(19) RÉPUBLIQUE FRANÇAISE

INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE

PARIS

(11) N° de publication : **2 642 882**
(à n'utiliser que pour les commandes de reproduction)

(21) N° d'enregistrement national : **89 01542**

(51) Int Cl[5] : G 10 L 5/04.

(12) DEMANDE DE BREVET D'INVENTION A1

(22) Date de dépôt : 7 février 1989.

(30) Priorité :

(43) Date de la mise à disposition du public de la demande : BOPI « Brevets » n° 32 du 10 août 1990.

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

(71) Demandeur(s) : *RIPOLL Jean-Louis.* — FR.

(72) Inventeur(s) : Jean-Louis Ripoll.

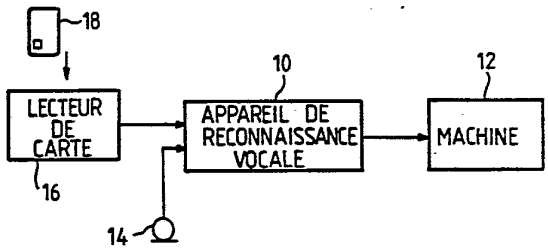
(73) Titulaire(s) :

(74) Mandataire(s) : Cabinet Ballot-Schmit.

(54) Appareil de traitement de la parole.

(57) L'invention concerne l'analyse et la synthèse de la parole, et plus généralement même le codage et le décodage de la parole.

Etant donné que la reconnaissance de parole multilocuteurs est très difficile du fait des différences de prononciation des mêmes phonèmes par des locuteurs différents, l'invention propose un système de reconnaissance utilisant des cartes portatives, et tout particulièrement des cartes à puces, dans lesquelles on enregistre des paramètres caractéristiques de la voix du locuteur titulaire de la carte. Ces paramètres sont lus par un lecteur 16, transmis à un appareil de reconnaissance de parole 10 qui adapte ses algorithmes ou circuits de traitement en fonction du contenu de la carte pour optimiser la reconnaissance en fonction d'un locuteur déterminé. L'appareil de reconnaissance 10 peut alors commander avec une fiabilité maximale une machine 12, en fonction d'un signal de parole transmis par un microphone 14.



FR 2 642 882 - A1

Vente des fascicules à l'IMPRIMERIE NATIONALE, 27, rue de la Convention — 75732 PARIS CEDEX 15

APPAREIL DE TRAITEMENT DE LA PAROLE

L'invention concerne l'analyse et la synthèse de la parole, et plus généralement même le codage et le décodage de la parole.

Les applications dans lesquelles on envisage de traiter électroniquement les signaux de voix humaine sont de plus en plus nombreuses. Il y a d'abord la reconnaissance et la synthèse de parole en vue de faciliter la communication homme-machine qui se fait jusqu'à maintenant principalement à travers un clavier de saisie et un écran de visualisation, ou à travers de boutons et manettes de commande. Il y a aussi la reconnaissance de parole en vue de l'identification d'une personne par ses caractéristiques vocales. Et il y a également des applications dans lesquelles le traitement sert à comprimer les informations émises oralement pour les transmettre à une plus grande vitesse ou avec une plus faible bande passante, etc.

Mais le traitement de la parole est une opération très difficile, à cause de la complexité des mécanismes physiologiques par lesquels la parole est produite et par lesquels elle est entendue et comprise.

Le support de transmission de l'information est une vibration acoustique de l'air. Cette vibration est constituée par une succession d'ondes acoustiques de formes complexes. Lorsqu'on enregistre ces formes d'onde, on constate qu'il est pratiquement impossible, par simple observation visuelle, de faire un lien entre telle ou telle partie du diagramme et le son qui a été prononcé.

Il en résulte qu'il est très difficile d'établir

des circuits électroniques ou programmes de traitement de données qui seraient capables de reconnaître autre chose que des sons isolés très simples. Les problèmes sont également difficiles en synthèse vocale si on veut reproduire des sons qui ressemblent suffisamment fidèlement au langage humain.

Pour donner une idée plus précise des difficultés rencontrées, on va rappeler ci-dessous quelques notions relatives à l'analyse, la reconnaissance et la synthèse de la parole.

