

Brevet N° **88189**
du 06.11.1992
Titre délivré 15 AVR. 1993

GRAND-DUCHÉ DE LUXEMBOURG



Monsieur le Ministre
de l'Économie et des Classes Moyennes
Service de la Propriété Intellectuelle
LUXEMBOURG

Demande de Brevet d'Invention

I. Requête

La société dite: KOREA TELECOMMUNICATION AUTHORITY, 100 Sejong-Ro, Chongro-Gu, Séoul, République de Corée, (2)

représentée par Weyland J J Pierre, agissant en qualité de mandataire (3)

dépose(nt) ce 6 novembre 1990 quatre-vingt-douze (4)

à 14.00 heures, au Ministère de l'Économie et des Classes Moyennes, à Luxembourg:

1. la présente requête pour l'obtention d'un brevet d'invention concernant:

Procédés de codage de segments de parole et de contrôle de hauteur de son (5)
pour des systèmes de synthèse de la parole

2. la description en langue française de l'invention en trois exemplaires:

3. 15 planches de dessin. en trois exemplaires:

4. la quittance des taxes versées au Bureau de l'Enregistrement à Luxembourg, le 30.10.1992 ;

5. la délégation de pouvoir, datée de Séoul le 7 octobre 1992 ;

6. le document d'ayant cause (autorisation):

déclare(nt) en assumant la responsabilité de cette déclaration, que l'(es) inventeur(s) est (sont): (6)

Chong Rak Lee, 13, Wolgye-Dong, Nowon-Gu, Séoul, République de Corée
Yong Kyu Park, 142-14, Chuneui-Dong, Puchun-City, Kyungki-Do, République de Corée

revendique(nt) pour la susdite demande de brevet la priorité d'une (des) demande(s) de (7)

brevet déposée(s) en (8) République de Corée

le (9) 6 novembre 1991

sous le N° (10) 91-19617

au nom de (11) la demanderesse

élit(élisent) domicile pour lui (elle) et, si désigné, pour son mandataire, à Luxembourg

MARKS & CLERK, 18-20 rue Glésener, L-1630 LUXEMBOURG (12)

solicite(nt) la délivrance d'un brevet d'invention pour l'objet décrit et représenté dans les annexes susmentionnées,

avec ajournement de cette délivrance à - mois. (13)

Le déposant/ mandataire: [Signature] (14)

II. Procès-verbal de Dépôt

La susdite demande de brevet d'invention a été déposée au Ministère de l'Économie et des Classes Moyennes,

Service de la Propriété Intellectuelle à Luxembourg, en date du: 6 novembre 1992

à 14.00 heures

Pr. le Ministre de l'Économie et des Classes Moyennes,

Le chef du service de la propriété intellectuelle,



A 68007

EXPLICATIONS RELATIVES AU FORMULAIRE DÉPOSÉ:
(1) s'il y a lieu "Demande de certificat d'addition au brevet principal, à la demande de brevet principal No du". (2) inscrire les nom, prénom, profession, adresse du demandeur, lorsque celui-ci est un particulier ou les dénomination sociale, forme juridique, adresse du siège social, lorsque le demandeur est une personne morale - (3) inscrire les nom, prénom, adresse du mandataire agréé, conseil en propriété industrielle, muni d'un pouvoir spécial, s'il y a lieu: "représenté par agissant en qualité de mandataire" - (4) date de dépôt en toutes lettres - (5) titre de l'invention - (6) inscrire les noms, prénoms, adresses des inventeurs ou l'indication "(voir) désignation séparée (suivra)", lorsque la désignation se fait ou se fera dans un document séparé, ou encore l'indication "ne pas mentionner", lorsque l'inventeur signe ou signera un document de non-mention à joindre à une désignation séparée présente ou future - (7) brevet, certificat d'addition, modèle d'utilité, brevet européen (CBE), protection internationale (PCT) - (8) Etat dans lequel le premier dépôt a été effectué ou, le cas échéant, Etats désignés dans la demande européenne ou internationale prioritaire - (9) date du premier dépôt - (10) numéro du premier dépôt complet, le cas échéant, par l'institution de l'office receveur - (11) PCT - (12) nom du titulaire du premier dépôt - (13) adresse du domicile effectif ou élu au Grand-Duché de Luxembourg - (14) 1, 2 ou 18 mois - (15)

Brevet N°

88189

du 06.11.1992

Titre délivré

GRAND-DUCHÉ DE LUXEMBOURG



Monsieur le Ministre
de l'Économie et des Classes Moyennes
Service de la Propriété Intellectuelle
LUXEMBOURG

Demande de Brevet d'Invention

I. Requête

La société dite: KOREA TELECOMMUNICATION AUTHORITY, 100 Sejong-Ro (2)
Chongro-Gu, Séoul, République de Corée,

représentée par Weyland J J Pierre, agissant en qualité de mandataire (3)

dépose(nt) ce 6 novembre 1990 quatre-vingt-douze (4)

à 14.00 heures, au Ministère de l'Économie et des Classes Moyennes, à Luxembourg:

1. la présente requête pour l'obtention d'un brevet d'invention concernant:

Procédés de codage de segments de parole et de contrôle de hauteur de son (5)
pour des systèmes de synthèse de la parole

2. la description en langue française de l'invention en trois exemplaires;

3. 15 planches de dessin, en trois exemplaires;

4. la quittance des taxes versées au Bureau de l'Enregistrement à Luxembourg, le 30.10.1992;

5. la délégation de pouvoir, datée de Séoul le 7 octobre 1992;

6. le document d'ayant cause (autorisation);

déclare(nt) en assumant la responsabilité de cette déclaration, que l'(es) inventeur(s) est (sont): (6)

Chong Rak Lee, 13, Wolgye-Dong, Nowon-Gu, Séoul, République de Corée
Yong Kyu Park, 142-14, Chuneui-Dong, Puchun-City, Kyungki-Do, République
de Corée

revendique(nt) pour la susdite demande de brevet la priorité d'une (des) demande(s) de (7)

brevet déposée(s) en (8) République de Corée

le (9) 6 novembre 1991

sous le N° (10) 91-196179/

au nom de (11) la demanderesse

élit(élisent) domicile pour lui (elle) et, si désigné, pour son mandataire, à Luxembourg

MARKS & CLERK, 18-20 rue Glesener, L-1630 LUXEMBOURG (12)

solicite(nt) la délivrance d'un brevet d'invention pour l'objet décrit et représenté dans les annexes susmentionnées,

avec ajournement de cette délivrance à - mois. (13)

Le déposant / mandataire: (14)

II. Procès-verbal de Dépôt

La susdite demande de brevet d'invention a été déposée au Ministère de l'Économie et des Classes Moyennes.
Service de la Propriété Intellectuelle à Luxembourg, en date du: 6 novembre 1992

à 14.00 heures

Pr. le Ministre de l'Économie et des Classes Moyennes,

Le chef du service de la propriété intellectuelle,

A 68007

EXPLICATIONS RELATIVES AU FORMULAIRE PRÉPRÉPARÉ
(1) s'il y a lieu "Demande de certificat d'addition au brevet principal, à la demande de brevet principal No. du" - (2) inscrire les nom, prénom, profession, adresse du demandeur, lorsque celui-ci est un particulier ou les dénomination sociale, forme juridique, adresse du siège social, lorsque le demandeur est une personne morale - (3) inscrire les nom, prénom, adresse du mandataire agréé, conseil en propriété industrielle, muni d'un pouvoir spécial, s'il y a lieu: "représenté par agissant en qualité de mandataire" - (4) date de dépôt en toutes lettres - (5) titre de l'invention - (6) inscrire les noms, prénoms, adresses des inventeurs ou l'indication "(voir) désignation séparée (suivra)", lorsque la désignation se fait ou se fera dans un document séparé, ou encore l'indication "ne pas mentionner", lorsque l'inventeur signe ou signera un document de non-mention à joindre à une désignation séparée présente ou future - (7) brevet, certificat d'addition, modèle d'utilité, brevet européen (CBE), protection internationale (PCT) - (8) Etat dans lequel le premier dépôt a été effectué ou, le cas échéant, Etats désignés dans la demande européenne ou internationale prioritaire - (9) date du premier dépôt - (10) numéro du premier dépôt complet, le cas échéant, par l'indication de l'office récepteur (CBE/PCT) - (11) nom du titulaire du premier dépôt - (12) adresse du domicile effectif ou élu au Grand-Duché de Luxembourg - (13) 2, 6, 12 ou 18 mois - (14)

Ref.: 92276-1

Revendication de la priorité
d'une demande de brevet déposée
en République de Corée, le 6 novembre 1991
sous le N° 91-19617

MEMOIRE DESCRIPTIF

déposé à l'appui d'une demande de

BREVET D'INVENTION

Société dite:

KOREA TELECOMMUNICATION AUTHORITY
100, Sejong-Ro, Chongro-Gu
Séoul, République de Corée

Désignation:

**PROCEDES DE CODAGE DE SEGMENTS DE PAROLE ET DE
CONTROLE DE HAUTEUR DE SON POUR DES SYSTEMES
DE SYNTHESE DE LA PAROLE**

La présente invention concerne un système de synthèse de la parole ainsi qu'un procédé de synthétisation de la parole et plus particulièrement, un procédé de codage de segments de parole et de contrôle de fréquence fondamentale (appelée par la suite hauteur de son) qui améliore de façon significative la qualité de la parole synthétisée.

Le principe de la présente invention peut s'appliquer directement non seulement à la synthèse de la parole mais également à la synthèse d'autres sons tels que les sons d'instruments de musique ou de chants dont chacun présente une propriété similaire à celle d'une parole, ou à un codage de parole ou conversion de vitesse de parole à vitesse très faible. La présente invention est décrite ci-après en se concentrant sur la synthèse de la parole.

Il existe des procédés de synthèse de la parole pour mettre en oeuvre un système de synthèse par transformation de texte en parole qui permet de synthétiser des vocabulaires sans limitation en convertissant un texte, c'est-à-dire des chaînes de caractères, en parole. Cependant, un procédé qui est facile à mettre en oeuvre et qui est le plus généralement utilisé est constitué par le procédé de synthèse segmentaire de la parole appelé également procédé de synthèse par concaténation dans lequel la voie humaine est échantillonnée et analysée selon des unités phonétiques telles que des demi-syllabes ou diphtonges afin d'obtenir des segments de parole courts qui sont ensuite codés et stockés dans une mémoire et ainsi, lorsque le texte est entré, il est converti en transcriptions phonétiques. Des segments de parole

correspondant aux transcriptions phonétiques sont ensuite retrouvés séquentiellement dans la mémoire et décodés pour synthétiser la parole correspondant au texte entré.

Dans ce type de procédé de synthèse de parole segmentaire, l'un des éléments les plus importants de ceux qui gouvernent la qualité de la parole synthétisée est constitué par le procédé de codage des segments de parole. Dans le procédé de synthèse segmentaire de parole de l'art antérieur du système de synthèse de parole, un procédé de codage de la voix (appelé communément procédé de vocodage) présentant une qualité de parole faible est principalement utilisé en tant que procédé de codage de parole pour stocker des segments de parole. Cependant, ceci constitue l'une des causes les plus importantes d'abaissement de la qualité de la parole synthétisée. Une brève description en relation avec le procédé de codage de segments de parole de l'art antérieur suit.

Le procédé de codage de parole peut être grossièrement éclaté en un procédé de codage de forme d'onde présentant une bonne qualité de parole et en un procédé de codage de la voix présentant une qualité de parole faible. Puisque le procédé de codage de forme d'onde est un procédé qui est destiné à transférer la forme d'onde de parole telle qu'elle est, il est très difficile de modifier la fréquence et la durée de la hauteur de son de telle sorte qu'il est impossible de régler l'intonation et la vitesse de la parole lorsque l'on effectue la synthèse de la parole. Par ailleurs, il est impossible de lier les segments de parole entre eux sans discontinuité de telle sorte que le procédé de codage de forme d'onde ne convient fondamentalement pas pour coder les segments de parole.

Au contraire, lorsque le procédé de codage de la voix (également appelé procédé d'analyse-synthèse) est

utilisé, le motif de hauteur de son et la durée du segment de parole peuvent être modifiés arbitrairement. En outre, puisque les segments de parole peuvent également être liés sans discontinuité en interpolant les paramètres d'estimation d'enveloppe spectrale de telle sorte que le procédé de codage de la voix convienne pour le moyen de codage destiné à la synthèse par transformation de texte en parole, des procédés de codage de parole tels qu'un codage à prédiction linéaire (LPC) ou qu'un codage de la voix par formants sont adoptés dans la plupart des systèmes actuels de synthèse de la parole. Cependant, puisque la qualité de la parole décodée est faible lorsque la parole est codée en utilisant le procédé de codage de la voix, la parole synthétisée obtenue en décodant les segments de parole stockés et en les concaténant ne peut pas présenter une qualité de parole meilleure que celle offerte par le procédé de codage de la voix.

Les tentatives faites jusqu'à ce jour pour améliorer la qualité de parole offerte par le procédé de codage de la voix remplacent le train d'impulsions utilisé par un signal d'excitation qui présente une forme d'onde moins artificielle. L'une de ces tentatives a consisté à utiliser une forme d'onde présentant une caractéristique de pic inférieure à celle de l'impulsion, par exemple une forme d'onde triangulaire ou une forme d'onde demi-circulaire ou une forme d'onde similaire à une impulsion de la glotte. Une autre tentative a été faite pour sélectionner une impulsion de hauteur de son échantillonnée sur une période de hauteur de son de signal résiduel ou sur certaines de ces périodes, l'impulsion étant obtenue par filtrage inverse, et pour utiliser en lieu et place de l'impulsion une impulsion d'échantillonnage pour la totalité de la période temporelle ou pour une période temporelle significativement longue. Cependant, ces

tentatives pour remplacer l'impulsion par une impulsion d'excitation d'autres formes d'onde n'ont pas amélioré la qualité de la parole ou ne l'ont amélioré éventuellement que légèrement et n'ont jamais permis d'obtenir une parole synthétisée présentant une qualité proche de celle d'une parole naturelle.

L'objet de la présente invention consiste à effectuer une synthèse de la parole de qualité élevée présentant un caractère naturel et une intelligibilité aussi bonne que celle de la parole humaine en utilisant un nouveau procédé de codage de segment de parole permettant d'obtenir une bonne qualité de parole et un contrôle de hauteur de son de bonne qualité. Le procédé de la présente invention combine les mérites du procédé de codage de forme d'onde qui permet d'obtenir une bonne qualité de parole sans permettre le contrôle de la hauteur de son et du procédé de codage de la voix qui permet de contrôler la hauteur de son mais qui présente une qualité de parole faible.

La présente invention utilise un procédé de décomposition de forme d'onde périodique qui est un procédé de codage qui décompose un signal contenu dans un secteur de son voisé de la parole originale en ondelettes équivalentes à des formes d'onde de parole d'une période constituées par des impulsions de la glotte afin de coder et de stocker le signal décomposé, et un procédé de relocalisation d'ondelette basé sur des modifications temporelles qui est un procédé de synthèse de forme d'onde permettant de régler arbitrairement la durée et la fréquence de hauteur de son du segment de parole tout en maintenant la qualité de la parole originale en sélectionnant des ondelettes les plus proches de positions où des ondelettes doivent être placées parmi des ondelettes stockées puis en décodant les ondelettes sélectionnées et en les

superposant. Pour la présente invention, des sons musicaux sont traités en tant que sons voisés.

Les objets précédents peuvent être considérés comme offrant essentiellement quelques-unes des caractéristiques et applications les plus pertinentes de l'invention. Bon nombre d'autres résultats avantageux peuvent être obtenus en appliquant l'invention décrite d'une manière différente ou en modifiant l'invention tout en restant dans le cadre de la description. Par conséquent, d'autres objets qui impliquent une compréhension plus complète de l'invention peuvent être obtenus par report à la fois au résumé de l'invention et à la description détaillée qui suit, lesquels décrivent le mode de réalisation particulier en plus du cadre de l'invention, en conjonction avec les dessins annexés.

Des procédés de codage de segment de parole et de contrôle de hauteur de son pour des systèmes de synthèse de la parole de la présente invention sont décrits au travers de modes de réalisation spécifiques représentés sur les dessins annexés. Pour résumer l'invention, on peut dire que l'invention concerne un procédé permettant de synthétiser la parole qui s'approche de la qualité d'une parole naturelle en réglant sa durée et sa fréquence de hauteur de son au moyen d'ondelettes de codage de forme d'onde de chaque période, en les stockant dans une mémoire et, à l'instant de la synthèse, en les décodant et en les localisant en des points temporels appropriés de telle sorte qu'elles présentent le motif de hauteur de son souhaité puis en les superposant afin de générer une parole naturelle, un chant, de la musique et autre.

La présente invention inclut un procédé de codage de segment de parole destiné à une utilisation avec un système de synthèse de la parole, le procédé comprenant la formation d'ondelettes en obtenant des paramètres

qui représentent une enveloppe spectrale pour chaque intervalle temporel d'analyse. Ceci est effectué en analysant un signal numérique périodique ou quasi-périodique tel qu'une parole voisée avec la technique d'estimation de spectre. Un signal original est tout d'abord soumis à une déconvolution pour obtenir une réponse impulsionnelle représentée par les paramètres d'enveloppe spectrale et un signal de train d'impulsions de hauteur de son périodique ou quasi-périodique présentant une enveloppe spectrale pratiquement plate. Un signal d'excitation est obtenu en rajoutant des échantillons de valeur zéro après un signal impulsionnel de hauteur de son d'une période obtenu en segmentant le signal de train d'impulsions de hauteur de son période par période de telle sorte qu'une impulsion de hauteur de son soit contenue dans chaque période et qu'une réponse impulsionnelle correspondant à un jeu de paramètres d'enveloppe spectrale, dans le même intervalle temporel que celui du signal d'excitation, soit soumise à une convolution pour former une ondelette pour cette période.

Les ondelettes, au lieu d'être formées au moyen d'un codage de forme d'onde et d'être stockées dans une mémoire à l'avance, peuvent être formées en faisant correspondre une information obtenue par codage de forme d'onde d'un signal impulsionnel de hauteur de son de chaque intervalle de période obtenu par segmentation à une information obtenue par codage d'un jeu de paramètres d'estimation d'enveloppe spectrale présentant le même intervalle temporel que celui de l'information mentionnée ci-avant, ou à une réponse impulsionnelle correspondant aux paramètres et en stockant l'information d'ondelette dans une mémoire. Il existe deux procédés de production d'une parole de synthèse en utilisant l'information d'ondelette stockée dans une mémoire. Le premier procédé consiste à former

chaque ondelette en effectuant la convolution du signal d'excitation obtenu en rajoutant des échantillons de valeur zéro après un signal impulsif de hauteur de son d'une période obtenu en décodant l'information et d'une réponse impulsionnelle correspondant aux paramètres d'enveloppe spectrale décodés dans le même intervalle temporel en tant que signal d'excitation puis à assigner les ondelettes à des points temporels appropriés de telle sorte qu'elles présentent un motif de hauteur de son souhaité et un motif de durée souhaité, à les localiser aux points temporels puis à les superposer.

Le second procédé consiste à former un signal d'excitation synthétique en assignant les signaux impulsifs de hauteur de son obtenus en décodant l'information d'ondelette à des points temporels appropriés de telle sorte qu'ils présentent un motif de hauteur de son souhaité et un motif de durée souhaité et en les localisant aux points temporels, à former un jeu de paramètres d'enveloppe spectrale synthétique, soit en comprimant, soit en dilatant temporellement le jeu de fonctions temporelles des paramètres sur une base sous-segment par sous-segment, selon que la durée d'un sous-segment contenu dans un segment de parole qui doit être synthétisé est plus courte ou plus longue que la durée d'un sous-segment correspondant du segment de parole original, ou en localisant le jeu de fonctions temporelles des paramètres d'une période de façon synchrone par rapport au signal impulsif de hauteur de son en correspondance d'une période localisée pour former le signal d'excitation synthétique, et à effectuer la convolution du signal d'excitation synthétique et de la réponse impulsionnelle correspondant au jeu de paramètres d'enveloppe spectrale synthétique en utilisant un filtre variable en fonction du temps ou en utilisant une technique de

convolution rapide basée sur une transformée de Fourier rapide (FFT). Dans le dernier procédé, un intervalle de suppression se produit lorsqu'une période de hauteur de son souhaitée est plus longue que la période de hauteur de son originale et un intervalle de chevauchement se produit lorsque la période de hauteur de son souhaitée est plus courte que la période de hauteur de son originale.

Dans l'intervalle de chevauchement, le signal d'excitation synthétique est obtenu en ajoutant les signaux impulsionnels de hauteur de son en chevauchement les uns aux autres ou en sélectionnant l'un d'eux et le paramètre d'enveloppe spectrale est obtenu en sélectionnant l'un quelconque des paramètres d'enveloppe spectrale en chevauchement ou en utilisant une valeur moyenne de deux paramètres en chevauchement.

Dans l'intervalle de suppression, le signal d'excitation synthétique est obtenu par remplissage avec des échantillons de valeur zéro et le paramètre d'enveloppe spectrale synthétique est obtenu par répétition des valeurs des paramètres d'enveloppe spectrale à des points de début et de fin des périodes précédente et suivante, avant et après le centre de l'intervalle de suppression, ou par répétition de l'une des deux valeurs ou d'une valeur moyenne des deux valeurs, ou par remplissage avec des valeurs et par connexion sans discontinuité des deux valeurs.

