenciée. Cette chaîne comprend deux blocs de traitements 400A et 400B. Le bloc 400A permet de calculer le module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation : EQ dB [F1-F2].

[0090] Le deuxième bloc 400B permet de calculer la réponse impulsionnelle du filtre égaliseur afin d'obtenir les coefficients eq(n) du filtre différenciés selon la classe du locuteur.

[0091] Un détecteur de trames d'activité vocale 401 permet de déclencher les différents traitements.

[0092] Le bloc de traitement 410 permet le classement du locuteur.

[0093] Le bloc de traitement 420 permet de calculer le spectre à long terme suivi du calcul du cepstre partiel de ce locuteur.

[0094] La sortie de ces deux blocs est appliquée sur l'opérateur 428a ou 428b. La sortie de cet opérateur fournit le module de la réponse fréquentielle de l'égaliseur adapté en dB restreinte à la bande d'égalisation F1-F2 via le bloc 429 pour 428a, via le bloc 440 pour 428b.

[0095] Les blocs de traitement 430 à 435 permettent de calculer les coefficients eq(n) du filtre.

[0096] La sortie x(n) du pré-égaliseur est analysée par trames successives d'une durée typique de 32 ms, avec un recouvrement inter-trames de 50 % typiquement. On ouvre pour cela une fenêtre d'analyse représentée par les blocs 402 et 403.

[0097] L'opération d'égalisation adaptée est mise en oeuvre par un filtre RIF 300 dont les coefficients sont calculés à chaque trame d'activité vocale par la chaîne de traitement illustrée sur les figures 16 et 17.

[0098] Le calcul de ces coefficients correspond au calcul de la réponse impulsionnelle du filtre à partir du module

de la réponse fréquentielle.

[0099] Le spectre à long terme de x(n), γ_x, est d'abord calculé (à partir de l'instant initial de fonctionnement) sur une fenêtre temporelle croissant de 0 à une durée T d'activité vocale (typiquement 4 s), puis ajusté récursivement à chaque trame d'activité vocale, ce qui se traduit par la formule générique suivante :

$$\gamma_x(f,n) = \alpha(n)|X(f,n)|^2 + (1-\alpha(n))\gamma_x(f,n-1), \quad (0.10)$$

où γ_x (f,n) est le spectre à long terme de x à la nième trame d'activité vocale, X(f,n) la transformée de Fourier de la nième trame d'activité vocale, et α(n) est défini par la relation (0.11). En notant N le nombre de trames dans la durée T,

$$\alpha(n) = \frac{1}{\min(n,N)} . \quad (0.11)$$

Ce calcul est opéré par les blocs 421, 422, 423.

[0100] On calcule ensuite, à partir de ce spectre à long terme le cepstre partiel Cp, selon la relation (0.4) mise en oeuvre par les blocs de traitement 424, 425, 426.

[0101] Le pitch moyen $\overline{F_0}$ est estimé par le bloc de traitement 412 à chaque trame voisée selon la formule :

$$\overline{F_0}(m) = \alpha(m)F_0(m) + (1-\alpha(m))\overline{F_0}(m-1), \quad (0.12)$$

où F0(m) est le pitch de la mième trame voisée et est calculé par le bloc 411 selon une méthode appropriée de l'état de l'art (par exemple méthode de l'autocorrélation, avec détermination du voisement par comparaison de l'auto-corrélation normalisée à un seuil [UIT-T/G.729, 1996]).

[0102] Ainsi, à chaque trame d'activité vocale, on dispose d'un nouveau vecteur x de composantes le pitch moyen et les coefficients 1 à L du cepstre partiel, auquel on applique la fonction discriminante $\underline{a}$ définie à partir du corpus d'apprentissage. Ce traitement est mis en oeuvre par le bloc 413. Le locuteur est alors affecté à la classe q de score discriminant minimal.

[0103] Le module en dB de la réponse fréquentielle de l'égaliseur adapté restreinte à la bande F1-F2, noté $|EQ|_{dB[F1-F2]}$, est calculé selon l'une des deux méthodes suivantes

[0104] :La première méthode (figure 16) consiste à calculer $|EQ|_{F1-F2}$ selon l'équation (0.3), où γ_{ref}(f) est le spectre de référence de la classe du locuteur (transformée de Fourier du centre de la classe). Cette méthode de calcul est mise en oeuvre dans cette variante représentée sur la figure 16 avec les opérateurs 414a, 428a, 427 et 429.

[0105] La deuxième méthode (figure 17) consiste à transcrire l'équation (0.3) dans le domaine du cepstre partiel, puisque l'on dispose du cepstre partiel de la sortie x du pré-égaliseur, nécessaire au classement du locuteur. Ainsi, l'équation (0.3) devient :

$$C^p_{eq} = C^p_{ref} - C^p_x - C^p_{s_rx} - C^p_{l_rx}, \quad (0.13)$$

où C^p_{eq}, C^p_x, $C^p_{s_rx}$ et $C^p_{l_rx}$ sont les cepstres partiels respectifs de l'égaliseur adapté, de la sortie x du pré-égaliseur, du système de réception et de la ligne de réception, C^p_{ref} étant le cepstre partiel de référence, centre de la classe du locuteur. Les cepstres partiels sont calculés comme indiqué précédemment, en sélectionnant la bande de fréquences F1-F2. Ce calcul est effectué uniquement pour les coefficients 1 à 20, les coefficients suivants étant inutiles car représentatifs d'une finesse spectrale qui sera éliminée par la suite.

[0106] Les 20 coefficients du cepstre partiel de l'égaliseur adapté sont obtenus par les opérateurs 414b et 428b selon la relation (0.13).

[0107] Le bloc de traitement 441 complète ces 20 coefficients par des zéros, les symétrise et calcule, à partir du vecteur ainsi formé, le module en dB de la réponse fréquentielle de l'égaliseur adapté restreinte à la bande F1-F2 en mettant en oeuvre la relation suivante :

$$EQ_{dB|F_1-F_2} = TFD^{-1}(C^p_{eq}) . \quad (0.14)$$

[0108] Cette réponse est décimée d'un facteur ¾ par l'opérateur 442.

[0109] Pour les deux variantes qui viennent d'être décrites, les valeurs de |EQ| hors de la bande F1-F2 sont calculées

par extrapolation linéaire de la valeur en dB de $|EQ|_{F1\text{-}F2}$, notée EQ_{dB} par la suite, par le bloc 430 et de la manière suivante :

[0110] Pour chaque indice de fréquence k, l'approximation linéaire de EQ_{dB} s'exprime par

$$\tilde{EQ}_{dB}(k) = a_1 + a_2 k \qquad (0.15)$$

[0111] Les coefficients a1 et a2 sont choisis de manière à minimiser l'erreur quadratique de l'approximation sur l'intervalle F1-F2, définie par

$$e = \sum_{k=k_1}^{k_2} \left(EQ_{dB}(k) - \tilde{EQ}_{dB}(k) \right)^2 \qquad (0.16)$$

[0112] Les coefficients a1 et a2 sont donc définis par :

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} k_2 - k_1 + 1 & \sum_{k=k_1}^{k_2} k \\ \sum_{k=k_1}^{k_2} k & \sum_{k=k_1}^{k_2} k^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{k=k_1}^{k_2} EQ_{dB}(k) \\ \sum_{k=k_1}^{k_2} kEQ_{dB}(k) \end{pmatrix} \qquad (0.17)$$

[0113] Les valeurs de |EQ|, en dB, hors de la bande F1-F2, sont alors calculées à partir de la formule (0.15).

[0114] La caractéristique en fréquence ainsi obtenue doit être lissée. Le filtrage étant réalisé dans le domaine temporel, le moyen permettant ce lissage est de multiplier par une fenêtre étroite la réponse impulsionnelle correspondante.

[0115] La réponse impulsionnelle est obtenue par une opération IFFT appliquée sur |EQ| réalisée par les blocs 431 et 432 suivie d'une symétrisation réalisée par le bloc de traitement 433, de manière à obtenir un filtre causal à phase linéaire. La réponse impulsionnelle résultante est multipliée, opérateur 435, par une fenêtre temporelle 434. La fenêtre utilisée est typiquement une fenêtre de Hamming de longueur 31 centrée sur le pic de la réponse impulsionnelle et est appliquée sur la réponse impulsionnelle au moyen de l'opérateur 435.

