(19) RÉPUBLIQUE FRANÇAISE

INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE

PARIS

(11) N° de publication : **2 579 357**
(à n'utiliser que pour les commandes de reproduction)

(21) N° d'enregistrement national : **86 03987**

(51) Int Cl⁴ : G 10 L 9/18.

(12) DEMANDE DE BREVET D'INVENTION A1

(22) **Date de dépôt :** 20 mars 1986.

(30) **Priorité :** US, 21 mars 1985, n° 714 517.

(43) **Date de la mise à disposition du public de la demande :** BOPI « Brevets » n° 39 du 26 septembre 1986.

(60) **Références à d'autres documents nationaux apparentés :**

(71) **Demandeur(s) :** *Société dite : AMERICAN TELEPHONE AND TELEGRAPH COMPANY.* — US.

(72) **Inventeur(s) :** Biing-Hwang Juang, Stephen Eliot Levińson, Lawrence Richard Rabiner et Man Mohan Sondhi.

(73) **Titulaire(s) :**

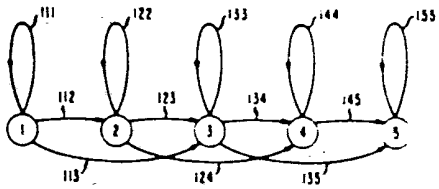
(74) **Mandataire(s) :** Cabinet Flechner.

(54) Procédé et dispositif d'analyse de la parole par modèle de Markov caché.

(57) L'invention concerne les techniques de reconnaissance de la parole.

On forme des gabarits de configurations de parole en utilisant un modèle de Markov qui comporte un certain nombre d'états 1, 2, ..., 5, et en analysant des configurations de parole identifiées pour générer des séquences d'intervalles de temps de signaux de caractéristiques acoustiques représentatives de ces configurations. On produit un gabarit en générant de façon itérative des ensembles successifs de signaux de modèle de Markov, en partant d'un ensemble initial. A chaque itération, on forme un ensemble de signaux représentatif de l'itération courante sous la dépendance des séquences d'intervalles de temps de signaux de caractéristiques acoustiques et de l'un des ensembles de signaux de modèle de Markov précédents, et on compare l'ensemble de l'itération courante avec l'ensemble précédent pour produire un signal correspondant à leur similitude. On arrête les itérations lorsqu'un degré de similitude prédéterminé est atteint.

Application à la commande vocale de machines.



FR 2 579 357 - A1

La présente invention concerne la reconnaissance de configurations, et elle porte plus particulièrement sur des dispositifs destinés à analyser des configurations en utilisant une modélisation du type de Markov.

Dans les systèmes de télécommunication, d'informatique et d'autres systèmes de manipulation d'information, il est souhaitable de prévoir une entrée de reconnaissance de la parole et une sortie vocale synthétisée pour des demandes, des ordres et l'échange d'information. De telles possibilités d'interface faisant usage de la parole permettent l'interaction avec un équipement informatique à partir d'emplacements éloignés, sans utiliser un terminal coûteux, et elles permettent à un utilisateur de communiquer avec des dispositifs du genre des ordinateurs d'une manière naturelle, sans employer de claviers actionnés manuellement. Bien que les avantages d'interfaces utilisant la parole soient bien connus, l'obtention de l'exactitude de reconnaissance de la parole qui est nécessaire pour une utilisation commerciale a présenté des problèmes techniques extrêmement importants. La reconnaissance précise de la parole est relativement difficile à réaliser à cause de la complexité des configurations de parole et des variations de celles-ci entre différents locuteurs. On a obtenu des résultats acceptables dans des applications spécialisées dans lesquelles la reconnaissance est restreinte à des personnes particulières utilisant des vocabulaires limités. Le succès des équipements de reconnaissance automatique de la parole est cependant très limité lorsqu'il n'y a pas de

restriction sur le nombre de locuteurs ou lorsque le vocabulaire de configurations de parole à identifier est étendu.

Les dispositifs de reconnaissance de la parole sont généralement conçus de façon à convertir une configuration de parole inconnue en une séquence de caractéristiques acoustiques fixées, qui est ensuite comparée à des ensembles enregistrés de séquences de caractéristiques acoustiques représentatives de configurations de parole identifiées précédemment. La comparaison permet d'identifier la configuration de parole inconnue comme étant l'ensemble enregistré ayant la séquence de caractéristiques acoustiques la plus similaire, sur la base de critères de reconnaissance prédéterminés. L'exactitude de la reconnaissance de tels systèmes dépend fortement des caractéristiques acoustiques qui sont fixées et des critères de reconnaissance utilisés.

La comparaison entre une configuration de parole inconnue et les ensembles de référence enregistrés peut être directe ou peut être ajustée de façon à tenir compte de différences dans la vitesse d'élocution et de différences d'articulation. Certains systèmes de reconnaissance de la parole utilisent la programmation dynamique pour déterminer l'alignement optimal entre des configurations. Un tel alignement temporel dynamique atténue les effets que les variations de la vitesse d'élocution et de l'articulation ont sur l'exactitude de la reconnaissance. Cependant, les dispositifs de traitement de signal prévus pour effectuer des comparaisons avec alignement temporel dynamique sont complexes et le temps nécessaire pour reconnaître une configuration de parole est fonction de l'étendue du vocabulaire de configurations de référence, ainsi que de la vitesse de fonctionnement de l'équipement de reconnaissance. Lorsque la reconnaissance est indépendante du locuteur, le nombre de configurations de référence est très grand, ce

qui fait que la reconnaissance en temps réel d'une configuration pour des vocabulaires de l'ordre de 50 mots est difficile à réaliser avec une exactitude acceptable.

Une autre approche pour la reconnaissance automatique de la parole utilise une modélisation probabiliste, par exemple au moyen de modèles de Markov, dans laquelle la séquence de caractéristiques acoustiques d'une configuration de parole est transformée en une série de transitions faisant passer par un ensemble d'états, sur la base d'estimations statistiques. On a conçu des dispositifs de reconnaissance dépendants du locuteur, comme celui décrit dans l'article "The DRAGON System-An Overview", par James K. Baker, paru dans IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-23, n° 1, février 1975, pages 24-29, dans lesquels on génère des séquences de caractéristiques acoustiques spectrales correspondant à des configurations de parole, et on les évalue dans une série de modèles de Markov hiérarchisés de caractéristiques acoustiques, de mots et de langue. On analyse des séquences de caractéristiques acoustiques dans des modèles de Markov d'éléments phonémiques. On enchaîne les modèles en éléments acoustiques plus grands, tels que des mots, et on traite ensuite les résultats dans une hiérarchie de modèles de Markov, portant par exemple sur la syntaxe et le contexte, pour obtenir une identification des configurations de parole. Cependant, l'utilisation de modèles d'éléments phonémiques enchaînés et la complexité qui intervient dans une modélisation de Markov hiérarchisée non restreinte, exigent de nombreuses heures d'apprentissage du système par chaque locuteur identifié, pour obtenir un nombre suffisant d'éléments de modèles pour rendre les modèles de Markov valides.

Un système de reconnaissance indépendant du locuteur, décrit dans l'article "On the Application of Vector Quantization and Hidden Markov Models to Speaker-Indepen-

dent, Isolated Word Recognition", par L. R. Rabiner, S.E. Levinson, et M. M. Sondhi, paru dans The Bell System Technical Journal, Vol. 62, n° 4, avril 1983, pages 1075-1105, utilise un modèle de Markov relativement simple comportant un nombre restreint d'états et de transitions entre états. Ce dispositif indépendant du locuteur réduit avantageusement la complexité du traitement de reconnaissance, ce qui fait que la vitesse d'identification d'une configuration de parole dépend moins de la taille du vocabulaire et des possibilités des dispositifs de traitement. Ceci permet d'obtenir une reconnaissance en temps réel.

Bien qu'on puisse simplifier le traitement de reconnaissance de la parole par l'utilisation d'une modélisation de Markov, la génération des signaux qui forment des modèles de configurations de référence auxquels on compare une configuration inconnue est complexe, elle demande beaucoup de temps et elle est sujette à erreurs. Ces facteurs ont empêché l'utilisation pratique de la reconnaissance de la parole par modèles de Markov. Un but de l'invention est de procurer une technique perfectionnée de reconnaissance automatique de la parole basée sur une modélisation de Markov, dans laquelle la formation de modèles est plus rapide et plus exacte.

L'invention comprend les opérations consistant à former une succession de modèles de Markov convergents à partir d'une base de données enregistrée de séquences de signaux de caractéristiques de configuration de parole, à comparer les modèles successifs jusqu'à l'obtention d'une similitude déterminée entre des modèles successifs, et à estimer à nouveau le dernier modèle successif, de façon à obtenir un gabarit de modèle de Markov très exact pour la configuration de parole.

L'invention sera mieux comprise à la lecture de la description qui va suivre d'un mode de réalisation et en se référant aux dessins annexés sur lesquels :

La figure 1 est un schéma d'un modèle de Markov caché et restreint d'une configuration, du type utilisé dans l'invention ;

La figure 2 montre des signaux illustrant la probabilité de caractéristiques et des approximations discrètes et continues de celle-ci pour un modèle de Markov ;

La figure 3 représente un organigramme général du procédé de génération d'un modèle de Markov qui illustre l'invention ;

Les figures 4-6 représentent un organigramme détaillé du procédé de génération de modèle de Markov de la figure 3 ;

La figure 7 représente un schéma synoptique d'un circuit de génération de modèle de Markov qui illustre l'invention ;

La figure 8 représente le dispositif d'enregistrement de séquences de caractéristiques acoustiques dans le circuit de la figure 7 ;

La figure 9 montre un organigramme détaillé représentant des opérations, accomplies sur la figure 4, qui consistent en un alignement d'intervalles de temps de configurations de parole représentées par des modèles de Markov ;

La figure 10 montre la disposition de signaux dans la mémoire d'états d'intervalles de temps du circuit de la figure 7 ;

La figure 11 est un organigramme représentant en détail la formation d'une séquence de symboles d'observation de la figure 6 ;

La figure 12 est un organigramme représentant en détail l'étape de nouvelle estimation du modèle de Markov de la figure 6 ;

La figure 13 représente un schéma synoptique d'un circuit de reconnaissance de la parole utilisant les

gabarits de référence de modèles de Markov de l'invention ; et

La figure 14 est un organigramme représentant le fonctionnement du circuit de reconnaissance de la figure 13.

Comme il est bien connu dans la technique, on peut utiliser un modèle de Markov caché pour évaluer la probabilité d'apparition d'une séquence d'observations $O_1, O_2, \ldots O_t, \ldots, O_T$, dans laquelle chaque observation O_t peut être soit un symbole discret soit un vecteur continu. On peut modéliser la séquence d'observations sous la forme d'une fonction porbabiliste d'une chaîne de Markov sous-jacente, ayant des transitions entre états qui ne sont pas directement observables. La figure 1 illustre un tel modèle. En considérant la figure 1, on note qu'il y a N, par exemple 5 états. Les transitions entre états sont représentées par une matrice de transitions $A=[a_{ij}]$. Chaque terme a_{ij} de la matrice de transitions est la probabilité d'apparition d'une transition vers l'état j alors que le modèle est dans l'état i. La probabilité de symbole de sortie du modèle est représentée par un ensemble de fonctions $B=[b_j(O_t)]$, dans lequel le terme $b_j(O_t)$ de la matrice de symboles de sortie est la probabilité d'apparition en sortie de l'observation O_t, alors que le modèle est dans l'état j. Le modèle de la figure 1 est soumis aux restrictions suivantes : l'état 1 est toujours l'état initial pour le premier intervalle de temps de l'émission vocale, seules sont possibles les transitions appartenant à un ensemble déterminé de transitions entre états de la gauche vers la droite, et on définit un état final prédéterminé à partir duquel il ne peut pas y avoir de transitions vers d'autres états. Ces restrictions sont représentées dans le diagramme d'états de modèle de la figure 1, dans lequel l'état 1 est l'état initial, l'état 5 est l'état final ou "d'absorption", et les transitions de gauche à droite déterminées sont indiquées par les lignes de direction reliant les états.

A partir de l'état 1 de la figure 1, les seules transitions possibles sont le retour à l'état 1 par le chemin 111, la progression vers l'état 2 par le chemin 112 ou la progression vers l'état 3 par le chemin 113. Les transitions sont restreintes au retour dans un état ou au passage à l'un des deux états suivants. De telles transitions sont définies dans le modèle sous la forme de probabilités de transition. A titre d'exemple, une configuration de parole ayant au moment considéré un intervalle de temps de signaux de caractéristiques dans l'état 2 présente une probabilité a_{22} de retour à l'état 2, une probabilité a_{23} de passage à l'état 3 et une probabilité $a_{24}=1-a_{21}-a_{22}$ de passage à l'état 4. La probabilité a_{21} de passage à l'état 1 ou la probabilité a_{25} de passage à l'état 5 est nulle, et la somme des probabilités a_{21} à a_{25} est égale à un.

Dans chaque état du modèle, on peut identifier l'intervalle de temps de caractéristiques courant à un symbole parmi un ensemble de symboles de sortie prédéfinis, ou bien on peut le désigner de manière probabiliste. Dans ce cas, la probabilité du symbole de sortie $b_j(O_t)$ correspond à la probabilité fixée par le modèle pour que le symbole de l'intervalle de temps de caractéristiques soit O_t. La structure de modélisation comprend ainsi une matrice $A=[a_{ij}]$ de probabilités de transitions et une technique de calcul de $B=b_j(O_t)$, c'est-à-dire la probabilité du symbole d'intervalle de temps de caractéristiques dans l'état j.