La présente invention inclut en outre un procédé de contrôle de hauteur de son d'un système de synthèse de la parole permettant de contrôler la durée et la hauteur de son d'un segment de parole au moyen d'un procédé de relocalisation d'ondelette basé sur des modifications temporelles qui rend possible la synthèse de la parole selon pratiquement la même qualité que celle de la parole naturelle, en codant des points temporels de limite importants tels que le point de

début, le point de fin et les points de régime permanent d'un segment de parole, et des positions d'impulsion de hauteur de son de chaque ondelette ou de chaque signal impulsionnel de hauteur de son et en les stockant dans une mémoire simultanément à l'instant de stockage de chaque segment de parole et, au moment de la synthèse, en obtenant une fonction de modification temporelle en comparant des points temporels de limite souhaités et des points temporels de limite originaux stockés correspondant aux points temporels de limite souhaités, en retrouvant les points temporels originaux correspondant à chaque position d'impulsion de hauteur de son souhaitée en utilisant la fonction de modification temporelle, en sélectionnant des ondelettes présentant des positions d'impulsion de hauteur de son les plus proches des points temporels originaux, en les localisant en des positions d'impulsion de hauteur de son souhaitées et en superposant les ondelettes.

Le procédé de contrôle de hauteur de son peut inclure en outre la production d'une parole de synthèse en sélectionnant des signaux impulsionnels de hauteur de son d'une période et des paramètres d'enveloppe spectrale correspondant aux signaux impulsionnels de hauteur de son au lieu des ondelettes et en les localisant, en effectuant la convolution des signaux impulsionnels de hauteur de son localisés et de la réponse impulsionnelle correspondant aux paramètres d'enveloppe spectrale afin de produire des ondelettes et en superposant les ondelettes produites ou en effectuant la convolution d'un signal d'excitation synthétique obtenu en superposant les signaux impulsionnels de hauteur de son localisés et d'une réponse impulsionnelle variable en fonction du temps correspondant à des paramètres d'enveloppe spectrale

synthétique obtenus par concaténation des paramètres d'enveloppe spectrale localisée.

Un dispositif de synthèse de la parole voisée d'un système de synthèse de parole est décrit et il inclut un sous-bloc de décodage 9 produisant une information d'ondelette en décodant des codes d'ondelette à partir du bloc de stockage de segment de parole 5. Un sous-bloc de contrôle de durée 10 produit des données de modification temporelle à partir de l'entrée de données de durée en provenance d'un sous-système de génération de prosodie 2 et à partir de points temporels de limite inclus dans une information d'en-tête en provenance du bloc de stockage de segment de parole 5. Un sous-bloc de contrôle de hauteur de son 11 produit une information de position d'impulsion de hauteur de son de telle sorte qu'elle présente un motif d'intonation tel qu'indiqué par des données de motif d'intonation à partir de l'entrée de l'information d'en-tête en provenance du bloc de stockage de segment de parole 5, à partir des données de motif d'intonation en provenance du système de génération de prosodie et à partir de l'information de modification temporelle en provenance du sous-bloc de contrôle de durée 10. Un sous-bloc de contrôle d'énergie 12 produit une information de gain de telle sorte que la parole synthétisée présente le motif de sollicitation tel qu'indiqué par des données de motif de sollicitation à partir de l'entrée des données de motif de sollicitation en provenance du sous-système de génération de prosodie 2, à partir de l'information de modification temporelle en provenance du sous-bloc de contrôle de durée 10 et à partir de l'information de position d'impulsion de hauteur de son en provenance du sous-bloc de contrôle de hauteur de son 11. Un sous-bloc d'assemblage de forme d'onde 13 produit un signal de parole voisé à partir de l'entrée de l'information

d'ondelette en provenance du sous-bloc de décodage 9, à partir de l'information de modification temporelle en provenance du sous-bloc de contrôle de durée 10, à partir de l'information de position d'impulsion de hauteur de son en provenance du sous-bloc de contrôle de hauteur de son 11 et à partir de l'information de gain en provenance du sous-bloc de contrôle d'énergie 12.

Ainsi, selon la présente invention, un texte est entré sur le sous-système de pré-traitement phonétique 1 où il est converti en symboles de transcription phonétique et en données d'analyse syntaxique. Les données d'analyse syntaxique sont émises en sortie vers un sous-système de génération de prosodie 2. Le sous-système de génération de prosodie 2 émet en sortie une information de prosodie pour le sous-système de concaténation de segments de parole 3. Les symboles de transcription phonétique émis en sortie depuis le sous-système de pré-traitement sont également entrés sur le sous-système de concaténation de segments de parole 3. Les symboles de transcription phonétique sont ensuite entrés dans le bloc de sélection de segment de parole 4 et les données de prosodie correspondantes sont entrées dans le bloc de synthèse de son voisé 6 ainsi que dans le bloc de synthèse de son non voisé 7. Dans le bloc de sélection de segment de parole 4, chaque symbole de transcription phonétique entré est mis en correspondance avec une unité de synthèse de segment de parole correspondante et une adresse de mémoire correspondante de l'unité de synthèse correspondante qui correspond à chaque symbole de transcription phonétique d'entrée est trouvée à partir d'une table de segments de parole contenue dans le bloc de stockage de segment de parole 5. L'adresse de l'unité de synthèse correspondante est alors émise en sortie dans le bloc de stockage de segment de parole 5 où le segment de

parole correspondant qui se présente sous la forme d'une ondelette codée est sélectionné pour chacune des adresses des unités de synthèse en correspondance. Le segment de parole sélectionné qui se présente sous la forme d'une ondelette codée est émis en sortie dans le bloc de synthèse de son voisé 6 pour le son voisé et dans le bloc de synthèse de son non voisé 7 pour le son non voisé. Le bloc de synthèse de son voisé 6 qui utilise le procédé de relocalisation d'ondelette basé sur la modification temporelle afin de synthétiser un son de parole et le bloc de synthèse de son non voisé 7 émettent en sortie des signaux de parole synthétique numériques pour le convertisseur numérique/analogique afin de convertir les signaux numériques d'entrée en signaux analogiques qui sont les sons de parole synthétisée.

Afin d'utiliser la présente invention, une parole et/ou une musique est tout d'abord enregistrée sur une bande magnétique. Le son résultant est ensuite converti de la forme signaux analogiques à la forme signaux numériques au moyen d'un filtrage passe-bas des signaux analogiques puis en appliquant les signaux filtrés à un convertisseur analogique/numérique. Les signaux de parole numérisés résultants sont ensuite segmentés en un certain nombre de segments de parole comportant des sons qui correspondent à des unités de synthèse telles que des phonèmes, des diphones, des demi-syllabes et autre, en utilisant les outils d'édition de parole connus. Chaque segment de parole résultant est ensuite différencié en segments de parole voisée et non voisée en utilisant des outils connus de détection et d'édition de parole voisée/non voisée. Les segments de parole non voisée sont codés au moyen de procédés de codage de la voix connus qui utilisent un bruit aléatoire en tant que source de parole non voisée. Les procédés de codage de la voix incluent le procédé de

codage à prédiction linéaire (LPC), le procédé homomorphique, le procédé de codage de la voix par formants et autre.

Les segments de parole voisée sont utilisés pour former des ondelettes $s_j(n)$ conformément au procédé décrit ci-après en relation avec la figure 4. Les ondelettes $s_j(n)$ sont ensuite codées en utilisant un procédé de codage de forme d'onde approprié. Les procédés de codage de forme d'onde connus incluent le procédé de modulation par impulsions et codage (MIC), le procédé de modulation par impulsions et codage différentielle adaptable (MICDA), le procédé de codage prédictif adaptatif (APC) et autre. Les segments de parole voisée codés résultants sont stockés dans le bloc de stockage de segment de parole 5, comme représenté sur les figures 6A et 6B. Les segments de parole non voisée codés sont également stockés dans le bloc de stockage de segment de parole 5.

Les caractéristiques les plus pertinentes et importantes de la présente invention ont été mises en exergue ci-avant afin que la description détaillée de l'invention qui suit soit mieux comprise et que la présente contribution à l'art puisse être pleinement appréciée. Des caractéristiques supplémentaires de l'invention décrite ci-après constituent le contenu de l'invention. L'homme de l'art peut apprécier que la conception et que le mode de réalisation spécifique décrits ici puissent être facilement utilisés en tant que base pour modifier ou concevoir d'autres structures afin de mettre en oeuvre les mêmes objets de la présente invention. En outre, l'homme de l'art peut réaliser que ces constructions équivalentes ne s'écartent ni du cadre ni de l'esprit de l'invention.

Pour une meilleure compréhension de la nature et des objets de l'invention, référence doit être faite à

la description détaillée qui suit que l'on lira en relation avec les dessins annexés parmi lesquels :

la figure 1 représente le système de synthèse par transformation de texte en parole du procédé de synthèse de segment de parole ;

la figure 2 représente le sous-système de concaténation de segments de parole ;

les figures 3A à 3T représentent des formes d'onde destinées à expliquer le principe du procédé de décomposition de forme d'onde périodique et le procédé de relocalisation d'ondelette selon la présente invention ;

la figure 4 représente un schéma fonctionnel destiné à expliquer le procédé de décomposition de forme d'onde périodique ;

les figures 5A à 5E représentent des schémas fonctionnels destinés à expliquer la procédure du procédé de déconvolution aveugle ;

les figures 6A et 6B représentent des formats de code pour l'information de segment de parole voisée stockée au niveau du bloc de stockage de segment de parole ;

la figure 7 représente le bloc de synthèse de parole voisée selon la présente invention ; et

les figures 8A et 8B représentent des graphiques destinés à expliquer le procédé de contrôle de durée et de hauteur de son selon la présente invention.

Des caractères de référence similaires se rapportent à des parties similaires pour l'ensemble des vues des dessins.

La structure du système de synthèse par transformation de texte en parole du procédé de synthèse de segment de parole de l'art antérieur est constituée par trois sous-systèmes :

A. un sous-système de pré-traitement phonétique
(1) ;

B. un sous-système de génération de prosodie (2) ; et

C. un sous-système de concaténation de segments de parole (3), comme représenté sur la figure 1. Lorsque le texte est entré depuis un clavier, un calculateur ou tout autre système dans le système de synthèse par transformation de texte en parole, le sous-système de pré-traitement phonétique (1) analyse la syntaxe du texte puis transforme le texte en une chaîne de symboles de transcription phonétique en lui appliquant des règles de recodage phonétique. Le sous-système de génération de prosodie (2) génère des données de motif d'intonation et des données de motif de sollicitation en utilisant des données d'analyse syntaxique de telle sorte qu'une intonation et qu'une sollicitation appropriées puissent être appliquées à la chaîne de symboles de transcription phonétique puis il émet en sortie les données pour le sous-système de concaténation de segments de parole (3). Le sous-système de génération de prosodie (2) produit également les données en relation avec la durée de chaque phonème pour le sous-système de concaténation de segments de parole (3).

Les trois données de prosodie mentionnées ci-avant, c'est-à-dire les données de motif d'intonation, les données de motif de sollicitation et les données concernant la durée de chaque phonème, sont en général envoyées au sous-système de concaténation de segments de parole (3) avec la chaîne de symboles de transcription phonétique générée par le sous-système de pré-traitement phonétique (1) bien qu'elles puissent être transférées au sous-système de concaténation de segments de parole (3) indépendamment de la chaîne de symboles de transcription phonétique.

Le sous-système de concaténation de segments de parole (3) génère une parole continue en recherchant

séquentiellement des segments de parole appropriés qui sont codés et stockés dans sa mémoire en relation avec la chaîne de symboles de transcription phonétique (non représentée) et en les décodant. A cet instant, le sous-système de concaténation de segments de parole (3) peut générer une parole de synthèse présentant l'intonation, la sollicitation et la vitesse de parole telles qu'attendues par le sous-système de génération de prosodie (2) en contrôlant l'énergie (l'intensité), la durée et la période de hauteur de son de chaque segment de parole en relation avec l'information de prosodie.

La présente invention améliore notablement la qualité de la parole par comparaison avec une parole synthétisée de l'art antérieur en améliorant le procédé de codage destiné à stocker les segments de parole dans le sous-système de concaténation de segments de parole (3). Une description relative au fonctionnement du sous-système de concaténation de segments de parole (3) par report à la figure 2 suit.

Lorsque la chaîne de symboles de transcription phonétique formée par le sous-système de pré-traitement phonétique (1) est entrée dans le bloc de sélection de segment de parole (4), le bloc de sélection de segment de parole (4) sélectionne séquentiellement les unités de synthèse telles que les diphones et les demi-syllabes en inspectant en continu la chaîne de symboles de transcription phonétique qui arrive et il trouve les adresses des segments de parole correspondant aux unités de synthèse sélectionnées dans sa mémoire, comme décrit dans le tableau 1. Le tableau 1 représente un exemple d'un tableau de segments de parole maintenus dans le bloc de sélection de segment de parole (4) qui sélectionne des segments de parole sur la base de diphones. Ceci aboutit à la formation d'une adresse du

segment de parole sélectionné qui est émise en sortie pour le bloc de stockage de segment de parole (5).

Les segments de parole correspondant aux adresses du segment de parole sont codés conformément au procédé de la présente invention qui sera décrit ultérieurement et sont stockés au niveau des adresses de la mémoire du bloc de stockage de segment de parole (5).

Tableau 1

Symbole de transcription phonétique d'un segment de parole	Adresse mémoire (en hexadécimal)
/ai/	0000
/au/	0021
/ab/	00A3
/ad/	00FF
.	.
.	.
.	.

Lorsque l'adresse du segment de parole sélectionné dans le bloc de sélection de segment de parole (4) est entrée dans le bloc de stockage de segment de parole (5), le bloc de stockage de segment de parole (5) recherche les données de segment de parole correspondantes dans la mémoire contenue dans le bloc de stockage de segment de parole (5) et il l'envoie à un bloc de synthèse de son voisé (6) s'il s'agit d'un son voisé ou d'un son de fricative voisé, ou dans un bloc de synthèse de son non voisé (7) s'il s'agit d'un son non voisé. C'est-à-dire que le bloc de synthèse de son voisé (6) synthétise un signal de parole numérique correspondant aux segments de parole de son voisé ; et le bloc de synthèse de son non voisé (7) synthétise un signal de parole numérique

correspondant au segment de parole de son non voisé. Chaque signal de parole synthétisée numérique du bloc de synthèse de son voisé (6) et du bloc de synthèse de son non voisé (7) est ensuite converti en un signal analogique.

Ainsi, le signal de parole synthétisé numérique résultant émis en sortie depuis le bloc de synthèse de son voisé (6) ou depuis le bloc de synthèse de son non voisé (7) est envoyé à un bloc de conversion numérique/analogique N/A (8) constitué par un convertisseur numérique/analogique, par un filtre passe-bas analogique et par un amplificateur analogique, et il est converti en un signal analogique pour produire un son de parole synthétisée.

Lorsque le bloc de synthèse de son voisé (6) et le bloc de synthèse de son non voisé (7) concatènent les segments de parole, ils produisent la prosodie telle que prévue par le système de génération de prosodie (2) pour la parole synthétisée en réglant de manière appropriée la durée, l'intensité et la fréquence de hauteur de son du segment de parole sur la base de l'information de prosodie, c'est-à-dire les données de motif d'intonation, les données de motif de sollicitation et les données de durée.

La préparation du segment de parole en vue d'un stockage dans le bloc de stockage de segment de parole (5) est la suivante. Une unité de synthèse est tout d'abord sélectionnée. Ces unités de synthèse incluent des unités de phonème, d'allophone, de diphone, de syllabe, de demi-syllabe CVC, VCV, CV, VC ("C" représente un phonème de consonne et "V" représente un phonème de voyelle) ou des combinaisons de celles-ci. Les unités de synthèse qui sont le plus largement utilisées dans le présent procédé de synthèse de la parole sont les dipphones et les demi-syllabes.

Le segment de parole correspondant à chaque élément d'une agrégation des unités de synthèse est segmenté à partir des échantillons de parole qui sont réellement prononcés par un être humain. Par conséquent, le nombre d'éléments de l'agrégation d'unités de synthèse est le même que le nombre de segments de parole. Par exemple, dans le cas où des demi-syllabes sont utilisées en tant qu'unités de synthèse en anglais, le nombre de demi-syllabes est d'environ 1000 et par conséquent, le nombre de segments de parole est également d'environ 1000. En général, ces segments de parole sont constitués par l'intervalle de son non voisé et par l'intervalle de son voisé.

Dans la présente invention, le segment de parole non voisé et le segment de parole voisé obtenus en segmentant le segment de parole de l'art antérieur en intervalles de son non voisé et en intervalles de son voisé sont utilisés en tant qu'unité de synthèse de base. La partie de synthèse de parole de son non voisé est réalisée selon l'art antérieur tel que discuté ci-après. La synthèse de parole de son voisé est réalisée selon la présente invention.

Ainsi, les segments de parole non voisée sont décodés au niveau du bloc de synthèse de son non voisé (7) représenté sur la figure 2. Dans le cas du décodage du son non voisé, il a été noté dans l'art antérieur que l'utilisation d'un signal de bruit blanc aléatoire artificiel en tant que signal d'excitation pour un filtre de synthèse n'aggrave pas ou ne diminue pas la qualité de la parole décodée. Par conséquent, lors du codage et du décodage des segments de parole non voisée, le procédé de codage de la voix de l'art antérieur peut être appliqué tel quel, le bruit blanc étant utilisé dans ce procédé en tant que signal d'excitation. Par exemple, lors de la synthèse de l'art antérieur d'un son non voisé, le signal de bruit blanc

peut être généré par un algorithme de génération de nombres aléatoires et peut être utilisé ou le signal de bruit blanc généré et stocké à l'avance dans une mémoire peut être retrouvé dans la mémoire lors de la synthétisation, ou un signal résiduel obtenu en filtrant l'intervalle de son non voisé de la parole réelle en utilisant un filtre d'enveloppe spectrale inverse et stocké dans une mémoire peut être retrouvé dans la mémoire lors de la synthétisation. S'il n'est pas nécessaire de modifier la durée du segment de parole non voisé, un procédé de codage extrêmement simple peut être utilisé et dans ce procédé, la partie de son non voisée est codée conformément à un procédé de codage de forme d'onde tel qu'un procédé de modulation par impulsions et codage (MIC) ou qu'un procédé de modulation par impulsions et codage différentielle adaptable (MICDA) et est stockée. Elle est ensuite décodée pour être utilisée lors de la synthétisation.

La présente invention concerne un procédé de codage et de synthèse des segments de parole voisée qui gouverne la qualité de la parole de synthèse. Une description relative à un tel procédé qui met l'accent sur le bloc de stockage de segment de parole et sur le bloc de synthèse de son voisé (6) est représentée sur la figure 2.

Les segments de parole voisée pris parmi les segments de parole stockés dans la mémoire du bloc de stockage de segment de parole (5) sont décomposés en ondelettes de composantes périodiques de hauteur de son à l'avance conformément au procédé de décomposition de forme d'onde périodique de la présente invention et sont stockés dedans. Le bloc de synthèse de son voisé (6) synthétise une parole présentant la hauteur de son souhaitée et les motifs de durée souhaités en sélectionnant et en agençant de façon appropriée les

ondelettes conformément au procédé de relocalisation d'ondelette basé sur la modification temporelle. Le principe de ces procédés est décrit ci-après par report aux dessins.

Une parole voisée $s(n)$ est un signal périodique obtenu lorsqu'une onde de glotte périodique générée au niveau des cordes vocales traverse le filtre des voies vocales $V(f)$ constitué par la cavité buccale, par la cavité du pharynx et par la cavité nasale. On suppose ici que le filtre des voies vocales $V(f)$ inclut une caractéristique de fréquence due à un effet de lèvre. Un spectre $S(f)$ d'une parole voisée est caractérisé par :

1. une structure fine qui varie rapidement en relation avec une fréquence " f " ; et
2. une enveloppe spectrale qui varie lentement relativement à cette structure, la structure étant due à une périodicité du signal de parole voisée et l'enveloppe étant due à la réflexion du spectre d'une impulsion de glotte et à une caractéristique de fréquence du filtre des voies vocales.

Le spectre $S(f)$ de la parole voisée prend la même forme que la forme obtenue lorsque la structure fine d'un train d'impulsions due à des composantes harmoniques qui existent à des multiples entiers de la fréquence de hauteur de son F_0 est multipliée par une fonction d'enveloppe spectrale $H(f)$. Par conséquent, une parole voisée $s(n)$ peut être considérée comme étant un signal de sortie lorsqu'un signal de train d'impulsions de hauteur de son périodique $e(n)$ présentant une enveloppe spectrale plate et la même période que la parole voisée $s(n)$ est entré sur un filtre variable en fonction du temps présentant la même caractéristique de réponse en fréquence que la fonction d'enveloppe spectrale $H(f)$ de la parole voisée $s(n)$. Au vu de cela, dans le domaine temporel, la parole voisée

$s(n)$ est une convolution d'une réponse impulsionnelle $h(n)$ du filtre $H(f)$ et du signal de train d'impulsions de hauteur de son périodique $e(n)$. Puisque $H(f)$ correspond à la fonction d'enveloppe spectrale de la parole voisée $s(n)$, le filtre variable en fonction du temps qui a pour caractéristique de réponse en fréquence $H(f)$ est appelé filtre d'enveloppe spectrale ou filtre de synthèse.