Les sons du langage peuvent être émis de plusieurs manières : il y a d'abord une distinction entre les sons voisés et les sons non voisés. Les sons voisés sont émis à partir d'une vibration des cordes vocales et sont modulés à travers le pharynx et la cavité buccale (et notamment par la langue et les lèvres); certains sons utilisent également la cavité nasale. Les sons non voisés ne sont pas émis à partir des cordes vocales; ils sont directement produits à l'intérieur de la cavité buccale.

D'autre part, que ce soit parmi les sons voisés ou les sons non voisés, on peut faire la distinction entre les sons produits par des turbulences d'air (dans une ouverture étroite), et ceux qui correspondent plutôt à un écoulement régulier. Les consonnes sont en général produites par des turbulences. Les voyelles correspondent plutôt à des écoulements réguliers.

Les consonnes fricatives (s, f, z, v) sont produites respectivement par un flux d'air dans l'intervalle étroit entre les dents (s, z) ou entre les lèvres (f, v). Les consonnes s et f ne sont pas voisées. Mais les consonnes z et v sont voisées.

Les consonnes plosives font intervenir une occlusion complète du conduit vocal en un point ou un autre, suivie d'une libération brusque de la pression

accumulée dans le conduit. Le point de fermeture détermine le son produit. Ce son peut être, là encore, voisé ou non voisé. Les consonnes p (non voisée) et b (voisée) correspondent à une fermeture des lèvres; t (non voisée) et d (voisée) correspondent à une occlusion par la langue dans la partie antérieure du palais. Les consonnes k (non voisée) et g (voisée) correspondent à une occlusion par la langue vers l'arrière du palais.

On peut ainsi décrire comment sont produits la plupart des phonèmes correspondant à une langue donnée. Le phonème est le plus petit élément sonore permettant de distinguer un mot d'un autre ou plus précisément de modifier sa signification. Il n'y a guère que quelques dizaines de phonèmes différents dans une langue donnée. On considère qu'il y en a une quarantaine dans la langue française.

Mais c'est un chiffre théorique. Dans la pratique on s'aperçoit que les phonèmes sont prononcés différemment selon les phonèmes qui les précèdent ou les suivent. C'est le phénomène de coarticulation entre phonèmes, qui complique sérieusement les problèmes de reconnaissance ou synthèse car il multiplie par 4 ou 5 le nombre de phonèmes pratiquement émis. Il est d'ailleurs souvent plus simple de fonder la reconnaissance de parole ou la synthèse non pas sur les phonèmes mais soit sur des "diphonémes" qui sont des couples de phonèmes associés incluant la transition entre ces phonèmes, soit sur des "diphones" qui sont des segments sonores débutant au milieu d'un phonème et s'arrêtant au milieu du phonème suivant (incluant donc la transition entre deux phonèmes mais pas la totalité de chacun des deux phonèmes).

L'oreille humaine les distingue très bien les uns des autres, mais les formes d'onde acoustique qui les distinguent ne semblent pas être suffisamment

caractéristiques pour qu'une machine puisse facilement les reconnaître, surtout dans une parole en continu.

Les ondes acoustiques correspondant aux voyelles ont un spectre de fréquences plus simple et plus étroit que les consonnes. Les voyelles représentent en effet plutôt une partie stable du signal vocal, tandis que les consonnes représentent plutôt des transitions. Les plosives par exemple représentent des transitions brutales, avec un spectre de fréquences très large durant la transition.

C'est pourquoi on a essayé de proposer des méthodes de traitement de la parole fondées essentiellement sur l'analyse fréquentielle des signaux acoustiques.

Par ces analyses fréquentielles on arrive mieux à discerner des paramètres correspondant aux différents phonèmes ou diphones émis.

A titre d'exemple, une méthode d'analyse fréquentielle qui a déjà prouvé son efficacité aussi bien en reconnaissance vocale qu'en synthèse vocale est la méthode des formants. On va rappeler en quelques paragraphes ce que sont les formants, pour mieux faire comprendre l'invention, bien que l'invention ne soit pas limitée aux systèmes utilisant une analyse ou une synthèse à formants.