Conformément à l'invention, on forme un modèle de Markov pour une configuration de référence à partir d'un ensemble de séquences de configurations d'apprentissage identifiées, et les probabilités de symbole de sortie sont des densités de probabilité de fonctions gaussiennes à plusieurs variables. On utilise le modèle de Markov caché restreint pour élaborer un ensemble de gabarits de configurations de référence, chaque gabarit représentant une configuration identifiée dans un vocabulaire de configurations

de référence. On utilise ensuite les gabarits de référence de modèle de Markov pour classer une séquence d'observations comme étant l'une des configurations de référence, sur la base de la probabilité de génération des observations à partir de chaque gabarit de configuration de référence de modèle de Markov. On peut alors identifier la configuration inconnue comme étant la configuration de référence ayant la probabilité la plus élevée.

Pour la reconnaissance de la parole, on analyse une configuration de parole d'entrée inconnue pour générer une séquence d'intervalles de temps de caractéristiques acoustiques. Chaque caractéristique peut être une caractéristique spectrale, un vecteur de coefficients de prédiction linéaire, ou d'autres caractéristiques acoustiques bien connues dans la technique. La séquence de caractéristiques qui est générée pour la configuration de parole inconnue, par exemple un mot ou un membre de phrase, forme la séquence d'observations $\underline{O}$, et on détermine la probabilité que $\underline{O}$ ait été générée par un gabarit de référence de modèle de Markov λ d'un mot ou d'un membre de phrase d'un vocabulaire, par la relation :

$$L(\underline{O}|\lambda) = \sum_{i_1 i_2 .. i_T} \pi_{i_1} b_{i_1}(\underline{O}_1) a_{i_1 i_2} \ldots a_{i_{T-1} i_T} b_{i_T}(\underline{O}_T) \qquad (1)$$

dans laquelle $i_1, i_2, \ldots, i_T$ est la séquence d'états de modèle de Markov et $\underline{O}_1, \underline{O}_2, \ldots, \underline{O}_T$ est la séquence observée de caractéristiques d'intervalle de temps de configuration de parole. On peut écrire l'équation (1) en faisant intervenir la probabilité partielle en sens direct $\alpha_t(i)$, définie par :

$$\alpha_t(i) = L(\underline{O}_1, \underline{O}_2, \ldots \underline{O}_t \text{ et l'état } \underline{i} \text{ à l'instant } t|\lambda) \qquad (2)$$

on peut ensuite évaluer $\alpha_{t+1}(j)$ par la relation :

$$\alpha_{t+1}(j) = \sum_{i=1}^{N} \alpha_t(i) a_{ij} \; b_j(\underline{O}_{t+1}) \qquad (3)$$
$$t = 1, 2, \ldots, T-1$$

ce qui fait que l'équation (1) devient :

$$L(\underline{O}|\lambda) = \sum_{j=1}^{N} \alpha_T(j) \qquad (4)$$

Une fois qu'on a généré le signal de probabilité pour chaque configuration de référence de modèle de Markov, on peut identifier la configuration de parole d'entrée comme étant le modèle de configuration de référence $\underline{r}$ qui correspond au signal $L(\underline{O}|\lambda^r)$ ayant la probabilité la plus élevée ($r=1,2,\ldots,R$), en désignant par R le nombre de mots dans le vocabulaire.

On connaît l'utilisation dans des techniques de reconnaissance d'une quantification vectorielle pour générer des signaux de probabilité de symbole de sortie de modèle de Markov $b(O_t)$. Bien que des techniques de quantification vectorielles permettent l'utilisation de tables de probabilité enregistrées de façon permanente, les probabilités de modèle résultantes ne sont qu'une approximation de la probabilité réelle. Le graphique de la figure 2 illustre l'effet de l'approximation obtenue par quantification vectorielle.

Sur la figure 2, la courbe 201 représente la probabilité réelle $b(\underline{O}_t)$ en fonction de caractéristiques acoustiques, et les lignes verticales correspondent aux valeurs de probabilité discrètes des caractéristiques discrètes quantifiées de façon vectorielle. On quantifie tout d'abord une caractéristique acoustique d'entrée extraite d'une configuration de parole, de façon à la faire correspondre à la caractéristique prototype la plus proche, et on prend $b(O_t)$ pour la probabilité de la caractéristique prototype. On voit aisément qu'une caractéristique d'entrée en x_1 sur l'axe des caractéristiques correspond à une probabilité y_1 d'après la courbe 201, mais que la probabilité pour le vecteur prototype le plus proche, 205, est y_2. La différence entre y_1 et y_2 est l'erreur due

à la quantification, et l'erreur affecte l'exactitude de la reconnaissance de la configuration de parole.

Il est donc souhaitable de réduire l'écart par rapport à la fonction de densité de probabilité réelle 201, en utilisant une approximation par une fonction gaussienne à plusieurs variables de la forme :

$$b_j(x) = \sum_{k=1}^{M} c_{jk} \quad N\ (x,\ \mu_{jk},\quad \sigma_{jk}) \qquad (5)$$

dans laquelle $N\ (x, \mu, \sigma)$ désigne une fonction de densité normale à D dimensions du vecteur moyen μ et de la matrice de covariance σ, M est le nombre de mélanges, et C est la matrice de gain des mélanges. La courbe 210 sur la figure 2 représente une telle fonction gaussienne à plusieurs variables. La fonction gaussienne à plusieurs variables de l'équation 5 est une combinaison linéaire de M fonctions gaussiennes ajustées conformément aux coefficients de mélange C, de façon qu'on puisse approcher la densité de probabilité du modèle de Markov avec n'importe quel degré de précision. La combinaison est analogue à l'approximation d'une fonction en utilisant des fonctions sinus et/ou cosinus en analyse de Fourier. Conformément à l'invention, la différence entre le signal de probabilité $b(\underline{O}_t)$ obtenu à partir de la courbe 210 (déduite de l'équation (5)) et la courbe de densité de probabilité réelle 201 est faible, ce qui fait que l'exactitude de la reconnaissance de configurations de parole à partir des gabarits à modèle de Markov utilisant $b(\underline{O}_t)$, est notablement améliorée.

L'organigramme de la figure 3 illustre le procédé général de formation d'un gabarit à modèle de Markov à partir d'un ensemble d'apprentissage de configurations de parole. On analyse initialement un grand nombre de configurations de parole pour générer des séquences d'intervalles de temps de caractéristiques acoustiques. On supposera à titre d'exemple que les configurations de parole correspon-

dent à des émissions vocales des chiffres 0 à 9, et qu'on effectue une analyse par prédiction linéaire pour chaque émission vocale d'un chiffre. Une seule émission vocale du chiffre 1, par exemple, peut donner une séquence de signaux de caractéristiques acoustiques :

$$O_1^1, O_2^1, \ldots, O_{n_1}^1 \qquad (6)$$

pour n_1 intervalles de temps. Chaque signal de caractéristique acoustique O est un vecteur multidimensionnel de coefficients de prédiction linéaire (CPL), bien connu dans la technique. Selon une variante, on peut utiliser des caractéristiques basées sur la transformée de Fourier du spectre de puissance de la parole, ou d'autres caractéristiques. Les séquences de caractéristiques acoustiques pour un grand nombre de telles émissions vocales de tous les chiffres sont enregistrées pour être utilisées dans la formation de gabarits à modèle de Markov. L'opération d'enregistrement de configurations de parole est indiquée par l'étape 301. Ensuite, à l'étape 305, on génère un modèle de Markov initial pour un chiffre sélectionné. Le modèle de Markov peut être sélectionné arbitrairement, ou bien il peut correspondre à un gabarit à modèle de Markov utilisé précédemment qui doit être amélioré. Le modèle pour le mot r consiste en un ensemble déterminé de signaux de matrice de probabilité de transition $A^r=[a^r_{ij}]$ et en un ensemble déterminé de signaux de probabilité de symbole de sortie, élaborés à partir d'un modèle de Markov :

$$B^r = \{c^r_{mj},\ \mu^r_{mjkl},\ \sigma^r_{mjk}\}$$

$$1 \leq m \leq M,\ 1 \leq j \leq N,\ 1 \leq k, 1 \leq D \qquad (7)$$

correspondant à une fonction gaussienne à plusieurs variables, dans laquelle M est le nombre de composants de mélange, N est le nombre d'états et D est le nombre de dimensions du vecteur de caractéristiques pour chaque intervalle de temps.

On combine ensuite les signaux de caractéristiques O pour le chiffre sélectionné avec les signaux de modèle de Markov courant, pour produire la meilleure séquence d'états de signaux de caractéristiques conformément au modèle de Markov courant pour le chiffre, comme il est indiqué à l'étape 310. On accomplit ceci en affectant des intervalles de temps de signaux de caractéristiques aux états du modèle de Markov de la figure 1, par l'emploi de techniques de correspondance de Viterbi classiques, comme décrit dans l'article précité : "On the Application of Vector Quantization and Hidden Markov Models to Speaker-Independent, Isolated Word Recognition", par L. R. Rabiner et col., de façon à obtenir un score de probabilité maximale.

Une telle procédure est décrite ci-après. On applique au modèle courant chaque séquence d'intervalles de temps de signaux de caractéristiques pour une émission vocale du chiffre sélectionné, et on évalue tous les chemins possibles des intervalles de temps parmi les états du modèle, pour déterminer le chemin ayant la probabilité maximale. Pour le modèle courant λ^r et la séquence d'observations O du signal de caractéristiques d'intervalles de temps, on peut représenter la procédure de détermination de score de la manière suivante :

$$\delta_1(i) = \ln\left[\pi_1 b_1^r(O_1)\right], \quad L=1$$

$$= -1.\ E\ 75 \qquad i = 2,3,\ldots,N \tag{8}$$

$$\delta_t(j) = \max\{\delta_{t-1}(i) + \ln\left[a_{ij}^r\right] + \ln\left[b_j^r(O_t)\right]$$

$$\max\ (1,\ j-2) \leq i \leq j$$

$$\text{pour } 2 \leq t \leq T, \quad i \leq j \leq N \tag{9}$$

$$p = \delta_T(N) \tag{10}$$

On trie ensuite les intervalles de temps du chiffre sélectionné conformément à la séquence d'états optimale, de façon à affecter un ensemble d'intervalles de temps de signaux de vecteurs de CPL à chaque état du modèle de Mar-

kov, ce qui correspond à l'étape 315. On répète cette procédure pour chacun des mots dans l'ensemble d'apprentissage pour le chiffre r. Lorsque ceci est terminé, pour chaque état, on rassemble en M groupes les signaux de vecteurs de caractéristiques de CPL des Q intervalles de temps de toutes les émissions vocales de chiffres sélectionnées qui sont affectées à l'état, en utilisant une procédure de groupement, conformément à l'étape 320. On peut utiliser la procédure de groupement décrite dans l'article "Considerations in applying clustering techniques to speaker-independent word recognition", par L. R. Rabiner et J. G. Wilpen, paru dans le Journal of the Acoustical Society of America, 66(3), septembre 1979, pages 663-673, ou d'autres techniques de groupement.

A l'étape 325, on forme des signaux correspondant à un nouveau modèle de Markov, à partir des signaux de vecteurs groupés de l'étape 320. On produit les nouveaux coefficients c_{mj} conformément à la relation :

$$\hat{c}^r_{mj} = \frac{\text{nombre de vecteurs de caractéristiques dans le groupe m pour l'état j}}{Q} \quad (11)$$

On forme les nouveaux signaux moyens μ_{mjk} conformément à la relation :

$$\hat{\mu}^r_{mjk} = \frac{\sum_{p=1}^{Q} \varepsilon_{r(p)=m} [FS(p)]_k}{\text{nombre de vecteurs de caractéristiques dans le groupe m, pour l'état j}} \quad (12)$$

dans laquelle r(p)=m correspond au vecteur de rang p, FS(p) se trouve dans le groupe m, et $[FS(p)]_k$ correspond à la composante de rang k du vecteur FS(p) ; et on génère les nouveaux signaux de covariance σ_{mjk} en utilisant la relation :

$$\hat{\sigma}^{r}_{mjkl} = \frac{\sum_{p=1}^{Q} [FS(p)]_k [FS(p)]_l}{\text{nombre de vecteurs de caractéristiques dans le groupe m pour l'état } \underline{j}} - (\mu^{r}_{mjk})(\mu^{r}_{mjl}) \quad (13)$$

dans laquelle $\underline{m}$ est le groupe de rang m et $\underline{k}$ est la composante de rang k du signal de vecteur de caractéristiques de CPL. La boucle allant de l'étape 315 à l'étape 325 est itérée sur l'ensemble d'états 1 à N du chiffre sélectionné.

A ce point, on a utilisé le modèle de Markov courant pour former un modèle de Markov immédiatement suivant du chiffre sélectionné, caractérisé par les signaux de coefficients $\hat{c}_{mj}$, les signaux moyens $\hat{\mu}_{mjk}$ et les signaux de covariance $\hat{\sigma}_{mjk}$ des équations (10), (11) et (12). On peut montrer de façon rigoureuse que les modèles de Markov successifs formés par ce procédé sont convergents. Conformément à l'invention, on forme un signal correspondant à chaque modèle de Markov de façon à pouvoir comparer quantitativement les modèles, jusqu'à ce que la différence entre des modèles de Markov successifs soit inférieure à une valeur prédéterminée.