Sur la figure 3A, un signal pour quatre périodes d'une forme d'onde de glotte est représenté. Classiquement, les formes d'onde des impulsions de glotte qui composent la forme d'onde de glotte sont similaires les unes aux autres mais pas complètement identiques et par ailleurs, les intervalles temporels qui séparent les impulsions de glotte adjacentes sont similaires les uns aux autres mais pas complètement égaux. Comme décrit ci-avant, la forme d'onde de parole voisée $s(n)$ de la figure 3C est générée lorsque la forme d'onde de glotte $g(n)$ représentée sur la figure 3A est filtrée par le filtre des voies vocales $V(f)$. La forme d'onde de glotte $g(n)$ est constituée par les impulsions de glotte $g_1(n)$, $g_2(n)$, $g_3(n)$ et $g_4(n)$ qui se distinguent les unes des autres en termes d'instant et lorsqu'elles sont filtrées par le filtre des voies vocales $V(f)$, les ondelettes $s_1(n)$, $s_2(n)$, $s_3(n)$ et $s_4(n)$ représentées sur la figure 3B sont générées. La forme d'onde de parole voisée $s(n)$ représentée sur la figure 3C est générée en superposant ces ondelettes.

Un concept fondamental de la présente invention est constitué par le fait que si l'on peut obtenir les ondelettes qui composent un signal de parole voisée en décomposant le signal de parole voisée, on peut synthétiser la parole avec un motif d'accent et d'intonation arbitraire en modifiant l'intensité des ondelettes et les intervalles temporels qui les séparent.

Du fait que la forme d'onde de parole voisée $s(n)$ représentée sur la figure 3C a été générée en superposant les ondelettes qui se chevauchent les unes les autres dans le temps, il est difficile de ré-obtenir les ondelettes à partir de la forme d'onde de parole $s(n)$.

Afin que les formes d'onde de chaque période ne se chevauchent pas les unes les autres dans le domaine temporel, les formes d'onde doivent être des formes d'onde en pic dans lesquelles l'énergie est concentrée autour d'un point temporel, comme on peut le voir sur la figure 3F.

Une forme d'onde en pic est une forme d'onde qui présente une enveloppe spectrale pratiquement plate dans le domaine des fréquences. Lorsqu'une forme d'onde de parole voisée $s(n)$ est donnée, un signal de train d'impulsions de hauteur de son périodique $e(n)$ présentant une enveloppe spectrale plate comme représenté sur la figure 3F peut être obtenu en tant que sortie en estimant l'enveloppe du spectre $S(f)$ de la forme d'onde $s(n)$ et en l'entrant dans un filtre d'enveloppe spectrale inverse $1/H(f)$ qui a pour caractéristique de fréquence l'inverse de la fréquence de la fonction d'enveloppe $H(f)$. Les figures 4, 5A et 5B concernent cette étape.

Du fait que les formes d'onde d'impulsions de hauteur de son de chaque période qui composent le signal de train d'impulsions de hauteur de son périodique $e(n)$ comme représenté sur la figure 3F ne se chevauchent pas les unes les autres dans le domaine temporel, elles peuvent être séparées. Le principe du procédé de décomposition de forme d'onde périodique consiste en ce que, du fait que les "signaux impulsionnels de hauteur de son pour une période" séparés $e_1(n)$, $e_2(n)$, ... présentent un spectre sensiblement plat, s'ils sont ré-entrés sur le filtre

d'enveloppe spectrale $H(f)$ de telle sorte que les signaux présentent le spectre original, alors les ondelettes $s_1(n)$, $s_2(n)$, etc... telles que représentées sur la figure 3B peuvent être obtenues.

La figure 4 est un schéma fonctionnel du procédé de décomposition de forme d'onde périodique de la présente invention selon lequel le segment de parole voisée est analysé en ondelettes. La forme d'onde de parole voisée $s(n)$ qui est un signal numérique est obtenue en limitant la bande du signal de parole voisée analogique ou du signal sonore d'instrument musical avec un filtre passe-bas, en convertissant les signaux résultants en signaux analogiques/numériques et en les stockant sur un disque magnétique sous la forme du format de code du procédé de modulation par impulsions et codage (MIC) par groupage de plusieurs bits à un instant donné puis en les retrouvant pour les traiter lorsque nécessaire.

La première étape d'un processus de préparation d'ondelette selon le procédé de décomposition de forme d'onde périodique est constituée par une déconvolution aveugle dans laquelle la forme d'onde de parole voisée $s(n)$ (signal périodique $s(n)$) subit une déconvolution pour obtenir une réponse impulsionnelle $h_c(n)$ qui est une fonction du domaine temporel de la fonction d'enveloppe spectrale $H(f)$ du signal $s(n)$ et un signal de train d'impulsions de hauteur de son périodique $e(n)$ présentant une enveloppe spectrale plate et la même période que le signal $s(n)$. Voir figures 5A et 5B ainsi que la discussion afférente.

Comme décrit, pour la déconvolution aveugle, la technique d'estimation spectrale au moyen de laquelle la fonction d'enveloppe spectrale $H(f)$ est estimée à partir du signal $s(n)$ est essentielle.

Les techniques d'estimation spectrale de l'art antérieur peuvent être classées selon trois procédés :

1. un procédé d'analyse de bloc ;
2. un procédé d'analyse de hauteur de son synchrone ; et
3. un procédé d'analyse séquentielle dépendant de la longueur d'un intervalle d'analyse.

Le procédé d'analyse de bloc est un procédé dans lequel le signal de parole est divisé en blocs de durée constante de l'ordre de 10 à 20 ms (millisecondes) puis l'analyse est effectuée par rapport au nombre constant d'échantillons de parole existant dans chaque bloc, en obtenant un jeu (classiquement 10-16 paramètres) de paramètres d'enveloppe spectrale pour chaque bloc, procédé pour lequel un procédé d'analyse homomorphique et un procédé d'analyse à prédiction linéaire de bloc sont typiques.

Le procédé d'analyse de hauteur de son synchrone obtient un jeu de paramètres d'enveloppe spectrale pour chaque période en effectuant une analyse sur chaque signal de parole périodique qui a été obtenu en divisant le signal de parole par la période de hauteur de son telle que l'unité (comme représenté sur la figure 3C), procédé pour lequel le procédé d'analyse par synthèse et le procédé d'analyse à prédiction linéaire de hauteur de son synchrone sont typiques.

Selon le procédé d'analyse séquentielle, un jeu de paramètres d'enveloppe spectrale est obtenu pour chaque échantillon de parole (comme représenté sur la figure 3D) en estimant le spectre pour chaque échantillon de parole et pour ce procédé, la méthode des moindres carrés et la méthode des moindres carrés récursive qui sont un type de procédé de filtrage adaptatif sont typiques.

La figure 3D représente une variation au cours du temps des quatre premiers coefficients de réflexion pris parmi quatorze coefficients de réflexion k_1, k_2, \dots, k_{14} qui constituent un jeu de paramètres

d'enveloppe spectrale obtenu au moyen du procédé d'analyse séquentielle. (Se reporter S.V.P à la figure 5A). Comme on peut le voir sur le dessin, les valeurs des paramètres d'enveloppe spectrale varient en continu du fait du déplacement continu des organes d'articulation, ce qui signifie que la réponse impulsionnelle $h(n)$ du filtre d'enveloppe spectrale varie en continu. Ici, par souci de commodité de l'explication, on suppose que $h(n)$ ne varie pas dans un intervalle d'une période, $h(n)$ pendant les première, seconde et troisième périodes étant respectivement représenté par $h(n)_1$, $h(n)_2$, $h(n)_3$, comme représenté sur la figure 3E.

Un jeu de paramètres d'enveloppe obtenu au moyen de diverses techniques d'estimation de spectre tel qu'un cepstrum $CL(i)$ qui est un jeu de paramètres obtenu au moyen du procédé d'analyse homomorphique et tel qu'un jeu de coefficients de prédiction $[a_i]$ ou un jeu de coefficients de réflexion $[k_i]$ ou qu'un jeu de paires de spectres en lignes, etc... qui est obtenu en appliquant la méthode des moindres carrés récurrente ou le procédé de prédiction linéaire est traité également en tant que $H(f)$ ou que $h(n)$ du fait qu'il peut prendre la caractéristique de fréquence $H(f)$ ou la réponse impulsionnelle $h(n)$ du filtre d'enveloppe spectrale. Par conséquent, la réponse impulsionnelle est ci-après également appelée jeu de paramètres d'enveloppe spectrale.

Les figures 5A et 5B représentent des procédés de déconvolution aveugle.

La figure 5A représente un procédé de déconvolution aveugle effectué en utilisant le procédé d'analyse à prédiction linéaire ou en utilisant la méthode des moindres carrés récurrente, ce procédé et cette méthode se rapportant tous deux à l'art antérieur. Etant donnée la forme d'onde de parole

voisée $s(n)$, comme représenté sur la figure 3C, les coefficients de prédiction $[a_1, a_2, \dots, a_N]$ ou les coefficients de réflexion $[k_1, k_2, \dots, k_N]$ qui sont les paramètres d'enveloppe spectrale représentant la caractéristique de fréquence $H(f)$ ou la réponse impulsionnelle $h(n)$ du filtre d'enveloppe spectrale sont obtenus en utilisant le procédé d'analyse à prédiction linéaire ou la méthode des moindres carrés récurrente. Normalement, 10 à 16 coefficients de prédiction suffisent pour l'ordre de prédiction "N". En utilisant les coefficients de prédiction $[a_1, a_2, \dots, a_N]$ et les coefficients de réflexion $[k_1, k_2, \dots, k_N]$ en tant que paramètres d'enveloppe spectrale, un filtre d'enveloppe spectrale inverse (simplement appelé filtre inverse) présentant une fréquence caractéristique de $1/H(f)$ qui est l'inverse de la caractéristique de fréquence $H(f)$ du filtre d'enveloppe spectrale peut facilement être construit par l'homme de l'art. Si la forme d'onde de parole voisée est l'entrée du filtre d'enveloppe spectrale inverse appelé également filtre d'erreur de prédiction linéaire dans le procédé d'analyse à prédiction linéaire ou dans la méthode des moindres carrés récurrente, le signal de train d'impulsions de hauteur de son périodique du type de la figure 3F présentant l'enveloppe spectrale plate appelé signal d'erreur de prédiction ou signal résiduel peut être obtenu en tant que sortie en provenance du filtre.

Les figures 5B et 5C représentent le procédé de déconvolution aveugle qui utilise le procédé d'analyse homomorphique, lequel est un procédé d'analyse de bloc, tandis que la figure 5B représente le procédé réalisé au moyen d'une division de fréquence (et la figure 5C représente le procédé effectué par filtrage inverse).

Une description de la figure 5B suit. Des échantillons de parole pour l'analyse d'un bloc sont obtenus en multipliant le signal de parole voisée $s(n)$

par une fonction de fenêtrage avec fenêtre à flancs inclinés telle qu'une fonction de fenêtrage de Hamming présentant une durée d'environ 10 à 20 ms. Une séquence spectrale $c(i)$ est ensuite obtenue en traitant les échantillons de parole en utilisant une série de procédures de traitement homomorphique constituée par une transformation de Fourier discrète (FFD), par un logarithme complexe et par une transformée de Fourier discrète inverse, comme représenté sur la figure 5D. Le "cepstrum" est une fonction de la "quéfrence" qui est une unité similaire à un temps.

Un cepstrum basse quéfrence $CL(i)$ qui se situe autour d'une origine qui représente l'enveloppe spectrale de la parole voisée $s(n)$ et un cepstrum haute quéfrence $CH(i)$ qui représente un signal de train d'impulsions de hauteur de son périodique $e(n)$ peuvent être séparés l'un de l'autre dans le domaine des quéfrences. C'est-à-dire qu'en multipliant respectivement le cepstrum $c(i)$ par une fonction de fenêtrage basse quéfrence et par une fonction de fenêtrage haute quéfrence, on obtient respectivement $CL(i)$ et $CH(i)$. Le fait de les récupérer par l'intermédiaire d'une procédure de traitement homomorphique inverse, comme représenté sur la figure 5E, donne la réponse impulsionnelle $h(n)$ et le signal de train d'impulsions de hauteur de son $e(n)$. Dans ce cas, le fait de récupérer $CH(i)$ par l'intermédiaire de la procédure de traitement homomorphique inverse ne donne pas directement le signal de train d'impulsions de hauteur de son $e(n)$ mais donne le signal de train d'impulsions de hauteur de son d'un bloc multiplié par une fonction de fenêtrage temporelle $w(n)$, $e(n)$ pouvant être obtenu en multipliant à nouveau le signal de train d'impulsions de hauteur de son par une fonction de fenêtrage temporelle inverse $1/w(n)$ correspondant à l'inverse de $w(n)$.

Le procédé de la figure 5C est le même que celui de la figure 5B à ceci près que $CL(i)$, en lieu et place de $CH(i)$, est utilisé sur la figure 5C pour obtenir le signal de train d'impulsions de hauteur de son périodique $e(n)$. C'est-à-dire que dans ce procédé, on utilise la propriété qui consiste en ce qu'une réponse impulsionnelle $h^{-1}(n)$ correspondant à $1/H(f)$ qui est l'inverse de la caractéristique de fréquence $H(f)$ peut être obtenue en traitant $-CL(i)$ qui est obtenu en prenant la valeur négative de $CL(i)$ par l'intermédiaire de la procédure de traitement homomorphique inverse, le signal de train d'impulsions de hauteur de son périodique $e(n)$ pouvant être obtenu en tant que sortie en construisant un filtre à réponse impulsionnelle de durée finie (FIR) qui présente une réponse impulsionnelle $h^{-1}(n)$ et en entrant sur le filtre un signal de parole originale $s(n)$ qui n'est pas multiplié par une fonction de fenêtrage. Ce procédé est un procédé de filtrage inverse qui est fondamentalement le même que celui de la figure 5A à ceci près que tandis que dans l'analyse homomorphique de la figure 5C le filtre d'enveloppe spectrale inverse $1/H(f)$ est construit en obtenant une réponse impulsionnelle $h^{-1}(n)$ du filtre d'enveloppe spectrale inverse, sur la figure 5A, le filtre d'enveloppe spectrale inverse $1/H(f)$ peut être directement construit au moyen des coefficients de prédiction $[a_i]$ ou des coefficients de réflexion $[k_i]$ obtenus au moyen du procédé d'analyse à prédiction linéaire.

Dans la déconvolution aveugle basée sur l'analyse homomorphique, la réponse impulsionnelle $h(n)$ ou le cepstrum basse fréquence $CL(i)$ représenté par des lignes en pointillés sur les figures 5B et 5C peuvent être utilisés en tant que jeu de paramètres d'enveloppe

spectrale. Lorsque l'on utilise la réponse impulsionnelle $[h(0), h(1), \dots, h(N-1)]$, un jeu de paramètres d'enveloppe spectrale est normalement constitué par un nombre important de paramètres de l'ordre de N , N valant entre 90 et 120, tandis que le nombre de paramètres peut être diminué jusqu'à 50 à 60 lorsque N vaut entre 25 et 30 si l'on utilise le cepstrum $[CL(-N)m, CL(-N+1), \dots, 0, \dots, CL(N)]$.

Comme décrit ci-avant, la forme d'onde de parole voisée $s(n)$ est soumise à une déconvolution pour obtenir la réponse impulsionnelle $h(n)$ du filtre d'enveloppe spectrale et le signal de train d'impulsions de hauteur de son périodique $e(n)$, conformément à la procédure de la figure 5.

Lorsque le signal de train d'impulsions de hauteur de son et les paramètres d'enveloppe spectrale ont été obtenus conformément à la procédure de déconvolution aveugle, alors des positions d'impulsions de hauteur de son P_1, P_2 , etc... sont obtenues à partir du signal de train d'impulsions de hauteur de son périodique $e(n)$ ou du signal de parole $s(n)$ en utilisant un algorithme de détection de position d'impulsion de hauteur de son dans le domaine temporel tel que l'algorithme de détection epoch. Puis les signaux impulsionnels de hauteur de son $e_1(n)$, $e_2(n)$ et $e_3(n)$ représentés respectivement sur les figures 3H, 3K, 3N sont obtenus en segmentant périodiquement le signal de train d'impulsions de hauteur de son $e(n)$ de telle sorte qu'une impulsion de hauteur de son soit incluse dans un intervalle périodique, comme représenté sur la figure 3F. Les positions de la segmentation peuvent être décidées en mettant des points centraux qui se situent entre les impulsions ou points de hauteur de son qui se situent à l'avant de chaque impulsion de hauteur de son, d'un temps constant. Cependant, du fait que la position de chaque impulsion

de hauteur de son, du point de vue du temps, coïncide avec la partie de fin de chaque impulsion de glotte, comme on peut l'apprécier pleinement en comparant les figures 3A et 3F, il est préférable de choisir un point qui se situe à un temps constant derrière chaque impulsion de hauteur de son en tant que position de la segmentation, comme indiqué par la ligne en pointillés sur la figure 3F. Cependant, du fait que l'impulsion de hauteur de son présente l'effet le plus conséquent sur l'audibilité, il n'y a pas de différence significative dans la parole synthétisée entre les deux cas.

Les signaux impulsionnels de hauteur de son $e_1(n)$, $e_2(n)$, $e_3(n)$, etc... obtenus au moyen de ce procédé sont respectivement soumis à nouveau à une convolution avec les $h_1(n)$, $h_2(n)$, $h_3(n)$ de la figure 3E qui sont des réponses impulsionnelles pendant l'intervalle périodique des signaux impulsionnels de hauteur de son $e_1(n)$, $e_2(n)$, $e_3(n)$, etc... et les ondelettes attendues telles que celles représentées sur les figures 3I, 3L, 3O sont obtenues. Cette convolution peut être effectuée de façon commode en entrant chaque signal de train d'impulsions de hauteur de son dans le filtre d'enveloppe spectrale $H(f)$ qui utilise les paramètres d'enveloppe spectrale en tant que coefficients de filtre, comme représenté sur la figure 4. Par exemple, dans les cas où les coefficients de prédiction linéaire ou les coefficients de réflexion ou les paires spectrales de lignes sont utilisées en tant que paramètres d'enveloppe spectrale comme dans le procédé d'analyse à prédiction linéaire, un filtre à réponse impulsionnelle de durée infinie (IIR) comportant les coefficients de prédiction linéaire ou les coefficients de réflexion ou les paires de lignes en tant que coefficients de filtre est composé. Dans les cas où la réponse impulsionnelle est utilisée pour les paramètres d'enveloppe spectrale, comme dans le

procédé d'analyse homomorphique, un filtre FIR présentant une réponse impulsionnelle en tant que coefficients est composé. Puisque le filtre de synthèse ne peut pas être directement composé si le paramètre d'enveloppe spectrale est un rapport d'aire logarithmique ou le cepstrum, les paramètres d'enveloppe spectrale doivent être retransformés en coefficients de réflexion ou en réponse impulsionnelle pour être utilisés en tant que coefficients du filtre IIR ou FIR. Si le signal impulsionnel de hauteur de son pour une période est l'entrée du filtre d'enveloppe spectrale composé comme décrit ci-avant avec les coefficients de filtre qui varient au fil du temps conformément aux paramètres d'enveloppe spectrale correspondant se situant aux mêmes instants que chaque échantillon du signal impulsionnel de hauteur de son, alors l'ondelette pour cette période est émise en sortie.

Pour cette raison, "les formes d'onde en fonction du temps" des paramètres d'enveloppe spectrale sont coupées au même point qu'au point où $e(n)$ a été coupé pour obtenir le signal impulsionnel de hauteur de son pour chaque période. Par exemple, dans le cas de l'analyse séquentielle, les paramètres d'enveloppe spectrale de première période $k_1(n)_1$, $k_2(n)_1$, etc... tels que représentés sur la figure 3G, sont obtenus en coupant les paramètres d'enveloppe spectrale correspondant à la même période temporelle que le signal impulsionnel de hauteur de son de première période $e_1(n)$ représenté sur la figure 3H à partir des fonctions temporelles $k_1(n)$, $k_2(n)$, etc... des paramètres d'enveloppe spectrale, tel que représenté sur la figure 3D. Les paramètres d'enveloppe spectrale de seconde et troisième périodes indiquées par une ligne en traits pleins sur les figures 3J et 3M peuvent également être obtenus d'une façon similaire à celle

mentionnée ci-avant. Sur la figure 4, les coefficients de réflexion k_1, k_2, \dots, k_N et la réponse impulsionnelle $H(0), H(1), \dots, H(N-1)$ sont représentés en tant que jeux de paramètres d'enveloppe spectrale typiques ou ils sont indiqués par $k_1(n), k_2(n), \dots, k_n(n)$ et $H(0n), H(1n), \dots, H(N-1, N)$ afin d'accentuer le fait qu'il s'agit de fonctions du temps. Pareillement, dans les cas où le cepstrum $CL(i)$ est utilisé en tant que jeu de paramètres d'enveloppe spectrale, il est indiqué par $CL(i, n)$.

Du fait qu'à la différence du procédé d'analyse séquentielle, les fonctions du temps des paramètres d'enveloppe spectrale ne sont pas obtenues dans le cas du procédé d'analyse à hauteur de son synchrone ou du procédé d'analyse par bloc mais que les valeurs de paramètres d'enveloppe spectrale qui sont constantes sur l'intervalle d'analyse sont obtenues, il est nécessaire de constituer les fonctions du temps des paramètres d'enveloppe spectrale à partir des valeurs des paramètres d'enveloppe spectrale puis de segmenter les fonctions du temps période par période afin d'obtenir les paramètres d'enveloppe spectrale pour une période. Cependant, dans la réalité, il est commode d'effectuer un traitement comme suit au lieu de composer les fonctions du temps. C'est-à-dire que dans le cas du procédé d'analyse par hauteur de son synchrone, puisqu'un jeu de paramètres d'enveloppe spectrale présentant des valeurs constantes correspond à chaque intervalle de période de hauteur de son, comme représenté par des hachures sur la figure 8B, les paramètres d'enveloppe spectrale ne présentent aucune modification même lorsque leurs fonctions du temps sont segmentées période par période. Par conséquent, les paramètres d'enveloppe spectrale pour une période qui doivent être stockés dans un tampon ne sont pas des

fonctions du temps mais des constantes indépendantes du temps.