Les formants sont les fréquences correspondant à des pics d'énergie du signal vocal : on voit clairement que le spectre de fréquences résultant de l'analyse du signal acoustique correspondant à une voyelle est un spectre comprenant des creux et des bosses. Les bosses sont les formants; et on distingue en général plusieurs formants successifs dans le spectre correspondant à un phonème déterminé.

Les formants sont repérés par leur position dans le spectre de fréquences. On parlera de premier formant pour le pic de plus basse fréquence, de deuxième formant

pour le pic suivant, etc.

Ces pics correspondent physiquement à des résonances de la cavité buccale, et la parole humaine consiste justement à moduler la forme de la cavité buccale de manière à modifier les différentes fréquences de résonance de cette cavité.

Il y a un lien direct entre la prononciation d'un phonème et la forme du conduit vocal : l'émission du phonème est en effet liée à des positions bien précises des différents éléments mobiles de la cavité buccale (position des lèvres, de la langue, du voile du palais, etc.); et il y a un lien entre les fréquences de formant et la forme du conduit vocal; on comprend donc qu'il y a aussi un lien direct entre un phonème émis et les fréquences de formant détectées dans le spectre de fréquences du signal acoustique correspondant à ce phonème.

L'analyse et la synthèse à formants sont fondés sur cette notion. Effectivement, on constate que la présence de certains formants est tout-à-fait caractéristique de l'émission de tel ou tel phonème. Pour les voyelles, dont le spectre de fréquences est relativement stable, on peut très bien caractériser une voyelle déterminée par la position (sur l'axe des fréquences) des trois premiers formants, c'est-à-dire des trois premiers pics du spectre du signal acoustique correspondant.

A titre indicatif, on peut donner l'exemple suivant: la voyelle A est un signal acoustique dont le premier formant est situé entre 500 et 800 hertz, le deuxième est situé entre 1000 et 1600 hertz mais n'est pas écarté du premier de plus de 600 à 900 hertz, et le troisième formant est situé entre 2300 et 3200 hertz.

Un autre exemple : la voyelle I aurait un premier formant entre 200 et 400 hertz, un deuxième formant situé entre 2100 et 2400 hertz, mais espacé d'au moins

2000 hertz du premier. Le troisième formant est à une fréquence plus élevée encore.

Avec un vecteur mathématique composé de trois nombres qui sont les fréquences des trois premiers formants on peut assez bien caractériser toutes les voyelles et certaines consonnes. Pour d'autres consonnes l'utilisation des formants est plus malaisée, mais d'autres méthodes peuvent être utilisées, et notamment une évaluation du sens et de la rapidité de variation des fréquences de formant dans les diphones comportant une transition par consonne.

Cependant, un problème supplémentaire vient de la diversité des prononciations des mêmes phonèmes par des personnes différentes. L'oreille humaine rétablit automatiquement la signification du phonème, même prononcé par plusieurs personnes différentes. Mais une machine de reconnaissance vocale confrontée à plusieurs vecteurs de formants aura beaucoup de mal à reconnaître ces différents vecteurs comme représentant un seul et même phonème si les vecteurs sont assez différents les uns des autres du fait qu'ils émanent de personnes différentes. C'est d'ailleurs d'autant plus vrai qu'on a déjà envisagé de réaliser des machines d'identification de personnes dont le fonctionnement repose sur la reconnaissance vocale, ce qui montre que dans une certaine mesure il peut y avoir des différences très significatives dans l'émission des mêmes phonèmes par des personnes différentes.

A titre d'exemple, la figure 1 représente un tableau schématique des zones de prononciation de différentes voyelles phonétiques. Les lettres entre crochets représentent des phonèmes usuels en français, selon le code de phonétique de l'Association Internationale de Phonétique. Le tableau est un diagramme fréquentiel représentant les zones de valeur

du premier formant (en ordonnée) et du deuxième formant (en abscisse). On voit notamment que certaines zones se recoupent, ce qui veut dire que le même son émis par deux personnes différentes peut correspondre à deux phonèmes de signification différentes. Et plus généralement, les zones sont assez proches les unes des autres de sorte qu'il peut être difficile à une machine de reconnaître les phonèmes présents dans la parole humaine.