On effectue la comparaison des modèles en générant des séquences d'intervalles de temps de vecteurs de signaux de caractéristiques à partir du modèle de Markov courant et du modèle immédiatement suivant. Dans la formation d'une séquence de vecteurs de signaux de caractéristiques, le modèle commence dans l'état i=1. On choisit de façon aléatoire un vecteur d'intervalle de temps d'observation, O, conformément à la courbe de probabilité $b_1(O)$ (courbe 210 de la figure 2) pour le modèle. On sélectionne ensuite de façon aléatoire une probabilité de transition, et on effectue la transition entre états qui est donnée par a_{1j}. La transition résultante place le modèle dans l'état $\underline{j}$ (par exemple j=1,2,3). On génère ensuite de façon aléatoire un vecteur de signaux de caractéristiques, O, conformément à la probabilité $b_j(C)$, et on génère une autre tran-

sition entre états pour les caractéristiques. On poursuit cette procédure du type Monte Carlo pour produire une séquence de vecteurs de signaux de caractéristiques, et on détermine la probabilité d'apparition de la séquence conformément à l'équation (1) pour le modèle de Markov courant λ. On répète la procédure pour les signaux de probabilité $\underline{\hat{a}}$ et $\underline{\hat{b}}$ du modèle de Markov immédiatement suivant, $\hat{\lambda}$. On produit des signaux correspondant à la probabilité de l'ensemble de vecteurs de caractéristiques $L(O|\lambda)$ et $L(O|\hat{\lambda})$ pour les modèles λ et $\hat{\lambda}$ (étape 335) et on les compare (étape 340) au moyen d'un signal de distance :

$$d(\lambda^r, \hat{\lambda}^r) = \frac{\ln L(O|\lambda^r) - \ln L(O|\hat{\lambda}^r)}{T} \qquad (14)$$

dans lequel T est le nombre total d'observations de caractéristiques d'intervalles de temps qui sont générées par le modèle. Si le signal de distance du modèle de Markov de l'étape 335 est supérieur à une valeur prédéterminée ε, les signaux de modèle de Markov qui viennent d'être produits sont enregistrés en tant que modèle de Markov d'entrée (étape 350) pour l'itération suivante, et on retourne à l'étape 310 pour une autre itération de la boucle de génération du modèle de Markov, de l'étape 310 à l'étape 340.

Lorsque le signal de différence du modèle de Markov est réduit à une valeur inférieure à ε, on passe à l'étape 345 et on effectue une nouvelle estimation des signaux de modèle de Markov formés au cours de la dernière itération de la boucle de génération de modèle de Markov, pour améliorer leur précision. On effectue la nouvelle estimation conformément au procédé de Baum bien connu, décrit dans l'article "An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes", par L. E. Baum, paru dans Inequalities-III, pages 1-8, 1972. Plus précisément, pour un ensemble de S versions d'apprentissage d'un mot, on génère de façon itérative un signal de probabilité

$$L = \prod_{s=1}^{S} \alpha_{T_s}^{(s)} (N) \qquad (15)$$

pour la séquence d'observations de caractéristiques d'intervalles de temps du chiffre sélectionné, provenant de la mémoire 715 :

$$O=\underline{o}_1^{(1)}, \underline{o}_2^{(1)},\ldots,\underline{o}_{T_1}^{(1)},\underline{o}_1^{(2)}, \underline{o}_2^{(2)},\ldots,\underline{o}_{T_2}^{(2)},\ldots,\underline{o}_1^{(S)}, \underline{o}_2^{(S)},\ldots,\underline{o}_{T_S}^{(S)}$$

en désignant par $\underline{o}_t^{(S)}$ un vecteur de signaux de caractéristiques à D dimensions pour $1 \leq s \leq S$, $1 \leq t \leq T_s$.

$$\alpha_t^{(s)}(j)=L(\underline{o}_1^{(s)}\underline{o}_2^{(s)}\ldots\underline{o}_t^{s} \text{ et } j|\lambda) \qquad (16)$$

est la probabilité partielle en sens direct des $\underline{t}$ premières observations de la séquence d'observations O, basée sur le modèle de Markov courant qui se termine dans l'état $\underline{j}$, et λ est le modèle courant. On peut former la séquence d'observations à l'aide des émissions vocales d'entrée correspondant au chiffre et provenant d'un ou de plusieurs locuteurs.

On compare les valeurs successives du signal de probabilité L conformément à la relation :

$$1 - \frac{L}{\overline{L}} \leq \theta \qquad (17)$$

dans laquelle $\overline{L}$ est le signal de vraissemblance généré précédemment, et L est le signal de vraissemblance courant qui est généré. On poursuit les itérations jusqu'à ce que cette condition soit satisfaite ou jusqu'à ce qu'un nombre prédéterminé d'itérations ait été accompli. Chaque fois que l'inégalité de l'équation (17) n'est pas vérifiée, on modifie les signaux du modèle de Markov, a_{ij}, c_{jm}, μ_{jm}, et σ_{jm}, et on forme une nouvelle valeur du signal de probabilité. On peut montrer que la procédure d'estimation converge, ce qui fait que les valeurs des signaux des paramètres du modèle de Markov à la fin de chaque itération procurent une meilleure précision. La modification des signaux de paramètres au cours

de chaque itération est effectuée conformément aux relations suivantes :

$$\bar{a}_{ij} = \frac{\sum_{s=1}^{S}\sum_{t=1}^{Ts-1} \alpha_t^{(s)}(i) a_{ij} b_j(O_{t+1}^{(s)}) \beta_{t+1}^{(s)}(j)}{\sum_{s=1}^{S}\sum_{t=1}^{Ts-1} \alpha_t^{(s)}(i)\beta_t^{(s)}(i)} \quad (18)$$

$$\bar{c}_{jm} = \frac{\sum_{s=1}^{S}\sum_{t=1}^{Ts} \rho_t^{(s)}(j,m)\beta_t^{(s)}(j)}{\sum_{s=1}^{S}\sum_{t=1}^{Ts} \alpha_t^{(s)}(j)\beta_t^{(s)}(j)} \quad (19)$$

$$\bar{\mu}_{jm} = \frac{\sum_{s=1}^{S}\sum_{t=1}^{Ts} \rho_t^{(s)}(j,m)\beta_t^{(s)}(j)(O_t^{(s)})}{\sum_{s=1}^{S}\sum_{t=1}^{Ts} \rho_t^{(s)}(j,m)\beta_t^{(s)}(j)} \quad (20)$$

$$\bar{\sigma}_{jm} = \frac{\sum_{s=1}^{S}\sum_{t=1}^{Ts} \rho_t^{(s)}(j,m)\beta_t^{(s)}(j)(O_t^{(s)}-\mu_{jm})(O_t^{(s)}-\mu_{jm})}{\sum_{s=1}^{S}\sum_{t=1}^{Ts} \rho_t^{(s)}(j,m)\beta_t^{(s)}(j)} \quad (21)$$

avec $1 \leq J \leq N$, $1 \leq m \leq M$

$$\beta^{(s)}(j) = L(O_{it+1}^{(s)} O_{it+2}^{(s)} \ldots O_{iTs}^{(s)} \mid j \text{ et } \lambda) \quad (22)$$

et

$$\rho_t^{(s)}(j,m) = \sum_{n=1}^{N} \alpha_{t-1}^{(s)}(n) a_{nj} c_{jm} \frac{\delta b_j}{\delta c_{jm}} O_t^{(s)} \quad (23)$$

Après que la comparaison de valeurs de probabilité successives à l'étape 340 a fourni un signal inférieur à Θ, ou après l'accomplissement d'un nombre déterminé d'itérations, les signaux de modèle de Markov qui correspondent au dernier signal de probabilité sont enregistrés en tant que gabarit pour le chiffre courant.

On va maintenant passer à une description détaillée en considérant les figures 4, 5 et 6 qui montrent des

organigrammes détaillés illustrant la formation de gabarits de modèle de Markov conformément à l'invention, et la figure 7 qui est un schéma synoptique d'un circuit de traitement qu'on peut utiliser pour mettre en oeuvre le procédé des figures 4, 5 et 6. Sur la figure 7, un transducteur électroacoustique 701 reçoit des émissions vocales identifiées, et des signaux électriques produits par le transducteur sous l'effet de chaque émission vocale sont appliqués à un générateur de séquences de caractéristiques acoustiques 705. Le générateur de caractéristiques accomplit une analyse par prédiction linéaire de l'émission vocale d'entrée, ce qui permet d'obtenir une séquence de signaux de vecteurs de coefficients de CPL qui sont caractéristiques de l'émission vocale. La séquence de signaux de vecteurs est transférée dans la mémoire de séquences de caractéristiques acoustiques 715 par l'intermédiaire de l'interface 710 et du bus 735. Le générateur de séquences de caractéristiques 705 peut comprendre un microprocesseur tel que le microprocesseur du type MC68000 fabriqué par Motorola Inc. et décrit dans la publication MC68000 16 Bit Microprocessor User's Manual, seconde édition, Motorola Inc., 1980.

Les séquences de signaux de vecteurs de CPL provenant du générateur de signaux de caractéristiques 705 peuvent représenter les émissions vocales de chiffres identifiés d'un ou de plusieurs locuteurs, et elles fournissent une base de données de signaux de caractéristiques de parole pour les chiffres, qui est suffisante pour former des modèles de Markov. La figure 8 montre la manière selon laquelle on enregistre les séquences de signaux de vecteurs de chiffres. Chaque rubrique dans la mémoire 715 comprend la séquence d'intervalles de temps de signaux de vecteurs de caractéristiques $O_1, O_2, \ldots, O_{T_i}$ qui résultent de l'émission vocale, un signal w_i qui identifie le chiffre qui est prononcé et un signal T_i qui correspond au nombre d'inter-

valles de temps dans l'émission vocale du chiffre. A titre d'exemple, la première rubrique est une séquence comprenant T_i signaux de vecteurs de CPL d'intervalles de temps, O_1, $O_2,\ldots,O_{T_i}$, l'identification du chiffre, w_1, et le nombre de signaux d'intervalles de temps, T_i. On peut enregistrer les rubriques dans n'importe quel ordre et on peut les adresser au moyen des codes d'identification de chiffres w_i.

Le processeur de signal 725 est conçu de façon à accomplir les opérations exigées par les organigrammes des figures 4-6, pour former les gabarits de modèle de Markov, sous la commande des instructions enregistrées de façon permanente dans la mémoire de programme de formation de gabarits de configurations de parole, 720. Les gabarits de modèle de Markov résultants sont transférés vers la mémoire de gabarits 728 et ils sont dirigés à partir de celle-ci vers le dispositif d'utilisation 740. Le dispositif 740 peut être un dispositif de reconnaissance automatique de la parole, qui peut être du type représenté sur les figures 13 et 14, ou bien il peut s'agir d'un dispositif conçu de façon à enregistrer de manière permanente des gabarits de modèle de Markov, pour l'utilisation dans un équipement de reconnaissance de la parole. Le processeur 725 peut être un processeur universel bien connu dans la technique, ou bien un microprocesseur tel que le MC68000. Selon une variante, le processeur 725, le bus 735, la mémoire de programme 720 et les mémoires vives 715 et 728 peuvent être constitués par l'ordinateur en une seule carte du type OB68K1A MC68000/MULTIBUS, fabriqué par Omnibyte Corporation, West Chicago, Illinois. On pourrait également utiliser une structure de bus Q. La mémoire 720 peut être constituée par une mémoire morte enregistrant de façon permanente des instructions qui réalisent les opérations représentées dans les organigrammes des figures 4-6.

En considérant l'organigramme de la figure 4, on

note que la séquence d'intervalles de temps de signaux de caractéristiques acoustiques est générée dans le générateur de signaux de caractéristiques 705 pour des configurations identifiées de chiffres prononcés, D_L, et les signaux de caractéristiques de CPL qui en sont issus sont placés dans des positions successives de la mémoire 715, conformément à l'étape 401. Lorsque Lmax configurations, par exemple 1000, ont été analysées et lorsque les séquences de signaux de caractéristiques correspondantes ont été enregistrées, le modèle de Markov initial est produit (étape 405) et il est enregistré dans la mémoire de gabarits 728. Selon une variante, des signaux de modèles de Markov peuvent avoir été enregistrés précédemment dans la mémoire de gabarits 728. L'index de configuration de parole (chiffre) est fixé à r=1 à l'étape 410, et le processus de formation de gabarit de modèle de Markov est déclenché à l'étape 415 pour la configuration de chiffre sélectionnée courante. Les signaux du modèle de Markov initial pour le chiffre courant sont fournis par la mémoire de gabarits. Cet ensemble initial de signaux peut être sélectionné arbitrairement ; on peut par exemple avoir : $a_{ij}=1/3$ pour j=i, i+1, i+2 ; i=1,2,...,N-2, $a_{N-1,N}=1/2$, $a_{N-1,N-1}=1/2$, $a_{NN}=1,0$; tout autre $a_{ij}=0$, $C_{mj}=1/M$, $\mu_{mj}=0$, $\sigma_{mj}=1$, ou bien il peut être constitué par des signaux de gabarit formés précédemment qui doivent être améliorés par l'utilisation d'une base de données différente ou développée de signaux de caractéristiques d'émissions vocales.

Les étapes 420 à 465 sont des opérations qui sont destinées à aligner les séquences d'intervalles de temps de signaux de caractéristiques enregistrés pour le chiffre courant avec le modèle de Markov courant λ^r, et à affecter les intervalles de temps des signaux de caractéristiques aux états du modèle. Au début de l'alignement, l'index de mémoire d'intervalles de temps FS(n,p) et l'index de comptage d'intervalles de temps CS(n) sont fixés à zéro

pour la séquence de comptage p=1,2,...,PM, par exemple PM=5000, et pour la séquence d'états de modèle de Markov n=1,2,...,N (étape 420). Les index de comptage et de symbole de mot p et L sont fixés à 1 (étape 425), et les signaux de caractéristiques correspondant à f(L,j), T_L, et W_L pour la configuration de chiffre D_L sont transférés de la mémoire 715 vers le processeur de signal 725 par le bus 735.