Revendications

1. Procédé de correction des déformations spectrales de la voix, introduites par un réseau de communication, comprenant une opération d'égalisation sur une bande de fréquence [F1-F2], adaptée à la distorsion réelle de la chaîne de transmission, cette opération étant réalisée au moyen d'un filtre numérique ayant une réponse fréquentielle fonction du rapport entre un spectre de référence et un spectre correspondant au spectre à long terme du signal voix des locuteurs, **caractérisé en ce qu'**il comprend :

   * Préalablement à l'opération d'égalisation du signal voix d'un locuteur en communication:

     - La constitution de classes de locuteurs avec une référence de voix par classe,

   * Puis, pour un locuteur en communication donné :

     - Le classement de ce locuteur c'est-à-dire son affectation à une classe à partir de critères de classement prédéfinis pour lui faire correspondre une référence de voix qui est la plus proche de la sienne,
     - L'égalisation du signal numérisé de la voix du locuteur opérée avec comme spectre de référence, la référence de voix de la classe à laquelle ledit locuteur a été affecté.

2. Procédé de correction des déformations spectrales de la voix selon la revendication 1, **caractérisé en ce que** :

* La constitution de classes de locuteurs comprend :

  - Le choix d'un corpus de N locuteurs enregistrés dans des conditions non dégradées et la détermination de leur spectre de fréquence à long terme,
  - La classification des locuteurs du corpus selon leur cepstre partiel c'est à dire le cepstre calculé à partir du spectre à long terme restreint à la bande d'égalisation [F1-F2] et en appliquant un critère de classification prédéfini sur ces cepstres pour obtenir K classes,
  - Le calcul du spectre de référence associé à chaque classe de manière à obtenir une référence de voix correspondant à chacune des classes.

3. Procédé de correction des déformations spectrales de la voix selon la revendication 2, **caractérisé en ce que** le spectre de référence sur la bande de fréquences d'égalisation [F1-F2], associé à chaque classe, est calculé par transformée de Fourier du centre de la classe défini par son cesptre partiel.

4. Procédé de correction des déformations spectrales de la voix selon la revendication 1, **caractérisé en ce que** :

* La classification d'un locuteur comprend :

  - L'utilisation du pitch moyen du signal voix et du cepstre partiel de ce signal comme paramètres de classement,
  - L'application d'une fonction discriminante à ces paramètres pour classer ledit locuteur.

5. Procédé de correction des déformations spectrales de la voix selon l'une quelconque des revendications précédentes, **caractérisé en ce qu'**il comprend en outre une étape de pré-égalisation du signal numérique par un filtre fixe ayant une réponse fréquentielle dans la bande de fréquence [F1-F2], correspondant à l'inverse d'une déformation spectrale de référence introduite par la liaison téléphonique.

6. Procédé de correction des déformations spectrales de la voix selon l'une quelconque des revendications précédentes, **caractérisé en ce que** l'égalisation du signal numérisé de la voix d'un locuteur comprend :

- la détection d'une activité vocale sur la ligne pour déclencher un enchaînement de traitements comprenant le calcul du spectre à long terme, la classification du locuteur, le calcul du module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] et le calcul des coefficients du filtre numérique différenciés selon la classe du locuteur, à partir de ce module,
- la commande du filtre avec les coefficients obtenus,
- le filtrage du signal sortant du pré-égaliseur par ledit filtre.

7. Procédé de correction des déformations spectrales de la voix selon la revendication 6, **caractérisé en ce que** le calcul du module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] est réalisé par la mise en oeuvre de la relation suivante :

$$|EQ(f)| = \frac{1}{|S_RX(f).L_RX(f)|} \sqrt{\frac{\gamma_{ref}(f)}{\gamma_x(f)}}, \qquad (0.3)$$

dans laquelle γ_{ref} (f) est le spectre de référence de la classe à laquelle appartient ledit locuteur,
et dans laquelle L_RX est la réponse fréquentielle de la ligne de réception, S_RX la réponse fréquentielle du système de réception et $\gamma x(f)$ le spectre à long terme du signal x d'entrée du filtre.

8. Procédé de correction des déformations spectrales de la voix selon la revendication 6, **caractérisé en ce que** le calcul du module [EQ] de la réponse fréquentielle du filtre égaliseur restreinte à la bande d' égalisation [F1-F2] est réalisé par la mise en oeuvre de la relation suivante :

$$C^p_{eq} = C^p_{ref} - C^p_x - C^p_{s_rx} - C^p_{l_rx}, \qquad (0.13)$$

dans laquelle C^p_{eq}, C^p_x, $C^p_{s_rx}$ et $C^p_{l_rx}$ sont les cepstres partiels respectifs de l'égaliseur adapté, du signal

d'entrée x du filtre égaliseur, du système de réception et de la ligne de réception, C^p_{ref} étant le cepstre partiel de référence, centre de la classe du locuteur ; le module [EQ] restreint à la bande F1-F2 étant calculé par Transformée de Fourier Discrète de C^p_{eq}.

9. Système de correction des déformations spectrales de la voix, introduites par un réseau de communication, comprenant des moyens d'égalisation adaptée dans une bande de fréquence [F1-F2] qui comportent un filtre numérique (300) dont la réponse fréquentielle est fonction du rapport entre un spectre de référence et un spectre correspondant au spectre à long terme d'un signal voix, **caractérisé en ce que** ces moyens comprennent en outre :

- des moyens de traitement du signal (400) pour le calcul des coefficients du filtre numérique munis :
  - d'un premier bloc de traitements (400A) du signal pour calculer le module de la réponse fréquentielle du filtre égaliseur restreinte à la bande d'égalisation [F1-F2] selon la relation suivante

$$|EQ(f)| = \frac{1}{|S_RX(f).L_RX(f)|}\sqrt{\frac{\gamma_{ref}(f)}{\gamma_x(f)}}\ , \qquad (0.3)$$

dans laquelle $\gamma_{ref}(f)$ est le spectre de référence, qui peut être différent d'un locuteur à un autre et qui correspond à une référence de classe prédéterminée à laquelle appartient ledit locuteur, et dans laquelle L_RX est la réponse fréauentielle de la ligne de réception, S_RX la réponse fréquentielle du système de réception et $\gamma_x(f)$ le spectre à long terme du signal x d'entrée du filtre;

  - d'un deuxième bloc de traitements (400B) pour le calcul de la réponse impulsionnelle à partir du module de réponse fréquentielle ainsi calculé, afin de déterminer les coefficients du filtre différenciés selon la classe du locuteur

10. Système de correction des déformations spectrales de la voix selon la revendication 9, **caractérisé en ce que** le premier bloc de traitements (400A) comprend des moyens (414b, 428b) de calcul du cepstre partiel du filtre égaliseur selon la relation :

$$C^p_{eq} = C^p_{ref} - C^p_x - C^p_{s_rx} - C^p_{l_rx}, \qquad (0.13)$$

dans laquelle C^p_{eq}, C^p_x, $C^p_{s_rx}$ et $C^p_{l_rx}$ sont les cepstres partiels respectifs de l'égaliseur adapté, du signal d'entrée x du filtre égaliseur, du système de réception et de la ligne de réception, C^p_{ref} étant le cepstre partiel de référence, centre de la classe du locuteur ; le module [EQ] restreint à la bande F1-F2 étant calculé par Transformée de Fourier Discrète de $C^p_{eq.}$.

11. Système de correction des déformations spectrales de la voix selon la revendication 9 ou 10, **caractérisé en ce que** le premier bloc de traitements comprend un sous-ensemble (420) pour calculer les coefficients du cepstre partiel d'un locuteur en communication et un deuxième sous-ensemble (410) pour opérer le classement de ce locuteur, ce deuxième sous ensemble comprenant un bloc (411) de calcul du pitch F_0, un bloc (412) d'estimation du pitch moyen à partir du pitch calculé F_0, et un bloc (413) de classement appliquant une fonction discriminante sur le vecteur x ayant pour composantes le pitch moyen et les coefficients du cepstre partiel pour classer ledit locuteur.

12. Système de correction des déformations spectrales de la voix selon l'une quelconque des revendications 9 à 11, **caractérisé en ce qu'**il comprend un pré-égaliseur (200) et **en ce que** le signal égalisé à partir de spectres de référence différenciés selon la classe du locuteur est le signal x de sortie du pré-égaliseur.

