



DEMANDE INTERNATIONALE PUBLIEE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

<p>(51) Classification internationale des brevets <sup>6</sup> : <b>G10L 5/06</b></p>	<p><b>A1</b></p>	<p>(11) Numéro de publication internationale: <b>WO 97/42625</b> (43) Date de publication internationale: 13 novembre 1997 (13.11.97)</p>
<p>(21) Numéro de la demande internationale: PCT/FR97/00753 (22) Date de dépôt international: 25 avril 1997 (25.04.97) (30) Données relatives à la priorité: 96/05566 3 mai 1996 (03.05.96) FR (71) Déposant (pour tous les Etats désignés sauf US): UNIVERSITE PIERRE ET MARIE CURIE [FR/FR]; Tour Centrale, 4, place Jussieu, F-75005 Paris (FR). (72) Inventeurs; et (75) Inventeurs/Déposants (US seulement): MONTACIE, Claude [FR/FR]; 10, rue de Vintimille, F-75009 Paris (FR). CARATY, Marie-José [FR/FR]; 7, rue Emile Goeury, F-94140 Alfortville (FR). LE FLOCH, Jean-Luc [FR/FR]; 8, rue Xavier Bichat, F-95150 Taverny (FR). (74) Mandataires: KEIB, Gérard etc.; Novamark Technologies, Anciennement Brevets Rodhain &amp; Porte, Immeuble Victoria Michelet, 122, rue Edouard Vaillant, F-92593 Levallois Perret Cedex (FR).</p>		<p>(81) Etats désignés: CA, JP, KR, US, brevet européen (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  Publiée Avec rapport de recherche internationale.</p>

(54) Title: METHOD FOR THE VOICE RECOGNITION OF A SPEAKER USING A PREDICTIVE MODEL, PARTICULARLY FOR ACCESS CONTROL APPLICATIONS

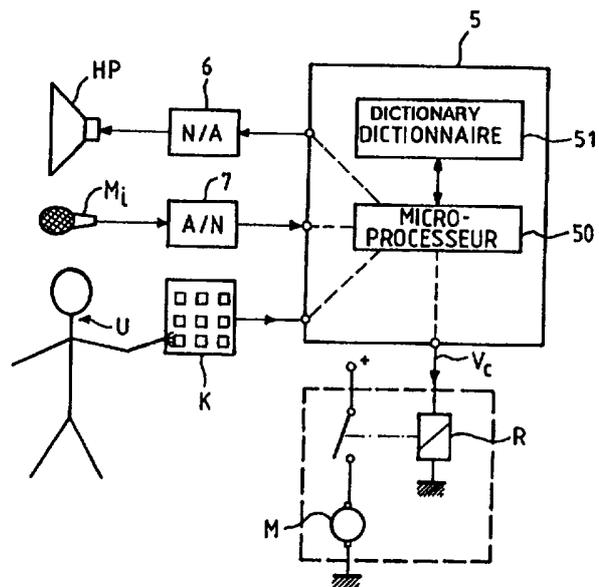
(54) Titre: PROCÉDE DE RECONNAISSANCE VOCALE D'UN LOCUTEUR METTANT EN OEUVRE UN MODELE PREDICTIF, NOTAMMENT POUR DES APPLICATIONS DE CONTROLE D'ACCES

(57) Abstract

The invention discloses a method for the voice recognition of a speaker using a q-order predictive model, comprising a step for extracting the statistical characteristics including a step of digital acquisition of the speaker's voice sample, corresponding to one or several utterances, a step of converting this voice sample into a sequence of vectors of size p, obtained from a series of analysis windows of average size T and an average spacing I, and a step of determining q+1 matrices from this vector sequence. The average size T is of a duration of less than 10 ms and the average spacing I of a duration of less than 4.5 ms. The invention is useful in a sound lock including an electroacoustical conversion system (HP, 6) and a recorded programme implementation system (5) for the method.

(57) Abrégé

L'invention concerne un procédé de reconnaissance vocale d'un locuteur mettant en oeuvre un modèle prédictif d'ordre q, comprenant une phase d'extraction de caractéristiques statistiques comportant une étape d'acquisition numérique d'un échantillon sonore du locuteur, correspondant à une ou plusieurs élocutions, une étape de transformation de cet échantillon sonore en une suite de vecteurs de dimension p, obtenus à partir d'une suite de fenêtres d'analyse de taille moyenne T et d'intervalle moyen I, et une étape de détermination de q+1 matrices de corrélation à partir de cette suite de vecteurs. Selon l'invention, la taille moyenne T a une durée inférieure à 10 ms et l'intervalle moyen I une durée inférieure à 4,5 ms. Application notamment à une serrure vocale comprenant un système de conversion électroacoustique (HP, 6) et un système (5) de mise en oeuvre du procédé à programme enregistré.



50...MICRO-PROCESSOR

Application notamment à une serrure vocale comprenant un système de conversion électroacoustique (HP, 6) et un système (5) de mise en oeuvre du procédé à programme enregistré.

**UNIQUEMENT A TITRE D'INFORMATION**

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	ES	Espagne	LS	Lesotho	SI	Slovénie
AM	Arménie	FI	Finlande	LT	Lituanie	SK	Slovaquie
AT	Autriche	FR	France	LU	Luxembourg	SN	Sénégal
AU	Australie	GA	Gabon	LV	Lettonie	SZ	Swaziland
AZ	Azerbaïdjan	GB	Royaume-Uni	MC	Monaco	TD	Tchad
BA	Bosnie-Herzégovine	GE	Géorgie	MD	République de Moldova	TG	Togo
BB	Barbade	GH	Ghana	MG	Madagascar	TJ	Tadjikistan
BE	Belgique	GN	Guinée	MK	Ex-République yougoslave de Macédoine	TM	Turkménistan
BF	Burkina Faso	GR	Grèce	ML	Mali	TR	Turquie
BG	Bulgarie	HU	Hongrie	MN	Mongolie	TT	Trinité-et-Tobago
BJ	Bénin	IE	Irlande	MR	Mauritanie	UA	Ukraine
BR	Bésil	IL	Israël	MW	Malawi	UG	Ouganda
BY	Bélarus	IS	Islande	MX	Mexique	US	Etats-Unis d'Amérique
CA	Canada	IT	Italie	NE	Niger	UZ	Ouzbékistan
CF	République centrafricaine	JP	Japon	NL	Pays-Bas	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norvège	YU	Yougoslavie
CH	Suisse	KG	Kirghizistan	NZ	Nouvelle-Zélande	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	République populaire démocratique de Corée	PL	Pologne		
CM	Cameroun	KR	République de Corée	PT	Portugal		
CN	Chine	KZ	Kazakstan	RO	Roumanie		
CU	Cuba	LC	Sainte-Lucie	RU	Fédération de Russie		
CZ	République tchèque	LI	Liechtenstein	SD	Soudan		
DE	Allemagne	LK	Sri Lanka	SE	Suède		
DK	Danemark	LR	Libéria	SG	Singapour		
EE	Estonie						

**Procédé de reconnaissance vocale d'un locuteur mettant en oeuvre un modèle prédictif, notamment pour des applications de contrôle d'accès**

5 L'invention concerne un procédé de reconnaissance vocale d'un locuteur mettant en oeuvre un modèle prédictif.

Elle concerne plus particulièrement un procédé de reconnaissance vocale d'un locuteur mettant en oeuvre un modèle prédictif du type autorégressif vectoriel.

10 Le procédé s'applique aussi bien à l'identification d'un locuteur ou à la détection des changements de locuteurs.

Elle trouve application dans de nombreux domaines et plus particulièrement dans le contrôle d'accès. Le contrôle d'accès s'effectue par la vérification d'une ou plusieurs élocutions d'un locuteur.

15 Elle trouve notamment application, bien que de façon non exclusive, dans les deux domaines suivants : autorisation d'accès faisant appel à une serrure vocale et autorisation d'accès à des informations confidentielles : validation de transactions et/ou opérations financières, accès sécurisé à des services d'information distants, etc.

20 Les procédés de l'art connu comportent habituellement un mode d'apprentissage et un mode de vérification. Ils comprennent en tout ou partie les phases et étapes suivantes : étapes de déclaration d'identité et de demande de services (phase 1), étapes d'autorisation d'apprentissage d'une ou plusieurs élocutions pour un locuteur (phase 2), étapes d'autorisation de vérification d'une ou  
25 plusieurs élocutions pour un locuteur (phase 3), étapes d'extraction de caractéristiques statistiques d'une ou plusieurs élocutions (phase 4), étapes de calcul de modèle prédictif à partir de caractéristiques statistiques d'une ou plusieurs élocutions (phase 5), étapes de vérification de l'identité par l'élocution (phase 6), phases d'identification du locuteur par l'élocution (phase 7), étapes d'autorisation  
30 d'accès à tout ou partie des ressources demandées en première phase (phase 8) et étapes de mise à jour du dictionnaire des caractéristiques statistiques et du modèle prédictif correspondant à tout ou partie des élocutions d'un locuteur (phase 9).

La première phase permet au locuteur de déclarer une identité et de requérir un mode de fonctionnement (apprentissage ou vérification) dans le but  
35 d'accéder à une ou plusieurs ressources.

La deuxième phase permet à un locuteur d'utiliser le mode d'apprentissage du dispositif d'accès.

La troisième phase permet à un locuteur d'utiliser le mode de vérification du dispositif d'accès.