Dans le cas du procédé d'analyse par bloc, puisqu'un jeu de paramètres d'enveloppe spectrale constants par bloc est obtenu, les valeurs d'un paramètre d'enveloppe spectrale pour une période appartenant à un bloc, par exemple $k_1(n)_1, k_1(n)_2, \dots, k_1(n)_L$ sont non seulement indépendantes de façon constante du temps mais sont également identiques. (Ici, $k_1(n)_J$ signifie la fonction du temps de k_1 pour le J -ième intervalle de période et M représente le nombre d'intervalles de période de hauteur de son appartenant à un bloc).

Il est à noter que dans le cas du procédé d'analyse par bloc, lorsque le signal impulsionnel de hauteur de son se situe sur la limite de deux blocs adjacents, les valeurs de paramètre d'enveloppe spectrale du bloc précédent et du bloc suivant doivent être utilisées respectivement pour les parties de signal précédente et suivante divisées par rapport à la limite des blocs.

Comme on peut le voir sur la figure 3I, la durée de l'ondelette n'est pas nécessairement égale à une période. Par conséquent, avant d'appliquer le signal impulsionnel de hauteur de son et les paramètres d'enveloppe spectrale d'une longueur de période obtenue au moyen de la segmentation périodique pour le filtre d'enveloppe spectrale, les processus de rajout de 0 et de traînée de paramètre représentés sur la figure 4 sont nécessaires pour la durée du signal impulsionnel de hauteur de son et les paramètres d'enveloppe spectrale doivent présenter une durée au moins aussi longue que la durée effective de l'ondelette. Le processus de rajout de 0 consiste à rendre la durée totale du signal impulsionnel de hauteur de son aussi longue que la longueur nécessaire en rajoutant les

échantillons qui ont la valeur 0 après le signal impulsif de hauteur de son d'une période. Le processus de traînée de paramètre consiste à rendre la durée totale du paramètre d'enveloppe spectrale aussi longue que la longueur nécessaire en rajoutant le paramètre d'enveloppe spectrale pour les périodes suivantes après le paramètre d'enveloppe spectrale d'une longueur de période. Cependant, même si un procédé simple de rajout répété de la valeur finale du paramètre d'enveloppe spectrale d'une période ou de la première valeur du paramètre d'enveloppe spectrale de la période suivante est utilisé, la qualité de la parole synthétisée n'est pas dégradée de façon significative.

Le fait que la durée effective de l'ondelette qui doit être générée par le filtre d'enveloppe spectrale dépend des valeurs des paramètres d'enveloppe spectrale rend difficile son estimation à l'avance. Cependant, du fait que des erreurs significatives ne sont pas obtenues lors d'une mise en oeuvre pratique dans la plupart des cas si l'on considère que la durée effective de l'ondelette vaut deux périodes depuis la position d'impulsion de hauteur de son dans le cas d'une voix masculine et trois périodes depuis la position d'impulsion de hauteur de son dans le cas d'une voix féminine ou d'un enfant, il est commode de décider que la durée du "signal impulsif de hauteur de son traîné" doit être constituée par un rajout de 0 et que "les paramètres d'enveloppe spectrale traînés" doivent être constitués par une traînée de paramètre présentant des durées de trois et quatre périodes respectivement pour des voix d'homme et de femme dans le cas où cette segmentation périodique est effectuée tout de suite après les impulsions de hauteur de son. Sur la figure 3G, des paramètres d'enveloppe spectrale traînés pour la première période de l'intervalle à

trois périodes "ad" obtenus en rajoutant les paramètres d'enveloppe spectrale pour l'intervalle à deux périodes "bd" indiqués par une ligne en pointillés à la suite du paramètre d'enveloppe spectrale de l'intervalle à une période "ab" obtenu par la segmentation périodique sont représentés à titre d'exemple. Sur la figure 3H, un signal impulsif de hauteur de son traîné pour la première période de l'intervalle à trois périodes "ad" obtenu par rajout des échantillons de valeur zéro à l'intervalle à deux périodes "bd" qui fait suite au signal impulsif de hauteur de son du premier intervalle de période "AB" obtenu au moyen de la segmentation périodique est représenté à titre d'exemple.

Dans le cas décrit ci-avant, du fait que la durée après le rajout de zéro et la traînée de paramètre est augmentée jusqu'à trois ou quatre périodes tandis que la durée du signal impulsif de hauteur de son et du paramètre d'enveloppe spectrale avant le rajout de zéro et la traînée de paramètre est d'une période, des tampons sont prévus entre la segmentation périodique et la traînée de paramètre, comme représenté sur la figure 4, et le signal impulsif de hauteur de son ainsi que les paramètres d'enveloppe spectrale obtenus au moyen de la segmentation périodique sont alors stockés dans les tampons et sont retrouvés lorsque nécessaire de telle sorte qu'un tamponnage temporel soit réalisé.

Si le signal impulsif de hauteur de son traîné et les paramètres d'enveloppe spectrale traînés sont obtenus au moyen du rajout de zéro et de la traînée de paramètre de la figure 4, le "signal d'ondelette" $sl(n)$ pour la première période de la longueur de l'intervalle à trois périodes telle que l'intervalle "ad", comme représenté sur la figure 3I, peut être pour finir obtenu en entrant le signal impulsif de hauteur de son traîné de la première

période telle que l'intervalle "ad" de la figure 3H dans le filtre d'enveloppe spectrale $H(f)$ et en faisant varier de façon synchrone les coefficients de la même manière que le paramètre d'enveloppe spectrale traîné de la première période telle que l'intervalle "ad" de la figure 3G. Les signaux d'ondelette $s_2(n)$ et $s_3(n)$ peuvent être respectivement pareillement obtenus pour les seconde et troisième périodes.

Comme décrit ci-avant, la forme d'onde de parole voisée $s(n)$ est pour finir décomposée en ondelettes qui composent la forme d'onde $s(n)$ au moyen de la procédure de la figure 4. A l'évidence, le ré-agencement des ondelettes de la figure 3I, de la figure 3L et de la figure 3O obtenues par décomposition en retour pour obtenir les points originaux conduit à la figure 3B et si les ondelettes sont superposées, la forme d'onde de parole originale $s(n)$ telle que représentée sur la figure 3C est obtenue à nouveau. Si les ondelettes de la figure 3I, de la figure 3L et de la figure 3O sont ré-agencées en faisant varier les inter-espaces puis sont superposées comme représenté sur la figure 3P, l'ondelette de parole présentant un motif de hauteur de son différent comme représenté sur la figure 3Q, est obtenue. Ainsi, le fait de faire varier de façon appropriée l'intervalle temporel entre les ondelettes obtenues par décomposition permet la synthèse de la parole présentant le motif de hauteur de son souhaité arbitraire, c'est-à-dire l'intonation. De façon similaire, le fait de faire varier de façon appropriée l'énergie des ondelettes permet la synthèse de la parole présentant le motif de sollicitation souhaité arbitraire.

Dans le bloc de stockage de segment de parole représenté sur la figure 2, chaque segment de parole voisée décomposé en autant d'ondelettes que le nombre d'impulsions de hauteur de son selon le procédé

représenté sur la figure 4 est stocké selon le format représenté sur la figure 6A qui se rapporte à l'information de segment de parole. Dans une zone d'en-tête qui constitue une partie avant de l'information de segment de parole, les points temporels de limite B_1, B_2, \dots, B_L qui sont des points temporels importants du segment de parole et des positions d'impulsions de hauteur de son P_1, P_2, \dots, P_M de chaque signal impulsif de hauteur de son utilisé dans la synthèse de chaque ondelette sont stockés, le nombre d'échantillons correspondant à chaque point temporel étant enregistré en prenant la position de premier échantillon du premier signal impulsif de hauteur de son $e_1(n)$ en tant que 0. Le point temporel de limite est la position temporelle des points de limite situés entre les sous-segments obtenus lorsque le segment de parole est segmenté en plusieurs sous-segments. Par exemple, la voyelle encadrée avant et après par des consonnes peut être considérée comme étant constituée par trois sous-segments pour la parole à vitesse faible du fait que la voyelle peut être divisée en un intervalle de régime permanent de partie médiane et en deux intervalles de transition qui se situent avant et après l'intervalle de régime permanent, et trois points d'extrémité des sous-segments sont stockés en tant que points temporels de limite dans la zone d'en-tête du segment de parole. Cependant, dans le cas où l'échantillonnage est effectué selon une vitesse de parole plus rapide, du fait que l'intervalle de transition devient un point de telle sorte que le segment de parole de la voyelle puisse être considéré comme étant constitué par deux sous-segments, deux points temporels de limite sont stockés dans l'information d'en-tête.

Dans la zone de code d'ondelette qui est la dernière partie de l'information de segment de parole,

des codes d'ondelette qui sont des codes obtenus en codant en forme d'onde l'ondelette correspondant à chaque période sont stockés. Les ondelettes peuvent être codées au moyen du procédé de codage de forme d'onde simple tel que le procédé MIC mais du fait que les ondelettes présentent une corrélation de court terme et de long terme significatives, la taille de mémoire nécessaire pour le stockage peut être diminuée de façon significative si les ondelettes sont codées en forme d'onde de façon efficace en utilisant le procédé MICDA qui comporte une boucle de prédiction de hauteur de son, un procédé de codage prédictif adaptatif ou un procédé de modulation delta adaptatif numérique. Le procédé selon lequel les ondelettes obtenues par décomposition sont codées en forme d'onde, les codes résultants étant stockés et, à l'instant de la synthèse, les codes étant décodés, ré-agencés et superposés pour produire une parole de synthèse, est appelé "procédé de stockage de code de forme d'onde".

Le signal impulsionnel de hauteur de son et les paramètres d'enveloppe spectrale correspondants peuvent être considérés comme étant identiques à l'ondelette du fait que ce sont des éléments dont l'ondelette est faite. Par conséquent, le procédé est également possible et dans ce procédé, les "codes source" obtenus en codant les signaux impulsionnels de hauteur de son et les paramètres d'enveloppe spectrale sont stockés et les ondelettes sont constituées avec les signaux impulsionnels de hauteur de son et les paramètres d'enveloppe spectrale obtenus en décodant les codes source puis les ondelettes sont ré-agencées et superposées pour produire la parole synthétisée. Ce procédé est appelé "procédé de stockage de code source". Ce procédé correspond au procédé dans lequel le signal impulsionnel de hauteur de son et les paramètres d'enveloppe spectrale stockés dans les

tampons, au lieu des ondelettes obtenues en tant que sortie de la figure 4, sont mis en correspondance les uns avec les autres dans le même intervalle de période puis sont stockés dans le bloc de stockage de segment de parole. Par conséquent, dans le procédé de stockage de code source, les procédures qui se situent après le tampon de la figure 4, c'est-à-dire la procédure de traînée de paramètre, la procédure de rajout de zéro et la procédure de filtrage par le filtre de synthèse $H(f)$, sont effectuées dans le sous-bloc d'assemblage de forme d'onde de la figure 7.

Dans le procédé de stockage de code source, le format de l'information de segment de parole est tel que représenté sur la figure 6B qui est la même que la figure 6A à l'exception du contenu de la zone de code d'ondelette. C'est-à-dire que les signaux impulsionnels de hauteur de son et les paramètres d'enveloppe spectrale nécessaires pour la synthèse des ondelettes au lieu des ondelettes sont codés et stockés aux positions où l'ondelette pour chaque période doit être stockée sur la figure 6A.

Les paramètres d'enveloppe spectrale sont codés conformément au procédé de quantification de l'art antérieur des paramètres d'enveloppe spectrale et sont stockés au niveau de la zone de code d'ondelette. A cet instant, si les paramètres d'enveloppe spectrale sont transformés de façon appropriée avant quantification, le codage peut être effectué de manière efficace. Par exemple, il est préférable de transformer les coefficients de prédiction en paramètres de la paire de spectres en lignes et les coefficients de réflexion en rapports d'aires logarithmiques et de les quantifier. En outre, puisque la réponse impulsionnelle présente une corrélation étroite pour des échantillons adjacents et pour des réponses impulsionnelles adjacentes, si les données sont codées en forme d'onde selon un procédé de

codage différentiel, la quantité de données nécessaires pour le stockage peut être réduite de manière significative. Dans le cas des paramètres de spectre, un procédé de codage est connu et dans ce procédé, le paramètre de spectre est transformé de telle sorte que la quantité de données puisse être réduite de manière significative.

Par ailleurs, le signal impulsionnel de hauteur de son est codé conformément à un procédé de codage de forme d'onde approprié et le code résultant est stocké au niveau de la zone de code d'ondelette. Les signaux impulsionnels de hauteur de son présentent une corrélation de court terme faible mais une corrélation de long terme significative les uns par rapport aux autres. Par conséquent, si le procédé de codage de forme d'onde tel que le codage MIC adaptatif prédictif de hauteur de son qui comporte la boucle de prédiction de hauteur de son est utilisé, une parole synthétisée de qualité élevée peut être obtenue même lorsque la taille de mémoire nécessaire pour le stockage est réduite jusqu'à trois bits par échantillon. Le coefficient de prédiction d'un prédicteur de hauteur de son peut être une valeur obtenue pour chaque période de hauteur de son conformément à un procédé d'auto-corrélation ou peut être une valeur constante. Au niveau de la première étape du codage, l'effet de prédiction de hauteur de son peut être augmenté par l'intermédiaire d'une normalisation en divisant le signal impulsionnel de hauteur de son qui doit être codé par la racine carrée de l'énergie moyenne par échantillon "G". Le décodage est effectué dans le bloc de synthèse de parole voisée et le signal impulsionnel de hauteur de son est restauré à son amplitude originale au moyen d'une multiplication par "G" à nouveau, lors de l'étape de fin du décodage.

Sur la figure 6B, l'information de segment de parole est représentée pour le cas où un procédé d'analyse à prédiction linéaire est adopté, lequel procédé utilise quatorze coefficients de réflexion en tant que paramètres d'enveloppe spectrale. Si l'intervalle d'analyse pour l'analyse à prédiction linéaire est la période de hauteur de son, quatorze coefficients de réflexion correspondent à chaque signal impulsionnel de hauteur de son et sont stockés. Si l'intervalle d'analyse est un bloc présentant une certaine longueur, les coefficients de réflexion pour plusieurs impulsions de hauteur de son d'un bloc présentent les mêmes valeurs de telle sorte que la taille de mémoire nécessaire pour le stockage des ondelettes est réduite. Dans ce cas, comme discuté ci-avant, puisque les coefficients de réflexion du bloc avant ou du dernier bloc sont utilisés au moment de la synthèse pour le signal impulsionnel de hauteur de son qui se situe dans la limite des deux blocs, selon que les échantillons du signal se situent avant ou après le point limite, la position du point limite situé entre les blocs peut être stockée en plus dans la zone d'en-tête. Si le procédé d'analyse séquentielle tel que la méthode des moindres carrés récursive est utilisé, les coefficients de réflexion k_1, k_2, \dots, k_{14} deviennent des fonctions continues de l'index temps " n ", comme représenté sur la figure 3D, et un lot de mémoires est nécessaire pour stocker la fonction du temps $k_1(n), k_2(n), \dots, k_{14}(n)$. En prenant le cas de la figure 3 à titre d'exemple, les formes d'onde pour l'intervalle "ab" de la figure 3G et de la figure 3H en tant que première période, et les formes d'onde pour l'intervalle "bc" de la figure 3J et de la figure 3K en tant que seconde période et les formes d'onde pour l'intervalle "cd" de la figure 3M et de la figure 3N en tant que troisième période de la zone de code

d'ondelette sont stockées dans la zone de code d'ondelette.

Le procédé de stockage de code de forme d'onde et le procédé de stockage de code source sont essentiellement le même procédé et de fait, le code de forme d'onde obtenu lorsque les ondelettes sont codées conformément au procédé de codage de forme d'onde efficace tel que le procédé APC (codage prédictif adaptatif) du procédé de stockage de code de forme d'onde devient pratiquement le même que le code source obtenu lors du procédé de stockage de code source du point de vue des contenus. Le code de forme d'onde du procédé de stockage de code de forme d'onde et le code source du procédé de stockage de code source sont appelés en combinaison le code d'ondelette.

La figure 7 représente la configuration interne du bloc de synthèse de parole voisée de la présente invention. Les codes d'ondelette stockés dans la zone de code d'ondelette de l'information de segment de parole reçue depuis le bloc de stockage de segment de parole sont décodés, selon la procédure inversée par rapport à la procédure selon laquelle ils avaient été codés, par un sous-bloc de décodage 9. Les signaux d'ondelette obtenus lorsque les codes de forme d'onde sont décodés selon le procédé de stockage de code de forme d'onde ou les signaux impulsionnels de hauteur de son obtenus lorsque les codes source sont décodés selon le procédé de stockage de code source et les paramètres d'enveloppe spectrale correspondant aux signaux impulsionnels de hauteur de son sont appelés information d'ondelette et sont produits pour le sous-bloc d'assemblage de forme d'onde. Par ailleurs, l'information d'en-tête stockée dans la zone d'en-tête de l'information de segment de parole est l'entrée du sous-bloc de commande de durée 10 et d'un sous-bloc de commande de hauteur de son 11.

Le sous-bloc de commande de durée de la figure 7 reçoit en tant qu'entrée les données de durée contenues dans l'information de prosodie et les points temporels de limite inclus dans l'information d'en-tête de segment de parole et il produit l'information de modification temporelle en utilisant les données de durée et les points temporels de limite et il applique l'information de modification temporelle produite au sous-bloc d'assemblage de forme d'onde 13, au sous-bloc de commande de hauteur de son et au sous-bloc de commande d'énergie. Si la durée totale du segment de parole devient plus longue ou plus courte, la durée des sous-segments qui constituent le segment de parole devient plus longue ou plus courte en conséquence, le rapport de dilatation ou de compression dépendant de la propriété de chaque sous-segment. Par exemple, dans le cas de la voyelle entourée par des consonnes qui se trouvent avant et après elle, la durée de l'intervalle de régime permanent qui se situe dans le milieu présente un taux de variation sensiblement plus important que celle des intervalles de transition situés sur les deux côtés de la voyelle. Le sous-bloc de commande de durée compare la durée BL du segment de parole original qui a été stocké et la durée du segment de parole qui doit être synthétisé et qui est indiqué par les données de durée et il obtient la durée de chaque sous-segment qui doit être synthétisé en correspondance avec la durée de chaque sous-segment original en utilisant leur taux de variation ou la règle de durée, d'où l'obtention des points temporels de limite de la parole synthétisée. Les points temporels de limite originaux B1, B2, etc... et les points temporels de limite B'1, B'2, etc... de la parole synthétisée mis en correspondance avec les points temporels de limite originaux sont en combinaison appelée information de modification

temporelle et sur la base de ceci, dans le cas de la figure 8 par exemple, l'information de modification temporelle peut être représentée par $[(B1, B'1), (B1, B'2), (B2, B'3), (B3, B'3), (B4, B'4)]$.

La fonction du sous-bloc de commande de hauteur de son de la figure 7 consiste à produire l'information de position d'impulsion de hauteur de son de telle sorte que la parole synthétique présente le motif d'intonation indiqué par les données de motif d'intonation et à appliquer ce motif au sous-bloc d'assemblage de forme d'onde et au sous-bloc de commande d'énergie. Le sous-bloc de commande de hauteur de son reçoit en tant qu'entrée les données de motif d'intonation qui sont les valeurs des fréquences de hauteur de son cible pour chaque phonème et il produit un contour de hauteur de son représentant la variation continue de la fréquence de hauteur de son par rapport au temps en connectant les valeurs des fréquences de hauteur de son cible sans discontinuité. Le sous-bloc de commande de hauteur de son peut refléter un phénomène de micro-intonation dû à un élément qui gêne le contour de hauteur de son. Cependant, dans ce cas, le contour de hauteur de son devient une fonction discontinue dans laquelle la valeur de fréquence de hauteur de son varie brutalement par rapport au temps au niveau du point limite qui se situe entre le phonème gênant et l'autre phonème adjacent. La fréquence de hauteur de son est obtenue en échantillonnant le contour de hauteur de son pour la première position d'impulsion de hauteur de son du segment de parole et la période de hauteur de son est obtenue en prenant l'inverse de la fréquence de hauteur de son puis le point traité par la période de hauteur de son est déterminé en tant que seconde position d'impulsion de hauteur de son. La période de hauteur de son suivante est alors obtenue à partir de la fréquence de hauteur

de son à ce point et la position d'impulsion de hauteur de son suivante est obtenue à son tour, et la répétition de cette procédure permet d'obtenir toutes les positions d'impulsion de hauteur de son de la parole synthétisée. La première position d'impulsion de hauteur de son du segment de parole peut être appelée premier échantillon ou ses échantillons voisins dans le cas du premier segment de parole d'une série de segments de parole voisée continus de la parole synthétisée et la première position d'impulsion de hauteur de son pour le segment de parole suivant peut être appelée point correspondant à la position de l'impulsion de hauteur de son qui suit la dernière impulsion de hauteur de son du segment de parole précédent et ainsi de suite. Le sous-bloc de commande de hauteur envoie les positions d'impulsion de hauteur de son $P'1$, $P'2$, etc... de la parole de synthèse obtenues en tant que positions d'impulsion de hauteur de son originales $P1$, $P2$, etc... incluses dans l'information d'en-tête de segment de parole ainsi que dans un "joint" du sous-bloc d'assemblage de forme d'onde et du sous-bloc de commande d'énergie, auquel cas elles sont également appelées information de position d'impulsion de hauteur de son. Dans le cas de la figure 8 par exemple, l'information de position d'impulsion de hauteur de son peut être représentée par $[(P1, P2, \dots, P9), (P'1, P'2, \dots, P'8)]$.