Les machines de reconnaissance vocale proposées jusqu'à maintenant sont habituellement capables de reconnaître seulement un petit nombre de mots isolés, prononcés par un locuteur bien déterminé qui a enregistré dans la machine les mots à reconnaître (qu'il a prononcé lui-même).

On a proposé de rendre ces machines capables de reconnaître les mêmes mots, prononcés par plusieurs locuteurs différents. Mais alors, le passage d'un locuteur à un autre nécessite d'abord une phase d'apprentissage de la machine : le deuxième locuteur doit prononcer devant la machine la succession des différents mots qu'elle doit pouvoir reconnaître, de manière que la machine enregistre en mémoire la manière dont ces mots sont prononcés, et qu'elle puisse ensuite les reconnaître. Cette phase d'apprentissage est très lourde; d'autant plus lourde que la machine doit pouvoir reconnaître plus de mots. Si elle doit reconnaître 1000 mots, il faudra les prononcer tous; il faudra même peut-être les prononcer chacun plusieurs fois pour établir une prononciation moyenne (car la prononciation d'un mot par une personne n'est pas quelque chose de figé et invariable). Pendant la phase d'apprentissage, la machine sera indisponible pour exécuter sa fonction de reconnaissance; l'opérateur sera aussi contraint de réserver un temps pour cette opération. Mais cette

opération est a priori indispensable car la probabilité est très faible pour que la machine reconnaisse d'une manière fiable les mots prononcés par un locuteur autre que celui qui a enregistré les mots de référence.

Il est inutile de préciser que si la machine est destinée par exemple à une utilisation par le public dans un lieu public, il est hors de question de procéder à une phase d'apprentissage pour chaque utilisateur qui se présente devant la machine. On peut penser par exemple à une cabine téléphonique dans laquelle la composition du numéro appelé est faite oralement. Pour de telles machines, on est actuellement obligé de limiter au maximum le nombre de mots à reconnaître, pour augmenter la certitude de reconnaître le mot prononcé quelle que soit la personne qui le prononce.

La présente invention a entre autres pour but de proposer un moyen simple permettant de rendre plus facile l'utilisation d'une machine de reconnaissance par plusieurs locuteurs différents, sans réduire excessivement les possibilités de la machine.

Un autre but de l'invention est de proposer un moyen simple permettant d'améliorer la synthèse vocale en adaptant aussi étroitement que possible la voix synthétisée à la voix d'un locuteur bien déterminé, de sorte que par exemple si la voix d'un locuteur est codée, puis transmise sur une ligne téléphonique, puis resynthétisée avant d'être restituée à un auditeur, la voix synthétisée puisse se rapprocher aussi près que possible de la voix du locuteur initial.

Pour atteindre ces buts, la présente invention propose un système de traitement de parole comprenant un appareil de codage ou décodage de parole adapté à un codage ou un décodage multilocuteurs, caractérisé en ce que des paramètres spécifiques d'un locuteur déterminé sont contenus dans une carte portative personnelle que

le locuteur conserve avec soi, le système comportant un lecteur de carte adapté à lire le contenu de la carte et à communiquer ce contenu à l'appareil de codage ou décodage, pour l'adapter instantanément, sans phase d'apprentissage, à ce locuteur.

On comprend qu'avec ce système, on peut aller jusqu'à installer dans des lieux publics des machines complexes utilisant la reconnaissance ou la synthèse de parole, et que toute personne possédant une carte personnelle contenant les paramètres propres de sa voix, pourra communiquer avec cette machine ou à travers cette machine, alors qu'elle ne pourrait le faire autrement.

La carte pourrait contenir sous forme de données codées une prononciation d'un certain nombre de mots par le titulaire de la carte (autant de mots que la machine doit pouvoir reconnaître ou synthétiser par exemple). Mais il est plus avantageux que la carte contienne plutôt des paramètres de la voix indépendamment des mots à reconnaître ou synthétiser, car cela élargit les possibilités de reconnaissance ou synthèse.