La quatrième phase comporte une étape d'acquisition numérique et de filtrage d'une ou plusieurs élocutions, une étape d'extraction de vecteurs de dimension  $p$ , une étape de calcul de  $q+1$  matrices de corrélation de taille  $p \times p$  de tout ou partie des vecteurs calculés. Ces  $q+1$  matrices constituent les caractéristiques statistiques de l'élocution du locuteur.

La cinquième phase comporte une étape de calcul de  $q$  matrices de prédiction de taille  $p \times p$  à partir des matrices de corrélation, une étape de calcul de l'inverse de la matrice d'erreur associée. Ces  $q+1$  matrices constituent le modèle prédictif de l'élocution du locuteur. Les références de l'élocution d'un locuteur sont constituées des caractéristiques statistiques et du modèle prédictif associé.

La sixième phase comporte une étape de calcul des mesures de ressemblance entre les caractéristiques statistiques d'une ou plusieurs élocutions et tout ou partie des références des élocutions du dictionnaire, et une étape de calcul de la vraisemblance de la vérification d'identité.

La septième phase comporte une étape de calcul des mesures de ressemblance entre les caractéristiques statistiques d'une ou plusieurs élocutions et tout ou partie des références du dictionnaire, une étape de recherche des références les plus proches de l'élocution, et une étape de calcul de la vraisemblance de l'identification du locuteur.

La huitième phase autorise l'accès à tout ou partie des ressources demandées en première phase.

La neuvième phase permet la mise à jour des références de l'élocution d'un locuteur du dictionnaire ou l'ajout dans le dictionnaire des références d'un nouveau locuteur.

La vérification automatique du locuteur consiste à vérifier l'identité d'une personne en fonction d'un échantillon sonore de sa voix. Deux décisions sont possibles, selon un schéma binaire : "authentification" ou "non-authentification de l'identité".

Parmi le grand nombre de documents de l'état de l'art relatifs à des procédés de vérification du locuteur, on peut citer, à titre d'exemple non exhaustif, l'article de de Claude Montacié et Jean-Luc Le Floch : "Discriminant AR-Vector Models for Free-Text Speaker Verification" paru dans "Congrès EuroSpeech 1993", pages 161-164. Cet article divulgue un procédé de vérification automatique du locuteur, mais n'enseigne pas les conditions d'extraction des paramètres permettant

d'obtenir un système de vérification automatique du locuteur performant, rapide et fonctionnant dans des environnements bruités.

Les procédés précités s'appliquant, comme il a été indiqué, aussi bien à l'identification d'un locuteur ou à la détection des changements de locuteurs, il y lieu de tenir compte, entre autres, des caractéristiques physiologiques de la voix humaine. Notamment les périodes fondamentales des signaux vocaux complexes, selon qu'il s'agit d'un homme ou d'une femme, correspondent respectivement à des fréquences d'environ 100 Hz et 200 Hz. Lors de la phase quatre précitée, on utilise des fenêtres de temps qui seront définies ci-après. Dans l'art connu du traitement de la parole, il est admis que ces fenêtres de temps doivent être plus importantes que les périodes fondamentales précitées. En d'autres termes, l'analyse s'effectue sur une période plus importante que la période fondamentale. Il s'ensuit que les fenêtres habituellement utilisées sont comprises typiquement dans la gamme suivante : 15 à 40 ms. Des essais montrent d'ailleurs que lorsque l'on diminue cet intervalle de temps, les performances commencent à baisser.

D'autre part, on utilise habituellement plusieurs fenêtres qui se recouvrent. Il est également admis que l'espacement entre deux fenêtres consécutives, cet espacement étant défini comme la distance temporelle entre les centres des fenêtres, doit être de l'ordre de 10 ms environ, voire plus.

Or, de façon surprenante, il a été déterminé que lorsque l'on descendait à des valeurs bien inférieures aux valeurs précitées, les performances augmentaient, les résultats obtenus étaient meilleurs.

De façon plus précise, selon l'invention la durée de la fenêtre doit être inférieure à 10 ms.

De façon préférentielle, la période fondamentale étant, environ, de 5 ms pour les femmes et de 10 ms pour les hommes, on choisit une valeur de fenêtre égale à la période moyenne fondamentale, soit 7,5 ms.

De même, la distance temporelle entre fenêtre est choisie inférieure à 4,5 ms.

On choisit préférentiellement des valeurs bien inférieures à cette valeur, par exemple 2 ms.

La présente invention concerne donc un procédé de reconnaissance vocale d'un locuteur mettant en oeuvre un modèle prédictif qui tout en conservant les caractéristiques techniques et les avantages des procédés de l'art connu, présente cependant de meilleures performances.

Le procédé selon l'invention s'adapte tout particulièrement aux modèles prédictif du type autorégressif vectoriel.

L'invention a donc pour objet un procédé de reconnaissance vocale d'un locuteur mettant en oeuvre un modèle prédictif d'ordre  $q$ , comprenant au moins une  
5 phase d'extraction de caractéristiques statistiques comportant au moins une étape d'acquisition numérique d'un échantillon sonore d'une durée déterminée  $D$  du locuteur, correspondant à au moins une élocution du locuteur, une étape de transformation de cet échantillon sonore en une suite de vecteurs de dimension déterminée  $p$ , obtenus à partir d'une suite de fenêtres d'analyse de taille moyenne  $T$   
10 et d'intervalle moyen  $I$ , et une étape de détermination de  $q+1$  matrices de corrélation à partir de cette suite de vecteurs, avec  $p$  et  $q$  nombres entiers différents de zéro, caractérisé en ce que ladite taille moyenne  $T$  a une durée inférieure à 10 ms.

Elle a encore pour objet l'application d'un tel procédé à l'identification d'un locuteur ou à la détection des changements de locuteurs.

15 Elle a encore pour objet l'application d'un tel procédé au contrôle d'accès à l'aide d'une serrure vocale.

Elle a enfin pour objet l'application d'un tel procédé au contrôle d'accès à des informations confidentielles, notamment à la validation de transactions et/ou opérations financières, accès sécurisé à des services d'information distants.

20 L'invention sera mieux comprise et d'autres caractéristiques et avantages apparaîtront à la lecture de la description qui suit en référence aux figures annexées, et parmi lesquelles :

- La figure 1 illustre un exemple de signal de fenêtre ;
- La figure 2 est un organigramme permettant la description détaillée du  
25 procédé selon l'invention ;
- La figure 3 illustre l'application du procédé selon l'invention à un contrôle d'accès mettant en oeuvre une serrure vocale ;
- La figure 4 illustre l'application du procédé selon l'invention au contrôle d'accès à distance à des informations confidentielles.

30 Comme il a été indiqué précédemment, le procédé selon l'invention conserve les caractéristiques générales des procédés de l'art connu. Pour fixer les idées, sans que cela soit limitatif en quoique ce soit de la portée de l'invention, on se placera dans ce qui suit dans le cadre particulier d'un modèle prédictif du type autorégressif vectoriel.

35 Le procédé selon l'invention concerne plus particulièrement trois grandes familles d'opérations : l'extraction de caractéristiques statistiques d'un

échantillon sonore spécialement adapté aux modèles prédictifs autorégressifs vectoriels, le calcul de mesures de ressemblance entre échantillons sonores et la mise à jour du dictionnaire des références des locuteurs. Ces grandes familles d'opérations vont maintenant être décrites.

5 Cette partie du procédé caractérise un échantillon sonore par  $q+1$  matrices de corrélation d'ordre  $p \times p$ , calculées à partir d'un ensemble de vecteurs de dimension  $p$ . La méthode d'obtention des vecteurs est liée à la phase 4 précitée et comporte des spécificités propres à l'invention.

Avant de détailler les opérations liées à l'extraction de caractéristiques statistiques d'un échantillon sonore, il convient de définir les termes utilisés.

Un échantillon sonore est une représentation quelconque d'un signal acoustique (par exemple : acoustique, électrique, physique, magnétique, optique, etc.). Toute sous-partie d'un échantillon sonore est un échantillon sonore. Toute concaténation d'échantillons sonores est un échantillon sonore.

15 L'acquisition numérique de l'échantillon sonore est la transformation du signal analogique représentant l'échantillon sonore de durée  $D$  en une suite de  $N$  valeurs numériques  $\{x_i\}$  ( $i=1, \dots, N$ ) à des instants d'échantillonnage  $\{t_i\}$  ( $i=1, \dots, N$ ) croissants.

Le filtrage numérique est la transformation d'une suite de  $N$  valeurs numériques  $\{x_i\}$  ( $i=1, \dots, N$ ), aux instants d'échantillonnage  $\{t_i\}$  ( $i=1, \dots, N$ ), en une suite de  $N'$  valeurs numériques  $\{x'_i\}$  ( $i=1, \dots, N'$ ) aux instants d'échantillonnage  $\{t'_i\}$  ( $i=1, \dots, N'$ ) croissants. Les suites de valeurs numériques  $\{x_i\}$  ( $i=1, \dots, N$ ) peuvent subir, le cas échéant, de nombreux types de filtrages. Dans ce cas, on utilise les nouvelles suites numériques.

25 La fenêtre d'analyse  $F$  est une fonction discrète  $f$  de valeur  $\{f_i\}$  ( $i=1, \dots, N$ ) aux instants  $\{t_i\}$  ( $i=1, \dots, N$ ). Cette fenêtre d'analyse  $F$  est associée à l'acquisition numérique (et au filtrage éventuel) de l'échantillon sonore. On peut par exemple citer les fenêtres d'analyse rectangulaires, les fenêtres de Hamming, etc.