Le processeur 725 lit à l'étape 430 les rubriques d'émissions vocales de chiffres présentes dans la mémoire 715, et le code d'identification de chiffre W_L est comparé au chiffre courant qui est traité (étape 425). Si la séquence de signaux de caractéristiques correspond au chiffre courant qui est traité, on passe à l'étape 440 dans laquelle les intervalles de temps de signaux de caractéristiques de la rubrique sont alignés avec le modèle de Markov courant λ^r par l'utilisation des équations (7), (8) et (9). La figure 9 montre un organigramme détaillé du processus d'alignement.

De façon générale, l'organigramme de la figure 9 suit les équations (8), (9) et (10), et il montre les opérations qu'accomplit le processeur de signal 725 pour former un signal δ (N) pour chaque séquence de signaux d'observations de vecteurs de caractéristiques et des signaux γ(i) représentatifs du chemin le plus probable parmi les états du modèle de Markov qui est représenté sur la figure 1. La structure de la figure 9 applique les signaux de probabilité de transition a_{ij} et les signaux de probabilité d'observation $b_j(O_t)$ du modèle de Markov courant λ^r aux intervalles de temps t des séquences d'intervalles de temps de signaux de caractéristiques qui sont enregistrées dans la mémoire de caractéristiques acoustiques 715, pour trier les intervalles de temps de signaux de caractéristiques de façon à les affecter à des états. Les intervalles de temps triés résultants sont ensuite enregistrés dans la mémoire d'intervalles de temps/états

728.

En considérant la figure 9, on note qu'on passe à l'étape 901 à partir de l'étape de décision 435, et que le signal de probabilité initial pour le premier état (i=1) est généré. L'index d'état est alors placé à 1 (étape 905), et la boucle comprenant les étapes 907, 909 et 911 est itérée pour les états 2,3,...,N, de façon à fixer initialement à $-\infty$ les signaux de probabilité $\delta_1(i)$, et à 0 les signaux de chemins entre états d'intervalles de temps, $\Upsilon_1(i)$.

Après avoir restauré les signaux de probabilité et de chemins pour l'état N, on fixe à 1 l'index d'intervalles de temps de séquence et on l'incrémente aux étapes 913 et 915. Les étapes 915 à 940 correspondent aux équations (8) et (9), dans lesquelles on évalue la probabilité de chaque séquence d'intervalles de temps d'observations de caractéristiques, et on détermine les chemins pour la séquence, sur la base du modèle de Markov courant. On effectue l'évaluation sur toute la plage d'intervalles de temps d'observations de signaux de caractéristiques, depuis l'intervalle de temps t=2 jusqu'au dernier intervalle de temps T_L, comme il est indiqué à l'étape de décision 917. Chaque état de destination j dans la plage comprise entre j=1 et j=N est sélectionné jusqu'à ce que la trame T_L ait été traitée (étapes 920, 922 et 924). Pour chaque état de destination j, les états de source i dans la plage permise $\max(1,j-2) \leq i \leq j$ sont sélectionnés séquentiellement (étapes 927 et 929), et un signal représentatif de la probabilité que la séquence d'observations de signaux de caractéristiques suive le chemin déterminé à l'étape 935 est généré.

Le signal de chemin de probabilité maximale pour l'équation (9) est sélectionné aux étapes 937 et 940. Lorsque $i > j$ (étape 931), le signal de probabilité maximale $\delta_t(j)$ est enregistré pour l'état j (étape 945). La boucle allant de l'étape 922 à l'étape 945 est itérée en passant par les états de destination $1 \leq j \leq N$, de façon à emprunter

tous les chemins d'états d'intervalles de temps possibles pour les intervalles de temps allant de t=2 à $t=T_L$. Lorsque $\underline{t} > T_L$, la formation du signal de probabilité est terminée, et on passe à l'étape 950 à partir de l'étape de décision 917.

Les signaux q(t) dans l'organigramme de la figure 9 représentent les estimations de probabilité maximale des états des intervalles de temps d'observations de caractéristiques, basées sur le modèle de Markov courant. Le signal d'état estimé $q(T_L)$ est fixé initialement au dernier état N à l'étape 950, et l'index d'intervalle de temps $\underline{t}$ est fixé au dernier intervalle de temps T_L de la séquence d'observations de caractéristiques à l'étape 955. La boucle allant de l'étape 960 à l'étape 975 est ensuite itérée de façon que les intervalles de temps $Y_t(q(t))$ de la séquence d'observations obtenue à l'étape 940, dans la formation du signal de probabilité de chemins d'états d'intervalles de temps, soient triés conformément à l'état (étape 960) et soient enregistrés dans la mémoire d'états d'intervalles de temps 738 (étape 965). Le traitement du signal de probabilité et le traitement de recherche en arrière de la figure 9 sont accomplis dans le processeur de signal 725, de la manière commandée par les instructions enregistrées dans la mémoire de programme 720.

Après l'achèvement du processus d'alignement et de tri de l'étape 440, les intervalles de temps des signaux de caractéristiques des rubriques correspondant aux chiffres ont été affectés aux états q(j) du modèle λ^r. L'index d'intervalle de temps $\underline{j}$ est fixé à 1 (étape 450), et la boucle comprenant les étapes 455, 460 et 465 est itérée sur tous les intervalles de temps des configurations de parole. A l'étape 455, l'observation de vecteur de signal de caractéristiques f(L,j) de chaque intervalle de temps est enregistrée sous la forme d'un signal FS(q(j),p) dans la mémoire d'états d'intervalles de temps 738, adressée par l'état du

modèle de Markov, et le signal de compte d'intervalles de temps CS(q(j)) dans chaque état est incrémenté. Les index de compte et d'intervalle de temps p et j sont incrémentés à l'étape 460, et l'intervalle de temps à traiter est testé pour déterminer si tous les intervalles de temps d'une rubrique ont été triés (étape 465). Lorsque j est supérieur à T_L, on passe à l'étape 467 à laquelle l'index de rubrique L est incrémenté. Tant que l'index de rubrique L est supérieur au dernier numéro de rubrique Lmax, les rubriques contenues dans la mémoire 715 sont successivement adressées à l'étape 430, et les intervalles de temps des rubriques correspondant au chiffre courant qui est traité sont triés aux étapes 440 à 465.

Lorsque la dernière rubrique pour le chiffre courant a été traitée conformément à l'organigramme de la figure 4, les intervalles de temps des signaux de caractéristiques pour toutes les rubriques de chiffres ont été triés de façon à être affectés à des états, conformément aux signaux du modèle de Markov courant. La figure 10 illustre la disposition des signaux de caractéristiques dans la mémoire d'états d'intervalles de temps 738. Les intervalles de temps affectés à l'état 1 sont identifiés séquentiellement dans des positions de mémoire adressées par l'état 1, et le nombre d'intervalles de temps de l'état 1, CS(1) est inséré à la fin de la séquence de l'état 1. Les affectations d'intervalles de temps pour les autres états du modèle λ^r sont enregistrées de façon similaire dans la mémoire 738.

L'organigramme de la figure 5 illustre le fonctionnement du processeur 725 en ce qui concerne la séparation en groupes des signaux de caractéristiques d'intervalles de temps de chaque état de la mémoire d'états d'intervalles de temps 738, et la génération d'un nouvel ensemble de signaux de modèle de Markov $\hat{c}_{mj}$, $\hat{\mu}_{mjk}$, et $\hat{\sigma}_{mjk}$, à partir des signaux de caractéristiques groupés. On passe à l'étape

501 à partir de l'étape 469 de la figure 4, et on fixe respectivement à 1 et à CS(n) l'index d'état n et le nombre total de signaux de vecteurs de caractéristiques, Q, aux étapes 501 et 505. Q vecteurs de signaux de caractéristiques dans l'état de rang n qui sont enregistrés dans la mémoire d'états d'intervalles de temps 738 sont triés en M groupes à l'étape 510, en utilisant l'algorithme de groupement par quantification vectorielle qui est décrit dans l'article "Distortion Performance of Vector Quantization for LPC Voice Coding", par Bing-Hwang Juang, David Y. Wong et Augustine H. Gray, Jr., paru dans IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-30, n° 2, avril 1982, pages 294-303. Le groupement est effectué par le processeur de signal 725 sous la commande de la mémoire de programme 720.

Après le groupement des signaux de caractéristiques de l'état de rang n, les paramètres pour le modèle de Markov immédiatement suivant, $\hat{c}_{mj}$, $\hat{\mu}_{mjk}$, et $\hat{\sigma}_{mjk}$, sont générés conformément à l'étape 515. L'index de groupe m est incrémenté à l'étape 520 et l'apparition du dernier groupe M est contrôlée à l'étape 525. Lorsque le dernier groupe a été traité, l'index d'état m est incrémenté (étape 530), et l'apparition du dernier état N est contrôlée à l'étape 535. Les étapes 505 à 535 sont accomplies de façon itérative pour la séquence d'états 1,2,...,N, jusqu'à ce que le groupement et la formation des paramètres de modèle de Markov pour l'état N soient terminés.

Une fois que tous les signaux de modèles de Markov ont été formés conformément à la figure 5, le modèle de Markov courant λ^r et le modèle de Markov immédiatement suivant $\hat{\lambda}^r$ sont comparés par les opérations de l'organigramme de la figure 6. On passe à l'étape 605 de la figure 6 à partir de l'étape 535 de la figure 5, et on génère à partir du modèle de Markov courant λ^r un ensemble de K séquences de signaux d'observations :

$$O=O_1^1O_2^1\ldots O_{T_1}^1\,O_1^2O_2^2\ldots O_{T_2}^2$$

$$O_1^kO_2^k\ldots O_{T_k}^k\ldots O_1^KO_2^K\ldots O_{T_K}^K$$

$$O=O_1,O_2,\ldots,O_K \qquad (24)$$

ayant chacune une durée de séquence :

$$T = T_k, \text{ de } k=1 \text{ à } K \qquad (25)$$

et ayant une durée totale :

$$T = \sum_{k=1}^{K} T_k \qquad (26)$$

Les observations de signaux de caractéristiques de chaque séquence d'intervalles de temps sont produites par la sélection aléatoire de vecteurs d'observations, en utilisant les probabilités de vecteurs d'observations b, et par la sélection aléatoire des probabilités de transitions a du modèle λ^r. La figure 11 montre un organigramme détaillé illustrant la formation de la séquence d'observations de l'étape 605.

Initialement, on passe à la figure 11 à partir de l'étape 535 de la figure 5, et aux étapes 1101 et 1105 on fixe à 1 l'index de compte d'observations de signaux de caractéristiques q, ainsi que l'index d'état de modèle i. Aux étapes 1110 et 1115, on fixe à 0 le compte de durée d'état final n et on fixe à 1 l'index de temps d'observation t. On produit un signal de vecteur d'observations représentatif d'un signal de caractéristiques de vecteur de CPL, en sélectionnant de façon aléatoire une observation O conformément à la fonction de probabilité $b_i(O)$ du modèle de Markov à plusieurs variables λ^r, à l'étape 1120. On effectue une transition vers l'état suivant j, sur la base des transitions permises par le modèle et de la probabilité de chaque transition permise. Pour les modèles de Markov avec progression de la gauche vers la droite, seules sont permises les transitions vers l'état courant et vers

les deux états suivants. A l'étape 1130, l'état courant i devient l'état de destination j de l'étape 1125. Aussi longtemps que l'état courant i n'est pas l'état final N (étape 1135), on incrémente l'index de temps (étape 1142) et on retourne à l'étape de génération de symboles d'observations de signaux de caractéristiques 1120. Lorsque l'état final N est atteint (état 1135), l'index de durée d'état n est incrémenté (étape 1140), et la génération de symboles d'observations est itérée dans la boucle comprenant les étapes 1120, 1125, 1130, 1135, 1140, 1142 et 1145, jusqu'à ce que le compte de durée d'état atteigne une valeur prédéterminée (étape 1145). A ce moment, le compte de séquence d'observations est augmenté (étape 1150), et la séquence d'observations suivante est formée par l'intermédiaire de l'étape 1105. Une fois que les K séquences ont été générées et enregistrées, la formation de séquences de symboles d'observations représentée sur la figure 11 est accomplie pour le modèle de Markov $\hat{\lambda}^r$, en changeant les signaux de paramètres de modèle a et b dans les étapes de la figure 11. On passe à l'étape 610 de la figure 6 à partir de l'étape 1155 de la figure 11.

A ce point du processus de génération de gabarit, on a généré et enregistré un ensemble de séquences d'observations basées sur le modèle λ^r et un ensemble de séquences d'observations basées sur le modèle $\hat{\lambda}^r$. Du fait que la succession de modèles de Markov produits conformément à l'invention est convergente, une comparaison du modèle courant et du modèle immédiatement suivant montre s'il y a eu une amélioration dans la modélisation. Lorsque l'amélioration est inférieure à un seuil prédéterminé ε, le modèle immédiatement suivant $\hat{\lambda}^r$ est accepté en tant que gabarit pour la configuration de parole sélectionnée.