Patentansprüche

1. Verfahren zum Korrigieren spektraler Verzerrungen der Stimme, die durch ein Kommunikationsnetz eingeführt werden, wobei das Verfahren eine an die wirkliche Verzerrung der Übertragungskette angepasste Entzerrungsoperation in einem Frequenzband [F1-F2] umfasst, wobei diese Operation mittels eines digitalen Filters ausgeführt wird, das einen Frequenzgang besitzt, der vom Verhältnis zwischen einem Referenzspektrum und einem Spek-

trum, der dem Langzeitspektrum des Stimmsignals des Sprechers entspricht, abhängt, **dadurch gekennzeichnet, dass** es umfasst:

* vor der Operation der Entzerrung des Stimmsignals eines in Kommunikation befindlichen Sprechers:
  - Bilden von Klassen von Sprechern mit einer Stimmreferenz pro Klasse,
* dann für einen gegebenen in Kommunikation befindlichen Sprecher:
  - Klassifizieren dieses Sprechers, d. h. Zuordnung dieses Sprechers zu einer Klasse anhand von im Voraus definierten Klassifizierungskriterien, damit ihm eine Stimmreferenz entspricht, die der seinen am nächsten kommt,
  - Entzerren des digitalisierten Signals der verwendeten Stimme des Sprechers mit der als Referenzspektrum dienenden Stimmreferenz der Klasse, der der Sprecher zugeordnet worden ist.

2. Verfahren zum Korrigieren spektraler Verzerrungen der Stimme nach Anspruch 1, **dadurch gekennzeichnet, dass**:

* das Bilden von Klassen von Sprechern umfasst:
  - Wählen eines Körpers aus N Sprechern, die unter nicht verschlechterten Bedingungen aufgezeichnet worden sind, und Bestimmen ihres Langzeit-Frequenzspektrums,
  - Klassifizieren der Sprecher des Körpers entsprechend ihres Partialcepstrums, d. h. des Spektrums, das anhand des auf das Entzerrungsband [F1-F2] eingeschränkten Langzeitspektrums berechnet wird, wobei ein im Voraus definiertes Klassifizierungskriterium auf diese Cepstren angewendet wird, um K Klassen zu erhalten,
  - Berechnen des Referenzspektrums, das jeder Klasse zugeordnet ist, in der Weise, dass eine Stimmreferenz, die jeder der Klassen entspricht, erhalten wird.

3. Verfahren zum Korrigieren spektraler Verzerrungen der Stimme nach Anspruch 2, **dadurch gekennzeichnet, dass** das Referenzspektrum in dem Entzerrungsfrequenzband [F1-F2], das jeder Klasse zugeordnet ist, durch Fourier-Transformation des Zentrums der Klasse, die durch ihr Partialcepstrum definiert ist, berechnet wird.

4. Verfahren zum Korrigieren spektraler Verzerrungen der Stimme nach Anspruch 1, **dadurch gekennzeichnet, dass**:

* die Klassifizierung eines Sprechers umfasst:
  - Verwenden der mittleren Tonhöhe des Stimmsignals und des Partialcepstrums dieses Signals als Klassifizierungsparameter,
  - Anwenden einer diskriminierenden Funktion auf diese Parameter, um den Sprecher zu klassifizieren.

5. Verfahren zum Korrigieren spektraler Verzerrungen der Stimme nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, dass** es außerdem einen Schritt umfasst, bei dem das digitale Signal durch ein festes Filter entzerrt wird, das einen Frequenzgang in dem Frequenzband [F1-F2] besitzt, der der Umkehrung der durch die Telephonverbindung eingeführten spektralen Referenzverzerrung entspricht,.

6. Verfahren zum Korrigieren spektraler Verzerrungen der Stimme nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, dass** die Entzerrung des digitalisierten Signals der Stimme eines Sprechers umfasst:

- Erfassen einer Sprechaktivität auf der Leitung, um eine Verarbeitungskette auszulösen, die die Berechnung des Langzeitspektrums, die Klassifizierung des Sprechers, die Berechnung des Moduls des auf das Entzerrungsband [F1-F2] eingeschränkten Frequenzgangs des Entzerrungsfilters und anhand dieses Moduls die Berechnung der Filterkoeffizienten des digitalen Filters, die gemäß der Klasse des Sprechers unterschieden sind, umfasst,
- Steuern des Filters mit den erhaltenen Koeffizienten und
- Filtern des Signals, das den Vorentzerrer verlässt, durch das Filter.

7. Verfahren zum Korrigieren spektraler Verzerrungen der Stimme nach Anspruch 6, **dadurch gekennzeichnet, dass** die Berechnung des Moduls des auf das Entzerrungsband [F1-F2] eingeschränkten Frequenzgangs des Entzerrungsfilters durch Ausführen der folgenden Beziehung ausgeführt wird:

$$|EQ(f)| = \frac{1}{|S_RX(f).L_RX(f)|} \sqrt{\frac{\gamma_{ref}(f)}{\gamma_x(f)}}\ , \qquad (0.3)$$

worin $\gamma_{ref}(f)$ das Referenzspektrum der Klasse ist, zu der der Sprecher gehört,
und worin L_RX der Frequenzgang der Empfangsleitung ist, S_RX der Frequenzgang des Empfangssystems ist und $\gamma_x(f)$ das Langzeitspektrum des Eingangssignals x des Filters ist.

8. Verfahren zum Korrigieren spektraler Verzerrungen der Stimme nach Anspruch 6, **dadurch gekennzeichnet, dass** die Berechnung des Moduls [EQ] des auf das Entzerrungsband [F1-F2] einschränkten Frequenzgangs des Entzerrungsfilters durch Ausführen der folgenden Beziehung ausgeführt wird:

$$C^p_{eq} = C^p_{ref} - C^p_x - C^p_{s_rx} - C^p_{l_rx}, \qquad (0.13)$$

worin C^p_{eq}, C^p_x, $C^p_{s_rx}$ und $C^p_{l_rx}$ die Partialcepstren des angepassten Entzerrers bzw. des Eingangssignals x des Entzerrungsfilters bzw. des Empfangssystems bzw. der Empfangsleitung sind, wobei C^p_{ref} das Referenz-Partialcepstrum, das Zentrum der Klasse des Sprechers, ist; wobei das auf das Band F1-F2 eingeschränkte Modul [EQ] durch diskrete Fourier-Transformation von C^p_{eq} berechnet wird.

9. System zum Korrigieren spektraler Verzerrungen der Stimme, die durch ein Kommunikationsnetz eingeführt werden, wobei das System Mittel zur angepassten Entzerrung in einem Frequenzband [F1-F2] umfasst, die ein digitales Filter (300) enthalten, dessen Frequenzgang vom Verhältnis zwischen einem Referenzspektrum und einem dem Langzeitspektrum eines Stimmsignals entsprechenden Spektrum abhängt, **dadurch gekennzeichnet, dass** diese Mittel außerdem umfassen:

- Signalverarbeitungsmittel (400) für die Berechnung der Koeffizienten des digitalen Filters, die versehen sind mit:

  * einem ersten Block (400A) für Verarbeitungen des Signals, um das Modul des auf das Entzerrungsband [F1-F2] eingeschränkten Frequenzgangs des Entzerrungsfilters gemäß der folgenden Beziehung zu berechnen:

$$|EQ(f)| = \frac{1}{|S_RX(f).\ L_RX(f)|} \sqrt{\frac{\gamma_{ref}(f)}{\gamma_x(f)}}\ , \qquad (0.3)$$

  worin $\gamma_{ref}(f)$ das Referenzspektrum ist, das von einem Sprecher zum nächsten unterschiedlich sein kann und einer vorgegebenen Klassenreferenz entspricht, der der Sprecher zugehört, und worin L_RX der Frequenzgang der Empfangsleitung ist, S_RX der Frequenzgang des Empfangssystems ist und $\gamma_x(f)$ das Langzeitspektrum des Eingangssignals x des Filters ist;
  * einem zweiten Block (400B) für Verarbeitungen, um die Impulsantwort anhand des somit berechneten Moduls des Frequenzgangs zu berechnen, um die Koeffizienten des Filters zu bestimmen, die gemäß der Klasse des Sprechers unterschieden sind.