Les paramètres enregistrés dans la carte peuvent alors être des signaux électriques codés représentant les formes d'onde temporelle ou les spectres de fréquence de phonèmes ou diphonèmes ou diphones prononcés par le titulaire de la carte. Mais on préférera utiliser comme paramètres des vecteurs correspondant à ces phonèmes ou diphonèmes ou diphones, par exemple des vecteurs de trois ou quatre formants; chaque vecteur de trois ou quatre formants comprendra donc trois ou quatre valeurs de fréquences (ou plus vraisemblablement trois ou quatre gammes de fréquences) représentant un phonème ou diphonème ou diphone déterminé. Ces vecteurs seront stockés dans la carte, et transférés à la machine au moment de l'utilisation, en remplacement des vecteurs que la machine aura pu

recevoir précédemment lors de l'utilisation par un autre locuteur disposant d'une autre carte personnelle.

On comprendra que si les formants semblent être les vecteurs les plus commodes pour représenter les voyelles, d'autres paramètres existent et peuvent être stockés pour d'autres phonèmes, diphonèmes ou diphones. Notamment, les consonnes ou les diphones incluant des consonnes s'exprimeront plus facilement par des paramètres relatifs à la manière dont les formants varient: chute plus ou moins rapide du premier formant et simultanément montée plus ou moins rapide du deuxième, etc.

Des coefficients de fonctions de transfert échantillonnées (fonction de transfert en z) pourraient également être stockés comme paramètres de la voix dans une carte personnelle portative.

La carte pourrait être une carte à piste magnétique, ou optique; mais elle sera de préférence une carte à puce incorporant une puce de circuit-intégré avec notamment une mémoire non volatile contenant les paramètres personnels de la voix. La carte peut être aussi un autre support d'information portable tel que par exemple : cartes magnétiques à haute densité de stockage, dont la surface magnétique couvre la totalité ou la quasi-totalité d'une des faces; mémoire de stockage de type EPROM ou EEPROM ou RAM non-volatile stockée dans un boîtier de forme très compacte et facilement transportable; clés à puce n'ayant pas spécialement la forme d'une carte plate, etc.

D'autres caractéristiques et avantages de l'invention apparaîtront à la lecture de la description qui suit et qui est faite en référence aux dessins annexés dans lesquels :

- la figure 1, déjà décrite, représente un diagramme de position de divers phonèmes dans l'espace des formants (deux premiers formants);

- la figure 2 représente schématiquement une application de l'invention à la commande vocale d'une machine;

- la figure 3 représente schématiquement une application de l'invention aux communications téléphoniques.

Une première application de l'invention est la reconnaissance de la parole, telle qu'on peut l'utiliser par exemple pour la commande d'un robot, d'une machine industrielle, d'un véhicule, etc., ou, dans une application plus sophistiquée, pour une machine à dicter ou une machine à traduire.

La figure 2 schématise cette application dans le cas de la commande d'un robot. Un appareil de reconnaissance 10 est connecté à un robot industriel 12 pour lui fournir des ordres de commande de marche, d'arrêt, de rotation, etc. L'appareil de reconnaissance est couplé à un microphone 14 de sorte que les ordres de commande peuvent être donnés oralement sous la forme de mots simples tels que "marche", "stop", "droite", "gauche", etc. L'appareil est par ailleurs couplé à un lecteur de carte à puces 16 dans lequel on peut introduire une carte à puce 18 qui contient dans une mémoire non volatile (mémoire EPROM ou EEPROM) des données personnalisées relatives à la voix d'un locuteur titulaire de cette carte.

Lors du fonctionnement, les données de la carte sont d'abord chargées dans l'appareil de reconnaissance; ces données servent à modifier soit des configurations de circuits électroniques dans l'appareil, soit des algorithmes de reconnaissance utilisés dans l'appareil.

Les configurations modifiées ou les algorithmes modifiés sont tels que l'appareil soit alors adapté de manière optimale à la reconnaissance des mots ou phrases prononcés par le locuteur titulaire de la carte.

