



(12) DEMANDE DE BREVET D'INVENTION

- (11) N° de publication : **MA 38738 A1** (51) Cl. internationale : **G10L 15/00; G10L 17/00; G10L 15/02**
- (43) Date de publication : **31.10.2017**

-
- (21) N° Dépôt : **38738**
- (22) Date de Dépôt : **29.12.2015**
- (71) Demandeur(s) : **UNIVERSITÉ MOHAMMED V DE RABAT, Angle avenue Allal El Fassi et Mfadel Cherkaoui, Alirfane 8007.N.U, Rabat Rabat-Chellah (MA)**
- (72) Inventeur(s) : **Mourad Raji ; Maazouzi Abderrahim**
- (74) Mandataire : **ZAOUI FATIMA**

-
- (54) Titre : **Système d'appariement pour la reconnaissance rapide de locuteur**
- (57) Abrégé : La majorité des systèmes de reconnaissance de locuteur sont basés sur le calcul de la distance ou la probabilité entre de vecteurs caractéristiques d'un locuteur inconnu et les modèles enregistrés dans la base de données du système. Le processus d'identification dépend du nombre de ces vecteurs caractéristiques, leurs dimensions et le nombre de locuteur à identifier. L'idée de cet algorithme est de développer un système capable d'identifier une personne en utilisant un échantillon de sa parole. Nous avons choisi de travailler avec les systèmes dépendant du mot prononcé (word-dependent system) en utilisant par exemple le mot correspondant au chiffre « one » en anglais comme mot de passe. En effet, le système identifie les locuteurs même s'ils prononcent le même mot de passe, en d'autres termes, nous cherchons à trouver la signature vocale de chaque personne dans la prononciation du mot « one ». Les caractéristiques du signal vocal sont extraites en utilisant l'estimation de la densité spectrale de puissance, et la reconnaissance est faite à l'aide de l'algorithme de RAJI (Journal of computational chemistry, 1997) basé sur le calcul de la distance euclidienne par rapport à une interpolation cubique des signaux constituant la base de données.

Abrégé

La majorité des systèmes de reconnaissance de locuteur sont basés sur le calcul de la distance ou la probabilité entre de vecteurs caractéristiques d'un locuteur inconnu et les modèles enregistrés dans la base de données du système. Le processus d'identification dépend du nombre de ces vecteurs caractéristiques, leurs dimensions et le nombre de locuteur à identifier. L'idée de cet algorithme est de développer un système capable d'identifier une personne en utilisant un échantillon de sa parole. Nous avons choisi de travailler avec les systèmes dépendant du mot prononcé (word-dependent system) en utilisant par exemple le mot correspondant au chiffre « one » en anglais comme mot de passe. En effet, le système identifie les locuteurs même s'ils prononcent le même mot de passe, en d'autres termes, nous cherchons à trouver la signature vocale de chaque personne dans la prononciation du mot « one ». Les caractéristiques du signal vocal sont extraites en utilisant l'estimation de la densité spectrale de puissance, et la reconnaissance est faite à l'aide de l'algorithme de RAJI (Journal of computational chemistry, 1997) basé sur le calcul de la distance cosinus par rapport à une interpolation cubique des signaux constituant la base de données.

TITRE : Système d'appariement pour la reconnaissance rapide de locuteur

Description :

La majorité des systèmes de reconnaissance de locuteur sont basés sur le calcul de la distance ou la probabilité entre de vecteurs caractéristiques d'un locuteur inconnu et les modèles enregistrés dans la base de données du système. Le processus d'identification dépend du nombre de ces vecteurs caractéristiques, leurs dimensions et le nombre de locuteur à identifier. L'idée de cet algorithme est de développer un système capable d'identifier une personne en utilisant un échantillon de sa parole. Nous avons choisi de travailler avec les systèmes dépendant du mot prononcé (word-dependent system) en utilisant par exemple le mot correspondant au chiffre « one » en anglais comme mot de passe. En effet, le système identifie les locuteurs même s'ils prononcent le même mot de passe, en d'autres termes, nous cherchons à trouver la signature vocale de chaque personne dans la prononciation du mot « one ». Les caractéristiques du signal vocal sont extraites en utilisant l'estimation de la densité spectrale de puissance, et la reconnaissance est faite à l'aide de l'algorithme de RAJI (Journal of computational chemistry, 1997) basé sur le calcul de la distance cosinus par rapport à une interpolation cubique des signaux constituant la base de données.