Le sous-bloc de commande d'énergie de la figure 7 produit une information de gain grâce à laquelle la parole synthétisée présente le motif de sollicitation tel qu'indiqué par les données de motif de sollicitation, et il envoie cette information de gain au sous-bloc d'assemblage de forme d'onde. Le sous-bloc de commande d'énergie reçoit en tant qu'entrée les données de motif de sollicitation qui sont les valeurs d'amplitude cible pour chaque phonème, et il produit un

contour d'énergie représentant la variation continue de l'amplitude par rapport au temps en les connectant sans discontinuité. On suppose que les segments de parole sont normalisés à l'avance au moment du stockage de telle sorte qu'ils présentent une énergie relative en relation avec la place du segment de parole afin de refléter la différence relative d'énergie pour chaque phonème. Par exemple, dans le cas des voyelles, une voyelle faible présente une énergie plus importante par temps unitaire qu'une voyelle haute et un son nasal présente environ la moitié de l'énergie par temps unitaire par comparaison à la voyelle. En outre, l'énergie pendant l'intervalle de fermeture du son plausible est très faible. Par conséquent, lorsque les segments de parole sont stockés, ils doivent avoir été codés après réglage à l'avance de telle sorte qu'ils présentent cette énergie relative. Dans ce cas, le contour d'énergie produit dans le sous-bloc de commande d'énergie devient un gain qui doit être multiplié pour la forme d'onde qui doit être synthétisée. Le sous-bloc de commande d'énergie obtient les valeurs de gain G_1 , G_2 , etc... à chaque position d'impulsion de hauteur de son P'_1 , P'_2 , etc... de la parole synthétique en utilisant le contour d'énergie et l'information de position d'impulsion de hauteur de son et il applique ces valeurs de gain au sous-bloc d'assemblage de forme d'onde, ces valeurs de gain étant appelées information de gain. Dans le cas de la figure 8 par exemple, l'information de gain peut être représentée par $[(P'_1, G_1), (P'_2, G_2), \dots, (P'_8, G_8)]$.

Le sous-bloc d'assemblage de forme d'onde de la figure 7 reçoit en tant qu'entrée l'information d'ondelette décrite ci-avant, l'information de modification temporelle, l'information de position d'impulsion de hauteur de son et l'information de gain et il produit pour finir le signal de parole voisée. Le

sous-bloc d'assemblage de forme d'onde produit la parole présentant le motif d'intonation, le motif de sollicitation et la durée tels qu'indiqués par l'information de prosodie en utilisant l'information d'ondelette reçue depuis le sous-bloc de décodage. A cet instant, certaines des ondelettes sont répétées et certaines sont omises. Les données de durée, les données de motif d'intonation et les données de motif de sollicitation incluses dans l'information de prosodie sont des informations indicatrices indépendantes les unes des autres bien qu'elles doivent être traitées avec des inter-relations du fait qu'il y a une inter-relation entre ces trois informations lorsque la forme d'onde est synthétisée avec l'information d'ondelette. L'un des problèmes les plus importants rencontrés lors de l'assemblage de forme d'onde est constitué par le fait de savoir quelle ondelette sélectionner en tant qu'ondelette qui doit être agencée à chaque position d'impulsion de hauteur de son de la parole synthétisée. Si les ondelettes appropriées ne sont pas sélectionnées et agencées, une parole synthétique de bonne qualité ne peut pas être obtenue. Est donnée ci-après une description du fonctionnement du sous-bloc d'assemblage de forme d'onde qui utilise le procédé de relocalisation d'ondelette basé sur la modification temporelle de la présente invention qui est un procédé de relocalisation d'ondelette permettant d'obtenir une qualité élevée lors de la synthèse de la parole synthétique en utilisant l'information de segment de parole reçue depuis le bloc de stockage de segment de parole.

La procédure de synthèse de forme d'onde de parole voisée du sous-bloc d'assemblage de forme d'onde est constituée par deux étapes, c'est-à-dire l'étape de relocalisation d'ondelette qui utilise la fonction de

modification temporelle et l'étape de superposition destinée à superposer les ondelettes relocalisées.

C'est-à-dire que dans le cas du procédé de stockage de code de forme d'onde, les ondelettes qui conviennent le mieux sont sélectionnées pour les positions d'impulsion de hauteur de son de la parole synthétique parmi les signaux d'ondelette reçus en tant qu'information d'ondelette et sont localisées au niveau de leurs positions d'impulsion de hauteur de son, et leurs gains sont réglés puis la parole synthétisée est produite en les superposant.

Selon le procédé de stockage de code source, le signal impulsionnel de hauteur de son et les paramètres d'enveloppe spectrale pour chaque période correspondant au signal impulsionnel de hauteur de son sont reçus en tant qu'information d'ondelette. Dans ce cas, deux procédés d'assemblage de parole synthétique sont possibles. Le premier procédé consiste à obtenir chaque ondelette en appliquant au filtre de synthèse les paramètres d'enveloppe spectrale et le signal impulsionnel de hauteur de son pour une longueur d'intervalle à 2-4 périodes obtenue en effectuant les procédures correspondant au côté droit du tampon de la figure 4, c'est-à-dire la traînée de paramètre décrite ci-avant et le rajout de zéro quant à l'information d'ondelette, puis à assembler la parole synthétique avec les ondelettes conformément à la procédure identique à celle du procédé de stockage de code de forme d'onde. Ce procédé est fondamentalement le même que l'assemblage de la parole synthétique du procédé de stockage de code de forme d'onde et par conséquent, sa description distincte est omise. Le second procédé consiste à obtenir un signal de train d'impulsions de hauteur de son synthétique ou un signal d'excitation synthétique présentant une enveloppe spectrale plate mais comportant un motif de hauteur de son différent de

celui du signal de train d'impulsions de hauteur de son périodique original en sélectionnant les signaux qui conviennent le mieux pour les positions d'impulsion de hauteur de son de la parole synthétique parmi les signaux impulsionnels de hauteur de son et en les localisant ainsi qu'en réglant leurs gains puis en les superposant, et à obtenir des paramètres d'enveloppe spectrale synthétique constitués en rapportant le paramètre d'enveloppe spectrale à chaque signal impulsionnel de hauteur de son constituant le signal de train d'impulsions de hauteur de son synthétique ou le signal d'excitation synthétique puis à produire la parole synthétisée en appliquant le signal d'excitation synthétique et les paramètres d'enveloppe spectrale synthétique au filtre de synthèse. Ces deux procédés sont essentiellement identiques à ceci près que la séquence qui se situe entre le filtre d'assemblage et la procédure de superposition, lors de l'assemblage de la parole de synthèse, est inversée.

Le procédé d'assemblage de parole synthétique décrit ci-avant est décrit ci-après par report à la figure 8. Le procédé de relocalisation d'ondelette peut être fondamentalement également appliqué à la fois au procédé de stockage de code de forme d'onde et au procédé de stockage de code source. Par conséquent, les procédures d'assemblage de forme d'onde de parole synthétique des deux procédés sont décrites simultanément par report à la figure 8.

Sur la figure 8A est représentée la corrélation entre le segment de parole original et le segment de parole qui doit être synthétisé. Les points temporels de limite originaux B1, B2, etc... indiqués par des lignes en pointillés, les points temporels de limite B'1, B'2, etc... du son synthétisé et la corrélation entre eux indiquée par les lignes hachurées sont inclus dans l'information de modification temporelle reçue

depuis le sous-bloc de commande de durée. En outre, les positions d'impulsion de hauteur de son originales P_1 , P_2 , etc... indiquées par les lignes en traits pleins et les positions d'impulsion de hauteur de son P'_1 , P'_2 , etc... du son synthétisé sont incluses dans l'information de position d'impulsion de hauteur de son reçue depuis le sous-bloc de commande de hauteur de son. Par commodité de l'explication de la figure 8, on suppose que la période de hauteur de son de la parole originale et que la période de hauteur de son du son synthétisé sont respectivement constantes et que la seconde vaut 1,5 fois la première.

Le sous-bloc d'assemblage de forme d'onde forme la fonction de modification telle que représentée sur la figure 8B en utilisant les points temporels de limite originaux, les points temporels de limite du son synthétisé et la corrélation entre eux. L'abscisse de la fonction de modification temporelle représente le temps " t " du segment de parole original et l'ordonnée représente le temps " t' " du segment de parole qui doit être synthétisé. Sur la figure 8A par exemple, du fait que le premier sous-segment et que le dernier sous-segment du segment de parole original doivent être comprimés respectivement selon un facteur de $2/3$ et doivent être dilatés respectivement selon un facteur de 2, leur corrélation apparaît respectivement en tant que lignes présentant une pente de $2/3$ et de 2 dans la fonction de modification temporelle de la figure 8B. Le second sous-segment ne varie pas du point de vue de sa durée de manière à apparaître en tant que ligne de pente 1 dans la fonction de modification temporelle. Le second sous-segment du segment de parole qui doit être synthétisé résulte de la répétition du point temporel de limite " B_1 " du segment de parole original et au contraire, le troisième sous-segment du segment de parole original va varier jusqu'au point temporel de

limite "B'3" du segment de parole qui doit être synthétisé. Les corrélations apparaissent dans ces cas respectivement en tant que ligne verticale et que ligne horizontale. La fonction de modification temporelle est donc obtenue en présentant le point temporel de limite du segment de parole original et le point temporel de limite du segment de parole qui doit être synthétisé correspondant au point temporel de limite du segment de parole original en tant que deux points et en les reliant par une ligne. Il peut être possible dans certains cas de représenter la corrélation entre les sous-segments comme étant plus proche de la réalité en reliant les points à l'aide d'une courbe lissée.

Selon le procédé de stockage de code de forme d'onde, le sous-bloc d'assemblage de forme d'onde recherche le point temporel original correspondant à la position d'impulsion de hauteur de son du son synthétique en utilisant la fonction de modification temporelle et recherche l'ondelette présentant la position d'impulsion de hauteur de son la plus proche du point temporel original puis localise l'ondelette à la position d'impulsion de hauteur de son du son synthétique.

Lors de l'étape suivante, le sous-bloc d'assemblage de forme d'onde multiplie chaque signal d'ondelette localisée par le gain correspondant à la position d'impulsion de hauteur de son du signal d'ondelette trouvée à partir de l'information de gain et pour finir, il obtient le son synthétique souhaité en superposant les signaux d'ondelette à gain réglé, simplement en les ajoutant. Sur la figure 3Q est représenté le son synthétique produit par une telle procédure de superposition dans le cas où les ondelettes des figures 3I, 3L et 3O sont relocalisées tels que sur la figure 3P.

De façon similaire, selon le procédé de stockage de code source, le sous-bloc d'assemblage de forme d'onde recherche le point temporel original correspondant à la position d'impulsion de hauteur de son du son synthétique en utilisant la fonction de modification temporelle et recherche le signal impulsional de hauteur de son présentant la position d'impulsion de hauteur de son la plus proche du point temporel original puis il localise le signal impulsional de hauteur de son à la position d'impulsion de hauteur de son du son synthétique.

Les nombres de signaux impulsionnels de hauteur de son ou les nombres d'ondelettes localisées de cette manière pour chaque position d'impulsion de hauteur de son du segment de parole qui doit être synthétisé sont représentés sur les figures 8A et 8B. On peut voir sur les dessins que certaines des ondelettes qui constituent le segment de parole original sont omises du fait de la compression des sous-segments et que certaines ondelettes sont utilisées de façon répétée du fait de la dilatation des sous-segments. Il est supposé sur la figure 8 que le signal impulsional de hauteur de son pour chaque période a été obtenu en effectuant une segmentation immédiatement après chaque impulsion de hauteur de son.

La superposition des ondelettes selon le procédé de stockage de code de forme d'onde est équivalente à la superposition des signaux impulsionnels de hauteur de son selon le procédé de stockage de code source. Par conséquent, dans le cas du procédé de stockage de code source, le sous-bloc d'assemblage de forme d'onde multiplie chaque signal impulsional de hauteur de son relocalisé par le gain correspondant à la position d'impulsion de hauteur de son du signal impulsional de hauteur de son relocalisé trouvé à partir de l'information de gain et pour finir, il obtient le

signal d'excitation synthétique souhaité en superposant les signaux impulsionnels de hauteur de son à gain réglé. Cependant, dans ce cas, du fait que la plus grande partie de l'énergie est concentrée sur l'impulsion de hauteur de son, il peut être possible de constituer le signal d'excitation synthétique en obtenant tout d'abord un signal d'excitation synthétique sans réglage de gain, en superposant les signaux impulsionnels de hauteur de son localisés, puis en multipliant le signal d'excitation synthétique sans réglage de gain par le contour d'énergie généré au niveau du sous-bloc de commande d'énergie au lieu de superposer les signaux impulsionnels de hauteur de son à gain réglé constant. La figure 3R représente le signal d'excitation synthétique obtenu lorsque les signaux impulsionnels de hauteur de son des figures 3H, 3K et 3N sont relocalisés conformément à cette procédure de telle sorte que le motif de hauteur de son devienne le même que pour le cas de la figure 3P.

Selon le procédé de stockage de code source, le sous-bloc d'assemblage de forme d'onde doit constituer les paramètres d'enveloppe spectrale synthétique et deux façons sont possibles, c'est-à-dire le procédé de compression et de dilatation temporelles représenté sur la figure 8A et le procédé de correspondance synchrone représenté sur la figure 8B. Si les paramètres d'enveloppe spectrale sont des fonctions continues du temps et s'ils représentent pleinement l'enveloppe spectrale de la parole, les paramètres d'enveloppe spectrale synthétique peuvent être obtenus simplement en comprimant ou en dilatant temporellement les paramètres d'enveloppe spectrale originaux sur une base sous-segment par sous-segment. Sur la figure 8A, le paramètre d'enveloppe spectrale obtenu au moyen du procédé d'analyse séquentielle est représenté par une courbe en pointillés et le paramètre d'enveloppe

spectrale codé en approximant la courbe au moyen de la connexion de plusieurs points tels que A, B, C, etc... au moyen de segments de lignes est représenté par une ligne en traits pleins. Puisque seulement la position temporelle de chaque point varie pour obtenir les points A', B', C', etc... en tant que résultat de la compression et de la dilatation temporelles, ce procédé de codage de segments de ligne convenant tout particulièrement pour le cas de la compression et de la dilatation temporelles. Cependant, dans le cas de l'utilisation du procédé d'analyse de bloc ou du procédé d'analyse synchrone de hauteur de son, puisque la correspondance spectrale n'est pas précise et que la variation temporelle du paramètre d'enveloppe spectrale est discontinue, le procédé de compression et de dilatation temporelles ne peut pas donner la qualité de son synthétique souhaitée et il est préférable d'utiliser le procédé de correspondance synchrone dans lequel les paramètres d'enveloppe spectrale synthétique sont assemblés en corrélant les paramètres d'enveloppe spectrale pour chaque intervalle de période de hauteur de son avec chaque signal impulsionnel de hauteur de son correspondant, comme représenté sur la figure 8B. C'est-à-dire que puisque l'ondelette selon le procédé de stockage de code de forme d'onde est équivalente au signal impulsionnel de hauteur de son et aux paramètres d'enveloppe spectrale correspondants pour le même intervalle de période de hauteur de son, les paramètres d'enveloppe spectrale synthétique peuvent être produits en localisant de façon synchrone les paramètres d'enveloppe spectrale pendant un intervalle de période, pour le même intervalle de période de chaque signal impulsionnel de hauteur de son localisé. Sur la figure 8B, k_1 qui est l'un des paramètres d'enveloppe spectrale et k'_1 qui est le paramètre d'enveloppe spectrale synthétique correspondant à k_1 assemblé au

moyen de ces procédés pour le procédé d'analyse de bloc et pour le procédé d'analyse synchrone de hauteur de son sont représentés respectivement par la ligne en traits pleins et par la ligne en pointillés. Bien entendu, comme mis en exergue ci-avant, moyennant le paramètre d'enveloppe spectrale obtenu au moyen du procédé d'analyse séquentielle, le paramètre d'enveloppe spectrale synthétique peut être assemblé conformément au procédé de la figure 8A. Par exemple, si le signal impulsionnel de hauteur de son pour chaque période a été relocalisé comme représenté sur la figure 3R, les paramètres d'enveloppe spectrale pour chaque période sont localisés comme représenté sur la figure 3S en relation avec les signaux impulsionnels de hauteur de son.

Au moment de l'assemblage du signal d'excitation synthétique et des paramètres d'enveloppe spectrale synthétique selon le procédé de stockage de code source, si la période de hauteur de son du son synthétisé est plus longue que la période de hauteur de son originale, un intervalle de suppression en résulte et il est situé entre deux intervalles de période de hauteur de son adjacents, comme représenté par des lignes obliques sur la figure 8. Si la période de hauteur de son du son synthétisé est plus courte que la période de hauteur de son originale, des intervalles de chevauchement dans lesquels deux intervalles de période de hauteur de son adjacents se chevauchent l'un l'autre se produisent. L'intervalle en chevauchement "fb" et l'intervalle de suppression "gh" sont représentés sur les figures 3R et 3S à titre d'exemple. Comme précédemment décrit, les signaux impulsionnels de hauteur de son relocalisés doivent être superposés à l'instant du chevauchement. Cependant, il est raisonnable que les paramètres d'enveloppe spectrale relocalisés en relation avec les signaux impulsionnels

de hauteur de son soient moyennés au lieu d'être superposés à l'instant du chevauchement. Par conséquent, le procédé d'assemblage du signal d'excitation synthétique et des paramètres d'enveloppe spectrale synthétique avec les intervalles de suppression et les intervalles de chevauchement pris en considération est le suivant.

Les échantillons de valeur zéro sont insérés dans l'intervalle de suppression à l'instant de l'assemblage du signal d'excitation synthétique. Dans le cas d'un son de fricative voisé, un son davantage naturel peut être synthétisé si le signal de bruit filtré passe-haut en lieu et place des échantillons de valeur zéro est inséré dans l'intervalle de suppression. Les signaux impulsionnels de hauteur de son relocalisés doivent être ajoutés dans l'intervalle de chevauchement. Puisque ce procédé d'addition est ennuyeux, il est commode d'utiliser un procédé de troncation dans lequel seulement un signal est sélectionné parmi deux signaux impulsionnels de hauteur de son qui se chevauchent dans l'intervalle de chevauchement. La qualité du son synthétisé en utilisant le procédé de troncation n'est pas dégradée de façon significative. Sur la figure 3R, l'intervalle de suppression "gh" est rempli avec des échantillons à zéro et le signal impulsionnel de hauteur de son de l'intervalle avant est sélectionné dans l'intervalle de chevauchement "fb". C'est-à-dire que dans le cas de la survenue d'un chevauchement, l'intervalle avant pris parmi les intervalles de chevauchement de chaque signal impulsionnel de hauteur de son est tronqué et ce procédé est physiquement davantage empreint de sens par comparaison avec le procédé selon lequel les signaux impulsionnels de hauteur de son sont constitués en effectuant une segmentation immédiatement à l'avant de l'impulsion de hauteur de son et selon lequel, à l'instant de la

synthèse, le dernier intervalle pris parmi les intervalles de chevauchement du signal impulsionnel de hauteur de son est tronqué s'il y a chevauchement, comme décrit précédemment. Cependant, dans la réalité, l'un ou l'autre procédé ne produit pas une différence significative au niveau de la qualité de son du son synthétisé.

A l'instant de l'assemblage du paramètre d'enveloppe spectrale synthétique, il est idéal que l'intervalle de suppression soit rempli par les valeurs qui varient linéairement depuis une valeur du paramètre d'enveloppe spectrale, au point terminal de l'intervalle de période précédent, jusqu'à une valeur du paramètre d'enveloppe spectrale, au point de début de la période suivante, et que, dans l'intervalle de chevauchement, le paramètre d'enveloppe spectrale varie progressivement depuis le paramètre d'enveloppe spectrale de la période précédente jusqu'à celui de la période suivante en utilisant le procédé d'interpolation selon lequel la moyenne de deux paramètres d'enveloppe spectrale en chevauchement est obtenue à l'aide de valeurs de pondération qui varient linéairement en fonction du temps. Cependant, puisque ces procédés sont peu commodes, le procédé suivant peut être utilisé du fait qu'il est davantage commode et qu'il ne dégrade pas de façon significative la qualité du son. C'est-à-dire que pour le paramètre d'enveloppe spectrale situé dans l'intervalle de suppression, la valeur du paramètre d'enveloppe spectrale au niveau du point terminal de l'intervalle de période précédent peut être utilisée de façon répétée comme sur la figure 8B ou la valeur du paramètre d'enveloppe spectrale au niveau du point de début de l'intervalle de période suivant peut être utilisée de façon répétée, la valeur moyenne arithmétique des deux paramètres d'enveloppe spectrale peut être utilisée ou les valeurs du

paramètre d'enveloppe spectrale aux points de fin et de début des intervalles de période précédent et suivant peuvent être utilisées respectivement avant et après le centre de l'intervalle de suppression qui constitue une limite. Pour le paramètre d'enveloppe spectrale situé dans l'intervalle de chevauchement, l'une ou l'autre partie correspondant à l'impulsion de hauteur de son sélectionnée peut être simplement sélectionnée. Sur la figure 3S par exemple, puisque le signal impulsif de hauteur de son pour l'intervalle de période précédent est sélectionné en tant que signal d'excitation synthétique dans l'intervalle de chevauchement "fb", les valeurs des paramètres pour l'intervalle de période précédent sont pareillement sélectionnées en tant que paramètres d'enveloppe spectrale synthétique. Dans l'intervalle de suppression "gh" de la figure 8B et de la figure 3S, les valeurs de paramètre du paramètre d'enveloppe spectrale à la fin de l'intervalle de période précédent sont utilisées de façon répétée. Bien entendu, dans le cas de la figure 3S pour lequel le paramètre d'enveloppe spectrale est une fonction continue du temps, le procédé selon lequel la dernière valeur de l'intervalle de période précédent ou la première valeur de l'intervalle de période suivant est utilisée de façon répétée pendant l'intervalle de suppression et le procédé selon lequel les deux valeurs varient linéairement pendant l'intervalle de suppression conduisent au même résultat.