Le maximum  $M$  d'une fenêtre d'analyse  $F$  est défini par la relation :

$$30 \quad M = \max_{i=1, \dots, N} |f_i|$$

La zone  $Z$  d'une fenêtre d'analyse  $F$  est la fonction discrète  $f_z$  de valeurs  $\{f_{z_i}\}$  ( $i=1, \dots, N_z$ ) avec  $f_{z_i} = f_{Z(i)}$  et  $N_z \leq N$ ,  $Z(i)$  étant une fonction discrète de la forme  $Z(i) = i + D_z$ ,  $D_z$  étant une valeur entière comprise entre 0 et  $N - N_z$ .

La zone  $Z$  de type A d'une fenêtre d'analyse  $F$  est la zone  $Z$  pour laquelle la relation suivante est satisfaite :

$$35 \quad \forall i \in \{1, \dots, N_z\} \quad |f_{z_i}| > M / 5$$

La zone Z de type B d'une fenêtre d'analyse F est la zone Z pour laquelle la relation suivante est satisfaite :

$$\forall i \in \{1, \dots, N_Z\} \left| f_{Z,i} \right| \leq M / 5$$

Une fenêtre d'analyse F peut être décomposée en zones de types A ou B.  
5 La durée d'une zone Z est égale à  $t_{Z(N_Z)} - t_{Z(1)}$ .

Une zone  $Z_F$  est une zone de type A de plus grande durée sur toutes les décompositions de F.

La durée de la fenêtre d'analyse F est la durée de  $Z_F$ .

10 Le centre de la fenêtre d'analyse F est défini par la relation :

$$C_F = \frac{\sum_{t_i \in Z_F} t_i \times f_i}{\sum_{t_i \in Z_F} f_i}$$

Un échantillon sonore  $\{x_i\}$  ( $i=1, \dots, N$ ), fenêtré par la fenêtre d'analyse F, est un ensemble de N valeurs numériques  $\{x_{f_i}\}$  ( $i=1, \dots, N$ ), tel que  $x_{f_i} = x_i \times f_i$ .

15 Un intervalle entre deux fenêtres d'analyse  $F_1$  et  $F_2$  est défini par la relation :

$$I_{F_1 F_2} = C_{F_2} - C_{F_1}$$

L'ensemble de fenêtres (T, I, N, D) est la suite de  $N_F$  fenêtres d'analyse successives  $\{F_j\}$  ( $j=1, \dots, N_F$ ) de longueur N, avec  $N_F \leq (D - T) / I$  où T est la durée moyenne de la fenêtre d'analyse et I l'intervalle moyen entre deux fenêtres d'analyse.

20 Cet ensemble de fenêtres (T, I, N, D) est associé à l'acquisition numérique (et au filtrage éventuel) de l'échantillon sonore.

L'extraction d'un vecteur de dimension p est la fonction calculant un vecteur  $y_F$  de dimension p à partir d'un échantillon sonore  $\{x_i\}$  ( $i=1, \dots, N$ ) fenêtré par la fenêtre d'analyse F. Tous les  $\{x_{f_i}\}$  ( $i=1, \dots, N$ ) non nuls doivent être utilisés dans le calcul de cette fonction, c'est-à-dire que la modification sensible d'une des valeurs non nulles de l'ensemble  $\{x_{f_i}\}$  ( $i=1, \dots, N$ ) doit modifier le résultat de la fonction de manière sensible. Les paramétrisations suivantes peuvent, par exemple, être utilisées : transformée de Fourier, cespstre, prédiction linéaire, etc.

30 L'ensemble de vecteurs ( $\{x_i\}$ , p, T, I, N, D) est une suite de  $N_F$  vecteurs  $\{y_{F_j}\}$  ( $j=1, \dots, N_F$ ) calculés à partir d'un ensemble de fenêtres (T, I, N, D).

Les matrices de corrélation d'un échantillon sonore ( $\{x_i\}$ , p, q, T, I, N, D) sont un ensemble de q+1 matrices  $\{R_m\}$  ( $m=0, \dots, q$ ) de taille p x p, tel que la relation suivante soit satisfaite :

$$R_m = \sum_{j=1}^{N_F - m} y_{F_j} y_{F_{j+m}}^T$$

35 Elles peuvent subir, le cas échéant, toutes sortes de normalisation.

Le procédé selon l'invention, dans la quatrième phase, consiste en la détermination des paramètres (p, q, T, I) permettant d'obtenir les meilleures caractéristiques statistiques, pour le modèle prédictif utilisé, d'un échantillon sonore en terme de vérification du locuteur.

5 p peut être quelconque, cependant ce paramètre est généralement compris entre 6 et 30.

Selon une caractéristique importante de l'invention, T est choisi inférieur à 10 ms et I inférieur à 4,5 ms.

10 Il est bien connu que de tels signaux ne sont jamais parfaits. La figure 1 illustre schématiquement un signal temporel SF(t) représentant une fenêtre, dans un repère orthonormé "Amplitude A - Temps t". La largeur de la fenêtre peut être définie, par exemple, par l'intervalle de temps T séparant les intersections d'une droite parallèle à l'axe des temps avec la courbe représentant le signal SF(t), à 20% de l'amplitude maximale  $A_{max}$ .

15 En outre q, T et I doivent être reliés par l'inéquation suivante :

$$I < T/\max(3, q)$$

A titre d'exemple, si on choisit q = 4 avec T = 8ms, I doit être inférieur à 2 ms.

20 De façon général, plus le paramètre q est élevé, plus les caractéristiques statistiques obtenues sont meilleures. Cependant, cette amélioration est obtenu au prix d'un coût de calcul plus important. Aussi, de façon pratique, q est généralement choisi dans la gamme 1 à 10, ce qui constitue un bon compromis.

L'invention concerne également des mesures de ressemblance entre échantillons sonores qui sont utilisées dans les phases 6 et 7 précitées. Le calcul du modèle prédictif va être détaillé à partir des caractéristiques statistiques d'une ou plusieurs élocutions du locuteur, ce qui constitue la phase 5.

Les mesures de ressemblance sont basées sur le calcul des modèles prédictifs et des matrices d'erreur associées.

30 Comme précédemment, il est utile tout d'abord de définir les termes utilisés.

La prédiction directe du vecteur  $y_{F_j}$  est égale à  $\hat{y}_{F_j}^D$

La prédiction rétrograde du vecteur  $y_{F_j}$  est égale à  $\hat{y}_{F_j}^R$

35 Le modèle prédictif vectoriel direct ( $\{x_i\}$ , p, q, T, I, N, D) d'un ensemble de vecteurs ( $\{x_i\}$ , p, T, I, N, D) est l'ensemble de q matrices  $\{A_m\}$  ( $m=1, \dots, q$ ) de taille p x p calculées à partir des matrices de corrélation  $\{R_m\}$  ( $m=0, \dots, q$ ). Le but est de prédire au mieux un vecteur  $y_{F_j}$  à partir des q vecteurs précédents.

$$\hat{y}_{F_j}^D = \sum_{m=1}^q A_m y_{F_{j-m}}$$

Le modèle prédictif vectoriel rétrograde ( $\{x_i\}$ ,  $p$ ,  $q$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ) d'un ensemble de vecteurs ( $\{x_i\}$ ,  $p$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ) est l'ensemble de  $q$  matrices  $\{B_m\}$  ( $m=1, \dots, q$ ) de taille  $p \times p$  calculées à partir des matrices de corrélation  $\{R_m\}$  ( $m=0, \dots, q$ ). Le but est de prédire au mieux un vecteur  $y_{F_j}$  à partir des  $q$  vecteurs suivants :

$$\hat{y}_{F_j}^R = \sum_{m=1}^q B_m y_{F_{j+m}}$$

L'erreur directe  $\varepsilon_j$  ( $\{x_i\}$ ,  $p$ ,  $q$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ,  $\{A_m\}$ ) sur un ensemble de vecteurs ( $\{x_i\}$ ,  $p$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ) par un modèle prédictif direct ( $\{x_i\}$ ,  $p$ ,  $q$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ) est égale au filtrage inverse d'un ensemble de vecteurs ( $\{x_i\}$ ,  $p$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ) par un modèle prédictif direct  $\{A_m\}$  ( $m=1, \dots, q$ ), tel que la relation suivante soit vérifiée :

$$\varepsilon_j = y_{F_j} - \sum_{m=1}^q A_m y_{F_{j-m}}$$

L'erreur rétrograde  $\xi_j$  ( $\{x_i\}$ ,  $p$ ,  $q$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ,  $\{B_m\}$ ) sur un ensemble de vecteurs ( $\{x_i\}$ ,  $p$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ) par un modèle prédictif rétrograde ( $\{x_i\}$ ,  $p$ ,  $q$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ) est égale au filtrage inverse d'un ensemble de vecteurs ( $\{x_i\}$ ,  $p$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ) par un modèle prédictif rétrograde  $\{B_m\}$  ( $m=1, \dots, q$ ) tel que la relation suivante soit vérifiée :

$$\xi_j = y_{F_j} - \sum_{m=1}^q B_m y_{F_{j+m}}$$

La matrice d'erreur directe  $E_A$  ( $\{x_i\}$ ,  $p$ ,  $q$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ,  $\{A_m\}$ ) de l'erreur directe ( $\{x_i\}$ ,  $p$ ,  $q$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ,  $\{A_m\}$ ) est égale à la matrice de covariance, de taille  $p \times p$ , de l'erreur directe ( $\{x_i\}$ ,  $p$ ,  $q$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ,  $\{A_m\}$ ) tel que la relation suivante soit vérifiée :