Par exemple, les modifications d'algorithme peuvent consister en modifications des valeurs moyennes et valeurs limites des fréquences de formants pour chaque phonème ou diphonème ou diphone susceptible d'être prononcé; ou encore des modifications de coefficients de polynômes dans des algorithmes de calcul fondés sur la transformée en z des signaux acoustiques échantillonnés. Des modifications de configurations de circuits électroniques pourraient par exemple consister en modifications de valeurs de capacités (par commutation d'interrupteurs) dans des filtres à capacités commutées utilisés pour déterminer des fréquences de formants.

Selon la sophistication de l'appareil de reconnaissance 10, on pourra reconnaître des mots ou phrases plus ou moins complexes. Si l'appareil 10 est très performant (et ses performances vis-à-vis de locuteurs multiples seront considérablement améliorées par l'invention), on peut envisager que la machine 12 commandée soit une machine de traitement de texte, voire même une machine de traduction automatique. Cela suppose bien entendu que l'appareil de reconnaissance soit capable de reconnaître non pas seulement des mots isolés mais des phrases continues.

Pour le choix des paramètres que l'on peut inscrire dans la carte pour représenter de manière personnalisée la voix du titulaire de la carte, on pourra utiliser d'une manière générale les théories de reconnaissance et synthèse de la voix telles qu'elles ont été formulées jusqu'à maintenant. On trouvera une indication des méthodes mathématiques permettant de faire ces choix dans le traité de René Boite et Murat Kunt : "Traitement

de la parole", complément au Traité d'Electricité, publié aux Presses Polytechniques Romandes, ainsi que les ouvrages référencés dans la bibliographie de ce traité.

Une autre application de l'invention est représentée à la figure 3. Dans cette application, on cherche à coder le signal de parole émis sur une ligne téléphonique, pour comprimer le signal et ainsi limiter le débit d'informations utile pour une communication. Pour cela, on code le signal reçu par le microphone du combiné téléphonique; le codage est un codage phonétique au lieu d'être un codage numérique des formes d'onde du signal de parole : on code la parole en la décomposant en phonèmes ou diphones successifs; c'est donc une opération de reconnaissance de parole. Puis on envoie sur la ligne téléphonique des vecteurs successifs de données, chaque vecteur comportant plusieurs données relatives au phonème qui vient d'être prononcé dans le combiné. A la réception, on reconvertit les vecteurs de données en phonèmes; c'est une opération de synthèse de parole. La compression réalisée peut être très importante : on peut envisager de limiter à 2 kilobits par seconde la quantité de données nécessaire pour transmettre une conversation normale. En effet, le nombre de phonèmes émis ne dépasse pas une dizaine par seconde. On dispose donc de 200 bits pour coder chaque phonème ou diphone ainsi que la prosodie (c'est-à-dire la mélodie engendrée par la variation de la fréquence fondamentale des cordes vocales au cours de la phrase).

Dans cette application, on utilisera selon l'invention un premier codeur/décodeur 20 interposé entre un premier appareil téléphonique 22 et une ligne téléphonique numérique 24. Ce premier codeur a pour fonction de coder la parole émise et de décoder la parole reçue. Il est couplé à un premier lecteur de

cartes à puces 26 dans lequel on pourra introduire une carte 28 comportant les données personnalisées sur la voix de la personne qui téléphone. On utilisera aussi un deuxième codeur/décodeur 30 semblable au premier, raccordé à l'autre bout de la ligne 24, interposé entre la ligne et un deuxième appareil téléphonique 32. Le deuxième codeur/décodeur est aussi couplé à un deuxième lecteur de cartes 36 dans lequel on peut insérer une carte 38 comportant les données personnalisées relatives à la voix du correspondant à l'autre bout de la ligne.

Les codeur/décodeurs, qui sont en fait des appareils complets de reconnaissance et synthèse vocale, reçoivent les données contenues dans les deux cartes, de sorte que la partie codage est adaptée à la reconnaissance de la voix de la personne située au même bout de la ligne que le codeur/décodeur, alors que la partie décodage est adaptée à la synthèse de la voix de la personne située à l'autre bout de la ligne.