Pour une reconnaissance optimale, la phase de prétraitement du signal est d'une importance primordiale. Les segments de la voix doivent être séparés des parasites pour atteindre une précision de reconnaissance élevée. L'objectif de l'algorithme Endpoint detection est de déterminer le début et la fin de chaque mot prononcé, et également de supprimer les régions de silence ou de bruit. L'algorithme est basé sur la détermination du niveau d'énergie et le taux zero-crossing.

L'estimation de la densité spectrale de puissance est l'une des applications les plus importantes dans le traitement de signal. Les problèmes de la reconnaissance de la voix utilisent l'analyse spectrale comme étant une phase préliminaire, pour déterminer le contenu du signal, ainsi que d'autres traitements acoustiques. L'algorithme de Welch est un algorithme non-paramétrique pour estimer la densité spectrale de puissance. Il est obtenu par le raffinement de la méthode de Barlett en deux aspects. Premièrement, les segments de données dans l'algorithme de Welch sont laissés chevauchés. En second temps, chaque segment de donnée subit un fenêtrage avant que le periodogramme (permettant d'identifier les périodes ou fréquences dominantes dans les séries temporelles) soit calculé. Généralement, les autres techniques d'extraction qui existent, nécessitent des traitements de matrices volumineuses donc plus de temps de réponse. Dans notre système, nous avons utilisé la fenêtre de hamming et une taille de 512 pour la transformée de fourier qui permettent de générer un vecteur de 257 éléments dans la bande de fréquence inférieure à 4Khz ce qui permet d'identifier le

locuteur d'une manière rapide. L'idée est de comparer les signaux dans cette bande contenant plus d'information sur le contenu du mot.

Après avoir extrait les caractéristiques du signal sous forme de vecteur (voir figure 1), le but est de comparer les vecteurs caractéristiques des personnes constituant la base de données. L'idée est de différencier entre les locuteurs même s'ils utilisent le même mot de passe. Autrement dit, on essaye de trouver une transformation qui permet de caractériser une personne par rapport à une autre même s'ils prononcent le même mot de passe. L'objectif de cette étape est de calculer d'une manière précise et rapide la similarité entre deux signaux (voir la figure 2 montrant l'architecture du système). Dans le processus de vérification, le signal test est comparé directement avec son correspondant pour confirmer l'identité prononcée par l'utilisateur, c'est une comparaison 1-1 (one-to-one). En revanche, l'identification consiste à comparer le signal test avec l'ensemble des signaux enregistrés dans la base de données afin de déterminer l'identité du locuteur. Généralement, la décision peut être affectée soit par la forte similarité qui peut exister entre certains signaux appartenant à différents locuteurs, soit par une fausse acceptation, c'est le cas d'un locuteur qui n'est pas enregistré dans la base de données. Du coup, l'utilisation à la fois d'un algorithme d'appariement efficace et rapide en fixant un seuil adéquat pour empêcher une fausse acceptation s'avère très utile.

La similarité est mesurée au sens de la distance cosinus et l'appariement des signaux est basé sur l'algorithme de Rami utilisant l'interpolation 'cubic spline' du signal (voix) stocké dans la base de données et une transformation T qui permet la meilleure superposition de la voix de la personne à reconnaître sur ladite interpolation cubique de la base de données. Cela doit permettre de trouver la meilleure représentation de chaque personne de la base de données (voir figure 3). Le choix de bande de fréquence pour estimer la densité spectrale et la transformation T adéquate ont permis d'obtenir des résultats très satisfaisants dépassant 91% de taux de reconnaissance. En outre, l'utilisation d'un vecteur de 257 éléments pour représenter le contenu de signal (au lieu d'une représentation matricielle utilisée par les autres algorithmes) ainsi que la distance cosinus au sein de la transformation ont aidé à diminuer le temps de réponse du système.