Une fois que le signal d'excitation synthétique et les paramètres d'enveloppe spectrale synthétique pour un segment ont été assemblés, le sous-bloc d'assemblage de forme d'onde lisse normalement les deux extrémités des paramètres d'enveloppe spectrale synthétique assemblés en utilisant le procédé d'interpolation de telle sorte que la variation du

paramètre d'enveloppe spectrale soit lissée entre des segments de parole adjacents. Si le signal d'excitation synthétique et les paramètres d'enveloppe spectrale synthétique assemblés comme mentionné ci-avant sont entrés en tant que signal d'excitation et en tant que coefficients de filtre respectivement sur le filtre de synthèse contenu dans le sous-bloc d'assemblage de forme d'onde, le son synthétique souhaité est pour finir émis en sortie depuis le filtre de synthèse. Le signal d'excitation synthétique obtenu lorsque les signaux impulsionnels de hauteur de son des figures 3H, 3K et 3N sont relocalisés de telle sorte que le motif de hauteur de son soit le même que sur la figure 3P est représenté sur la figure 3R et les paramètres d'enveloppe spectrale synthétique obtenus en faisant correspondre des paramètres d'enveloppe spectrale pour une période des figures 3G, 3J et 3N aux signaux impulsionnels de hauteur de son contenus dans le signal d'excitation synthétique de la figure 3R sont représentés sur la figure 3S. Le fait de constituer un filtre de synthèse variable en fonction du temps qui présente en tant que coefficients de filtre les coefficients de réflexion variables représentés sur la figure 3S et le fait d'entrer le signal d'excitation synthétique tel que représenté sur la figure 3R dans le filtre de synthèse variable en fonction du temps conduisent au son synthétisé de la figure 3T qui est pratiquement le même que le son synthétisé de la figure 3P.

Si l'on compare maintenant le procédé de stockage de code de forme d'onde et le procédé de stockage de code source, les deux procédés peuvent être considérés comme étant identiques du point de vue du principe. Cependant, lorsque l'on concatène les segments de parole qui présentent une mauvaise connexion les uns avec les autres, on observe une différence qui consiste

en ce qu'il est possible de synthétiser le son connecté sans discontinuité en lissant les paramètres d'enveloppe spectrale au moyen de l'utilisation du procédé d'interpolation dans le cas du procédé de stockage de code source mais que ceci est impossible dans le cas du procédé de stockage de code de forme d'onde. En outre, le procédé de stockage de code source nécessite une mémoire plus petite que le procédé de stockage de code de forme d'onde puisque la forme d'onde de seulement une longueur de période par ondelette doit être stockée pour le procédé de stockage de code source et ce procédé de stockage de code source a pour avantage qu'il est facile de lui intégrer la fonction du bloc de synthèse de son voisé et la fonction du bloc de synthèse de son non voisé décrites ci-avant. Dans le cas de l'utilisation du procédé d'analyse homomorphique, le spectre ou la réponse impulsionnelle peut être utilisée en tant que paramètre d'enveloppe spectrale établi selon le procédé de stockage de code de forme d'onde tandis qu'il est pratiquement impossible selon le procédé de stockage de code source d'utiliser le spectre nécessitant le calcul à base de bloc du fait que la durée du bloc de synthèse qui présente les valeurs de paramètre d'enveloppe spectrale synthétique constantes varie bloc par bloc, comme on peut le voir à partir du paramètre d'enveloppe spectrale synthétique de la figure 8B représenté par une ligne en traits pleins. Le procédé de stockage de code source selon la présente invention utilise l'impulsion de hauteur de son d'une période en tant qu'impulsion d'excitation. Cependant, ce procédé est différent du procédé d'excitation d'impulsion de hauteur de son régulier de l'art antérieur qui envisage de remplacer l'impulsion par une impulsion de hauteur de son d'échantillon en ce que, dans la présente invention, l'impulsion de hauteur de son de chaque

période et les paramètres d'enveloppe spectrale de chaque période correspondant à l'impulsion de hauteur de son sont joints pour produire l'ondelette de chaque période.

Comme il ressort de la description qui précède, la présente invention convient pour le codage et le décodage du segment de parole du système de synthèse par transformation de texte en parole du procédé de synthèse segmentaire de parole. En outre, puisque la présente invention est un procédé dans lequel les durées totale et partielle ainsi que le motif de hauteur de son des unités phonétiques arbitraires telles que le phonème, la demi-syllabe, le diphone et le sous-segment, etc... qui constituent la parole peuvent être modifiés librement et indépendamment, ce procédé peut être utilisé dans un système de conversion de vitesse de parole ou dans un système de modification d'échelle temporelle qui modifie la vitesse vocale selon un taux constant de manière à ce qu'elle soit plus rapide ou plus lente que la vitesse originale sans modifier le motif d'intonation de la parole, et il peut également être utilisé dans le système de synthèse de voix chantée ou dans un système de codage de parole à vitesse très faible tel qu'un vocodeur phonétique ou qu'un vocodeur de segment qui transfère la parole en modifiant la durée et la hauteur de son de segments de parole gabarits stockés à l'avance.

Un autre domaine d'application de la présente invention est constitué par le système de synthèse de son musical tel que l'instrument musical électronique du procédé d'échantillonnage. Puisque pratiquement la totalité du son contenu dans la gamme des instruments musicaux électroniques est codée en forme d'onde numérique, est stockée et est reproduite lorsque nécessaire depuis le clavier, etc... dans l'art antérieur pour les procédés d'échantillonnage destinés

à des instruments musicaux électroniques, un inconvénient est constitué par le fait qu'une mémoire importante est nécessaire pour le stockage du son musical. Cependant, si le procédé de décomposition en formes d'onde périodiques et de relocalisation d'ondelette de la présente invention est utilisé, la taille mémoire nécessaire peut être réduite de façon significative du fait que les sons de diverses hauteurs de son peuvent être synthétisés en échantillonnant les tons de seulement peu de sortes de hauteurs de son. Le son musical est constitué typiquement par trois parties, c'est-à-dire une attaque, un soutenu et une chute. Puisque l'enveloppe spectrale varie progressivement non seulement entre les trois parties mais également au sein même du soutenu, le timbre varie également en conséquence. Par conséquent, si les segments de son musical sont codés conformément au procédé de décomposition en formes d'onde périodiques décrit ci-avant et sont stockés en prenant des points appropriés auxquels le spectre varie sensiblement en tant que points temporels de limite, et si le son est synthétisé conformément au procédé de relocalisation d'ondelette basé sur la modification temporelle décrite ci-avant lorsque des requêtes sont entrées depuis le clavier, etc..., alors le son musical présentant une hauteur de son souhaitée arbitraire peut être synthétisé. Cependant, dans les cas où le signal de son musical est soumis à une déconvolution conformément au procédé d'analyse à prédiction linéaire, puisqu'il y a une tendance à ce que l'enveloppe spectrale ne soit pas obtenue et à ce que l'impulsion de hauteur de son ne soit pas brutale, il est recommandé de réduire le nombre de paramètres d'enveloppe spectrale utilisés pour l'analyse et de différencier le signal avant analyse.

Bien que l'invention ait été décrite dans sa forme particulière moyennant un certain degré de particularité, l'homme de l'art appréciera que la présente description de la forme particulière ait été présentée seulement à titre d'exemple et que de nombreuses modifications au niveau des détails de la construction, de la combinaison et de l'agencement des parties puissent être envisagées sans que l'on s'écarte ni de l'esprit ni du cadre de l'invention.

REVENDECATIONS

1. Procédé de codage de segment de parole destiné à une utilisation dans un système de synthèse de la parole, caractérisé en ce qu'il comprend :

la formation d'ondelettes en obtenant des paramètres qui représentent une enveloppe spectrale dans chaque intervalle temporel d'analyse en analysant un signal numérique périodique ou quasi-périodique en utilisant une technique d'estimation spectrale, en soumettant à une déconvolution le signal original pour obtenir une réponse impulsionnelle représentée par les paramètres d'enveloppe spectrale et un signal de train d'impulsions de hauteur de son périodique ou quasi-périodique présentant une enveloppe spectrale pratiquement plate, et en soumettant à une convolution un signal d'excitation obtenu en rajoutant des échantillons de valeur zéro après un signal impulsionnel de hauteur de son d'une période obtenu en segmentant ledit signal de train d'impulsions de hauteur de son pour chaque période de telle sorte qu'une impulsion de hauteur de son soit contenue dans chaque période et qu'une réponse impulsionnelle correspondant aux paramètres d'enveloppe spectrale soit contenue dans le même intervalle temporel que ledit signal d'excitation ; et

le codage de forme d'onde d'ondelettes de chaque période, le stockage de celles-ci dans une mémoire et au moment de la synthèse de la parole, le décodage des ondelettes, le réglage de la durée et de la fréquence de hauteur de son en assignant aux ondelettes des points temporels appropriés de telle sorte qu'elles présentent un motif de hauteur de son souhaité, leur localisation au niveau desdits points temporels et leur superposition pour ainsi synthétiser la parole.

2. Procédé de codage de segment de parole selon la revendication 1, caractérisé en ce que les ondelettes sont formées en faisant correspondre une information obtenue en codant en forme d'onde un signal impulsionnel de hauteur de son de chaque période obtenue par segmentation à une information obtenue en codant un jeu de paramètres d'estimation d'enveloppe spectrale d'une période pendant le même intervalle temporel et en stockant ladite information dans lesdites mémoires, et au moment de la synthèse, les ondelettes sont constituées en soumettant à une convolution un signal d'excitation obtenu en rajoutant des échantillons de valeur zéro après un signal impulsionnel de hauteur de son d'une période obtenu en décodant ladite information et une réponse impulsionnelle correspondant aux paramètres d'enveloppe spectrale décodés contenus dans le même intervalle temporel que ledit signal d'excitation.

3. Procédé de codage de segment de parole selon la revendication 2, caractérisé en ce que le son de parole synthétique est produit au moment de la synthèse en soumettant à une convolution un signal d'excitation synthétique formé en assignant des signaux impulsionnels de hauteur de son obtenus en décodant ladite information à des points temporels appropriés de telle sorte qu'ils présentent un motif de hauteur de son souhaité et en les localisant auxdits points temporels, en ce qu'un intervalle de suppression se produisant lorsqu'une période de hauteur de son souhaitée est plus longue qu'une période de hauteur de son originale est rempli par des échantillons de valeur zéro et en ce que dans un intervalle de chevauchement qui se produit lorsque ladite période de hauteur de son souhaitée est plus courte que ladite période de hauteur de son originale, les signaux impulsionnels de hauteur de son en chevauchement sont ajoutés les uns aux autres

ou un unique signal pris parmi ceux-ci est sélectionné, et en ce qu'une réponse impulsionnelle variable en fonction du temps correspondant à un jeu de paramètres d'enveloppe spectrale synthétique est formée soit en comprimant soit en dilatant temporellement le jeu de fonctions du temps desdits paramètres sur une base sous-segment par sous-segment en fonction du fait que la durée d'un sous-segment contenu dans un segment de parole qui doit être synthétisé est plus courte ou plus longue que celle d'un sous-segment correspondant contenu dans le segment de parole original, soit en localisant le jeu de fonctions du temps desdits paramètres d'une période d'une façon synchrone par rapport audit signal impulsionnel de hauteur de son correspondant d'une période localisée pour former ledit signal d'excitation synthétique, dans lequel, dans le dernier cas, un paramètre d'enveloppe spectrale synthétique contenu dans ledit intervalle de suppression est obtenu soit en répétant la valeur du paramètre d'enveloppe spectrale au point de fin de la période précédente, soit en répétant la valeur du paramètre d'enveloppe spectrale au premier point de la période suivante, soit en répétant une valeur moyenne desdites deux valeurs, soit en effectuant un remplissage avec des valeurs reliant sans discontinuité lesdites deux valeurs, soit en répétant les valeurs des paramètres d'enveloppe spectrale au point de fin et au premier point des périodes précédente et suivante, avant et après le centre de l'intervalle de suppression et ledit paramètre d'enveloppe spectrale synthétique contenu dans ledit intervalle de chevauchement est obtenu en sélectionnant l'un ou l'autre des paramètres d'enveloppe spectrale en chevauchement ou en utilisant une valeur moyenne desdits deux paramètres en chevauchement.

4. Procédé de commande de hauteur de son d'un système de synthèse de la parole, caractérisé en ce qu'il comprend :

le contrôle simultané de la durée et de la hauteur de son d'un segment de parole au moyen d'un procédé de relocalisation d'ondelette basé sur la modification temporelle en codant des points temporels de limite incluant un point de début, un point de fin et un point de régime permanent inclus dans un segment de parole et des positions d'impulsions de hauteur de son de chaque ondelette ou signal impulsionnel de hauteur de son d'une période et en les stockant dans des mémoires simultanément au moment du stockage de chaque segment de parole, et au moment de la synthèse, l'obtention d'une fonction de modification temporelle en comparant des points temporels de limite souhaités et des points temporels de limite originaux stockés correspondant auxdits points temporels de limite souhaités, la recherche des points temporels originaux correspondant à chaque position d'impulsion de hauteur de son souhaitée en utilisant ladite fonction de modification temporelle, la sélection des ondelettes présentant des positions de hauteur de son les plus proches desdits points temporels originaux, la localisation de celles-ci aux positions d'impulsion de hauteur de son souhaitées et la superposition desdites ondelettes.

5. Procédé de commande de hauteur de son d'un système de synthèse de la parole selon la revendication 4, caractérisé en ce qu'il comprend en outre la production d'une parole de synthèse en sélectionnant des signaux impulsionnels de hauteur de son d'une période et des paramètres d'enveloppe spectrale correspondant auxdits signaux impulsionnels de hauteur de son et en les localisant, la soumission à une convolution dudit signal impulsionnel de hauteur de son

localisé et de la réponse impulsionnelle correspondant auxdits paramètres d'enveloppe spectrale pour produire des ondelettes et la superposition desdites ondelettes produites.

6. Procédé de commande de hauteur de son d'un système de synthèse de la parole selon la revendication 4, caractérisé en outre en ce qu'il inclut la production d'une parole synthétique en sélectionnant des signaux impulsionnels de hauteur de son d'une période et des paramètres d'enveloppe spectrale correspondant auxdits signaux impulsionnels de hauteur de son et en les localisant, et la soumission à une convolution d'un signal d'excitation synthétique obtenu en superposant lesdits signaux impulsionnels de hauteur de son localisés conformément au procédé selon la revendication 3 et d'une réponse impulsionnelle variable en fonction du temps correspondant à des paramètres d'enveloppe spectrale synthétique obtenus en concaténant lesdits paramètres d'enveloppe spectrale localisés conformément au procédé de la revendication 3.

7. Dispositif de synthèse de la parole voisée destiné à une utilisation dans un système de synthèse de la parole, caractérisé en ce qu'il comprend :

un sous-bloc de décodage (9) qui produit une information d'ondelette en décodant des codes d'ondelette en provenance du bloc de stockage de segment de parole (5) ;

un sous-bloc de commande de durée (10) qui produit une information de modification temporelle à partir d'une entrée de données de durée en provenance d'un sous-système de génération de prosodie (2) et à partir de points temporels de limite inclus dans une information d'en-tête en provenance dudit bloc de stockage de segment de parole (5) ;

un sous-bloc de commande de hauteur de son (11) qui produit une information de position d'impulsion de hauteur de son de telle sorte qu'elle présente un motif d'intonation tel qu'indiqué par des données de motif d'intonation à partir d'une entrée de ladite information d'en-tête en provenance dudit bloc de stockage de segment de parole (5), à partir desdites données de motif d'intonation en provenance dudit sous-système de génération de prosodie et à partir de ladite information de modification temporelle en provenance dudit sous-bloc de commande de durée (10) ;

un sous-bloc de commande d'énergie (12) qui produit une information de gain de telle sorte qu'une parole synthétique présente un motif de sollicitation tel qu'indiqué par les données de motif de sollicitation à partir d'une entrée desdites données de motif de sollicitation en provenance dudit sous-système de génération de prosodie (2), à partir de ladite information de modification temporelle en provenance dudit sous-bloc de commande de durée (10) et à partir d'une information de position d'impulsion de hauteur de son en provenance dudit sous-bloc de commande de hauteur de son (11) ; et

un sous-bloc d'assemblage de forme d'onde (13) qui produit un signal de parole voisée à partir d'une entrée de ladite information d'ondelette en provenance du sous-bloc de décodage (9), à partir de ladite information de modification temporelle en provenance dudit sous-bloc de commande de durée (10), à partir de ladite information de position d'impulsion de hauteur de son en provenance dudit sous-bloc de commande de hauteur de son (11) et à partir de ladite information de gain en provenance dudit sous-bloc de commande d'énergie (12).