A l'étape 610, on produit conformément à l'équation (1) un signal $L(O|\lambda^r)$ représentatif de la probabilité qu'une séquence d'observations ait été générée par le modè-

le de Markov λ^r. On forme également un signal $L(O|\hat{\lambda}^r)$, et on évalue la similitude entre les probabilités de modèles, par une mesure de distance conforme à l'équation (14), ce qui correspond à l'étape 615. Si la mesure de distance $d(\lambda^r, \hat{\lambda}^r)$ n'est pas inférieure à un seuil ε, on considère que le modèle $\hat{\lambda}^r$ n'est pas une approximation suffisamment exacte pour être utilisée en tant que gabarit de configuration de référence. Dans ce cas, le modèle de sortie $\hat{\lambda}^r$ devient le modèle d'entrée courant λ^r, et la commande est transférée à l'étape 420 de la figure 4 pour la formation d'un nouveau modèle de sortie successif $\hat{\lambda}^r$. Lorsque la similitude entre les modèles λ^r et $\hat{\lambda}^r$ à l'étape 620 est meilleure que le seuil ε, on effectue une nouvelle estimation des signaux de modèle de Markov de sortie $\hat{\lambda}^r$ à l'étape 625, pour obtenir le modèle de Markov correspondant à la meilleure approximation.

Le procédé utilisé pour effectuer la nouvelle estimation est illustré par l'organigramme de la figure 12. En considérant la figure 12, on note que le signal de probabilité L est initialement fixé à 0 à l'étape 1201, et que l'index d'itération pour la nouvelle estimation, i1, est restauré à 1 à l'étape 1203. Le signal de probabilité de l'itération courante est généré à l'étape 1210 en utilisant la relation de l'équation (15) pour le paramètre α_{Ts} donné par l'équation (16), après que l'index i1 a été comparé à la limite d'itération I1 (étape 1205). Le signal de probabilité de l'itération précédente L est ensuite comparée au signal de probabilité de l'itération courante $\overline{L}$ à l'étape 1215. Aussi longtemps que la relation de l'étape 1215 n'est pas satisfaite, les signaux de paramètres de modèle de Markov sont modifiés aux étapes 1220, 1225, 1230 et 1235 conformément aux équations (18), (19), (20) et (21) de façon à permettre la formation de paramètres de modèle de Markov plus exacts.

A l'étape 1220, le numérateur représente le nom-

bre prévu de transitions de l'état i vers l'état j pour la séquence d'observations O, et le dénominateur correspond au nombre prévu de transitions à partir de l'état i. Le rapport est une estimation de la probabilité de transition $\overline{a}_{ij}$ conformément à l'équation (18). L'étape 1225 fournit une estimation du coefficient de mélange de rang m, c_{jm}, de l'état de rang j, conformément à l'équation (19), qui est obtenue à partir du rapport entre le nombre prévu d'observations provenant du terme de rang m du mélange, pour l'état de rang j, et le nombre prévu d'observations dans l'état de rang j.

L'étape 1230 correspond à la valeur moyenne $\overline{\mu}_{jm}$ des observations de l'équation (20) et cette valeur moyenne est formée à partir du rapport entre la valeur prévue des observations de caractéristiques sur la totalité de la séquence d'observations O, et la probabilité d'une observation de caractéristiques provenant du terme de mélange de rang m dans l'état j. De manière similaire, la variance $\overline{\sigma}_{jm}$ des observations à l'étape 1235 correspond à l'équation (21), et on la forme en déterminant le rapport entre la valeur prévue de la variance d'observations de caractéristiques, et la probabilité qu'une observation de caractéristiques provienne du terme de mélange de rang m dans l'état j.

Après les modifications des signaux de paramètres de modèle de Markov de l'itération courante aux étapes 1220 à 1235, les signaux de paramètres de modèle de l'itération précédente sont remplacés par ceux obtenus dans l'itération courante (étape 1240). L'index d'itération i1 est ensuite incrémenté à l'étape 1245, et le processus retourne à l'étape de décision 1205. Du fait que les signaux de paramètres de modèle de Markov d'itérations successives sont convergents, le rapport entre des signaux de probabilité L et $\overline{L}$ successifs tend vers l'unité au fur et à mesure que l'index d'itération augmente.

Lorsque la condition de l'étape 1215 est satisfaite ou lorsque I1 itérations ont été effectuées, la nouvelle estimation du modèle de Markov pour le gabarit de chiffre courant est terminée, et les signaux de modèle sont enregistrés dans la mémoire de gabarit 728 de la figure 7 (étape 1250). Une fois que les signaux de modèle de Markov pour le gabarit de rang r ont fait l'objet d'une nouvelle estimation conformément aux étapes de la figure 12, le processus passe de l'étape 1250 à l'étape 630, à laquelle l'index de chiffre r est incrémenté. L'index incrémenté est testé à l'étape 635 pour déterminer s'il correspond au dernier gabarit de configuration de parole. La formation de gabarits à modèle de Markov pour les configurations de parole successives se poursuit conformément aux organigrammes des figures 4, 5 et 6, jusqu'à ce que le dernier gabarit de configuration de parole I soit placé dans la mémoire 728. Les gabarits à modèle de Markov pour le vocabulaire de configurations de parole peuvent ensuite être transférés vers le dispositif d'utilisation 740 par l'intermédiaire du bus 735 et de l'interface 730.

On génère un ensemble de signaux de modèle de Markov pour chaque configuration de référence à reconnaître, et on l'enregistre dans une mémoire de modèles de Markov de configurations de référence d'un dispositif de reconnaissance de la parole tel que celui représenté sur la figure 13. Les configurations de référence peuvent être par exemple les chiffres prononcés 0 à 9. Lorsque le dispositif de reconnaissance reçoit en entrée une configuration de parole, par le microphone 1310, il fonctionne conformément à l'organigramme de la figure 14 pour fournir un signal r* qui identifie le chiffre prononcé correspondant à la configuration de parole d'entrée. Sur la figure 13, les ensembles de signaux de modèles de Markov de configurations de référence sont enregistrés de façon permanente dans une mémoire morte 1360, et le fonctionnement du dispositif de

reconnaissance qui est présenté sur la figure 4 est commandé par les instructions enregistrées de façon permanente dans la mémoire de programme de commande 1355. Le générateur de signaux de caractéristiques 1330 peut être constitué par un microprocesseur tel que celui du type MC68000 fabriqué par Motorola Inc., comportant des instructions enregistrées de façon permanente pour commander la génération de signaux de caractéristiques et la détection de fin, comme il est bien connu dans la technique. Le processeur 1340, le bus 1345, la mémoire de commande 1350 et la mémoire vive 1355 peuvent être constitués par l'ordinateur en une seule carte du type OB68K1A MC68000/MULTIBUS, fabriqué par Omnibyte Corporation, West Chicago, Illinois. On pourrait également utiliser une structure de bus Q.

En considérant les figures 13 et 14, on note qu'on passe à l'étape 1407 à partir de l'étape 1405 lorsque le dispositif de reconnaissance est disponible pour traiter une nouvelle configuration de parole d'entrée au niveau du transducteur 1300. Le circuit de filtrage et d'échantillonnage 1310 et le convertisseur analogique-numérique 1320 convertissent la configuration de parole en une séquence de codes numériques représentatifs des échantillons de parole successifs de la configuration, comme il est bien connu dans la technique, et ceci correspond aux étapes 1407 et 1409. Sous l'effet des codes numériques provenant du convertisseur 1320, le générateur de signaux de caractéristiques de CPL/détecteur de fin 1130 génère une séquence de signaux de caractéristiques de configurations de parole (étape 1411), et cette séquence est transférée vers la mémoire de signaux de caractéristiques 1370.

Lorsque la fin de la configuration est détectée (étape 1418), l'ensemble de signaux de modèle de Markov de la première configuration de référence enregistrée est sélectionné (étape 1425) dans le circuit processeur 1340, conformément aux instructions programmées de la mémoire de

programme de commande 1355. Les signaux de caractéristiques de configurations de parole provenant de la mémoire de signaux de caractéristiques 1370 sont combinés dans le processeur 1340 avec l'ensemble de signaux de modèle de Markov sélectionné, provenant de la mémoire 1360, pour former un signal représentatif de la probabilité que la configuration de parole corresponde à la configuration de référence sélectionnée (étape 1430). La boucle comprenant les étapes 1425, 1430 et 1435 est itérée en passant par tous les ensembles de signaux de modèle de Markov de configurations de référence, de façon à enregistrer un signal de probabilité pour chaque configuration de référence. La configuration de référence ayant le signal de probabilité maximal est sélectionnée à l'étape 1440, et le processeur 1340 produit à l'étape 1445 le signal d'index de configuration de référence qui identifie la configuration de parole d'entrée. La commande retourne ensuite à l'étape 1405 pour la reconnaissance de configurations de parole suivantes.

Il va de soi que de nombreuses modifications peuvent être apportées au procédé et au dispositif décrit et représenté, sans sortir du cadre de l'invention.

REVENDICATIONS

1. Procédé d'analyse de la parole, caractérisé en ce qu'il comprend les opérations suivantes : (a) on analyse au moins une configuration de parole identifiée pour générer une séquence d'intervalles de temps de signaux de caractéristiques acoustiques représentatifs de cette configuration ; (b) on génère un premier ensemble de signaux représentatifs d'un modèle de Markov de la configuration de parole identifiée ; et (c) on produit un gabarit de configuration de parole correspondant à la configuration de parole identifiée, sous la dépendance du premier ensemble de signaux de modèle de Markov et de la séquence d'intervalles de temps de signaux de caractéristiques acoustiques; cette opération de production d'un gabarit de configuration de parole comprenant la génération itérative d'ensembles successifs de signaux de modèle de Markov, et chaque itération comprenant les opérations suivantes : (d) on forme un ensemble de signaux représentatifs du modèle de Markov de l'itération courante de la configuration de parole identifiée, sous la dépendance de la séquence d'intervalles de temps de signaux de caractéristiques acoustiques, et de l'un des ensembles précédents de signaux de modèle de Markov ; (e) on compare l'ensemble de signaux de modèle de Markov de l'itération courante avec l'ensemble de signaux de modèle de Markov précédent, pour générer un signal correspondant à la similitude entre eux ; et (f) on répète les opérations (d) et (e) sous la dépendance de l'ensemble de signaux de modèle de Markov formé en dernier et de la séquence d'intervalles de temps de signaux de caractéristiques acoustiques pour les configurations de parole identifiées, jusqu'à ce que le signal de similitude soit inférieur ou égal à une valeur prédéterminée.

2. Procédé d'analyse de la parole selon la revendication 1, caractérisé en ce que l'opération de comparaison

de modèles de Markov comprend les opérations suivantes : on forme un premier signal correspondant à la probabilité d'apparition des séquences de signaux de caractéristiques acoustiques des configurations de parole identifiées, à partir de l'ensemble de signaux de modèle de Markov précédent ; on forme un second signal correspondant à la probabilité d'apparition des séquences de caractéristiques acoustiques des configurations de parole identifiées, à partir de l'ensemble de signaux de modèle de Markov de l'itération courante ; et on génère un signal représentatif de la correspondance entre les premier et second signaux de probabilité.

3. Procédé d'analyse de la parole selon la revendication 2, caractérisé en ce que l'opération de formation de l'ensemble de signaux de modèle de Markov de l'itération courante comprend le tri des signaux de caractéristiques acoustiques des configurations de parole pour les affecter à des états du modèle de Markov précédent, sous la dépendance de l'ensemble de signaux de modèle de Markov précédent et des séquences de signaux de caractéristiques acoustiques des configurations de parole identifiées ; et l'opération de formation de signaux de modèle de Markov de l'itération courante comprend la formation de signaux représentatifs de la densité de probabilité d'apparition de caractéristiques acoustiques, sous la dépendance des séquences de signaux de caractéristiques acoustiques triés des configurations de parole, et de l'ensemble de signaux de modèle de Markov précédent.

4. Procédé d'analyse de la parole selon la revendication 3, caractérisé en ce que le tri des signaux de caractéristiques acoustiques de configurations de parole pour les affecter à des états comprend en outre le groupement en groupes associés des signaux de caractéristiques des configurations de parole affectées à chaque état.

5. Procédé d'analyse de la parole selon la reven-

dication 3, caractérisé en ce que les signaux de densité de probabilité comprennent un signal représentatif du nombre moyen de signaux de caractéristiques acoustiques triés dans chaque groupe de chaque état, un signal représentatif de la valeur moyenne des signaux de caractéristiques acoustiques triés dans ce groupe et dans cet état, et un signal représentatif de la variance des signaux de caractéristiques acoustiques triés dans ce groupe et cet état.

6. Analyseur de parole caractérisé en ce qu'il comprend : des moyens destinés à analyser au moins une configuration de parole identifiée, pour générer une séquence d'intervalles de temps de signaux de caractéristiques acoustiques représentatifs de cette configuration ; des moyens destinés à enregistrer un premier ensemble de signaux représentatifs d'un modèle de Markov de la configuration de parole identifiée ; et des moyens fonctionnant sous la dépendance du premier ensemble de signaux de modèle de Markov et des séquences d'intervalles de temps de signaux de caractéristiques acoustiques, de façon à produire un gabarit de configuration de parole correspondant à la configuration de parole identifiée ; et ces moyens de génération de gabarit de configuration de parole comportent des moyens destinés à générer de façon itérative des ensembles successifs de signaux de modèle de Markov, comprenant des moyens destinés à former un ensemble de signaux représentatifs du modèle de Markov de la configuration de parole identifiée qui correspond à l'itération courante, sous la dépendance des séquences d'intervalles de temps de signaux de caractéristiques acoustiques, et de l'un des ensembles précédents de signaux de modèle de Markov, et des moyens destinés à comparer l'ensemble de signaux de modèle de Markov de l'itération courante avec l'un des ensembles de signaux de modèle de Markov précédents, jusqu'à ce que la différence entre l'ensemble de signaux de modèle de Markov de l'itération courante et l'ensemble de signaux de modèle

de Markov précédent soit inférieure à une valeur prédéterminée.