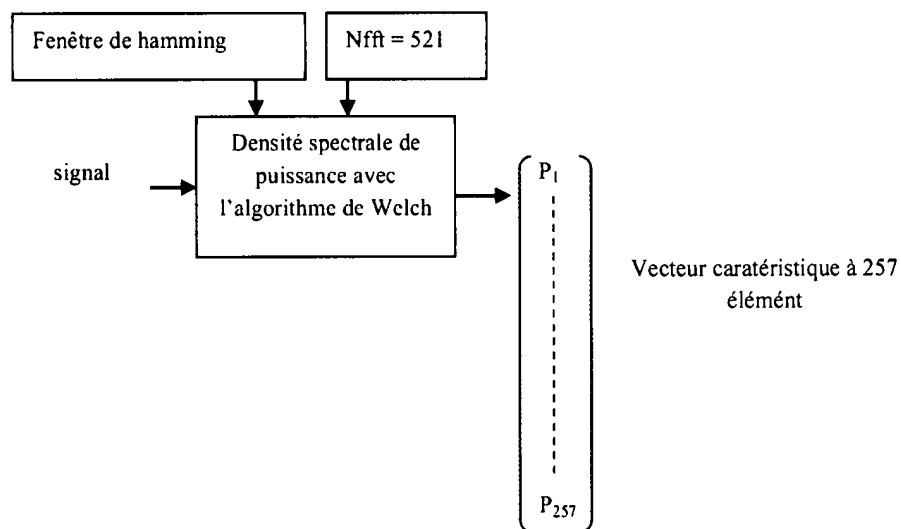
Figure 1 : méthode d'extraction du vecteur caractéristique du locuteur

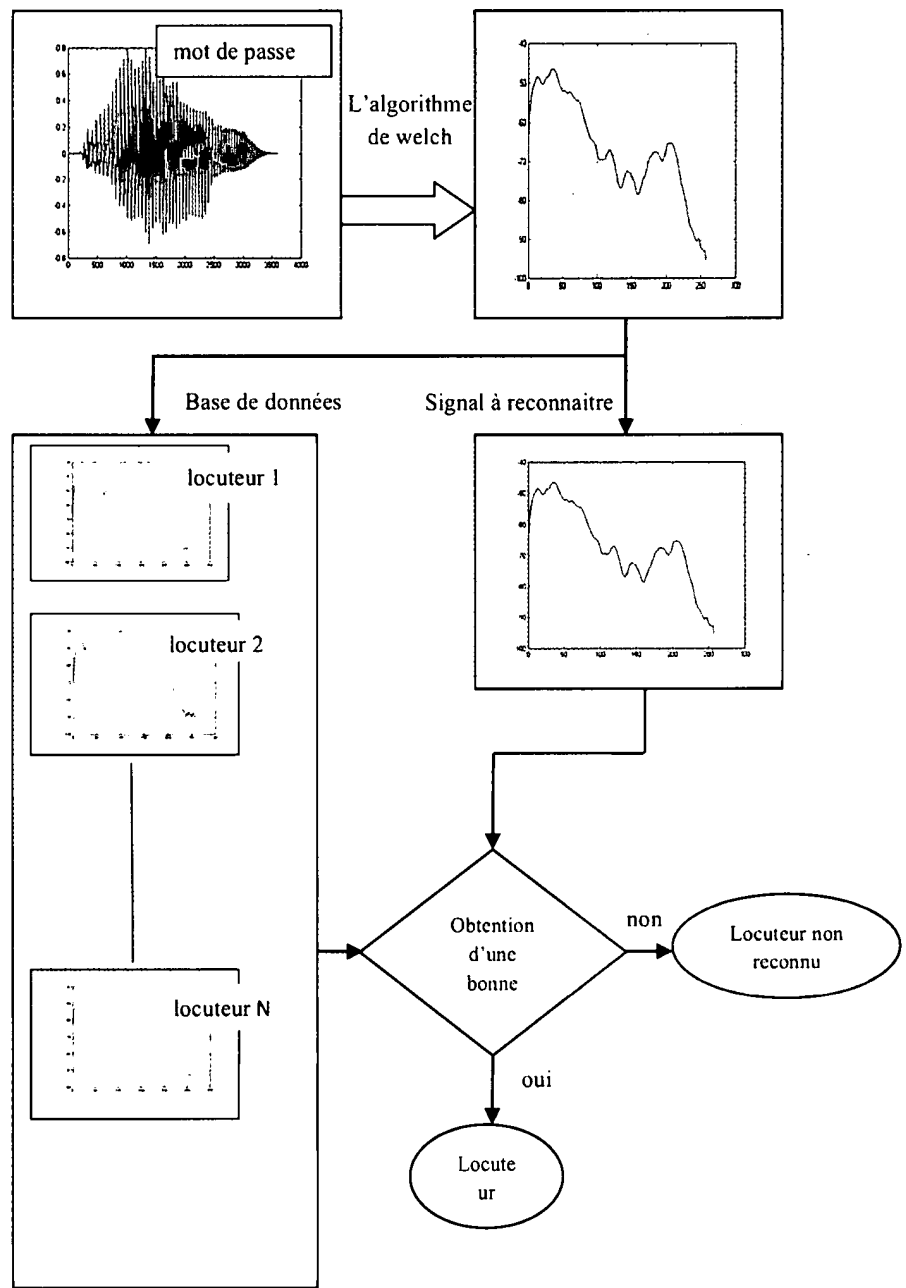
Figure 2 : architecture du système.