On prévoit donc en début de conversation téléphonique un protocole d'échanges de données pour envoyer dans les codeurs/décodeurs les données qui conviennent. Puis la conversation peut avoir lieu : l'une des personnes parle; sa voix est convertie en phonèmes codés, par le codeur qui a été spécialement adapté à la voix du locuteur; elle est envoyée sur la ligne; elle est reçue par le décodeur à l'autre bout de la ligne. Le décodeur a été lui aussi adapté à la voix du même locuteur; il synthétisera donc d'une manière optimale la voix de ce locuteur avant de la transmettre à l'écouteur du poste téléphonique. De même pour l'autre locuteur, codage et décodage sont spécialement adaptés à sa voix de sorte qu'à l'autre bout de la ligne le correspondant recevra une voix synthétisée d'une manière personnalisée.

Dans une autre application encore, on cherche à

interroger par téléphone une base de données. L'interrogation est faite par la parole et non par l'intermédiaire d'un clavier. Un exemple est la réservation téléphonique de transports aériens. L'utilisateur dispose, comme dans l'application précédente, d'un appareil téléphonique auquel est associé un lecteur de carte; la carte contient les paramètres de la voix de son titulaire. Les paramètres peuvent être utilisés de deux manières : d'une part ils peuvent être envoyés sur la ligne à titre d'éléments d'identification d'un titulaire autorisé; si les paramètres ne sont pas ceux d'un titulaire autorisé, la base de données n'est pas rendue accessible; d'autre part, après que les paramètres de la voix aient été transmis vers la base de données, un système de reconnaissance de parole utilise ces paramètres pour s'adapter au mieux à la voix de celui qui va parler sur la ligne téléphonique. L'utilisateur peut alors parler; sa voix est transmise normalement sur la ligne (contrairement à l'application précédente où elle est codée en vue d'une réduction du débit); une analyse de parole est faite à l'autre bout de la ligne, adaptée à la voix du locuteur, pour déterminer par machine le message transmis et instaurer le dialogue homme-machine via la ligne téléphonique.

Dans toutes les applications, on prévoira de préférence que les paramètres personnels de la voix, sont inscrits dans la carte d'un titulaire par une machine spécialisée dont la fonction principale est de déterminer et enregistrer ces paramètres. Le titulaire de la carte devra à cet effet prononcer devant la machine un certain nombre de mots caractéristiques qui serviront à faire cette détermination.

REVENDICATIONS

1. Système de traitement de la parole, comprenant un appareil de codage ou décodage de parole adapté à un codage ou un décodage multilocuteurs, caractérisé en ce que des paramètres spécifiques de la voix d'un locuteur déterminé sont contenus dans une carte portative personnelle que le locuteur conserve avec soi, le système comportant un lecteur de carte adapté à lire le contenu de la carte et à communiquer ce contenu à l'appareil de codage ou décodage pour l'adapter instantanément, sans phase d'apprentissage, à ce locuteur.

2. Système de traitement de parole selon la revendication 1, caractérisé en ce que les paramètres spécifiques du locuteur comprennent des vecteurs de données acoustiques correspondant à des phonèmes ou diphonèmes ou diphones, tels qu'ils sont prononcés par le locuteur titulaire de la carte.

3. Système de traitement de parole selon la revendication 2, caractérisé en ce que chaque vecteur est constitué par un ensemble de données acoustiques, parmi lesquelles on trouve des valeurs de fréquence de formants correspondant à un phonème ou diphonème ou diphone tel que prononcé par le locuteur titulaire de la carte.

4. Système de traitement de parole selon l'une des revendications 1 à 3, caractérisé en ce que les paramètres spécifiques contenus dans la carte comprennent des données relatives aux variations de fréquence de formants correspondant à des phonèmes ou diphonèmes ou diphones déterminés.

5. Système de traitement de parole selon l'une des revendications 1 à 4, caractérisé en ce que les

paramètres contenus dans la carte comprennent des coefficients de fonctions de transfert échantillonnées (fonction de transfert en z) de signaux acoustiques correspondant à des phonèmes ou diphonèmes ou diphones prononcés par le titulaire de la carte.