$$E_A = \sum_{j=1}^{N_f} \varepsilon_j \varepsilon_j^T$$

La matrice d'erreur rétrograde  $E_B$  ( $\{x_i\}$ ,  $p$ ,  $q$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ,  $\{B_m\}$ ) de l'erreur rétrograde ( $\{x_i\}$ ,  $p$ ,  $q$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ,  $\{B_m\}$ ) est égale à la matrice de covariance, de taille  $p \times p$ , de l'erreur rétrograde ( $\{x_i\}$ ,  $p$ ,  $q$ ,  $T$ ,  $I$ ,  $N$ ,  $D$ ,  $\{B_m\}$ ) tel que la relation suivante soit vérifiée :

$$E_B = \sum_{j=1}^{N_f} \xi_j \xi_j^T$$

Pour calculer les modèles prédictifs directs  $\{A_m\}$  ( $m=1, \dots, q$ ) ou rétrogrades  $\{B_m\}$  ( $m=1, \dots, q$ ), on peut citer, à titre d'exemple non limitatif,

l'algorithme de Levinson-Whittle-Robinson qui minimise la trace de  $E_A$  et  $E_B$  et qui s'énonce comme suit :

```

    A00 = Id; B00 = Id; EA0 = R0; EB0 = R0;
    for i = 0 to q-1 do
5      Fi = ∑m=0i Ami Ri+1-m
      KAi = -Fi EBi-1; KBi = -FiT EAi-1;
      EAi+1 = (Id - KAi KBi) EAi;
      EBi+1 = (Id - KBi KAi) EBi;
      A0i+1 = A0i; B0i+1 = KBi A0i;
10     Ai+1i+1 = KAi Bii; Bi+1i+1 = Bii;
      for m = 1 to i do
      Ami+1 = Ami + KAi Bm-1i
      Bmi+1 = Bm-1i + KBi Ami
      end for
15     end for
      for m = 1 to q do Am = Amq end for
      for m = 1 to q do Bm = Bmq end for

```

Soient les définitions et notations suivantes dans le cas du modèle prédictif direct :

- 20 a/  $\{x_{1i}\}$  ( $i=1, \dots, N_1$ ) est l'ensemble de  $N_1$  valeurs numériques de l'échantillon sonore 1 ;
- b/  $\{R_{1m}\}$  ( $m=0, \dots, q$ ) représente  $q+1$  matrices de corrélation pour ( $\{x_{1i}\}$ ,  $p, q, T, I, D$ ) ;
- 25 c/  $\{A_{1m}\}$  ( $m=1, \dots, q$ ) est le modèle prédictif direct d'ordre  $q$  pour ( $\{x_{1i}\}$ ,  $p, q, T, I, D$ ) ;
- d/  $\{x_{2i}\}$  ( $i=1, \dots, N_2$ ) est l'ensemble de  $N_2$  valeurs numériques de l'échantillon sonore 2 ;
- e/  $\{R_{2m}\}$  ( $m=0, \dots, q$ ) représente  $q+1$  matrices de corrélation pour ( $\{x_{2i}\}$ ,  $p, q, T, I, D$ ) ;
- 30 f/  $\{A_{2m}\}$  ( $m=1, \dots, q$ ) est le modèle prédictif direct d'ordre  $q$  pour ( $\{x_{2i}\}$ ,  $p, q, T, I, D$ ) ;
- g/  $E_{A11}$  est la matrice d'erreur directe ( $\{x_{1i}\}$ ,  $p, q, T, I, D, \{A_{1m}\}$ ) ;
- h/  $E_{A21}$  est la matrice d'erreur directe ( $\{x_{1i}\}$ ,  $p, q, T, I, D, \{A_{2m}\}$ ) ;
- 35 i/  $E_{A12}$  est la matrice d'erreur directe ( $\{x_{2i}\}$ ,  $p, q, T, I, D, \{A_{1m}\}$ ) ;
- j/  $E_{A22}$  : matrice d'erreur directe ( $\{x_{2i}\}$ ,  $p, q, T, I, D, \{A_{2m}\}$ ).

Ces matrices peuvent être calculées directement par l'algorithme suivant :

$$X_{A12} = \sum_{m=1}^q \left( \left( R2_m + \sum_{i=0}^{m-1} A1_{m-i} R2_m \right) + R2_m \right) A1_m^T$$

$$Y_{A12} = R2_0 + \sum_{m=1}^q A1_{m-i} R2_0 A1_m^T$$

$$E_{A12} = X_{A12} + X_{A12}^T + Y_{A12}$$

$$X_{A21} = \sum_{m=1}^q \left( \left( R1_m + \sum_{i=0}^{m-1} A2_m R1_{m-i} \right) + R1_m \right) A2_m^T$$

$$Y_{A21} = R1_0 + \sum_{m=1}^q A2_{m-i} R1_0 A2_m^T$$

$$E_{A11} = R1_0 + \sum_{m=1}^q A1_m R1_m^T$$

$$E_{A22} = R2_0 + \sum_{m=1}^q A2_m R2_m^T$$

Soit  $G(E)$  une fonction des valeurs propres d'une matrice  $E$  de taille  $p \times p$ . Par exemple,  $G(E)$  peut être égale à la somme des valeurs propres ( $\text{Trace}(E)$ ) ou au produit des valeurs propres ( $\text{Déterminant}(E)$ ).

La mesure de ressemblance entre deux échantillons sonores est une fonction de  $G(EA11EA12^{-1})$ ,  $G(EA12^{-1}EA11)$ ,  $G(EA22EA21^{-1})$  et  $G(EA21^{-1}EA22)$ . A titre d'exemple, on peut choisir une somme pondérée.

La mesure de ressemblance peut également être définie à partir des modèles rétrogrades, ou en mélangeant ces deux modélisations.

Le procédé selon l'invention concerne également la mise à jour du dictionnaire des références, c'est-à-dire des caractéristiques statistiques (matrices de corrélation) et des modèles prédictifs. La mise à jour des caractéristiques statistiques est réalisée par toute fonction possible permettant d'obtenir un nouvel ensemble de matrices de corrélation  $\{R_m\}$  à partir de deux ensembles de matrices de corrélation. Les deux jeux de matrices de corrélation sont celui du dictionnaire  $\{R1_m\}$  et celui provenant d'une ou plusieurs nouvelles élocutions  $\{R2_m\}$ .

A titre d'exemple non limitatif, le nouvel ensemble de matrices de corrélation  $\{R_m\}$  peut être obtenu par la relation suivante :

$$R_m = \alpha R1_m + \beta R2_m + \gamma \sqrt{R1_m R2_m} \quad (m = 0, \dots, q)$$

Le procédé selon l'invention va maintenant être décrit de façon détaillée en regard de l'organigramme de la figure 2. Celui-ci comprend deux branches principales, à partir d'un tronc de départ commun 1 : la branche de gauche concerne

l'apprentissage (branche 2), alors que la branche de droite concerne la vérification (branche 3). Les phases 1 à 9, référencées  $P_1$  à  $P_9$  sont relatives aux phases 1 à 9 précédemment rappelées. Un certain nombre de décisions intermédiaires sont prises lors du déroulement des phases du procédé et ont été référencées  $D_1$  à  $D_5$ .

5 Pendant la phase 1 ou  $P_1$ , le locuteur déclare une identité par les moyens mis à sa disposition. Il peut avoir recourt à la frappe d'un code connu de lui seul, utiliser une carte magnétique ou un moyen similaire. Il demande alors un des deux modes de fonctionnement : apprentissage ou vérification, ce dans le but d'accéder à une ou plusieurs ressources.

10 En fin de phase  $P_1$  (sortie 10) une première décision est prise ou décision  $D_1$ . Cette décision  $D_1$  porte sur l'acceptation de réaliser un apprentissage (branche 20) ou une vérification (branche 30), pour accéder aux ressources demandées dans la phase  $P_1$ . Par exemple un apprentissage peut être substitué à une vérification si l'apprentissage est jugé insuffisant, ou un refus peut être signifié  
15 (branche 11), si le mode de fonctionnement ou les ressources demandées sont indisponibles.

Les ressources précitées dépendent naturellement de l'application précise envisagée. Des exemples seront décrits ci-après.

Si on considère que la décision  $D_1$  a aiguillé le processus vers la  
20 branche "Apprentissage" 2 (branche 20 de  $D_1$ ), la phase 2 ou  $P_2$  est initiée.

Cette phase  $P_2$  déclenche toutes les actions permettant l'apprentissage d'une référence d'une ou plusieurs élocutions du locuteur dont l'identité est déclarée. Selon les applications précises, il va s'agir de l'activation d'entrée sonore, de la recherche de dictionnaires, etc.

25 En fin de phase  $P_2$  (sortie 21), le procédé, dans cette branche "Apprentissage" 2 passe directement à la phase 4 ou  $P_4$ .

Pendant cette phase  $P_4$ , l'extraction des caractéristiques statistiques se décompose en trois étapes comme il a été précédemment décrit. La première étape consiste en l'acquisition numérique et le filtrage éventuel de l'échantillon sonore  
30 correspondant à une ou plusieurs élocutions. La deuxième étape est sa transformation en une suite de vecteurs obtenus à partir d'une suite de fenêtres d'analyse de taille moyenne  $T$  et d'intervalle moyen  $I$ , conformément à l'une des caractéristiques principales de l'invention. La troisième étape est le calcul de  $q+1$  matrices de corrélation à partir de cette suite de vecteurs.