10. System zum Korrigieren spektraler Verzerrungen der Stimme nach Anspruch 9, **dadurch gekennzeichnet, dass** der erste Verarbeitungsblock (400A) Mittel (414b, 428b) umfasst, um das Partialcepstrum des Entzerrungsfilters gemäß der folgenden Beziehung zu berechnen:

$$C^p_{eq} = C^p_{ref} - C^p_x - C^p_{s_rx} - C^p_{l_rx}, \qquad (0.13)$$

worin C^p_{eq}, C^p_x, $C^p_{s_rx}$ und $C^p_{L_rx}$ die Partialcepstren des angepassten Entzerrers bzw. des Eingangssignals

x des Entzerrungsfilters bzw. des Empfangssystems bzw. der Empfangsleitung sind, wobei C^p_{ref} das Referenz-Partialcepstrum, das Zentrum der Klasse des Sprechers, ist; wobei das auf das Band F1-F2 eingeschränkte Modul [EQ] durch diskrete Fourier-Transformation von C^p_{eq} berechnet wird.

11. System zum Korrigieren spektraler Verzerrungen der Stimme nach Anspruch 9 oder 10, **dadurch gekennzeichnet, dass** der erste Verarbeitungsblock eine Untereinheit (420) umfasst, um die Koeffizienten des Partialcepstrums eines in Kommunikation befindlichen Sprechers zu berechnen, und eine zweite Untereinheit (410) umfasst, um die Klassifizierung dieses Sprechers vorzunehmen, wobei diese zweite Untereinheit einen Block (411) zum Berechnen der Tonhöhe F_0, einen Block (412) zum Schätzen der mittleren Tonhöhe anhand der berechneten Tonhöhe F_0 und einen Klassifizierungsblock (413), der auf den Vektor x eine diskriminierende Funktion anwendet, der als Komponenten die mittlere Tonhöhe und die Koeffizienten des Partialcepstrums hat, um den Sprecher zu klassifizieren, umfasst.

12. System zum Korrigieren spektraler Verzerrungen der Stimme nach einem der Ansprüche 9 bis 11, **dadurch gekennzeichnet, dass** es einen Vorentzerrer (200) umfasst und dass das Signal, das anhand der gemäß der Klasse des Sprechers unterschiedenen Referenzspektren entzerrt wird, das Ausgangssignal x des Vorentzerrers ist.

Claims

1. A method for correcting spectral voice distortions, introduced by a communications network, comprising an equalization operation on a frequency band [F1-F2], adapted to the actual distortion of the transmission channel, this operation being performed by means of a digital filter with a frequency response depending on the ratio between a reference spectrum and a spectrum matching the long term spectrum of the vocal signal of the speakers, **characterized in that** it comprises:

* before the equalization operation on the voice signal of a communicating speaker:
  - forming classes of speakers with a voice reference per class,
* and then, for a given communicating speaker:
  - classifying this speaker, i.e. assigning him/her to a class from predefined classification criteria so as to have him/her match a voice reference which is the closest to his/hers,
  - equalizing the digitized signal of the voice of the speaker performed with the voice reference as reference spectrum, of the class to which said speaker was assigned.

2. The method for correcting spectral voice distortions according to claim 1, **characterized in that**:

* forming the classes of speakers comprises:
  - selecting a corpus of N speakers recorded under non-deteriorated conditions, and determining their long term frequency spectrum,
  - classifying the speakers of the corpus, according to their partial cepstrum, i.e., the cepstrum calculated from the long term spectrum restricted to the equalization band [F1-F2] and by applying a predefined classification criterion on these cepstra, in order to obtain K classes,
  - calculating the reference spectrum associated with each class in order to obtain a voice reference matching each of the classes.

3. The method for correcting spectral voice distortions according to claim 2, **characterized in that** the reference spectrum on the equalization frequency band [F1-F2], associated with each class, is calculated by a Fourier Transform of the center of the class defined by its partial cepstrum.

4. The method for correcting spectral voice distortions according to claim 1, **characterized in that**:

* classifying a speaker comprises:
  - the use of the average pitch of the voice signal and of the partial cepstrum of this signal as classification

parameters,

- applying a discriminating function to these parameters in order to classify said speaker.

5. The method for correcting spectral voice distortions according to any of the preceding claims, **characterized in that** it further comprises a step for pre-equalizing the digital signal with a fixed filter with a frequency response in the frequency band [F1-F2], matching the reciprocal of a reference spectral distortion introduced by the telephone link.

6. The method for correcting spectral voice distortions according to any of the preceding claims, **characterized in that** the equalization of the digitized signal of the voice of a speaker comprises:

- detecting vocal activity on the line in order to trigger a sequence of processing operations comprising the long term spectrum calculation, the classification of the speaker, the calculation of the modulus of the frequency response of the equalizer filter restricted to the equalization band [F1-F2] and the calculation of the coefficients of the digital filter differentiated according to the class of the speaker, from this modulus,
- controlling the filter with the obtained coefficients,
- filtering the output signal from the pre-equalizer by said filter.

7. The method for correcting spectral voice distortions according to claim 6, **characterized in that** the calculation of the frequency response of the equalizer filter restricted to the equalization band [F1-F2] is achieved by implementing the following relationship:

$$|EQ(f)| = \frac{1}{|S_RX(f).L_RX(f)|} \sqrt{\frac{\gamma_{ref}(f)}{\gamma_x(f)}}, \qquad (0,3)$$

wherein $\gamma_{ref}(f)$ is the reference spectrum for the class to which belongs said speaker, and wherein L_RX is the frequency response of the receiving line, S_RX the frequency response of the receiving system and $\gamma_x(f)$ the long term spectrum of the input signal x of the filter.

8. The method for correcting spectral voice distortions according to claim 6, **characterized in that** calculation of the modulus [EQ] of the frequency response of the equalizer filter restricted to the equalization band [F1-F2] is achieved by implementing the following relationship:

$$C^p_{eq} = C^p_{ref} - C^p_x - C^p_{s_rx} - C^p_{l_rx}, \qquad (0,13)$$

wherein C^p_{eq}, C^p_x, $C^p_{s_rx}$, and $C^p_{l_rx}$, are the respective partial cepstra of the adapted equalizer, of the input signal x of the equalizer filter, of the receiving system and of the receiving line, C^p_{ref} being the reference partial cepstrum, center of the class of the speaker; the modulus [EQ] restricted to the band [F1-F2] being calculated by a Discrete Fourier Transform of C^p_{eq}.

9. A system for correcting spectral voice distortions introduced by a communications network, comprising adapted equalization means in a frequency band [F1-F2] which include a digital filter (300), the frequency response of which depends on the ratio between a reference spectrum and a spectrum matching the long term spectrum of a voice signal, **characterized in that** these means further comprise:

- signal processing means (400) for calculating the coefficients of the digital filter, provided:

  * with a first block of signal processing operations (400A) for calculating the modulus of the frequency response of the equalizer filter restricted to the equalization band [F1-F2] according to the following relationship:

$$|EQ(f)| = \frac{1}{|S_RX(f).L_RX(f)|} \sqrt{\frac{\gamma_{ref}(f)}{\gamma_x(f)}}, \qquad (0,3)$$

wherein $\gamma_{ref}(f)$ is the reference spectrum, which may be different from one speaker to another, and which

matches a reference from a predetermined class to which belongs said speaker, and wherein L_RX is the frequency response of the receiving line, S_RX the frequency response of the receiving system and $\gamma_x(f)$ the long term spectrum of the input signal x of the filter;

* with a second block of processing operations (400B) for calculating the impulse response from the thereby calculated frequency response modulus, in order to determine the differentiated coefficients of the filter according to the class of the speaker.

10. The system for correcting spectral voice distortions according to claim 9, **characterized in that** the first block of processing operations (400A) comprises means (414b, 428b) for calculating the partial cepstrum of the equalizer filter according to the relationship:

$$C^{p}_{eq} = C^{p}_{ref} - C^{p}_{x} - C^{p}_{s_rx} - C^{p}_{l_rx}, \qquad (0,13)$$

wherein C^{p}_{eq}, C^{p}_{x}, $C^{p}_{s_rx}$, and $C^{p}_{l\text{-}rx}$, are the respective partial cepstra of the adapted equalizer, of the input signal x of the equalizer filter, of the receiving system and of the receiving line, C^{P}_{ref} being the reference partial cepstrum, center of the class of the speaker; the modulus [EQ] restricted to the band [F1-F2] being calculated by a Discrete Fourier Transform of C^{p}_{eq}.