7. Analyseur de parole selon la revendication 6, caractérisé en ce qu'il comprend en outre des moyens qui réagissent au fait que la différence entre l'ensemble de signaux de modèle de Markov de l'itération courante et l'ensemble de signaux de modèle de Markov précédent est inférieure à ladite valeur prédéterminée, en sélectionnant l'ensemble de signaux de modèle de Markov formé en dernier en tant que gabarit de référence pour la configuration de parole identifiée.

8. Analyseur de parole selon la revendication 6, caractérisé en ce que les moyens de comparaison de modèles de Markov accomplissent les opérations suivantes : ils forment un premier signal correspondant à la probabilité d'apparition des séquences de signaux de caractéristiques acoustiques des configurations de parole identifiées, à partir de l'ensemble de signaux de modèle de Markov précédent ; ils forment un second signal correspondant à la probabilité d'apparition des séquences de caractéristiques acoustiques des configurations de parole identifiées, à partir de l'ensemble de signaux de modèle de Markov de l'itération courante ; et ils génèrent un signal représentatif de la correspondance entre les premier et second signaux de probabilité.

9. Analyseur de parole selon la revendication 8, caractérisé en ce que les moyens destinés à former l'ensemble de signaux de modèle de Markov de l'itération courante comprennent : des moyens destinés à trier les signaux de caractéristiques acoustiques des configurations de parole, de façon à les affecter à des états du modèle de Markov précédent sous la dépendance de l'ensemble de signaux de modèle de Markov précédent et de la séquence de signaux de caractéristiques acoustiques des configurations de parole ; et des moyens destinés à générer des signaux représentatifs

de la densité de probabilité d'apparition de caractéristiques acoustiques, sous la dépendance de la séquence de signaux de caractéristiques acoustiques triés des configurations de parole, et de l'ensemble de signaux de modèle de Markov précédent.

10. Analyseur de parole selon la revendication 9, caractérisé en ce que les moyens destinés à trier les signaux de caractéristiques acoustiques des configurations de parole pour les affecter à des états comprennent en outre des moyens qui groupent en groupes associés les signaux de caractéristiques des configurations de parole qui sont affectés à chaque état.

11. Analyseur de parole selon la revendication 10, caractérisé en ce que les signaux de densité de probabilité de modèle de Markov comprennent un signal représentatif du nombre moyen de signaux de caractéristiques acoustiques triés dans chaque groupe et chaque état, un signal représentatif de la valeur moyenne du nombre moyen de signaux de caractéristiques acoustiques triés dans chaque groupe et chaque état, et un signal représentatif de la variance du nombre moyen de signaux de caractéristiques acoustiques triés dans chaque groupe et chaque état.

FIG. 1

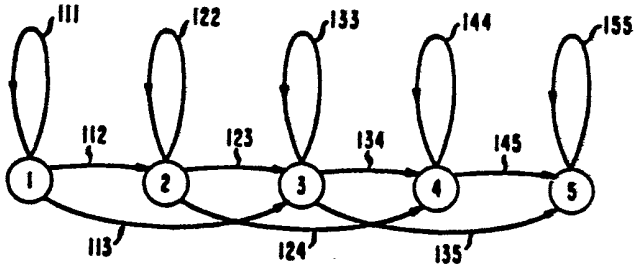

111
122
133
144
155
112
123
134
145
1
2
3
4
5
113
124
135

FIG. 2

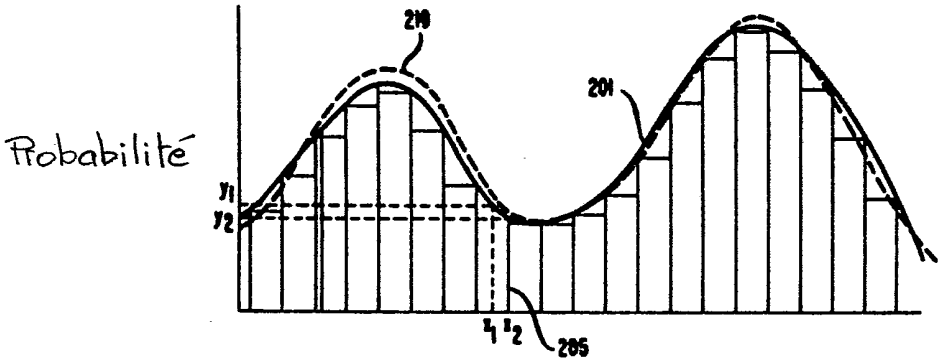
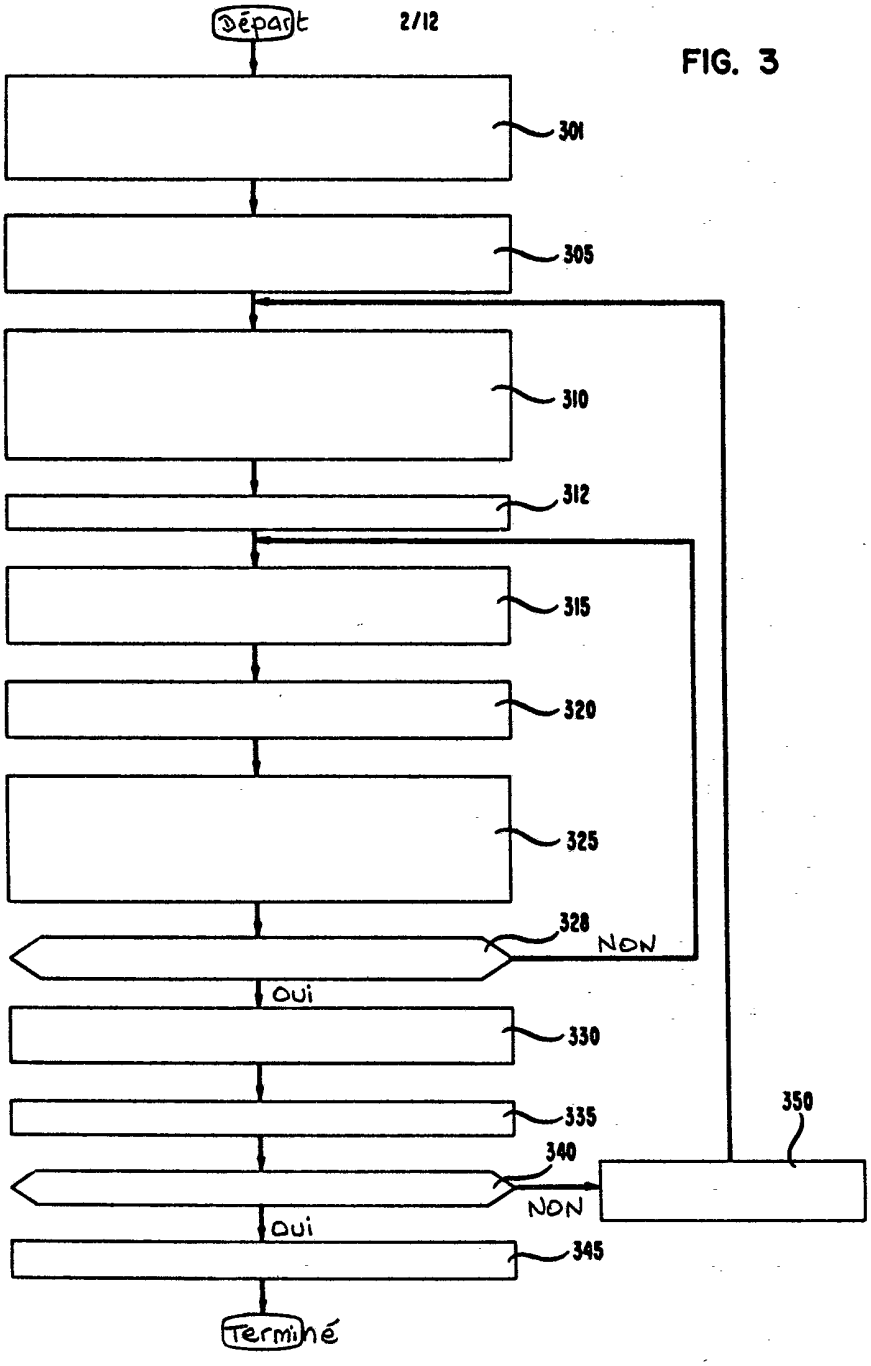

Probabilité
210
201
y1
y2
x1 x2
205
caractéristiques

FIG. 3


Départ
301
305
310
312
315
320
325
328
NON
oui
330
335
340
NON
oui
345
350
Terminé

3/12

FIG. 4

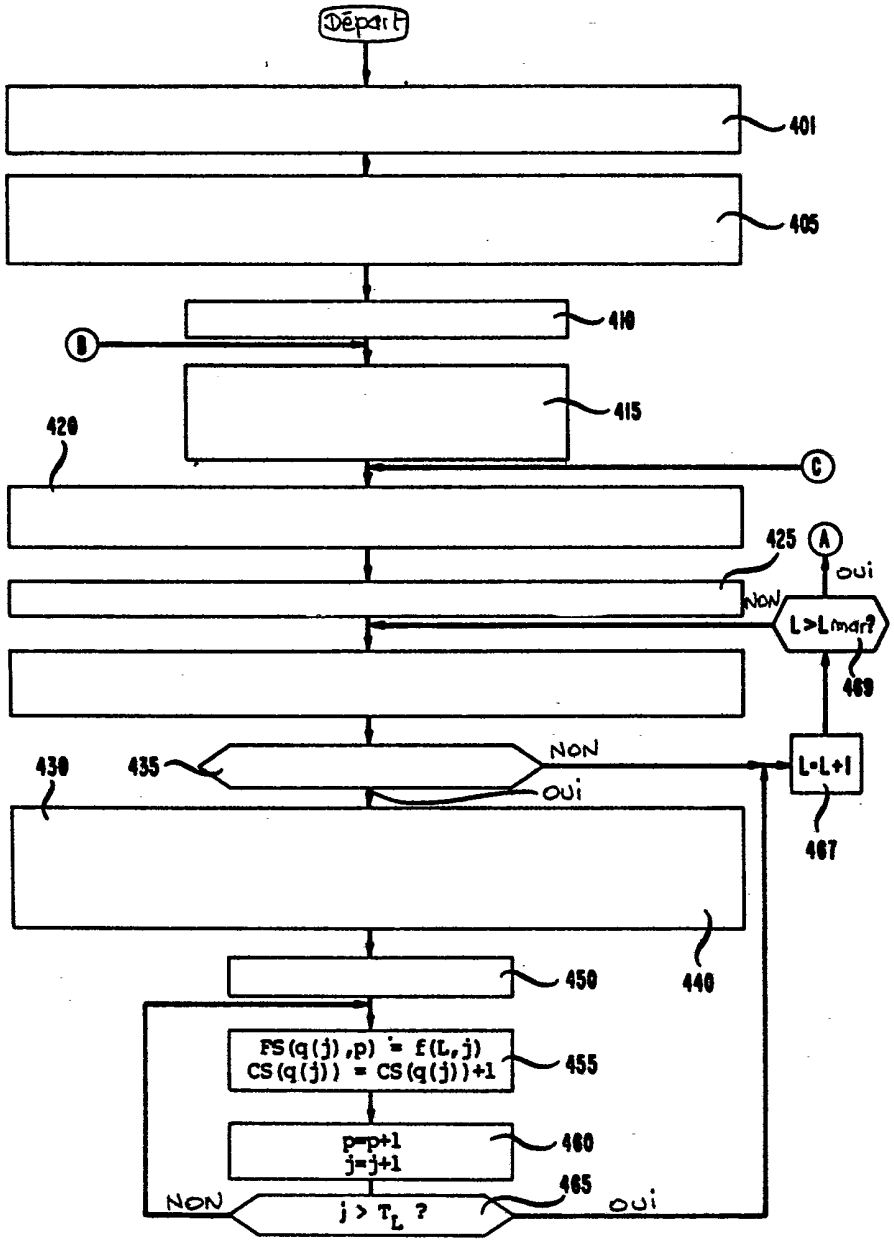

Départ
401
405
410
B
415
420
C
425
A
oui
NON
L>Lmax?
469
NON
435
oui
L=L+1
467
430
440
450
FS(q(j),p) = f(L,j)
CS(q(j)) = CS(q(j))+1
455
p=p+1
j=j+1
460
NON
j > T_L ?
465
oui