35 En fin de phase  $P_4$  (sortie 22), une deuxième décision  $D_2$  est prise. En effet, lors de l'apprentissage, une identification de l'identité du locuteur peut être

décidée pour s'assurer de son identité avant une mise à jour de ses références d'élocution. Si la décision est "non" (branche 23), la phase 9 ou P<sub>9</sub> est directement initiée. Cette phase P<sub>9</sub> sera décrite ci-après. Dans le cas contraire : décision "oui" (branche 24), la phase 5 ou P<sub>5</sub> est initiée.

5            La phase P<sub>5</sub> est réalisée en deux étapes. La première étape consiste à calculer, à partir des caractéristiques statistiques (matrices de corrélation) d'une ou plusieurs élocutions, q matrices de prédiction. Plusieurs choix sont possibles pour le calcul d'un modèle prédictif : un modèle prédictif direct, un modèle prédictif  
10            rétrograde ou une combinaison des deux modèles. La deuxième étape réalise le calcul de l'inverse de la matrice d'erreur entre la suite des vecteurs calculés à partir de l'échantillon sonore d'une ou plusieurs élocutions et le modèle prédictif calculé sur ces mêmes vecteurs. L'ensemble de ces q+1 matrices définit le modèle prédictif de l'élocution du locuteur. Les références de l'élocution d'un locuteur sont constituées des caractéristiques statistiques et du modèle prédictif associé.

15            Dans cette branche "d'Apprentissage" 2, à la sortie 25 de la phase P<sub>5</sub>, la phase 7 ou P<sub>7</sub> est directement initiée, sans passer par la phase 6 ou P<sub>6</sub>.

              La phase P<sub>7</sub> comporte quatre étapes. La première étape consiste en l'extraction d'un dictionnaire contenant les références de tous les locuteurs possibles. La deuxième étape consiste dans le calcul des matrices d'erreur entre les références  
20            précédemment sélectionnées et les références (ou seulement les caractéristiques statistiques) de l'échantillon sonore à identifier. La troisième étape consiste en un calcul des mesures de ressemblance à partir des matrices d'erreur calculées à l'étape précédente. La quatrième étape calcule la vraisemblance de l'identification du locuteur à partir des mesures de vraisemblance calculées lors de la troisième étape.

25            En sortie de cette phase P<sub>7</sub>, une troisième décision D<sub>3</sub> est prise. Cette décision D<sub>3</sub> porte sur l'adéquation de l'identité résultant de l'identification de la phase P<sub>7</sub> et de l'identité déclarée lors de la phase P<sub>1</sub>. S'il n'y a pas adéquation : branche "non" 27, le processus est arrêté. Selon les applications, une action peut alors être prise à ce niveau.

30            Dans le cas contraire : branche "oui" 28, le procédé se déroule sur un tronc 40 commun aux branches 23 (décision D<sub>2</sub>) et 28 (décision D<sub>3</sub>), ainsi qu'à la branche 39 (décision D<sub>5</sub>) qui sera explicité ultérieurement. La phase 9 ou P<sub>9</sub> est alors initiée.

              La phase 9 permet la mise à jour de références de l'élocution d'un  
35            locuteur du dictionnaire ou l'ajout dans le dictionnaire de celles d'un nouveau locuteur. La mise à jour des références comporte une étape d'extraction, du

dictionnaire, des  $q+1$  matrices de corrélation d'un locuteur, une étape de calcul et de stockage dans le dictionnaire de  $q+1$  nouvelles matrices de corrélation par accumulation des anciennes matrices et celles extraites d'une ou plusieurs élocutions, une étape de calcul et de stockage dans le dictionnaire du modèle  
5 prédictif correspondant aux nouvelles matrices de corrélation. L'ajout dans le dictionnaire des références de l'élocution d'un nouveau locuteur consiste en un stockage des  $q+1$  matrices de corrélation du locuteur et du modèle prédictif correspondant.

En sortie 41 de cette phase  $P_9$ , une action peut être prise, action qui  
10 dépend naturellement de l'application envisagée. Dans le cas le plus fréquent, le processus engagé s'arrête, le système étant mis à jour pour le locuteur courant ou prêt à le reconnaître, s'il s'agit d'un nouveau locuteur.

Si maintenant on considère la branche "Vérification" 3 (sortie 30 de la décision  $D_1$ ), le procédé initie directement la phase 3 ou  $P_3$ .

15 Cette phase  $P_3$  déclenche toutes les actions permettant la vérification du locuteur dont l'identité est déclarée. Comme dans le cas de la phase  $P_2$ , selon les applications précises, il va s'agir de l'activation d'entrée sonore, de la recherche de dictionnaires, etc.

En sortie 31 de la phase  $P_3$ , la phase  $P'_4$  va être initiée. Cette phase  $P'_4$   
20 est strictement identique à la phase  $P_4$  de la branche "Apprentissage" 2, et il est inutile de la redécrire.

De même, en sortie 32 de la phase  $P'_4$ , la phase  $P'_5$  va être initiée, phase  $P'_5$  identique également à la phase  $P_5$  de la branche "Apprentissage" 2.

A l'issu 33 de cette phase  $P'_5$ , la phase 6 ou  $P_6$  est initiée.

25 La phase  $P_6$  comporte quatre étapes. La première étape réalise l'extraction du dictionnaire des références de tout ou partie des locuteurs (par exemples les  $K$  plus proches références au sens d'une mesure de ressemblance). La deuxième étape est le calcul des matrices d'erreur entre tout ou partie des références précédemment sélectionnées et les références (ou seulement les caractéristiques  
30 statistiques) de l'échantillon sonore à vérifier. La troisième étape consiste en un calcul des mesures de ressemblance à partir de tout ou partie des matrices d'erreur calculées à l'étape précédente. La quatrième étape calcule la vraisemblance de la vérification du locuteur à partir des mesures de vraisemblance calculées lors de la troisième étape.

35 A l'issu 34 de cette phase  $P_6$ , une quatrième décision  $D_4$  est prise (la première pour cette branche "Vérification" 3). Cette décision porte sur

l'authentification ou la non-authentification de l'identité du locuteur en mode vérification.

S'il l'authentification est négative : branche "non" 35, le processus est arrêté ou une action, qui dépend de l'application précise, peut être effectuée. Il peut s'agir, par exemple, de "l'avalement" de la carte magnétique présentée en phase P<sub>1</sub>, d'une alarme sonore, etc.

Dans le cas contraire : branche "oui" 36, la phase 8 ou P<sub>8</sub> est initiée.

Cette phase P<sub>8</sub> autorise l'accès à tout ou partie des ressources demandées par le locuteur au cours de la première phase. Selon l'application précise, il peut s'agir, par exemple, de l'ouverture d'une porte, de l'accès à des informations confidentielles, etc.

Une cinquième décision D<sub>5</sub> peut également être prise à ce stade (sortie 37 de la phase P<sub>8</sub>).

Cette décision D<sub>5</sub> porte sur la mise à jour éventuelle des références d'élocution du locuteur en cas de très bonne authentification de son identité en mode vérification (branche 3).

Si le résultat de la décision D<sub>5</sub> est "non" (branche 38), le processus engagé va se terminer dans la majorité des cas.

Dans le cas contraire (branche "oui" 39), le procédé est de nouveau aiguillé vers la branche de sortie 4 (tronc commun 40) et la phase P<sub>9</sub> exécutée. Cette phase a déjà été explicitée, et il est inutile de la redécrire.

Comme pour la branche "Apprentissage" 2, en sortie 41 de cette phase P<sub>9</sub>, une action peut être prise, action qui dépend naturellement de l'application envisagée. Dans le cas le plus fréquent, le processus engagé s'arrête, le système étant mis à jour pour le locuteur courant et celui-ci ayant eu accès (phase P<sub>8</sub>) aux ressources demandées.

On va maintenant décrire un premier exemple d'application du procédé selon l'invention, par référence à la figure 3.

L'application visée concerne la protection par une serrure vocale de l'entrée d'un local ou la commande d'un appareil quelconque dont l'accès est réservé à des personnes autorisées.

On suppose que le locuteur U déclare son identité au moyen d'un clavier K type "portier". Pour ce faire, il doit saisir un code qui va être entré dans le système 5 mettant en oeuvre le procédé selon l'invention. D'autres moyens auraient pu être utilisés : badge magnétique, carte à puce, etc.

Le système comporte des circuits communs, en soi, à ceux de l'art connu. Il comprend notamment un dictionnaire 51, dans lequel sont enregistrées des données servant à la vérification des locuteurs connus, et un microprocesseur 50 ou tout système informatique approprié permettant d'effectuer les différents calculs et déterminations précédemment décrits. Dans l'état actuel des techniques, ces calculs s'effectuent sous la commande de programmes enregistrés dans des mémoires fixes (de type "ROM" ou "EPROM", par exemple) et/ou de masse (disque dur), non représentées sur la figure 3.

On supposera à ce stade, pour simplifier, qu'un seul mode de fonctionnement est possible : le mode-vérification.

Outre le clavier K, on peut également prévoir des moyens de guidage du locuteur, par exemple un haut-parleur HP commandé, via un convertisseur numérique-analogique 6, par le système 5. Des instructions peuvent être données par ce truchement, au locuteur U. On aurait pu également prévoir, en lieu et place, un écran de visualisation, par exemple à cristaux liquides affichant ces mêmes instructions et donnant des détails sur le déroulement des opérations en cours et des résultats des différentes décisions prises par le système 5.

Enfin, il est prévu un microphone  $M_j$ , ou tout transducteur électroacoustique approprié, destiné à capter la voix du locuteur U. Un convertisseur analogique-numérique 7 transforme les signaux électriques en sortie du microphone 7 en signaux numériques pouvant être lus par le microprocesseur 50. Pour ce faire, la sortie du convertisseur 7 peut être connectée à un port parallèle ou série, comme il est bien connu.