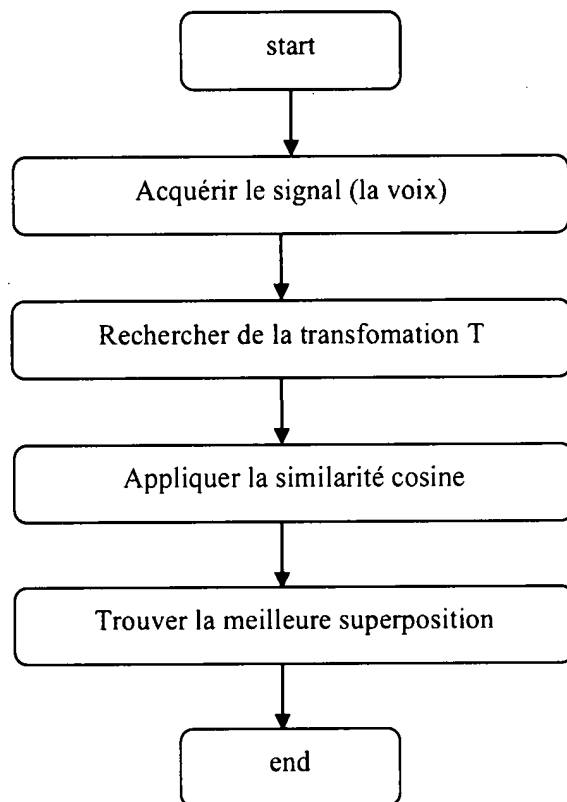
Figure 3: Etapes de l'algorithme d'appariement

Revendications

- 1- Un système de reconnaissance de locuteur, caractérisé en ce qu'il comprend :
 - Un prétraitement du signal vocal pour diminuer la contribution des parasites.
 - Une méthode d'extraction de vecteur caractéristique (257 éléments qui s'avère suffisant pour traiter le problème au lieu de taille plus grande comme suggérer par les algorithmes existants) du signal basée sur l'estimation de la densité spectrale de puissance au lieu de l'utilisation d'autres techniques d'extraction qui nécessitent des traitements de matrices volumineuses.
 - L'utilisation de l'algorithme dit de passage du discret au continu par interpolation cubique permettant de caractériser la personne en se basant sur la superposition de la représentation de la personne à partir de son vecteur caractéristique.
- 2- système selon la revendication 1, caractérisé en ce qu'il comprend un algorithme de détection du début et de la fin de l'enregistrement vocal pour éliminer le silence et le bruit, et que l'extraction des caractéristiques est effectuée à l'aide de l'algorithme de Welch, et que la reconnaissance est basée sur l'interpolation des splines cubiques avec une transformation donnant une meilleur de la représentation de la personne à partir de son vecteur caractéristique.
- 3- système selon la revendication 2, caractérisé en ce qu'il comprend un algorithme d'appariement utilisant une transformation spécifique permettant de caractériser chaque personne de la base donnée en fonction du mot de passe utilisé.
- 4- système selon l'une quelconque des revendications 1 à 3, caractérisé en ce qu'il comprend en outre la possibilité d'utiliser un mot de passe quelconque.