11. The system for correcting spectral voice distortions according to claim 9 or 10, **characterized in that** the first block of processing operations comprises a subset (420) for calculating the coefficients of the partial cepstrum of a communicating speaker and a second subset (410) for performing the classification of this speaker, this second subset comprising a block (411) for calculating the pitch F_0, a block (412) for estimating the average pitch from a calculated pitch F_0, and a classification block (413) applying a discriminating function onto vector x, the components of which are the average pitch and the coefficients of the partial cepstrum, for classifying said speaker.

12. The system for correcting spectral voice distortions according to any of claims 9 to 11, **characterized in that** it comprises a pre-equalizer (200) and **in that** the signal equalized from reference spectra differentiated according to the class of the speaker is the output signal x of the pre-equalizer.

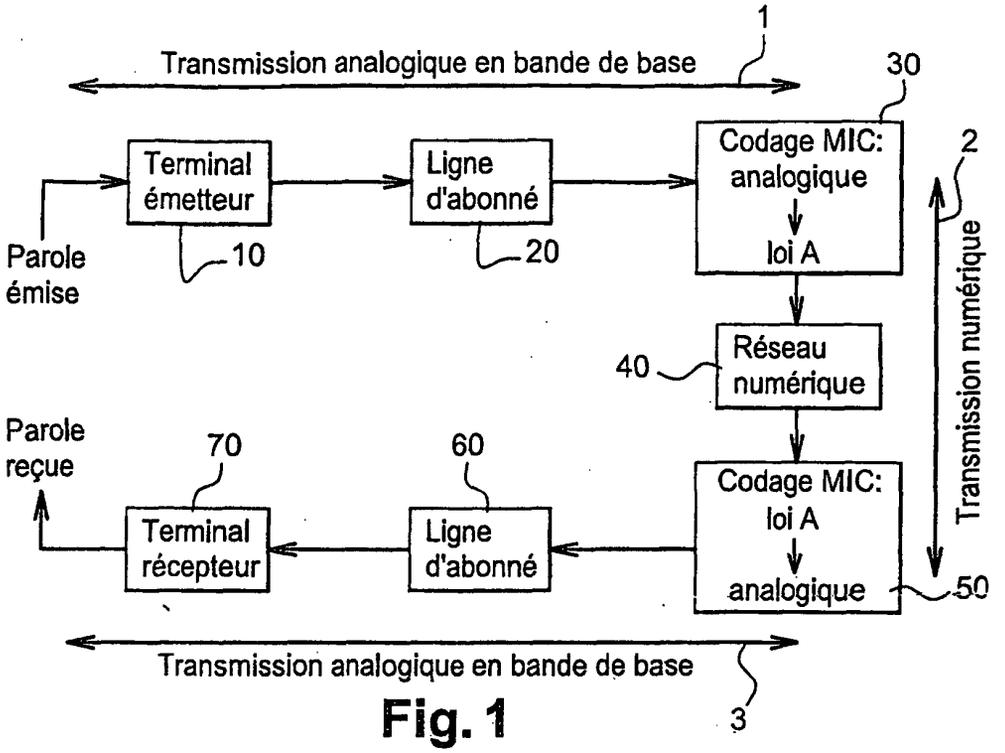


Fig. 1