Une fois le code saisi à l'aide du clavier K, le système 5 acquiert des données sur l'identité présumée du requérant, c'est-à-dire le locuteur U. La phase  $P_3$  et les phases subséquentes, notamment la phase  $P'_4$  (phase 4), peuvent être initiées et exécutées, ce jusqu'à ce qu'une décision soit prise par le système 5 : décision  $D_4$ , à l'issue de la phase  $P_6$ .

Si la décision est favorable (figure 2, branche 36), la phase  $P_8$  est exécutée et le locuteur a accès aux ressources demandées, par exemple celui-ci a accès à un local fermé par une porte commandée par une gâche électrique symbolisée, sur la figure 3, par un moteur M. On suppose, pour simplifier, que l'alimentation en énergie électrique de cette gâche, est commandée par un relais R sous le contrôle d'un signal électrique  $V_C$  délivré par le système 5.

A l'issu de la phase P<sub>8</sub>, le processus peut s'arrêter (figure 2, branche 38, suite au résultat de la décision D<sub>5</sub>) ou passer par la phase P<sub>9</sub>, pendant laquelle le dictionnaire 51 sera mis à jour pour le locuteur U.

Bien qu'il a été supposé que le système ne permettait qu'un seul mode de  
5 fonctionnement : le mode "Vérification" qui se prête plus à ce type d'application, le mode "Apprentissage" peut également être implanté. Il suffit, par exemple, de prévoir une touche spécialisée sur le clavier K, permettant le choix entre ces deux modes, ou de dupliquer, dans un autre lieu, les organes d'entrée de code et d'acquisition de la voix.

10 Dans ce cas il est nécessaire d'implémenter les circuits et instructions permettant d'aiguiller le procédé vers les branches 2 ou 3 (figure 2 : décision D<sub>1</sub>) et d'exécuter les phases (figure 2 : P<sub>2</sub>, P<sub>4</sub>, P<sub>5</sub> et P<sub>7</sub>) et décisions (figure 2 : D<sub>2</sub> et D<sub>3</sub>) propres à la branche 2.

15 Une deuxième série d'applications du procédé selon l'invention va maintenant être décrite en référence à la figure 4.

Les applications visées concernent l'accès sécurisé à des informations distantes. Il peut s'agir, par exemple, de valider des transactions et opérations financières (bancaires, boursières, etc.).

20 Pour ce type d'applications, il s'agit d'identifier avec certitude le demandeur ou le donneur d'ordre (cas précité des transactions bancaires, boursières, etc.).

Ce dernier n'étant pas, par hypothèse, en relation de proximité avec le centre de décision, va formuler ses requêtes à l'aide d'un appareil téléphonique standard, sauf à disposer de moyens de communication informatiques sécurisés, ce  
25 qui sort du cadre de la présente invention.

La figure 4 illustre schématiquement l'architecture d'un système permettant des transactions sécurisées et mettant en oeuvre le procédé selon l'invention. Le terme "transactions" est utilisé ici dans son sens le plus général et recouvre les diverses définitions rappelées ci-dessus.

30 Le locuteur U dispose d'un combiné téléphonique 9 pour appeler un "centre d'informations réservées" 8. Il peut s'agir, par exemple, de son agence bancaire ou de tout autre organisme similaire (agent en bourse, etc.).

Pour fixer les idées, on va supposer que le locuteur désire donner un ordre de transfert d'une somme donnée de son compte vers un autre compte. La  
35 donnée la plus importante à authentifier est naturellement l'identité du donneur d'ordre. On peut faire appel aux techniques habituelles telles que mots de passe ou

technique similaire, ou encore rappel du demandeur par le centre bancaire à un numéro prévu à l'avance pour s'assurer de l'identité de l'appelant. Cependant, ces techniques présentent une sécurité limitée.

5 Aussi, un appel du type précité sera commuté sur un système 5 de vérification de l'identité du locuteur U, similaire au système 5 de la figure 3. Dans une autre variante, illustrée par la figure 4, des numéros téléphoniques sont réservés à ce type d'appel et ce dernier aboutit directement sur un central spécialisé (non représenté) couplé au système 5.

A l'aide, par exemple, d'un synthétiseur de voix (non représenté), le système 5 peut "accueillir" le locuteur et le guider pour les opérations subséquentes.

Si on suppose que le téléphone comporte un clavier à touches numériques K, le locuteur U peut entrer un code secret d'identification à l'aide de celui-ci. Cette saisie initie la phase  $P_1$  du procédé.

15 L'entrée du code pourrait également s'effectuer à l'aide d'un appareil du type "MINITEL" muni d'un lecteur de carte.

Le synthétiseur de voix précité peut alors demander au locuteur de prononcer une ou plusieurs phrases, ce qui permet l'exécution des autres phases propres à la branche "Vérification" (figure 2 : 3), sauf s'il y a d'emblée refus (figure 2 : décision  $D_1$ , branche "Refus" 11). Cette éventualité existe si le code entré à l'aide du clavier n'est pas reconnu par le système : code erroné ou tentative frauduleuse. De façon classique, il pourra être accordé un nombre limité d'essais (par exemple trois). Chaque essai infructueux rebouclera sur la phase  $P_1$ . Au bout du  $N^{\text{ème}}$  essai infructueux (maximum permis), la liaison pourra être interrompue et, éventuellement, une trace de l'appel gardé en mémoire.

25 Si les phases de vérification se déroulent normalement, à la fin de la phase  $P_8$ , l'identité du locuteur aura été vérifiée formellement car sa voix aura été reconnue. Ceci est donc l'équivalent d'une "signature électronique", en l'occurrence sonore. Des essais ont montré que la sécurité est très élevée, car le taux de réussite (reconnaissance) est de l'ordre de 99 % sur plusieurs centaines de locuteurs. En outre, il doit être clair que le contenu du texte à prononcer est libre. Enfin, en cas de non-reconnaissance, le système ne donne pas accès aux ressources escomptées.

Dans l'exemple décrit, le système 5 peut donner accès à un système automatisé ou à une personne physique, sous la référence générale 8.

35 Dans le premier cas, le locuteur U pourra utiliser les touches de son clavier K pour transmettre ses ordres ou requêtes. Dans le second cas, il pourra directement converser avec son interlocuteur, ce dernier étant certain de l'identité de

l'appelant U. Le locuteur U pourra également être mis en communication avec un serveur vocal, par exemple pour obtenir des informations à diffusion restreinte (état de son compte, etc.).

5 Ce système permet donc d'effectuer un grand nombre d'opérations à distance ou d'obtenir toutes sortes d'informations, ce avec une sécurité maximale.

A la lecture de ce qui précède, on constate aisément que l'invention atteint bien les buts qu'elle s'est fixés.

10 Il doit être clair aussi que, bien que particulièrement adaptée aux deux séries d'applications qui ont été détaillées par référence aux figures 3 et 4, on ne saurait cantonner l'invention à ce seul type d'applications.

Elle trouve de nombreuses autres applications et notamment les suivantes :

- Indexation de données multimédias en terme de vérification du locuteur, d'identification du locuteur et de détection de changements de locuteurs ;
- 15 - Identification biométrique utilisée à des fins de preuve en justice, de vérification et validation de clauses contractuelles, de fabrication de carte d'identification ;
- Surveillance de canaux de communication.

## REVENDEICATIONS

1. Procédé de reconnaissance vocale d'un locuteur mettant en oeuvre un modèle prédictif d'ordre  $q$ , comprenant au moins une phase d'extraction de caractéristiques statistiques ( $P_4, P'_4$ ) comportant au moins une étape d'acquisition numérique d'un échantillon sonore d'une durée déterminée  $D$  du locuteur ( $U$ ), correspondant à au moins une élocution du locuteur ( $U$ ), une étape de transformation de cet échantillon sonore en une suite de vecteurs de dimension déterminée  $p$ , obtenus à partir d'une suite de fenêtres d'analyse de taille moyenne  $T$  et d'intervalle moyen  $I$ , et une étape de détermination de  $q+1$  matrices de corrélation à partir de cette suite de vecteurs, avec  $p$  et  $q$  nombres entiers différents de zéro, caractérisé en ce que ladite taille moyenne  $T$  a une durée inférieure à 10 ms.

2. Procédé selon la revendication 1, caractérisé en ce que ledit intervalle moyen  $I$  a une durée inférieure à 4,5 ms.

3. Procédé selon la revendication 2, caractérisé en ce que ledit intervalle moyen  $I$ , ladite taille moyenne  $T$  et ledit l'ordre  $q$  du modèle prédictif sont reliés par l'inéquation suivante :

$$I < T/\max(3, q)$$

4. Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que, lesdites matrices de corrélation et ledit modèle prédictif correspondant étant stockés dans au moins un dictionnaire (51), il comprend une phase ( $P_9$ ) de mise à jour d'un desdits dictionnaires (51) pour un locuteur déterminé ( $U$ ) comportant au moins une première étape d'extraction à partir d'un dictionnaire (51) desdites  $q+1$  matrices de corrélation de ce locuteur, une deuxième étape de détermination d'une nouvelle suite de  $q+1$  matrices de corrélation à partir de vecteurs de dimension  $p$ , obtenus par la transformation d'un nouvel échantillon sonore du locuteur en cette suite de vecteurs et une troisième étape de détermination d'une suite de  $q+1$  matrices de corrélation mises à jour consistant à l'accumulation et au stockage dans l'un desdits dictionnaires (51) des  $q+1$  matrices de corrélation extraites pendant la première étape et des  $q+1$  matrices de corrélation déterminées pendant la deuxième étape.