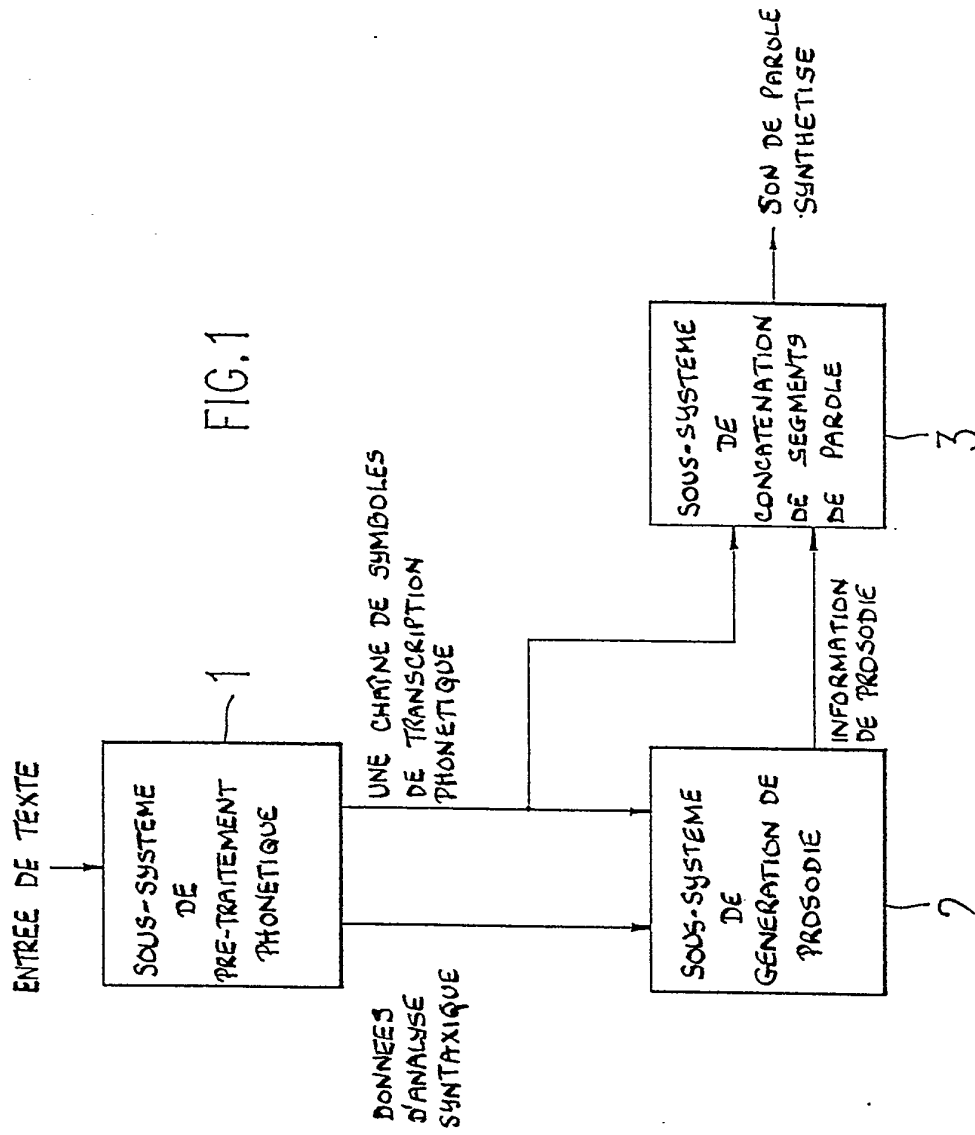
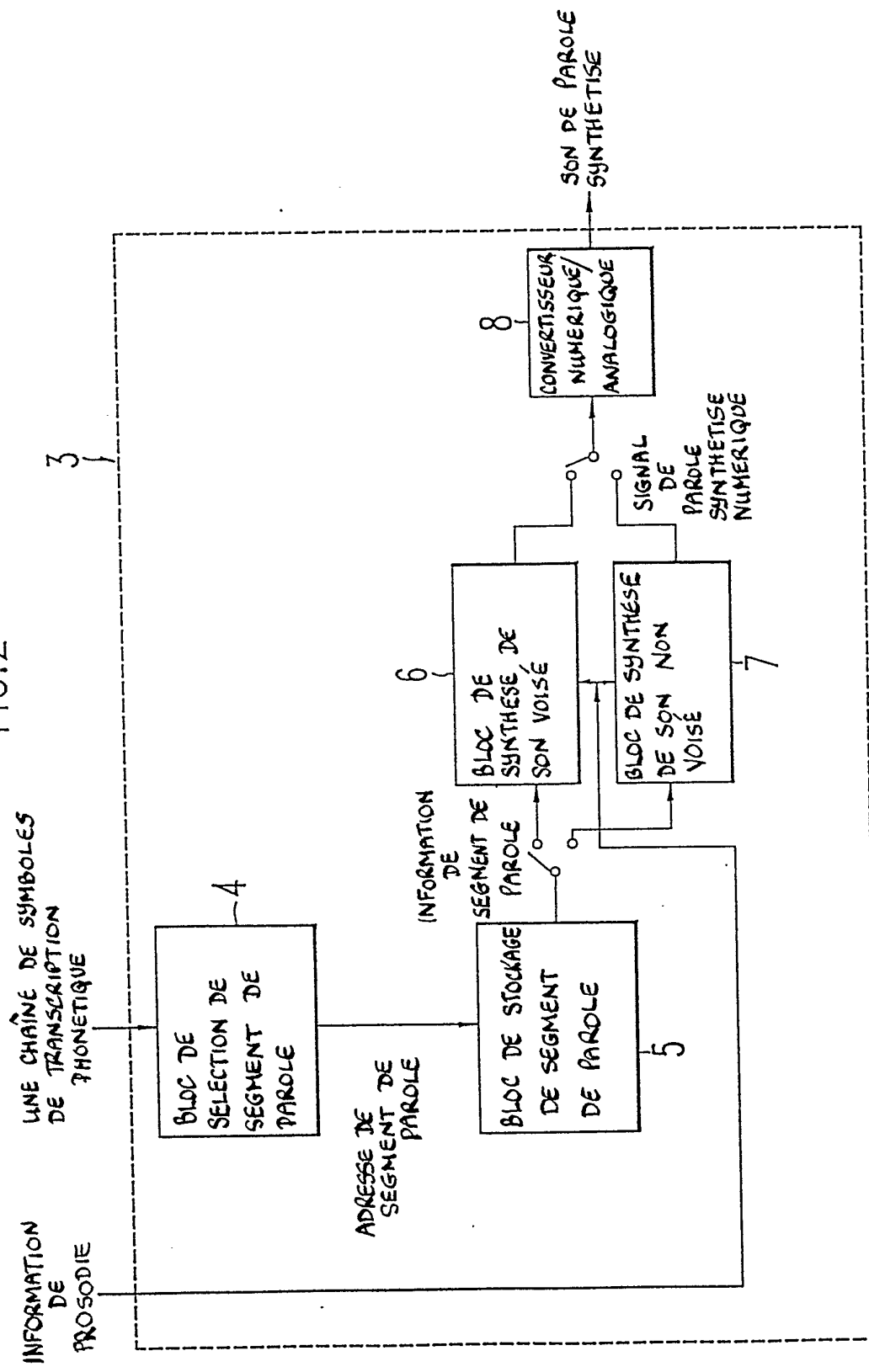
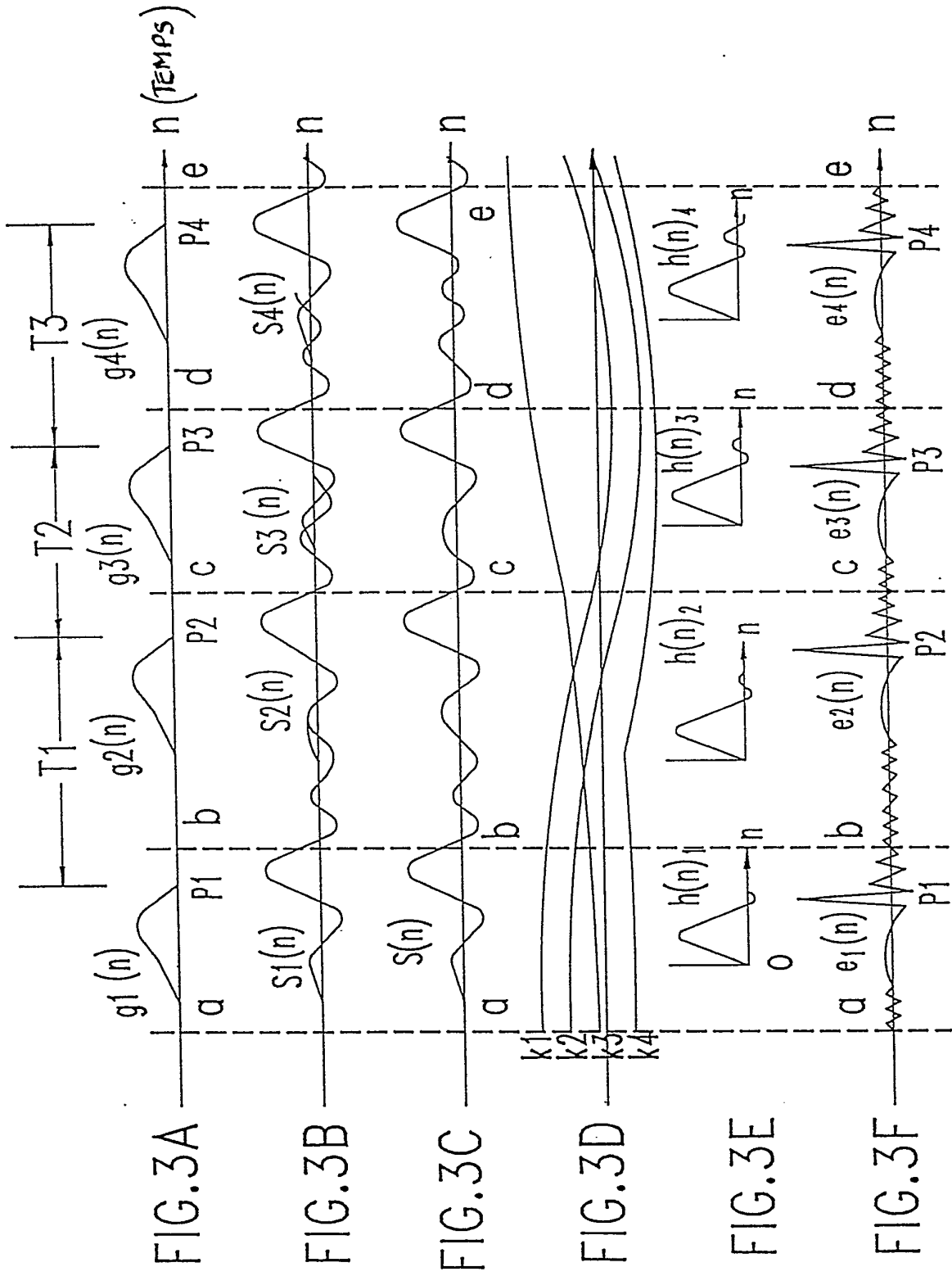
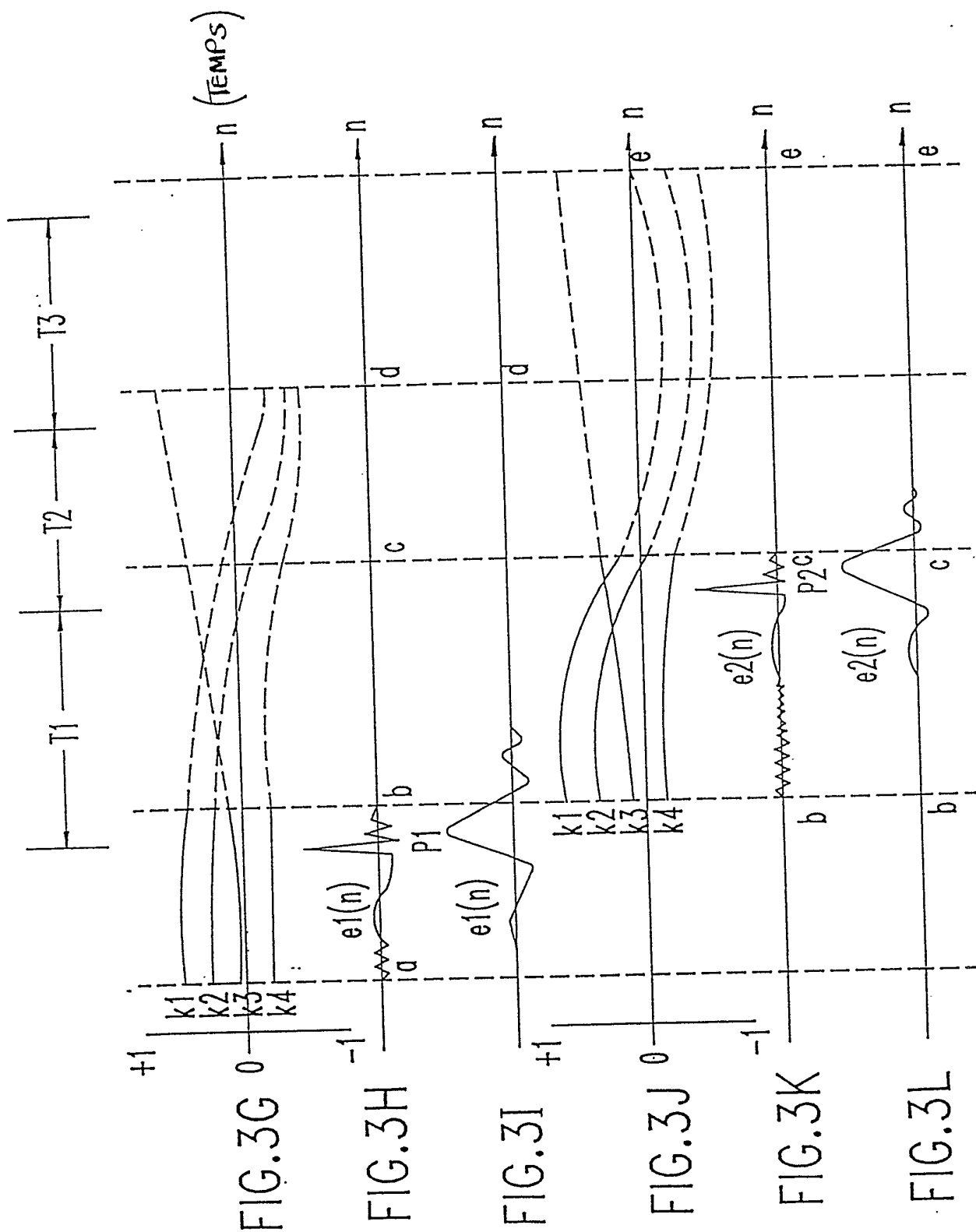
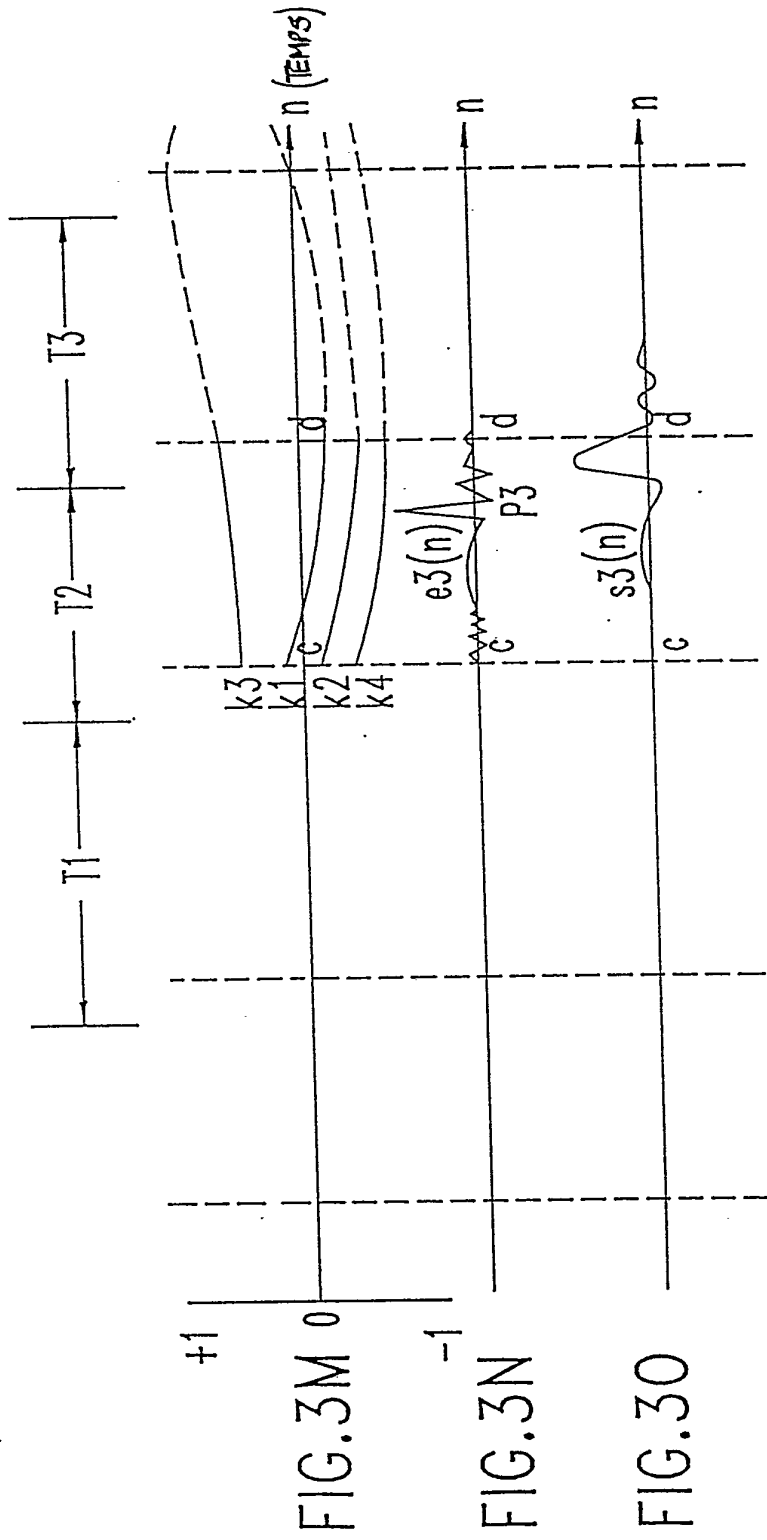