FIG. 5

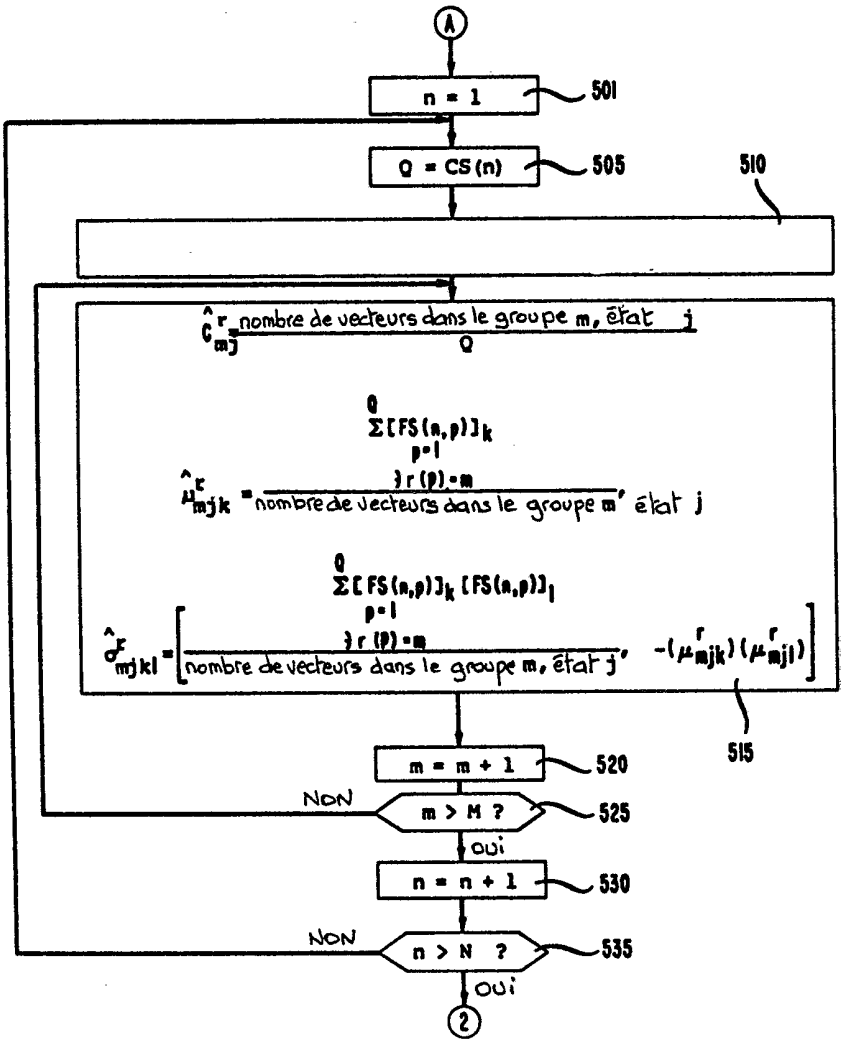

A
n = 1
501
Q = CS(n)
505
510
$\hat{C}^r_{mj} = \frac{\text{nombre de vecteurs dans le groupe m, état j}}{Q}$
$\hat{\mu}^r_{mjk} = \frac{\sum_{p=1}^{Q} [FS(n,p)]_k}{\text{nombre de vecteurs dans le groupe m' état j}}$
$\hat{\sigma}^r_{mjkl} = \left[\frac{\sum_{p=1}^{Q} [FS(n,p)]_k [FS(n,p)]_l}{\text{nombre de vecteurs dans le groupe m, état j'}} - (\mu^r_{mjk})(\mu^r_{mjl})\right]$
515
m = m + 1
520
NON
m > M ?
525
oui
n = n + 1
530
NON
n > N ?
535
oui
2

FIG. 6

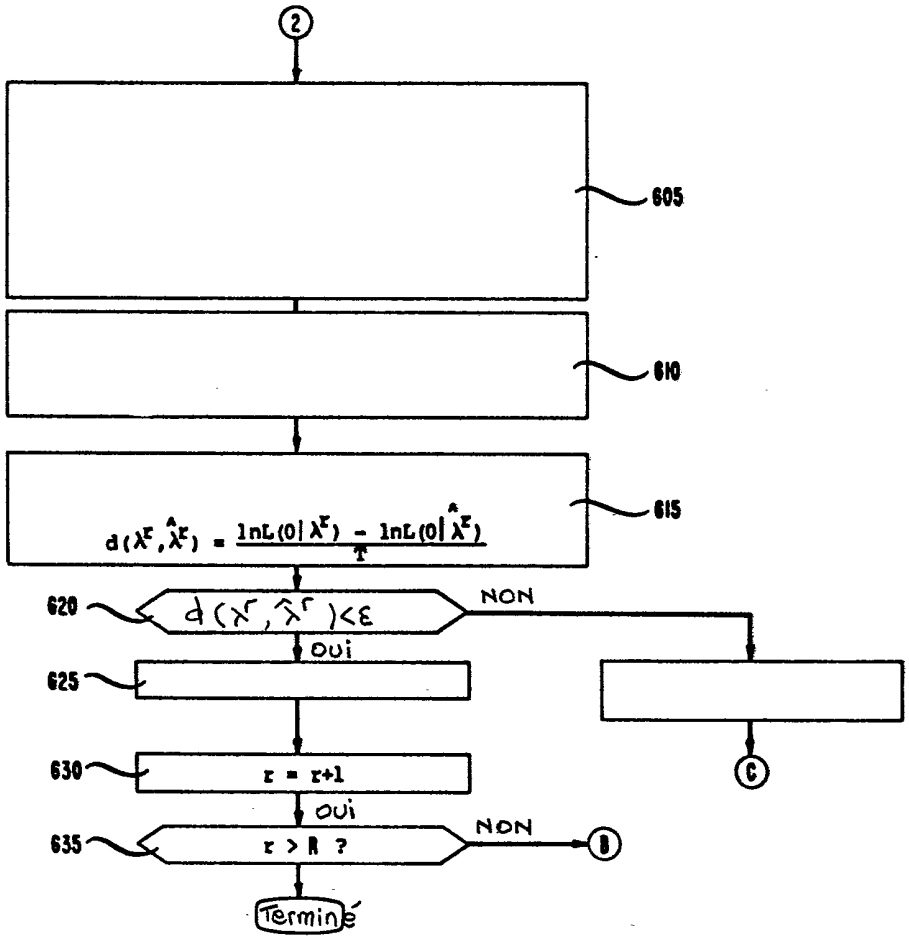

2
605
610
615
$d(\lambda^r, \hat{\lambda}^r) = \frac{\ln L(O|\lambda^r) - \ln L(O|\hat{\lambda}^r)}{T}$
620
$\exists p / d(\lambda^r, \hat{\lambda}^r) < \varepsilon$
NON
oui
625
630
r = r+1
oui
635
r > R ?
NON
B
C
Terminé

FIG. 7

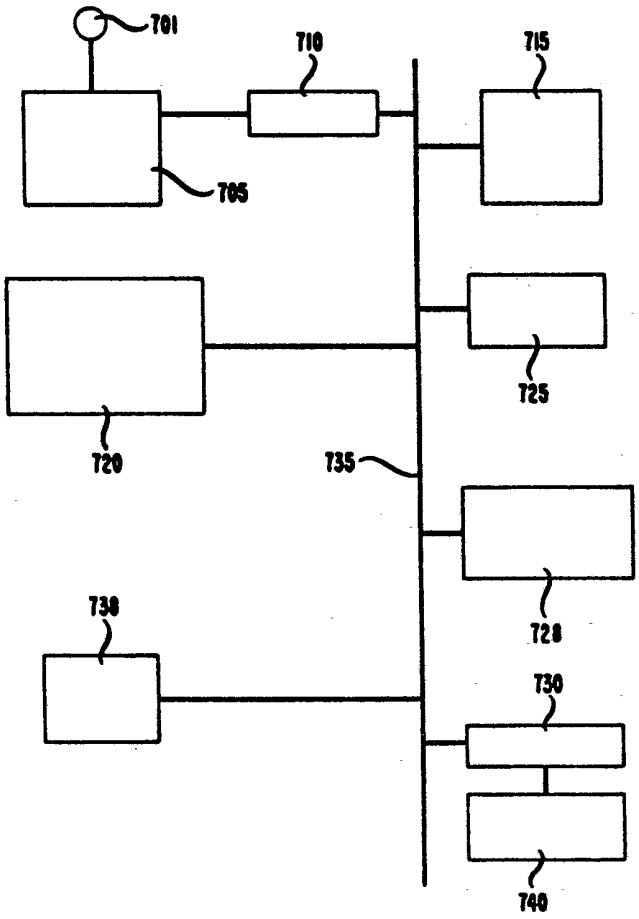


FIG. 8

Séquence d'intervalles de temps de caractéristiques acoustiques	Identification du mot	Nombre d'intervalles de temps
$o_1^1\ o_2^1$ ---------- o_{T11}^1	W_1	T_{11}
o_1^2 ---------- o_{T21}^2	W_2	T_{21}
o_1^3 ---------- o_{T31}^3	W_3	T_{31}
o_1^1 ---------- o_{T12}^1	W_1	T_{12}
o_1^9 ---------- o_{T91}^9	W_9	T_{91}
o_1^3 ---------- o_{T32}^3	W_3	T_{32}

FIG. 10

	Etat				
Intervalle de temps	1	2	3	4	5
1	o_{11}	o_{21}	o_{31}	o_{41}	o_{51}
2	o_{12}	o_{22}	o_{32}	o_{42}	
3	o_{13}		o_{33}		
4	o_{14}				
CS	$o_{1,CS(1)}$	$o_{2,CS(2)}$	$o_{3,CS(3)}$	$o_{4,CS(4)}$	$o_{5,CS(5)}$
compte	P_1	P_2	P_3	P_4	P_5

FIG. 9

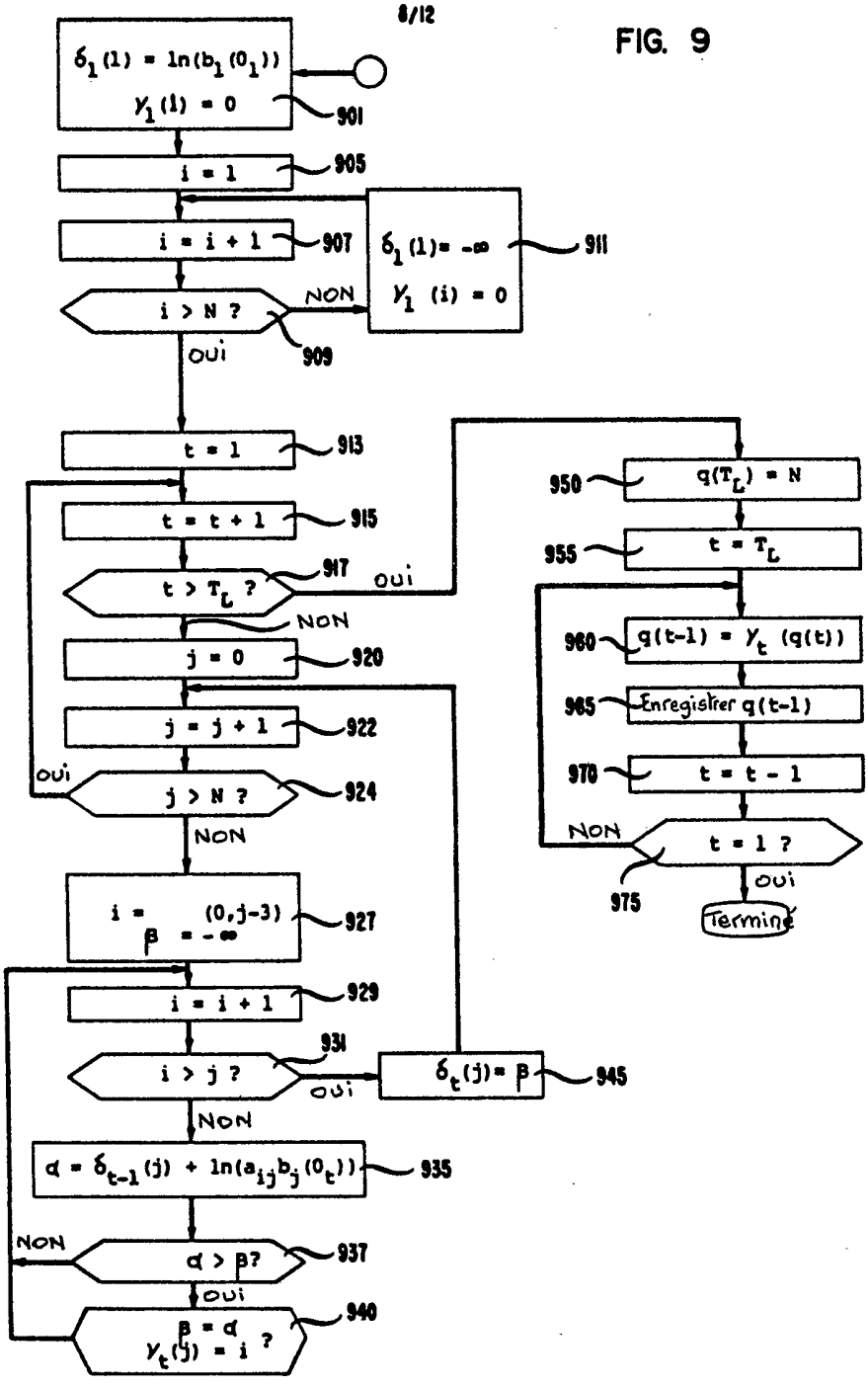
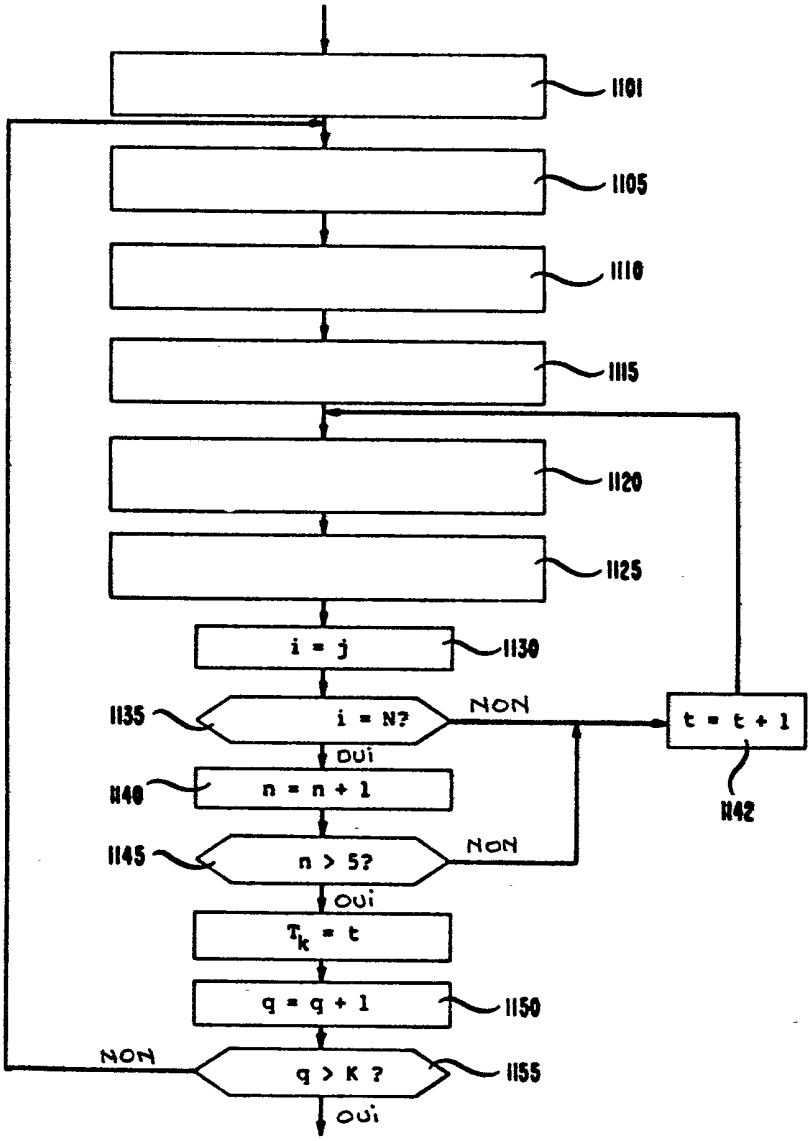