6. Système de traitement de parole selon l'une des revendications 1 à 5, caractérisé en ce que la carte est une carte à piste magnétique, ou optique, ou de préférence une carte à puce incorporant une puce de circuit-intégré avec notamment une mémoire non volatile contenant les paramètres personnels de la voix.

7. Système de traitement de parole selon l'une des revendications 1 à 5, caractérisé en ce que la carte est une carte magnétique à haute densité de stockage dont la surface magnétique couvre la totalité ou la quasi totalité d'une face, ou une clé à circuit intégré n'ayant pas spécifiquement une forme de carte plate.

8. Système de traitement de parole selon l'une des revendications 1 à 7, caractérisé en ce qu'il comprend un appareil de codage et décodage phonétique de parole interposé entre un appareil téléphonique et une ligne téléphonique, et capable de transmettre successivement sur la ligne des vecteurs de données correspondant à une succession de phonèmes ou diphonèmes ou diphones, et un lecteur de carte, l'appareil de codage et décodage étant apte à adapter sa fonction de codage en fonction de paramètres personnels de voix contenus dans une carte introduite dans le lecteur, et l'appareil étant apte par ailleurs à adapter sa fonction de décodage en fonction de paramètres personnels de voix reçus de la ligne téléphoniques.

9. Système de traitement de parole selon l'une des revendications 1 à 7, caractérisé en ce qu'il comporte un appareil téléphonique couplé à une ligne téléphonique, et un lecteur de carte associé à

l'appareil, des moyens pour transmettre sur la ligne les paramètres de la voix contenue dans la carte, et un système de reconnaissance de parole à l'autre bout de la ligne pour dans un premier temps recevoir de la ligne les dits paramètres et dans un deuxième temps recevoir un signal de parole en provenance de l'appareil téléphonique, le système de reconnaissance de parole étant apte à adapter son fonctionnement en fonction des paramètres de voix reçus.

FIG_1

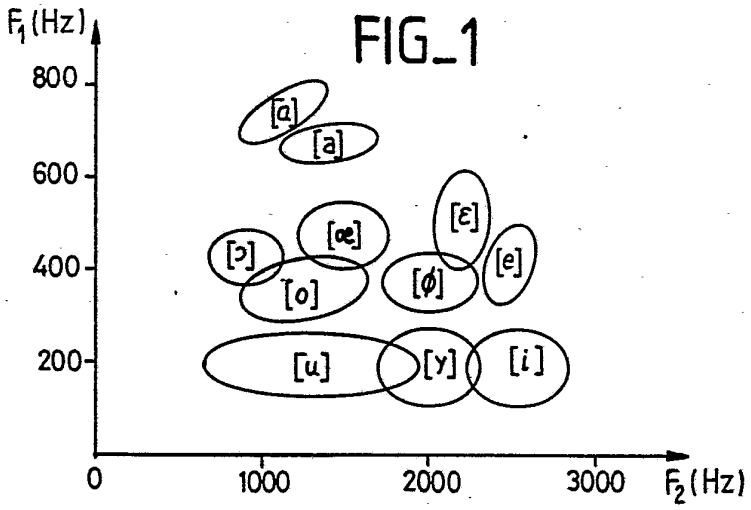


FIG_2

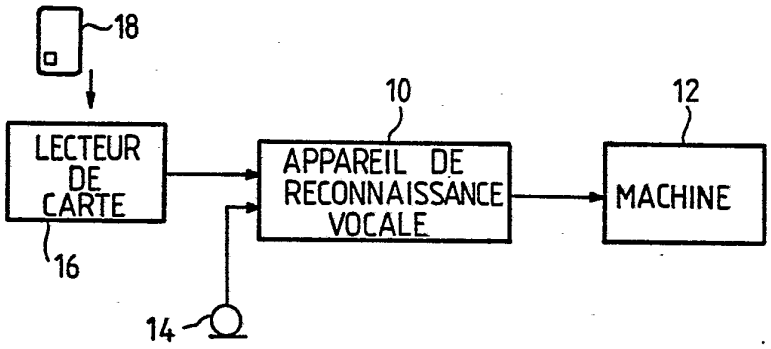


FIG_3