**RAPPORT DE RECHERCHE
AVEC OPINION SUR LA BREVETABILITE**
(Conformément aux articles 43 et 43.2 de la loi 17-97 relative à la
protection de la propriété industrielle telle que modifiée et
complétée par la loi 23-13)

Renseignements relatifs à la demande	
N° de la demande : 38738	Date de dépôt : 29/12/2015 ;
Déposant : UNIVERSITÉ MOHAMMED V DE RABAT	
Intitulé de l'invention : Système d'appariement pour la reconnaissance rapide de locuteur	
Le présent document est le rapport de recherche avec opinion sur la brevetabilité établi par l'OMPIC conformément aux articles 43 et 43.2, et notifié au déposant conformément à l'article 43.1 de la loi 17-97 relative à la protection de la propriété industrielle telle que modifiée et complétée par la loi 23-13.	
Les documents brevets cités dans le rapport de recherche sont téléchargeables à partir du site http://worldwide.espacenet.com , et les documents non brevets sont joints au présent document, s'il y en a lieu.	
Le présent rapport contient des indications relatives aux éléments suivants :	
Partie 1 : Considérations générales	
<input checked="" type="checkbox"/> Cadre 1 : Base du présent rapport <input type="checkbox"/> Cadre 2 : Priorité <input type="checkbox"/> Cadre 3 : Titre et/ou Abrégé tel qu'ils sont définitivement arrêtés	
Partie 2 : Rapport de recherche	
Partie 3 : Opinion sur la brevetabilité	
<input checked="" type="checkbox"/> Cadre 4 : Remarques de clarté <input checked="" type="checkbox"/> Cadre 5 : Déclaration motivée quant à la Nouveauté, l'Activité Inventive et l'Application Industrielle <input type="checkbox"/> Cadre 6 : Observations à propos de certaines revendications dont aucune recherche significative n'a pu être effectuée <input type="checkbox"/> Cadre 7 : Défaut d'unité d'invention	
Examineur: F.Belafkih	Date d'établissement du rapport : 14/11/2017
Téléphone: 212 5 22 58 64 14/00	



Partie 1 : Considérations générales		
Cadre 1 : base du présent rapport		
Les pièces suivantes de la demande servent de base à l'établissement du présent rapport :		
<ul style="list-style-type: none"> • <u>Description</u> 2 Pages • <u>Revendications</u> 4 • <u>Planches de dessin</u> 3 Pages 		
Partie 2 : Rapport de recherche		
Classement de l'objet de la demande :		
CIB : G10L 15/00 ; G10L 15/02 ; G10L 17/00		
Bases de données électroniques consultées au cours de la recherche :		
EPOQUE, Orbit		
Catégorie*	Documents cités avec, le cas échéant, l'indication des passages pertinents	N° des revendications visées
Y	WO2001035389 A1 ; Koninklijke Philips Electronics N.V. ; 17 Mai 2001 Tout le document ;	1-4
Y	Chapitre II : Reconnaissance De La Parole ; 25 Octobre 2002 http://www-primia.inrialpes.fr/Vaufreydaz/These/Reconnaissance.html Tout le document ;	1-4
Y	EP0508547 B1 ; Philips Electronics N.V., Philips Patentverwaltung Gmbh ; 26 Août 1998 Tout le document	1-4
A	CN101685634 B ; Shanghai Shengtao Intelligent Technology Co Ltd ; 21 Novembre 2012 Tout le document	1-4
*Catégories spéciales de documents cités :		
-« 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 -« A » document définissant l'état général de la technique, non considéré comme particulièrement pertinent -« P » documents intercalaires ; Les documents dont la date de publication est située entre la date de dépôt de la demande examinée et la date de priorité revendiquée ou la priorité la plus ancienne s'il y en a plusieurs -« E » Éventuelles demandes de brevet interférentes. Tout document de brevet ayant une date de dépôt ou de priorité antérieure à la date de dépôt de la demande faisant l'objet de la recherche (et non à la date de priorité), mais publié postérieurement à cette date et dont le contenu constituerait un état de la technique pertinent pour la nouveauté		