5. Procédé selon la revendication 4, caractérisé en ce que ladite troisième étape comprend en outre la détermination d'un modèle prédictif correspondant à la détermination de la suite de  $q+1$  matrices de corrélation mises à jour et à son stockage dans ledit dictionnaire.

6. Procédé selon l'une quelconque des revendications 1 à 3, caractérisé en ce que, lesdites matrices de corrélation et ledit modèle prédictif correspondant

étant stockés dans au moins un dictionnaire (51), il comprend une séquence de phases d'apprentissage ( $P_2, P_4, P_5, P_7, P_9$ ) des données statistiques associées à un nouveau locuteur (U), en ce que cette séquence comprend au moins une phase ( $P_9$ ) comportant au moins une première étape de détermination d'une nouvelle suite de  
5  $q+1$  matrices de corrélation à partir de vecteurs de dimension  $p$ , obtenus par la transformation d'un nouvel échantillon sonore du nouveau locuteur (U) en cette suite de vecteurs, à la détermination d'un modèle prédictif correspondant à cette suite de  $q+1$  matrices de corrélation et une deuxième étape consistant au stockage dans lesdits dictionnaires (51) des  $q+1$  matrices de corrélation et du modèle prédictif  
10 déterminées pendant la première étape.

7. Procédé selon l'une quelconque des revendications 1 à 3, caractérisé en ce qu'il comprend une séquence de phases ( $P_3, P'_4, P'_5, P_6, P_8$ ) permettant la vérification de l'identité d'un locuteur et en ce que cette séquence comprend au moins les phases suivantes :

15 - une phase ( $P'_5$ ) de détermination d'un modèle prédictif comportant une première étape consistant à déterminer, à partir desdites caractéristiques statistiques d'au moins une élocution,  $q$  matrices de prédiction et une deuxième étape réalisant le calcul de l'inverse de la matrice d'erreur entre une suite des vecteurs de dimension  $p$  calculés à partir de l'échantillon sonore de chacune des élocutions et le  
20 modèle prédictif calculé sur ces mêmes vecteurs, l'ensemble des  $q+1$  matrices définissant le modèle prédictif de l'élocution du locuteur et les références de l'élocution d'un locuteur étant constituées desdites caractéristiques statistiques et du modèle prédictif associé ;

25 - une phase ( $P_6$ ) de vérification de l'identité dudit locuteur comportant une première étape pendant laquelle des références d'élocution d'au moins une partie des locuteurs sont extraites d'un dictionnaire, une deuxième étape de calcul de matrices d'erreur entre au moins une partie des références d'élocution précédemment extraites et les références d'élocution de l'échantillon sonore à vérifier, une troisième étape consistant en un calcul des mesures de ressemblance à partir d'au moins une  
30 partie des matrices d'erreur calculées pendant la deuxième étape et une quatrième étape de calcul de la vraisemblance de la vérification dudit locuteur à partir des mesures de vraisemblance calculées lors de la troisième étape ;

et une phase ( $P_8$ ) autorisant l'accès à au moins une partie de ressources prédéterminées préalablement demandées par ledit locuteur au cours d'une phase  
35 préliminaire ( $P_1$ ), lorsque la vérification de ladite identité est positive.

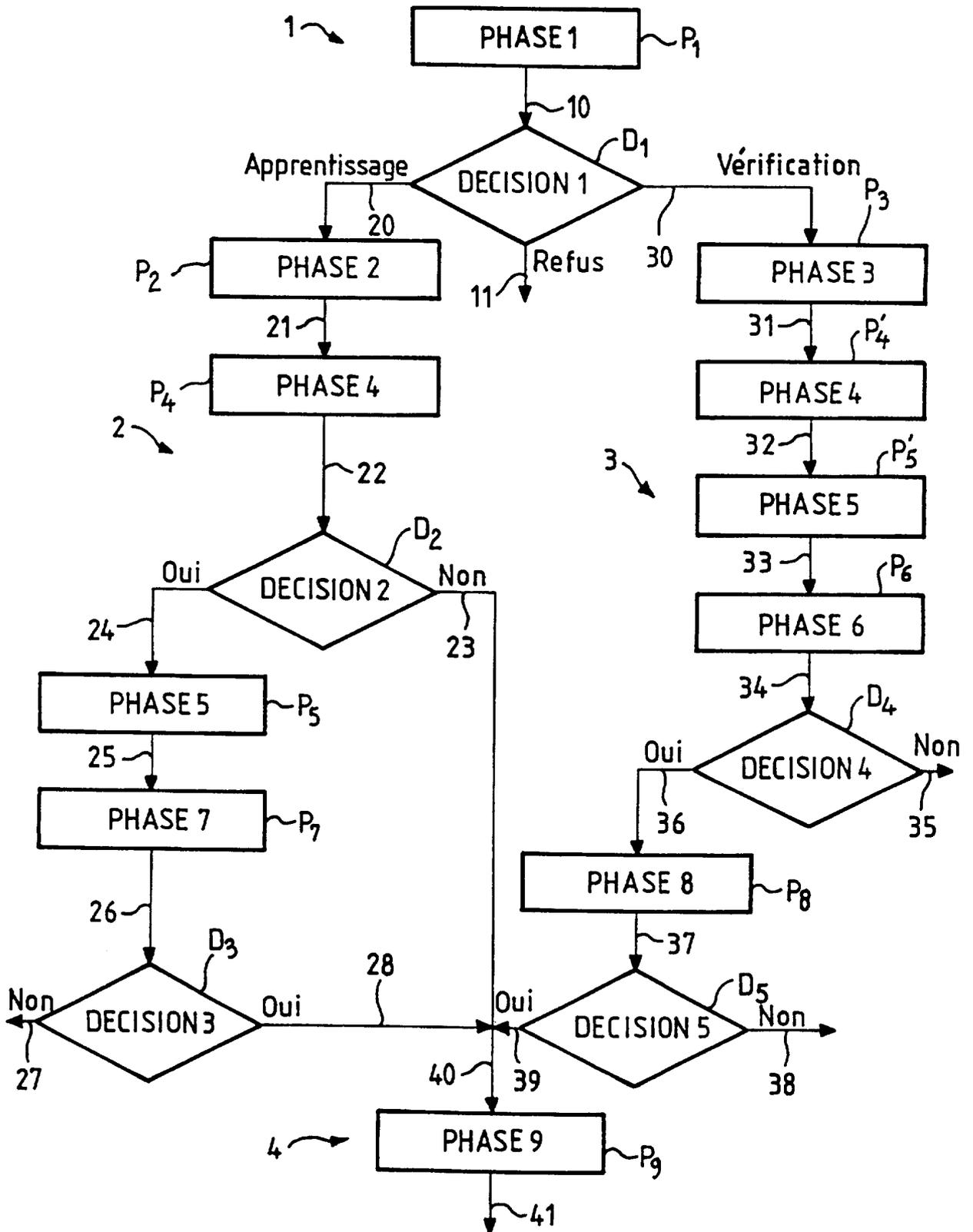
8. Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que ledit modèle prédictif est un modèle prédictif autorégressif vectoriel.