FIG.2









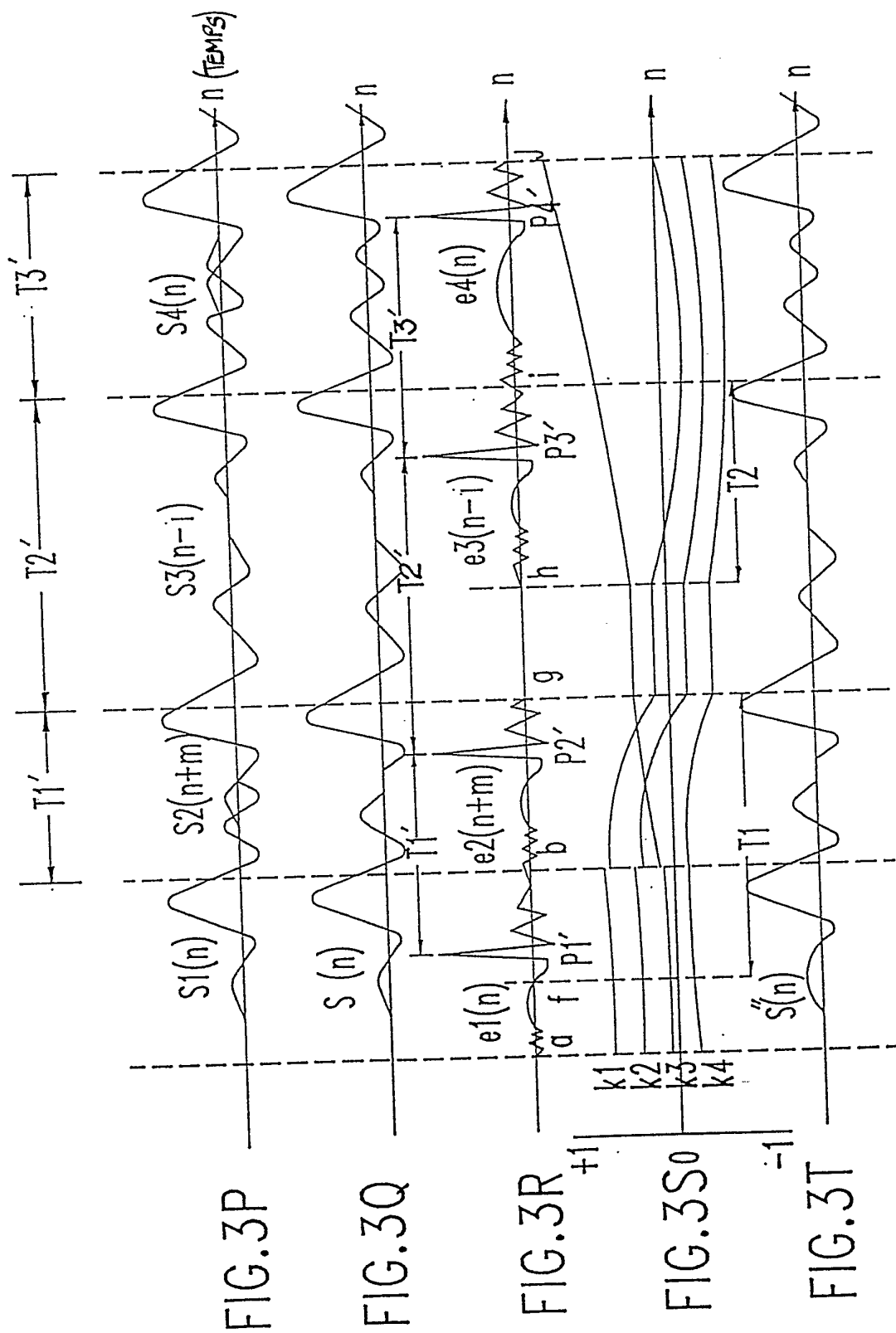


FIG. 4

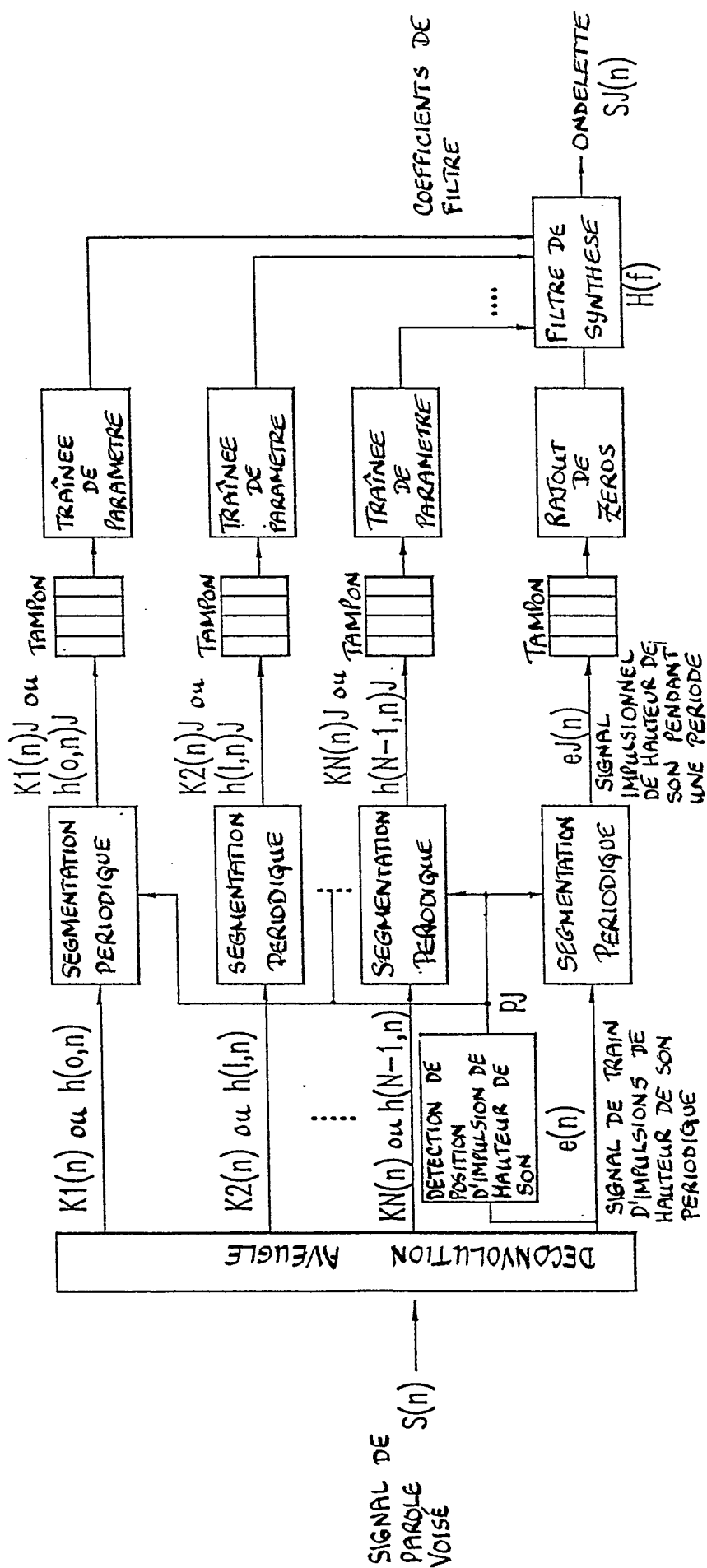


FIG.5A

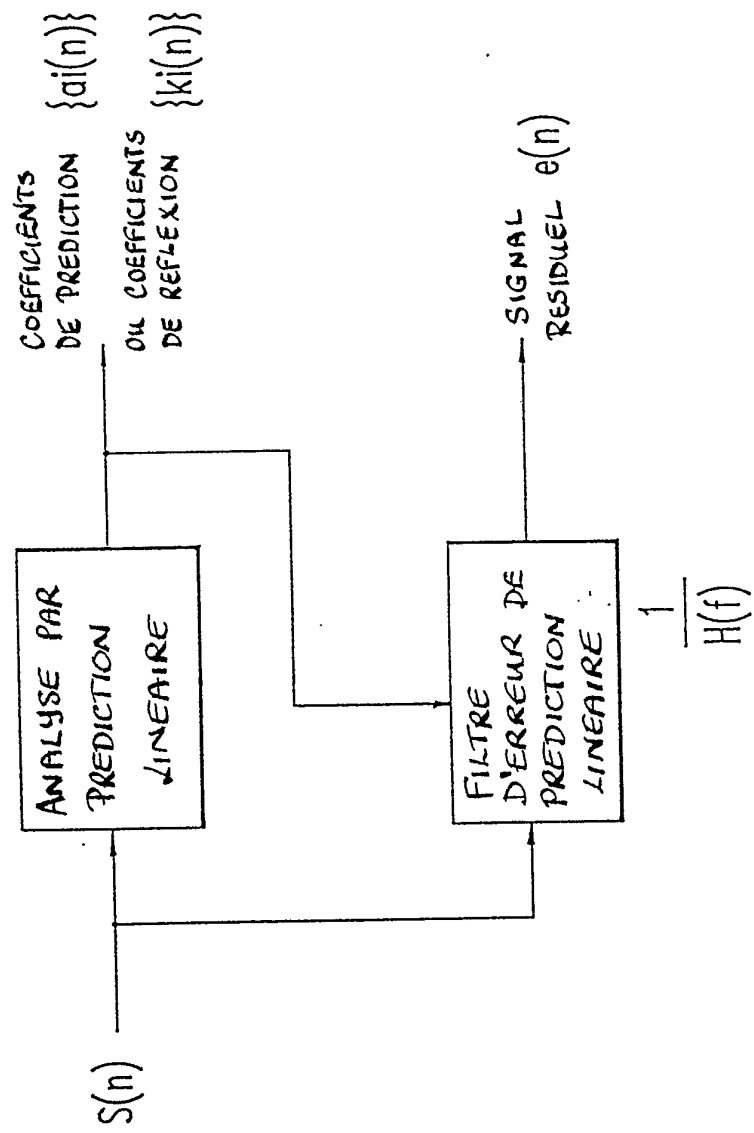


FIG.5B

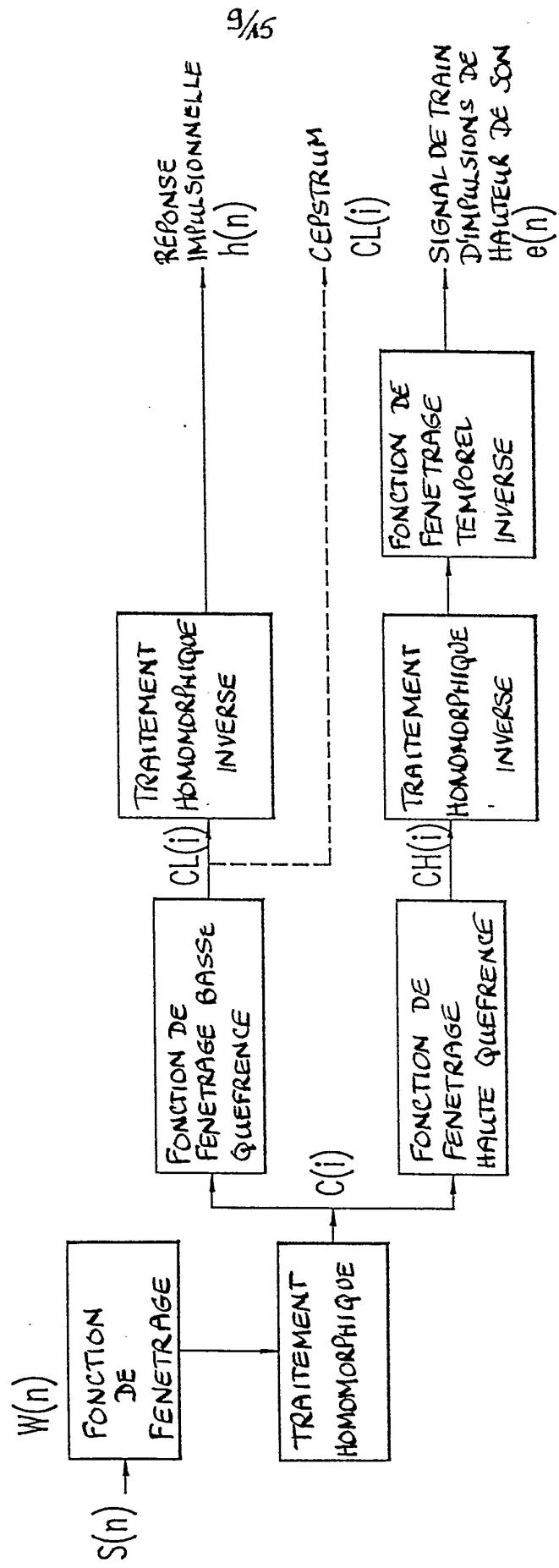


FIG.5C

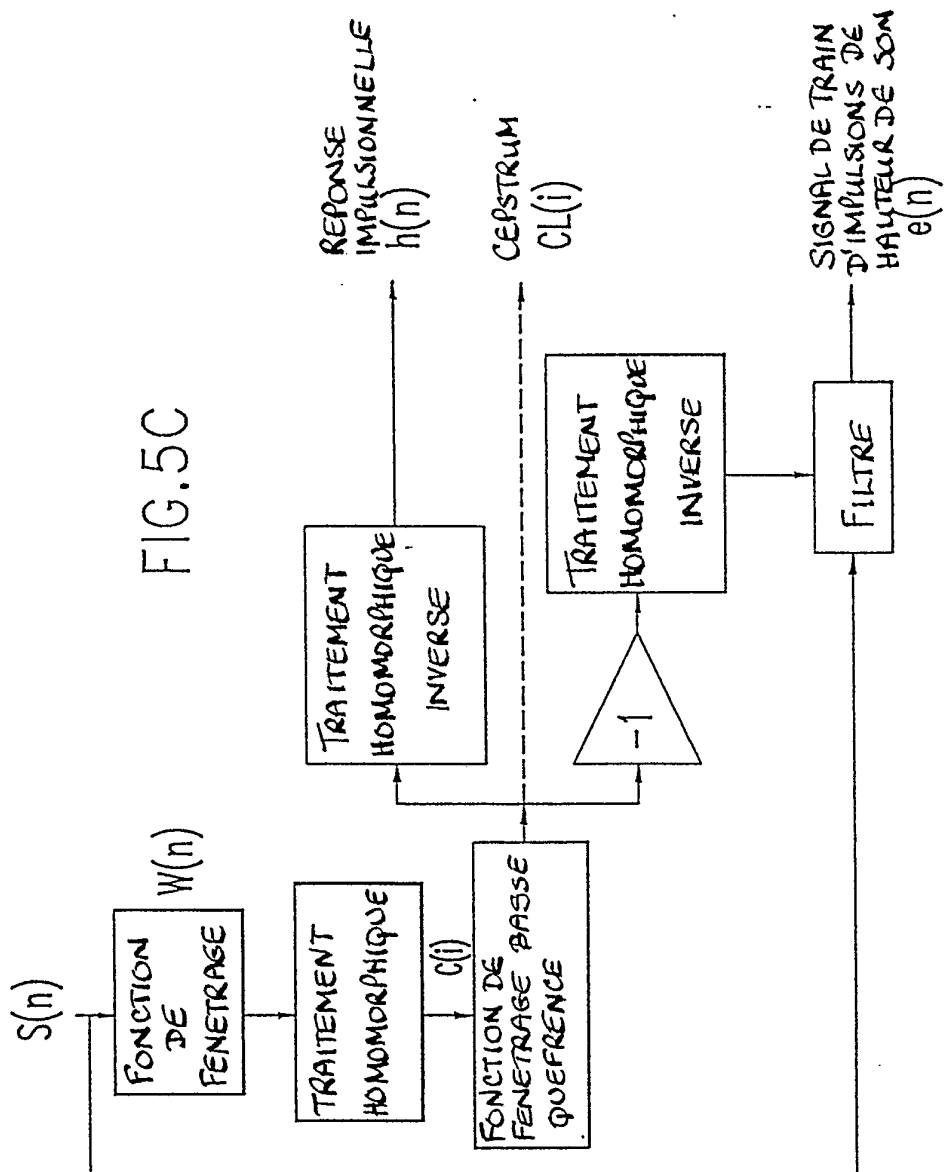


FIG.5D

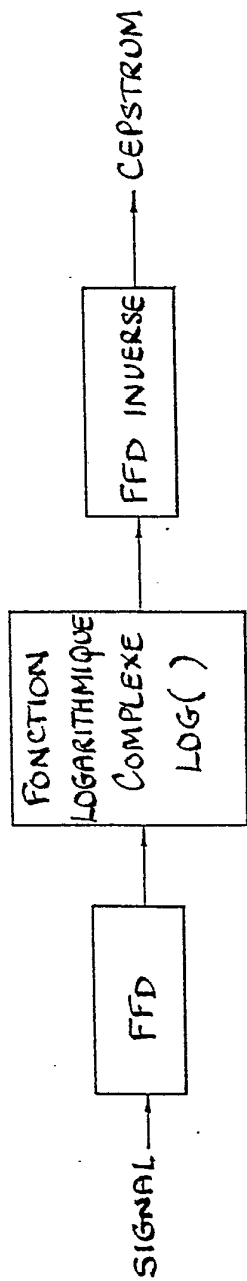


FIG.5E

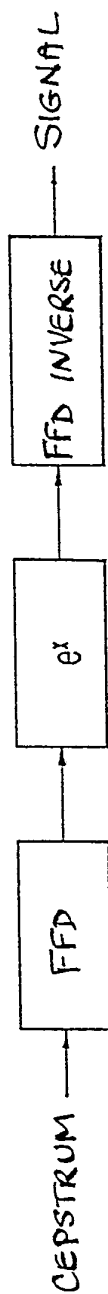


FIG. 6A

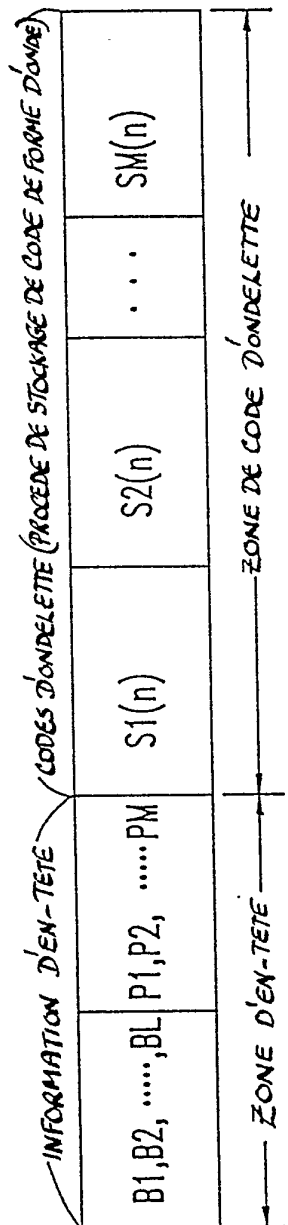


FIG. 6B

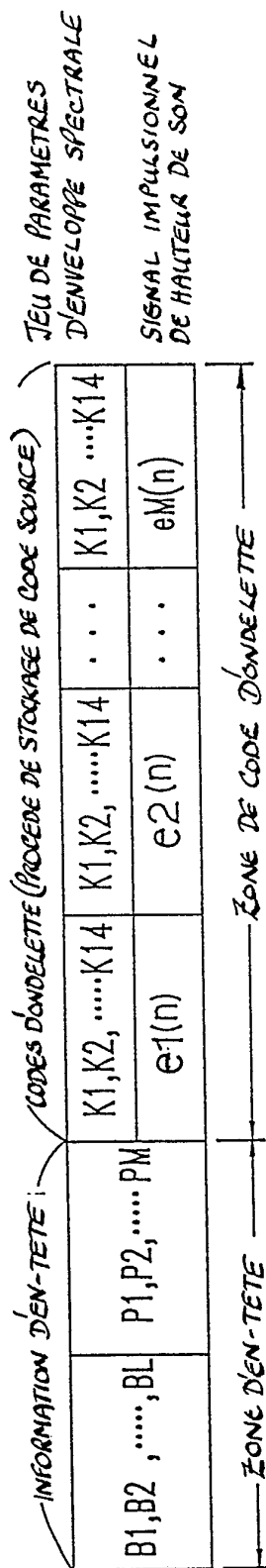
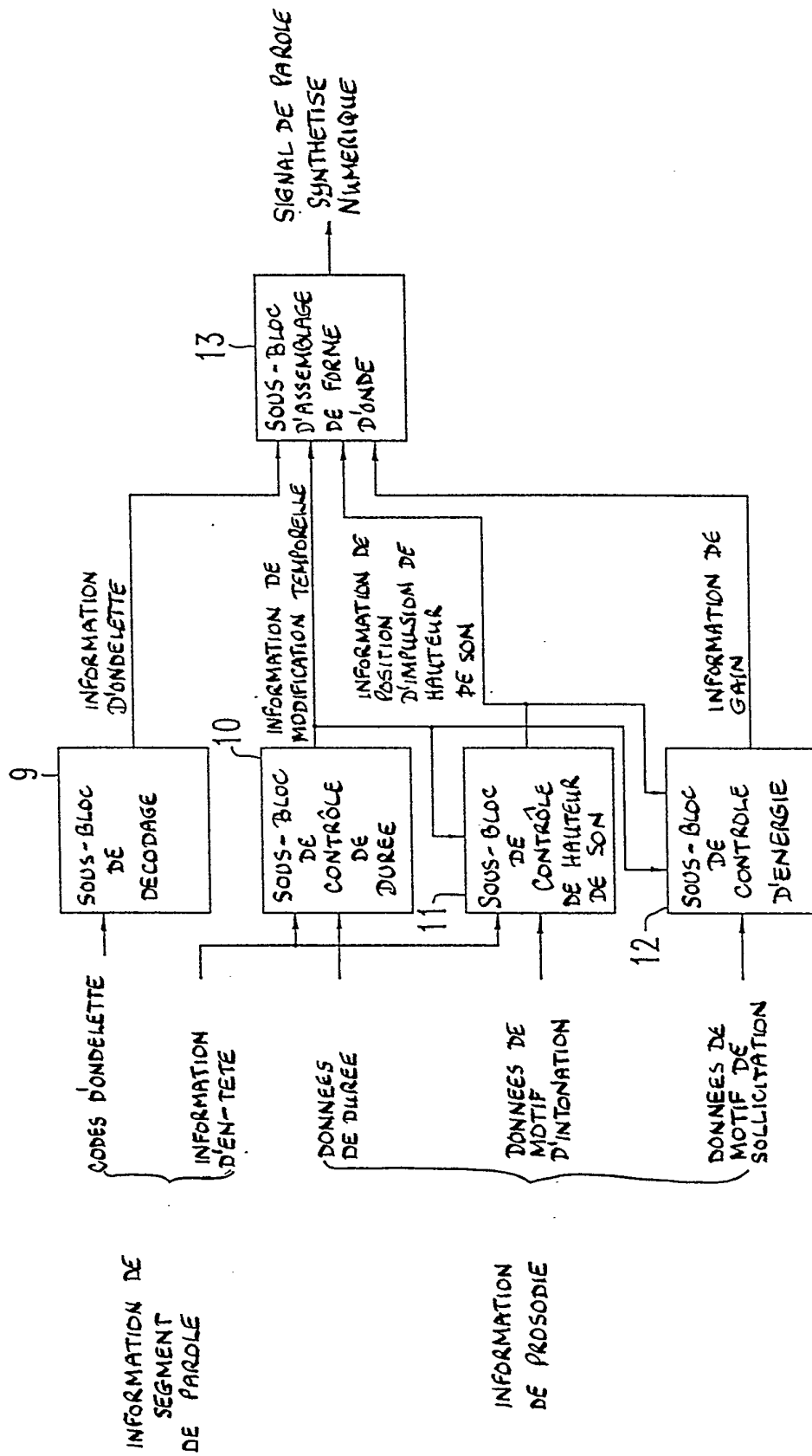


FIG. 7



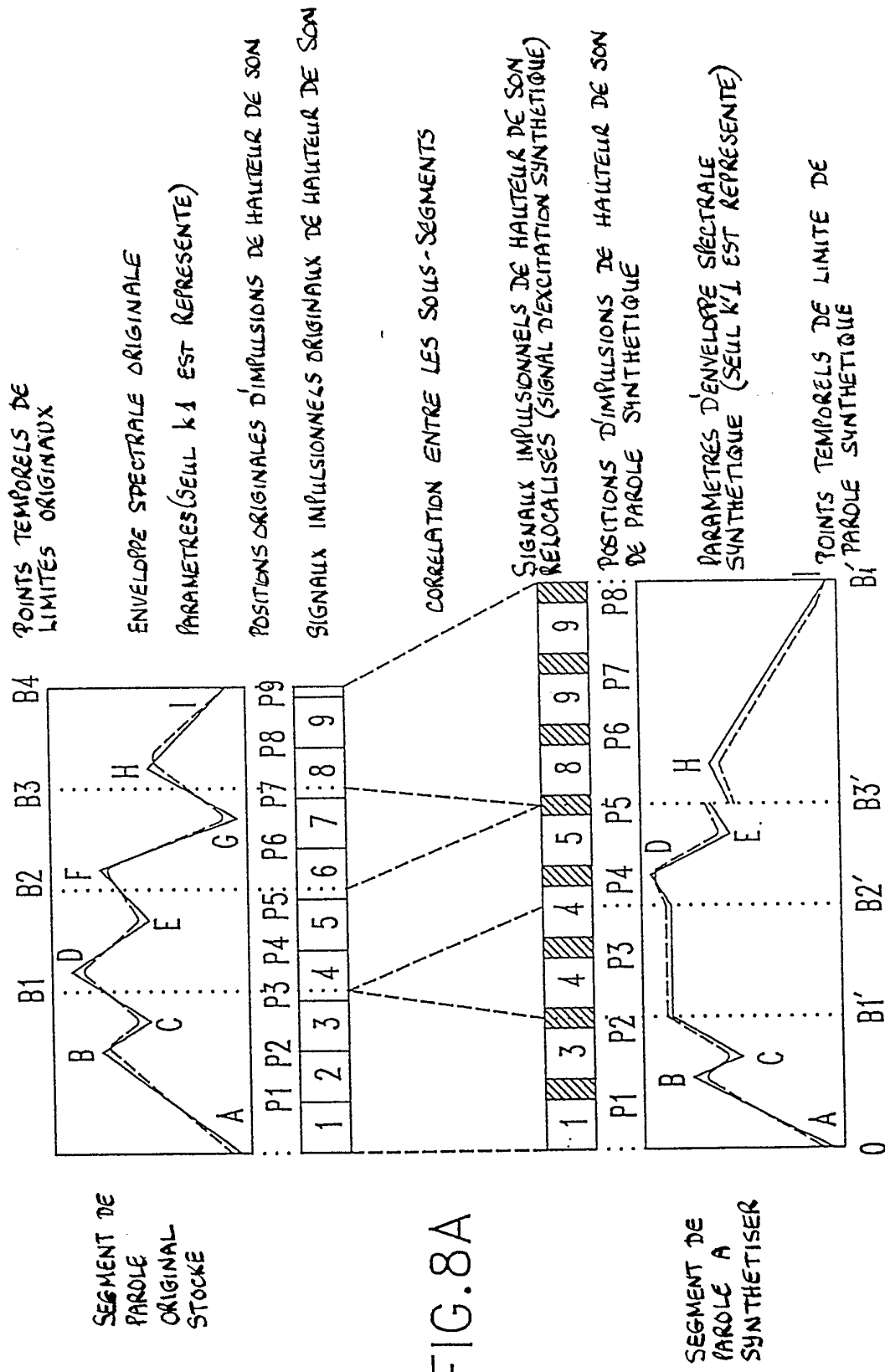


FIG:8B