Partie 3 : Opinion sur la brevetabilité*Cadre 4 : Remarques de clarté*

1. Il est à noter que les revendications de système sont considérées comme des revendications de dispositif, et non comme des revendications de méthode ou de procédé. Pour satisfaire aux exigences de clarté, le système faisant l'objet de la revendication 1 devrait par conséquent être défini en termes de caractéristiques de dispositif, et non par les étapes d'une méthode ou d'un procédé.
2. La revendication 1 ne satisfait pas aux exigences de clarté car l'objet de la protection demandée n'est pas défini. La revendication tente de définir l'objet par le résultat recherché « méthode d'extraction de vecteur caractéristique » au lieu de le définir en termes de caractéristiques techniques. En tout état de cause, cette formulation n'est pas acceptable en l'espèce, puisqu'il semble possible de définir l'objet en des termes plus concrets, c'est-à-dire en exposant comment l'effet peut être obtenu.

Cadre 5 : Déclaration motivée quant à la Nouveauté, l'Activité Inventive et l'Application Industrielle

Nouveauté (N)	Revendications 1-4 Revendications aucune	Oui Non
Activité inventive (AI)	Revendications 1-4 Revendications aucune	Oui Non
Possibilité d'application Industrielle (PAI)	Revendications 1-4 Revendications aucune	Oui Non

Il est fait référence aux documents suivants. Les numéros d'ordre qui leur sont attribués ci-après seront utilisés dans toute la suite de la procédure

D1 : WO2001035389 A1

D2 : <http://www-prima.inrialpes.fr/Vaufreydaz/These/Reconnaissance.html>

1. Nouveauté (N) :

Aucun des documents ci-dessus ne divulgue l'ensemble des caractéristiques techniques des revendications 1-4, d'où l'objet desdites revendications est nouveau au sens de l'article 26 de la loi 17-97 modifiée et complétée par la loi 23-13.

2. Activité inventive (AI) :

Le document D1 qui est considéré comme l'état de la technique le plus proche de l'objet de la demande décrit un système de reconnaissance vocale caractérisé par :

- Un prétraitement du signal vocal (D1 : Description Page 1 lignes 9-10) ;
- Extraction du vecteur caractéristique basée sur l'estimation de la densité spectrale de puissance (D1: Description Page 1 lignes 7-8) ;
- Passage du discret au continu par interpolation cubique (D1 : Description Page 6 lignes 24-32) ;

Par conséquent, l'objet de la revendication 1 diffère de l'état de la technique D1 en ce que : Le vecteur caractéristique possède 257 éléments.

L'effet technique lié à l'utilisation de 257 éléments (tel que revendiqué par le déposant) est de réduire le nombre d'éléments représentatifs pour le vecteur caractéristique et ainsi le temps de réponse du système. Toutefois l'état de la technique D1 propose l'utilisation de 24, 32 ou 63 éléments (D1 : Description Page 1 lignes 14-15), d'où l'effet est déjà obtenu dans l'état de la technique. Par conséquent, cette caractéristique n'implique également pas une activité inventive au sens de l'article 28 de la loi 17-97 modifiée et complétée par la loi 23-13.

Les revendications dépendantes 2-4 (dans la mesure où ces revendications sont compréhensibles) ne contiennent aucune caractéristique technique qui, en combinaison avec l'une quelconque des revendications auxquelles les revendications 2-4 sont liées, implique une activité inventive au sens de l'article 28 de la loi 17-97 modifiée et complétée par la loi 23-13.

3. Possibilité d'application industrielle (PAI) :

L'objet de la présente invention est susceptible d'application industrielle au sens de l'article 29 de la loi 17-97 telle que modifiée et complétée par la loi 23-13, parce qu'il présente une utilité déterminée, probante et crédible