$\delta_1(1) = \ln(b_1(0_1))$
$Y_1(1) = 0$
901
i = 1
905
i = i + 1
907
$\delta_1(1) = -\infty$
$Y_1(i) = 0$
911
i > N ?
NON
909
OUI
t = 1
913
t = t + 1
915
t > T_L ?
917
OUI
NON
j = 0
920
j = j + 1
922
OUI
j > N ?
924
NON
i = (0,j-3)
$\beta = -\infty$
927
i = i + 1
929
i > j ?
931
OUI
$\delta_t(j) = \beta$
945
NON
$\alpha = \delta_{t-1}(j) + \ln(a_{ij}b_j(0_t))$
935
NON
$\alpha > \beta$?
937
OUI
$\beta = \alpha$
$Y_t(j) = i$?
940
950
$q(T_L) = N$
955
$t = T_L$
960
$q(t-1) = Y_t(q(t))$
965
Enregistrer q(t-1)
970
t = t - 1
NON
t = 1 ?
975
OUI
Termine

FIG. 11


1101
1105
1110
1115
1120
1125
i = j
1130
1135
i = N?
NON
OUI
t = t + 1
1142
1140
n = n + 1
1145
n > 5?
NON
OUI
T_k = t
q = q + 1
1150
NON
q > K ?
1155
OUI

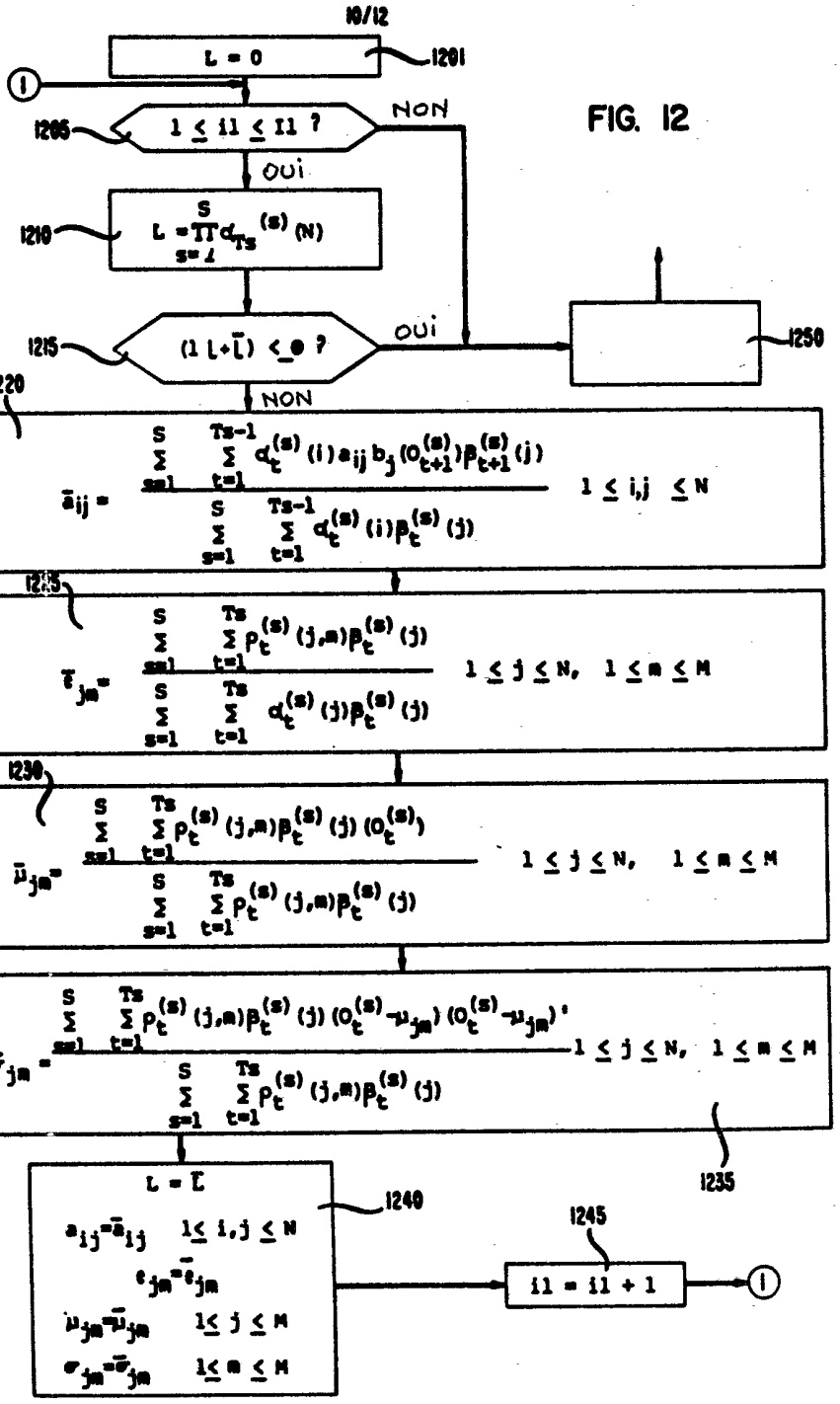

L = 0
1201
1205
1 ≤ i1 ≤ I1 ?
NON
OUI
1210
$L = \prod_{s=1}^{S} \alpha_{T_s}^{(s)}(N)$
1215
$(1 L+\bar{L}) \leq \theta$?
OUI
1250
NON
1220
$\bar{a}_{ij} = \frac{\sum_{s=1}^{S} \sum_{t=1}^{T_s-1} \alpha_t^{(s)}(i) a_{ij} b_j(O_{t+1}^{(s)}) \beta_{t+1}^{(s)}(j)}{\sum_{s=1}^{S} \sum_{t=1}^{T_s-1} \alpha_t^{(s)}(i) \beta_t^{(s)}(j)}$ $1 \leq i,j \leq N$
1225
$\bar{c}_{jm} = \frac{\sum_{s=1}^{S} \sum_{t=1}^{T_s} \rho_t^{(s)}(j,m) \beta_t^{(s)}(j)}{\sum_{s=1}^{S} \sum_{t=1}^{T_s} \alpha_t^{(s)}(j) \beta_t^{(s)}(j)}$ $1 \leq j \leq N, \quad 1 \leq m \leq M$
1230
$\bar{\mu}_{jm} = \frac{\sum_{s=1}^{S} \sum_{t=1}^{T_s} \rho_t^{(s)}(j,m) \beta_t^{(s)}(j) (O_t^{(s)})}{\sum_{s=1}^{S} \sum_{t=1}^{T_s} \rho_t^{(s)}(j,m) \beta_t^{(s)}(j)}$ $1 \leq j \leq N, \quad 1 \leq m \leq M$
$\bar{\sigma}_{jm} = \frac{\sum_{s=1}^{S} \sum_{t=1}^{T_s} \rho_t^{(s)}(j,m) \beta_t^{(s)}(j) (O_t^{(s)}-\mu_{jm}) (O_t^{(s)}-\mu_{jm})'}{\sum_{s=1}^{S} \sum_{t=1}^{T_s} \rho_t^{(s)}(j,m) \beta_t^{(s)}(j)}$ $1 \leq j \leq N, \quad 1 \leq m \leq M$
1235
L = L̄
1240
$a_{ij} = \bar{a}_{ij}$ $1 \leq i,j \leq N$
$c_{jm} = \bar{c}_{jm}$
$\mu_{jm} = \bar{\mu}_{jm}$ $1 \leq j \leq M$
$\sigma_{jm} = \bar{\sigma}_{jm}$ $1 \leq m \leq M$
1245
i1 = i1 + 1
1

FIG. 12

FIG. 13

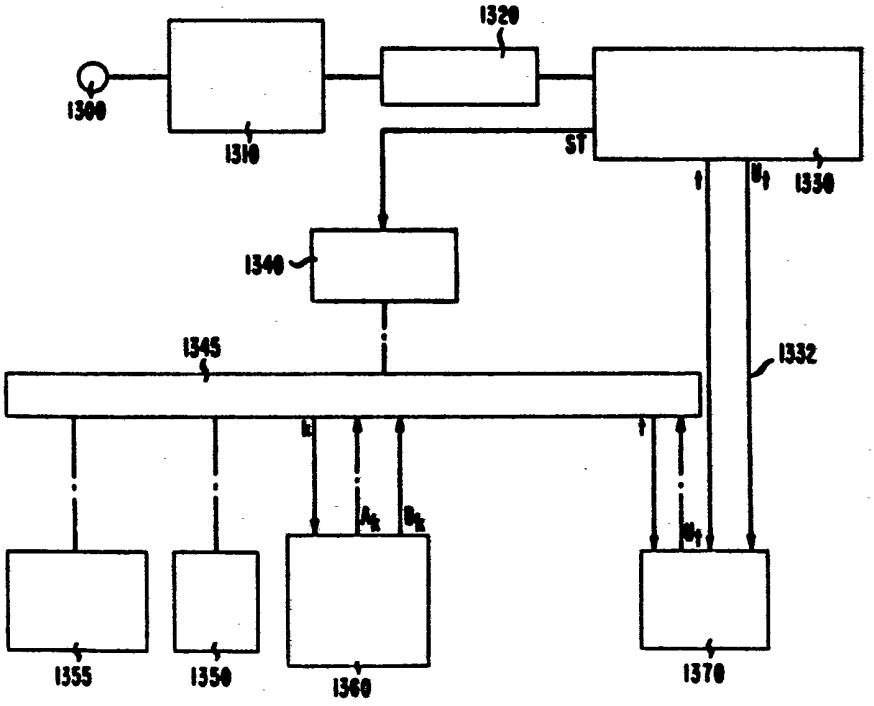
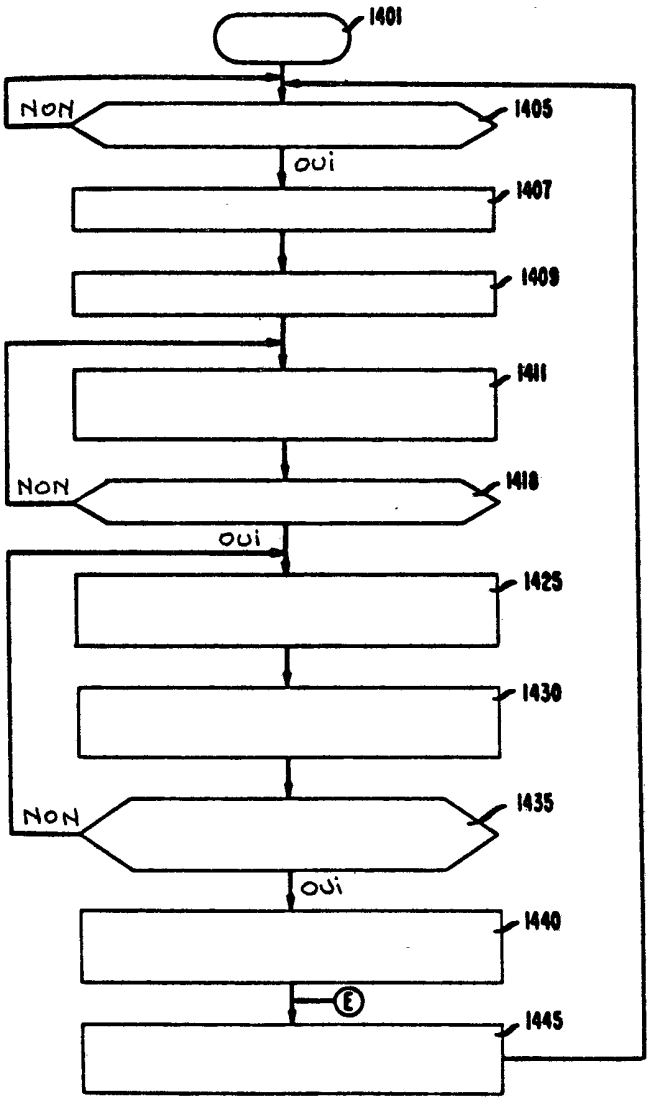


FIG. 14


1401
NON
1405
OUI
1407
1409
1411
NON
1418
OUI
1425
1430
NON
1435
OUI
1440
E
1445