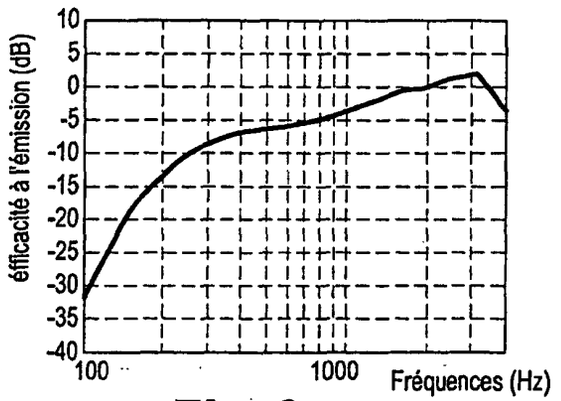


Fig. 2

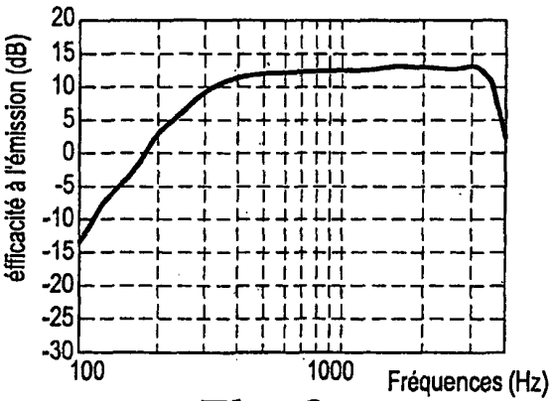


Fig. 3

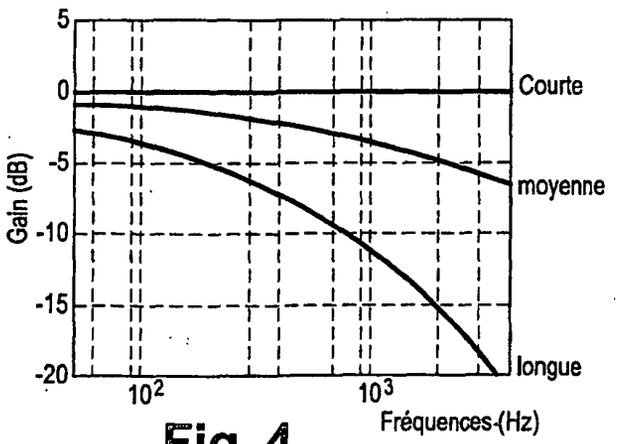


Fig. 4

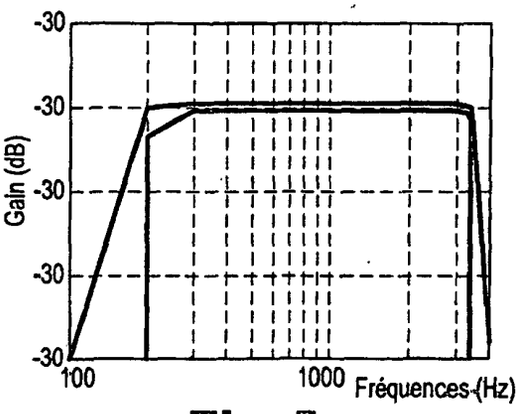


Fig. 5

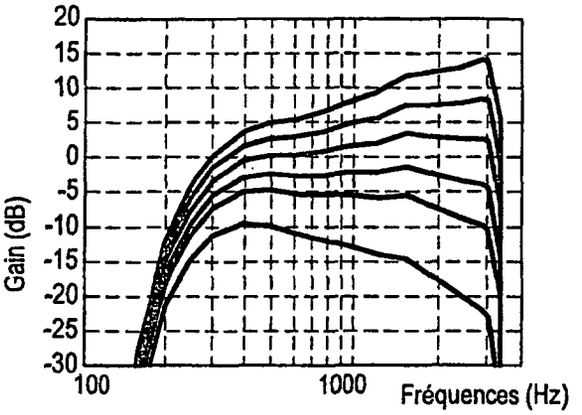


Fig. 6

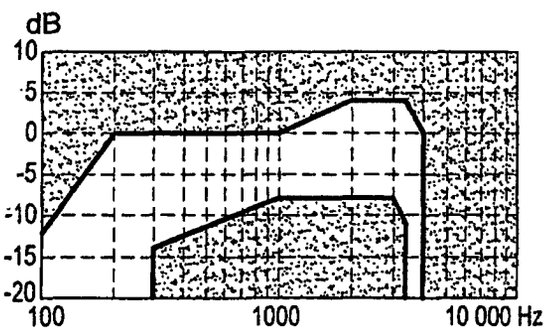


Fig. 7

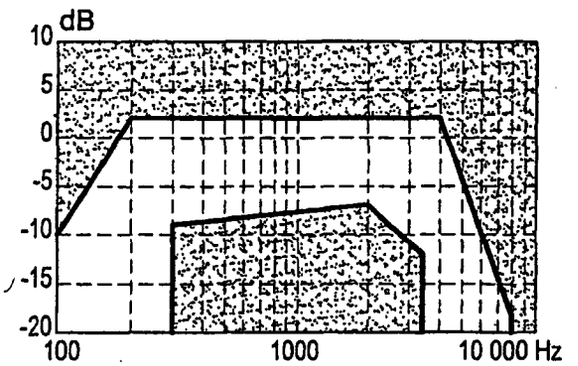


Fig. 8

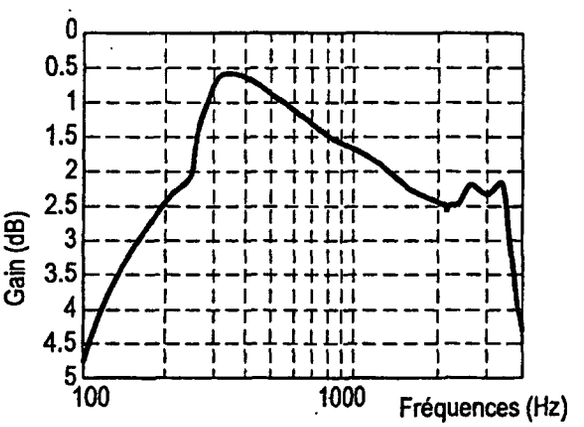


Fig. 9

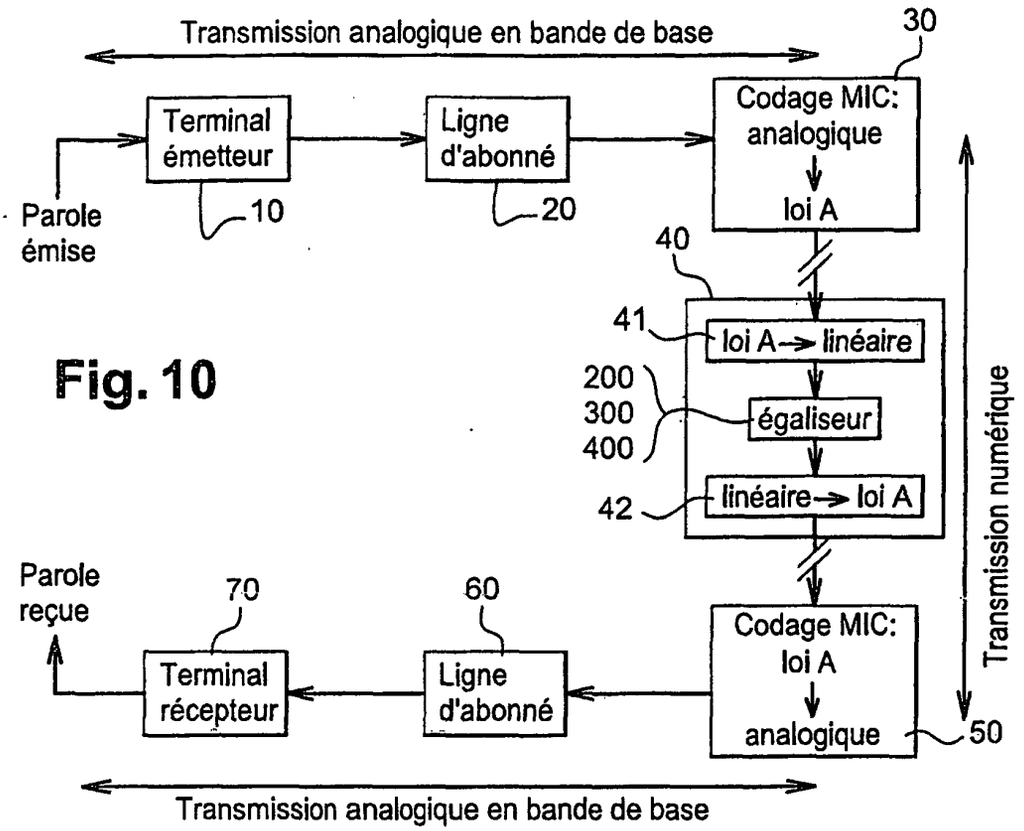

Transmission analogique en bande de base
30
Terminal émetteur
Ligne d'abonné
Codage MIC: analogique
loi A
Parole émise
10
20
40
41
loi A→ linéaire
200
300
400
égaliseur
42
linéaire→ loi A
Transmission numérique
Parole reçue
70
60
Terminal récepteur
Ligne d'abonné
Codage MIC: loi A
analogique
50
Transmission analogique en bande de base

Fig. 10

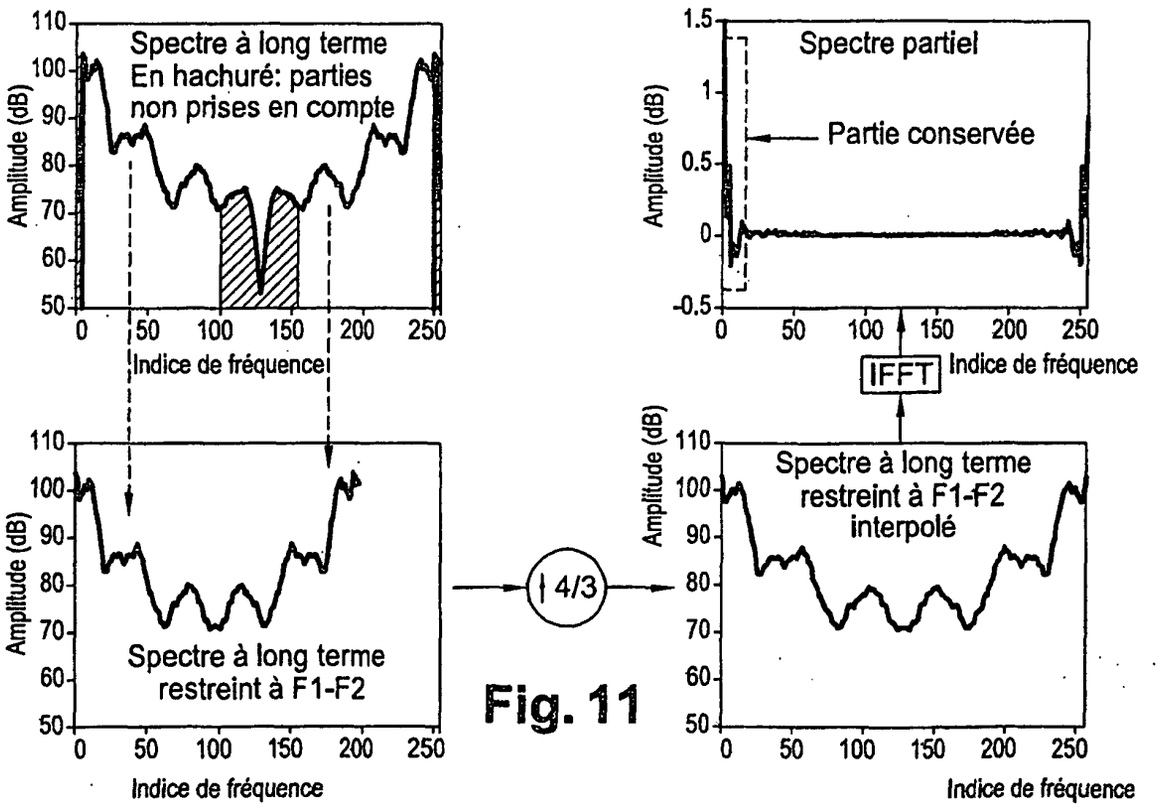

Spectre à long terme
En hachuré: parties non prises en compte
Amplitude (dB)
Indice de fréquence
Spectre partiel
Partie conservée
IFFT
Spectre à long terme restreint à F1-F2
Spectre à long terme restreint à F1-F2 interpolé
4/3