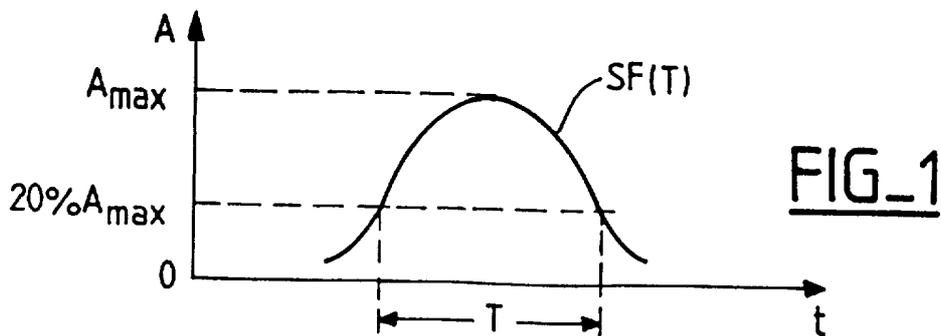
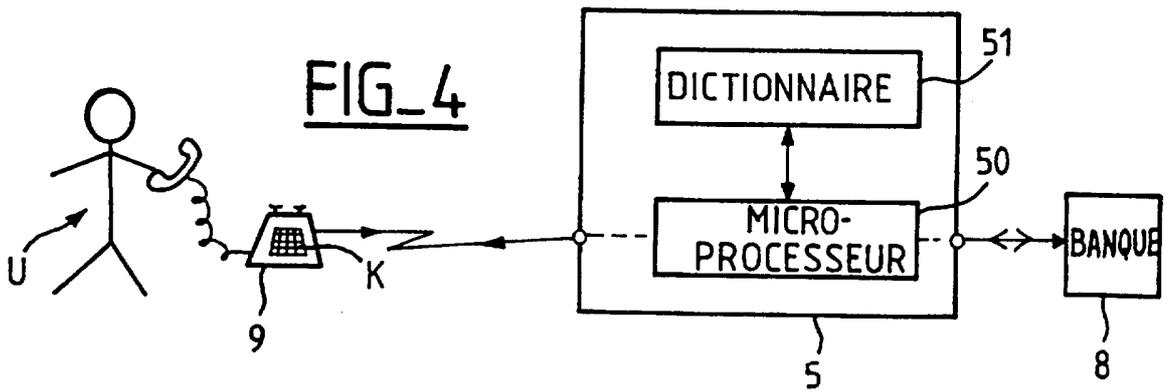
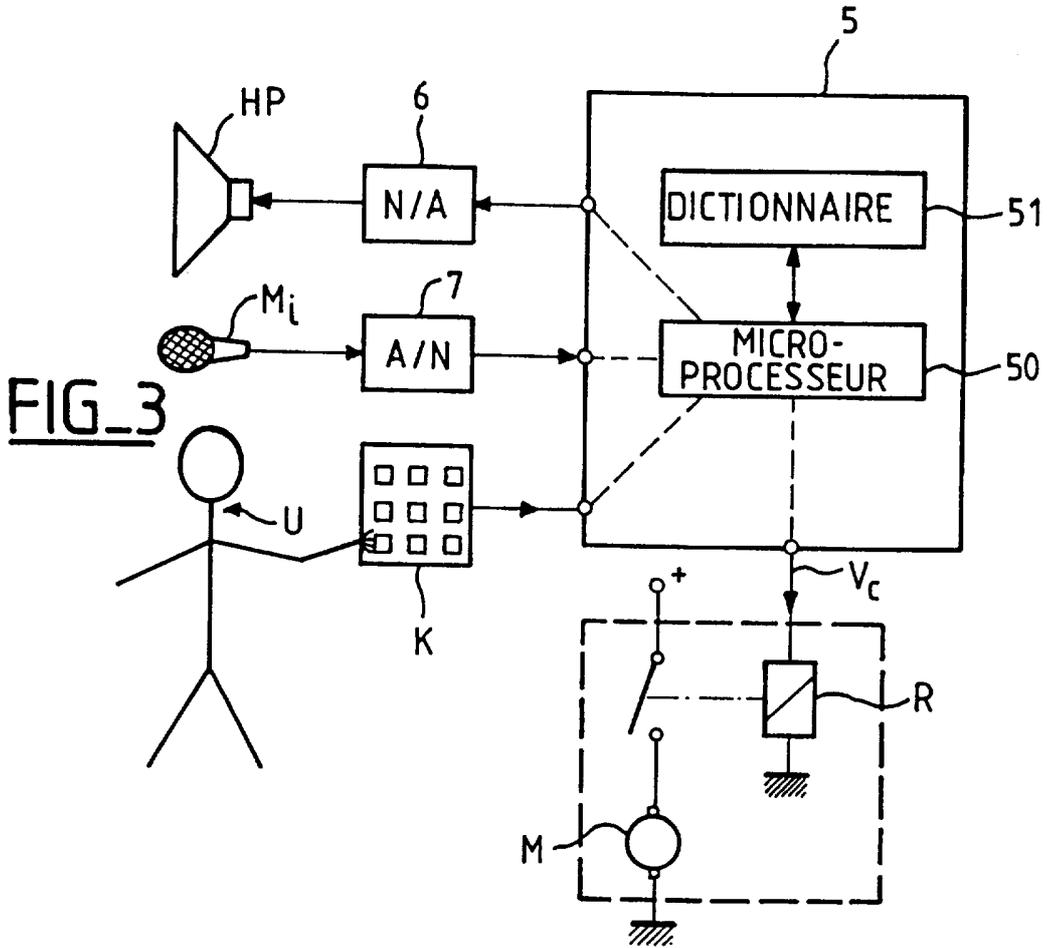
5 9. Système d'accès sécurisé à une ressource réservée à au moins un locuteur autorisé, système mettant en oeuvre le procédé selon les revendications 7 ou 8, caractérisé en ce qu'il comprend une serrure vocale sensible à un échantillon sonore d'un locuteur (U) à vérifier, la serrure comportant au moins un transducteur électroacoustique ( $M_i$ ) couplé à un dispositif de conversion (7) de l'échantillon sonore dudit locuteur en un signal numérique, un système de calcul (5, 50) à  
10 programme enregistré, couplé à au moins un dictionnaire (51), recevant ce signal numérique pour déterminer lesdites références d'élocution de l'échantillon sonore à vérifier, pour vérifier à partir de ces références et de références d'élocutions stockées dans le dictionnaire (51) l'identité du locuteur (U), et pour délivrer un signal de commande ( $V_C$ ) présentant au moins deux états, un état étant représentatif d'une  
15 vérification positive de l'identité d'un locuteur autorisé permettant l'accès à ladite ressource.

10. Système d'accès sécurisés à une transaction réservée à au moins un locuteur autorisé, système mettant en oeuvre le procédé selon les revendications 7 ou 8, caractérisé en ce qu'il comprend au moins un dispositif téléphonique (9)  
20 comportant au moins un transducteur électroacoustique, le dispositif téléphonique transmettant (9), par une ligne téléphonique, au moins un échantillon sonore d'un locuteur (U) à vérifier à un système de calcul (5, 50) à programme enregistré, couplé à au moins un dictionnaire (51), le système déterminant lesdites références d'élocution de l'échantillon sonore à vérifier, vérifiant à partir de ces références et de  
25 références d'élocutions stockées dans le dictionnaire (51) l'identité du locuteur (U) et autorisant ladite transaction sur vérification positive de l'identité d'un locuteur autorisé.

1/2

FIG\_2





# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/FR 97/00753

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 6 G10L5/06

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY (EUROSPEECH), vol. 1, 21 September 1993, pages 161-164, XP000576662 MONTACIE C ET AL: "DISCRIMINANT AR-VECTOR MODELS FOR FREE-TEXT SPEAKER VERIFICATION" cited in the application see the whole document ---	1
A	US 5 054 084 A (TANAKA ATSUO ET AL) 1 October 1991 see column 2, line 58 - line 63 ---	1
A	US 4 956 865 A (LENNING ET AL.) 11 September 1990 see column 2, line 9 - line 17 ---	1
	-/--	

Further documents are listed in the continuation of box C.       Patent family members are listed in annex.

- \* Special categories of cited documents :
- "A" document defining the general state of the art which is not considered to be of particular relevance
  - "E" earlier document but published on or after the international filing date
  - "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
  - "O" document referring to an oral disclosure, use, exhibition or other means
  - "P" document published prior to the international filing date but later than the priority date claimed
  - "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
  - "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
  - "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
  - "&" document member of the same patent family

Date of the actual completion of the international search  <b>18 July 1997</b>	Date of mailing of the international search report  <b>- 1. 08. 97</b>
--	--

Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Authorized officer  <b>Van Doremalen, J</b>
---	---

# INTERNATIONAL SEARCH REPORT

Int'l. Application No  
PCT/FR 97/00753

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 097 509 A (LENNIG MATTHEW) 17 March 1992 see column 3, line 1 - line 3 -----	1

1

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter. Application No

PCT/FR 97/00753

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5054084 A	01-10-91	JP 62235998 A	16-10-87
US 4956865 A	11-09-90	CA 1232686 A	09-02-88
		CN 1009320 B	22-08-90
		EP 0192898 A	03-09-86
		JP 7097279 B	18-10-95
		JP 61177494 A	09-08-86
		KR 9402437 B	24-03-94
US 5097509 A	17-03-92	NONE	

# RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No

PCT/FR 97/00753

A. CLASSEMENT DE L'OBJET DE LA DEMANDE  
CIB 6 G10L5/06

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)  
CIB 6 G10L

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY (EUROSPEECH), vol. 1, 21 Septembre 1993, pages 161-164, XP000576662 MONTACIE C ET AL: "DISCRIMINANT AR-VECTOR MODELS FOR FREE-TEXT SPEAKER VERIFICATION" cité dans la demande voir le document en entier ---	1
A	US 5 054 084 A (TANAKA ATSUO ET AL) 1 Octobre 1991 voir colonne 2, ligne 58 - ligne 63 ---	1
A	US 4 956 865 A (LENNING ET AL.) 11 Septembre 1990 voir colonne 2, ligne 9 - ligne 17 ---	1
	-/--	

Voir la suite du cadre C pour la fin de la liste des documents

Les documents de familles de brevets sont indiqués en annexe

\* Catégories spéciales de documents cités:

\*A\* document définissant l'état général de la technique, non considéré comme particulièrement pertinent

\*E\* document antérieur, mais publié à la date de dépôt international ou après cette date

\*L\* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

\*O\* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

\*P\* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

\*T\* document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

\*X\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

\*Y\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

\*Z\* document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

18 Juillet 1997

Date d'expédition du présent rapport de recherche internationale

- 1. 08. 97

Nom et adresse postale de l'administration chargée de la recherche internationale

Office Européen des Brevets, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+ 31-70) 340-3016

Fonctionnaire autorisé

Van Doremalen, J

# RAPPORT DE RECHERCHE INTERNATIONALE

Dem. Internationale No  
PCT/FR 97/00753

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	US 5 097 509 A (LENNIG MATTHEW) 17 Mars 1992 voir colonne 3, ligne 1 - ligne 3 -----	1

**RAPPORT DE RECHERCHE INTERNATIONALE**

Renseignements relatifs aux membres de familles de brevets

Den. de l'Internationale No

PCT/FR 97/00753

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
US 5054084 A	01-10-91	JP 62235998 A	16-10-87
-----			
US 4956865 A	11-09-90	CA 1232686 A	09-02-88
		CN 1009320 B	22-08-90
		EP 0192898 A	03-09-86
		JP 7097279 B	18-10-95
		JP 61177494 A	09-08-86
		KR 9402437 B	24-03-94
-----			
US 5097509 A	17-03-92	AUCUN	
-----			