Fig. 11

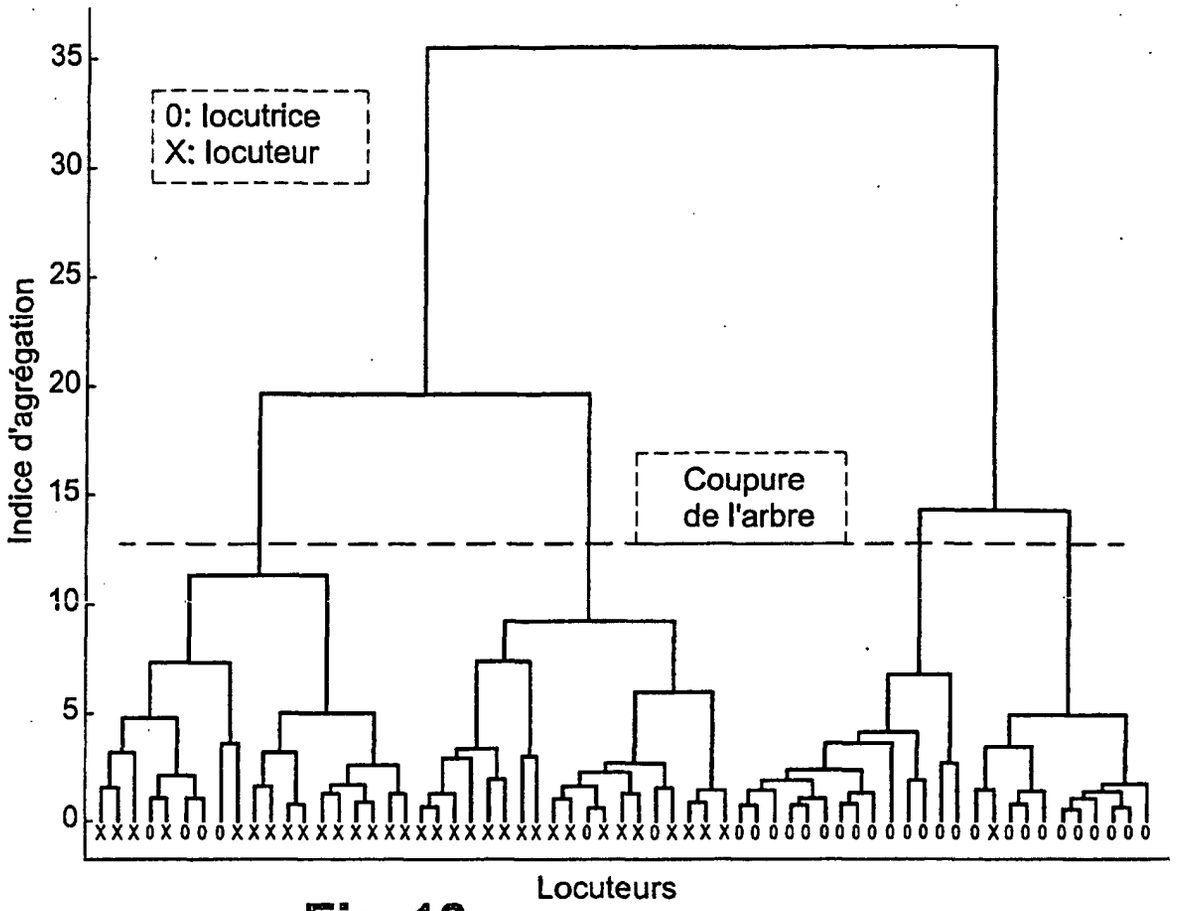


Fig. 12

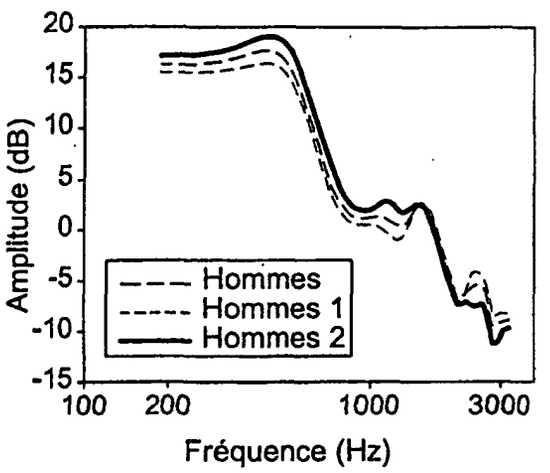


Fig. 13a

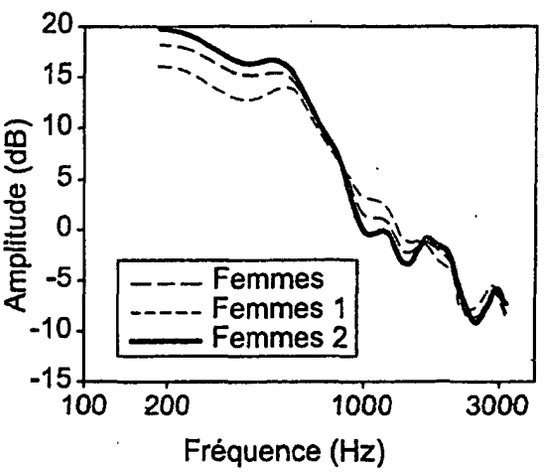


Fig. 13b

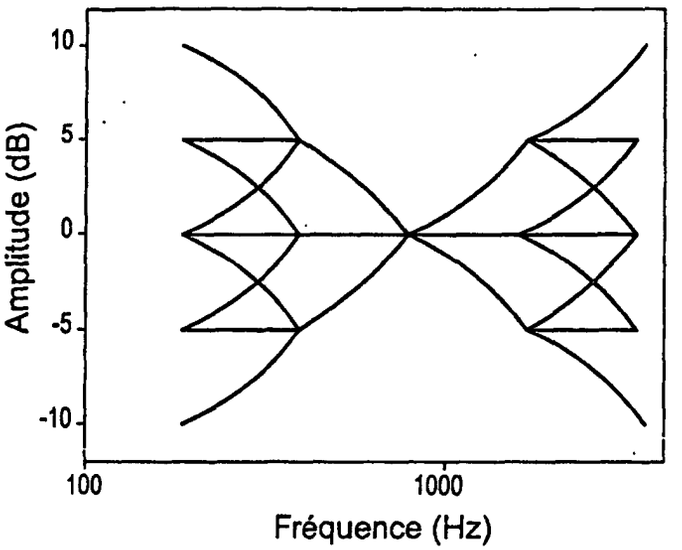

10
5
0
-5
-10
Amplitude (dB)
100
1000
Fréquence (Hz)

Fig. 14

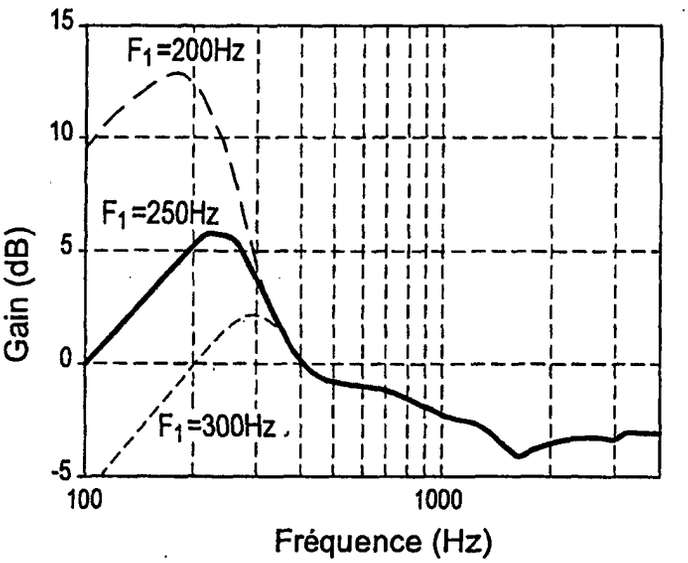

15
10
5
0
-5
Gain (dB)
F1=200Hz
F1=250Hz
F1=300Hz
100
1000
Fréquence (Hz)

Fig. 15

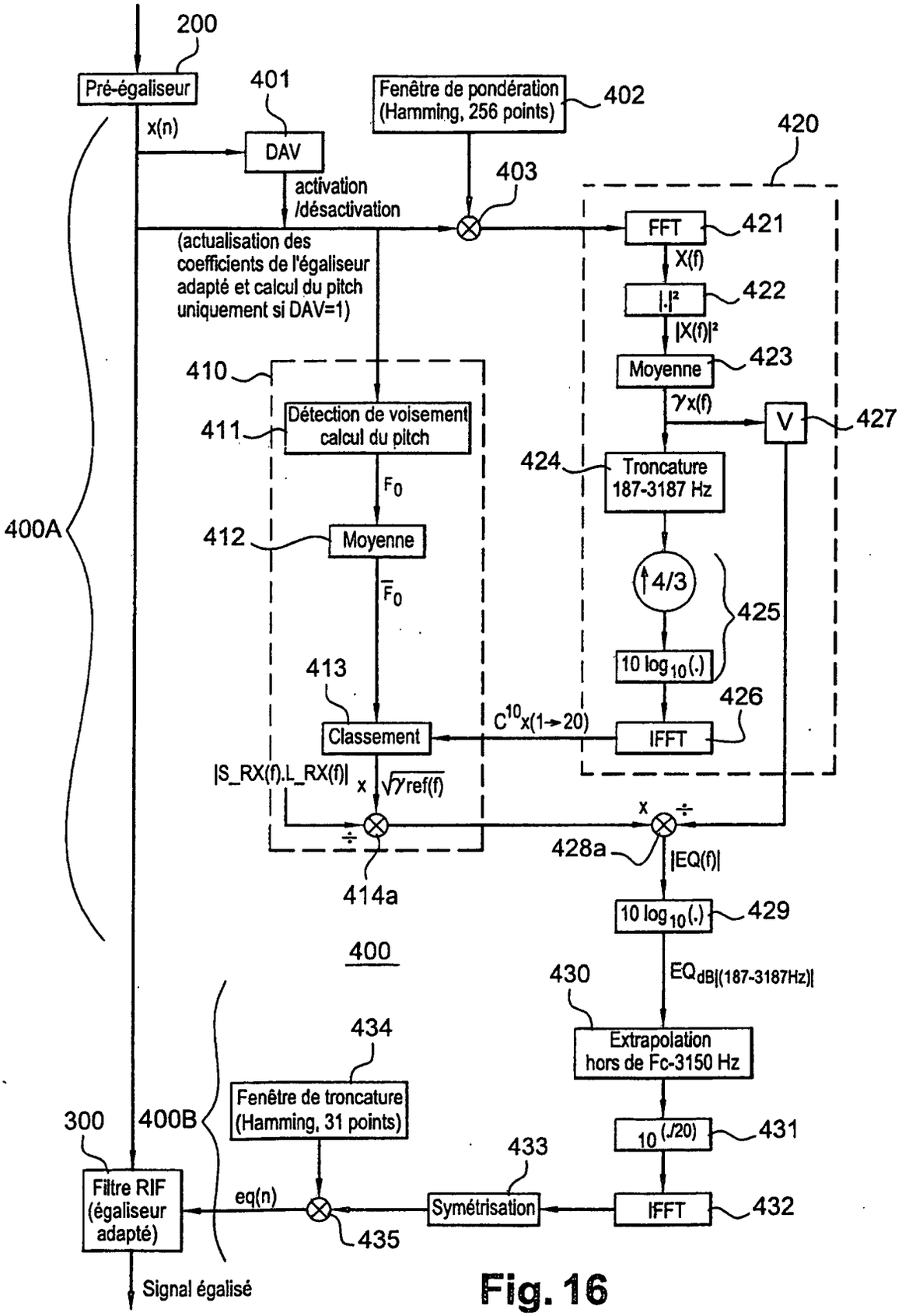

200
Pré-égaliseur
x(n)
401
DAV
activation
/désactivation
(actualisation des coefficients de l'égaliseur adapté et calcul du pitch uniquement si DAV=1)
Fenêtre de pondération (Hamming, 256 points)
402
403
420
FFT
421
X(f)
|.|²
422
|X(f)|²
Moyenne
423
γx(f)
V
427
424
Troncature 187-3187 Hz
4/3
425
10 log10(.)
426
IFFT
C10x(1→20)
410
411
Détection de voisement calcul du pitch
F0
412
Moyenne
F0
413
Classement
|S_RX(f).L_RX(f)|
x √γref(f)
400A
414a
x
÷
428a
|EQ(f)|
10 log10(.)
429
400
430
EQdB|(187-3187Hz)|
Extrapolation hors de Fc-3150 Hz
434
Fenêtre de troncature (Hamming, 31 points)
300
400B
10 (./20)
431
433
Filtre RIF (égaliseur adapté)
eq(n)
Symétrisation
IFFT
432
435
Signal égalisé

Fig. 16

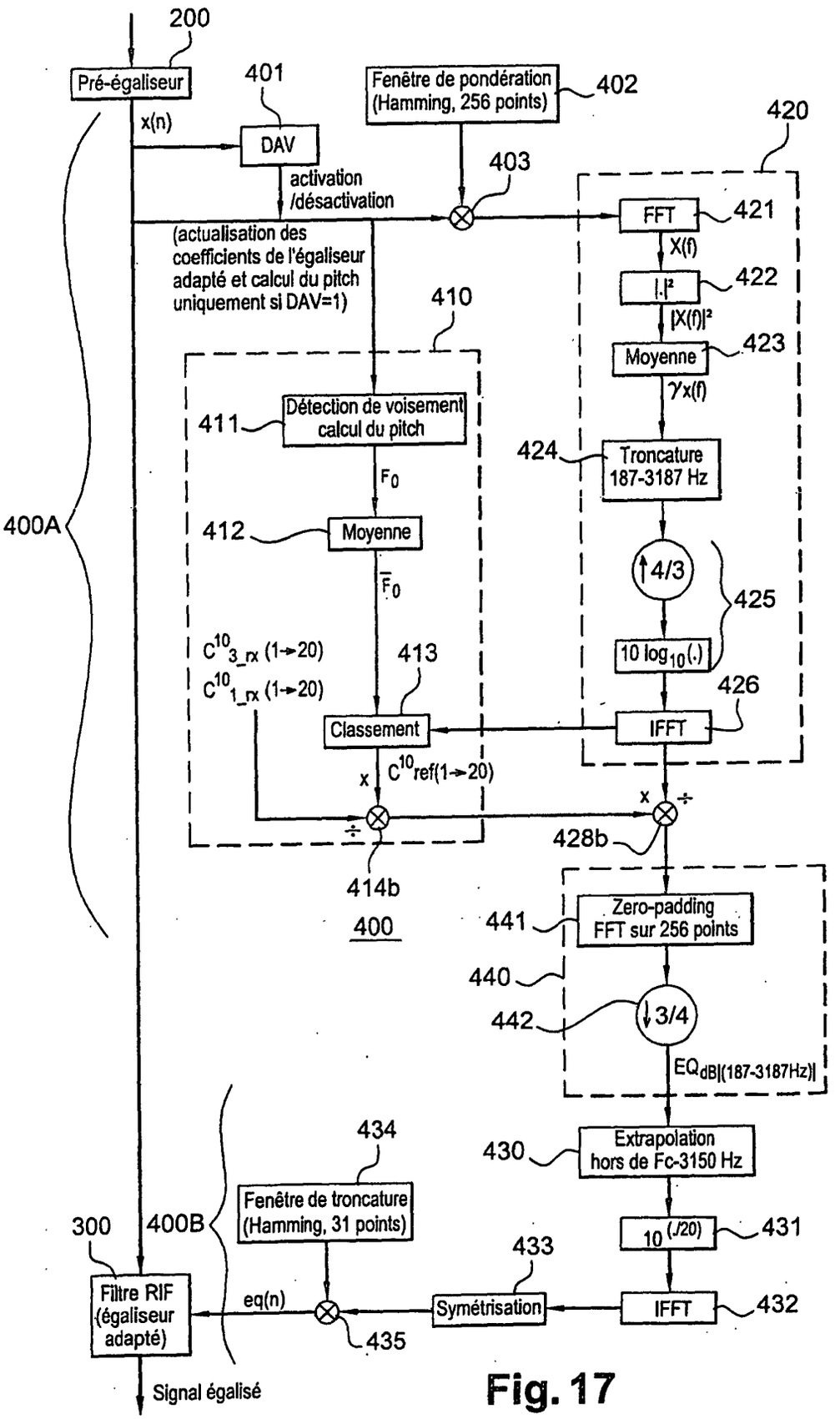

200
Pré-égaliseur
401
DAV
x(n)
activation
/désactivation
(actualisation des coefficients de l'égaliseur adapté et calcul du pitch uniquement si DAV=1)
Fenêtre de pondération (Hamming, 256 points)
402
403
420
FFT
421
X(f)
|.|²
422
|X(f)|²
Moyenne
423
γx(f)
Troncature 187-3187 Hz
424
410
411
Détection de voisement calcul du pitch
F0
412
Moyenne
F̄0
400A
C10 3_rx (1→20)
C10 1_rx (1→20)
413
Classement
4/3
425
10 log10(.)
426
IFFT
x
C10 ref(1→20)
÷
428b
414b
400
441
Zero-padding FFT sur 256 points
440
442
3/4
EQdB|(187-3187Hz)|
434
430
Extrapolation hors de Fc-3150 Hz
Fenêtre de troncature (Hamming, 31 points)
10 (./20)
431
433
300
400B
Filtre RIF (égaliseur adapté)
eq(n)
Symétrisation
IFFT
432
435
Signal égalisé

Fig. 17